Thomas Vogl MSc BSc

Synthetic biology to improve protein expression in Pichia pastoris

DOCTORAL THESIS

to achieve the university degree of

Doktor der Naturwissenschaften

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder

Institute of Molecular Biotechnology

Graz, January 2015

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

19.1.2015

Date

Signature

Abstract

Biopharmaceuticals, industrial biocatalysts and fine chemicals are widely produced by recombinant gene expression of single genes and entire metabolic pathways in microorganisms. The methylotrophic yeast *Pichia pastoris* is one of the most commonly used production hosts for single proteins and has recently also been adapted as a chassis for metabolic engineering endeavors. *P. pastoris* provides tightly regulated, exceptionally strong methanol inducible promoters, however their molecular regulation and exact properties were incompletely characterized.
In this work, synthetic biology concepts were applied for understanding and altering transcriptional regulation in *P. pastoris*. Insights on regulatory regions allowed the design of more efficient synthetic promoters and basic transcriptional activator circuits were generated. These novel tools are powerful enabling technologies for recombinant single gene expression and the functional expression and optimization of entire heterologous pathways in *P. pastoris*.

Zusammenfassung

Biopharmazeutika, industriell verwendete Biokatalysatoren und Feinchemikalien werden gängiger Weise durch rekombinante Gen Expression von einzelnen Genen oder ganzen Stoffwechselwegen in Mikroorgansimen produziert. Die methylotrophe Hefe *Pichia pastoris* ist eines der am häufigsten verwendeten mikrobiellen Expressionssysteme und wurde unlängst auch als Plattform für Metabolic Engineering Anwendungen adaptiert. *P. pastoris* verfügt über stringent regulierte, außergewöhnlich starke Methanol-induzierbare Promotoren. Jedoch sind deren molekulare Regulation und genaue Eigenschaften unvollständig erforscht.
In dieser Arbeit wurden Synthetic Biology Konzepte angewandt um die transkriptionelle Regulation in *P. pastoris* zu modifizieren. Tieferes Verständnis von dem Zusammenspiel regulatorischer Regionen ermöglichte das Design von effizienteren Promotoren und ein simpler Transkription aktivierender Schaltkreis wurde generiert. Diese neuartigen molekularen Werkzeuge haben das Potential die rekombinante Produktion von einzelnen Proteinen und die Expression von ganzen Stoffwechselwegen in *P. pastoris* drastisch zu verbessern.

Table of contents

Acknowledgments

This work would have never been possible to this extent without the dedicated participation and efforts of many people. Several projects were perused in close collaboration and equal contribution with outstanding colleagues and also contributions by highly motivated internship students were invaluable for completing the work. I would like to thank all people involved (roughly in chronological order).

I would like to thank my supervisor Prof. Dr. Anton Glieder for giving me the opportunity to work on these projects, for his guidance, for the freedom to pursue own ideas and for the tremendous support to carry out all the projects in parallel.

Big thanks to Luki Sturmberger, who started with an internship, stayed for a project lab and a master thesis. Thanks for the great work from beginning on, your dedication really pushed the MUT promoter and activator projects!

Also huge thanks to Thomas Kickenweiz for his devotion and extra hard work during his master thesis on bidirectional promoters. I think you've set the record for most constructs/*Pichia* transformations per time. All the follow up projects on BDPs were only possible because of your outstanding initial work.

Big thanks to Clemens Farnleitner for all the work during your apprenticeship and special thanks to T.K. for also supervising Clemens. Also big thanks to Christian Schmid for managing the place and allowing us to focus on the science. Aside the everyday assistance, special thanks for characterizing the monodirectional terminators, doing the first series of bidirectionalized promoters, assembling the bidirectional terminator screening vector and optimizing expression of the HNLs. Thanks to Alexander Korsunsky for the work on BDPs, their applications and the P_{CAT1}-ARS during the internships. Also huge thanks to Amneris Schlager-Weidinger for making the daily lab work so easy and helping with everything.

Big thanks to the 'old' PhD- students Martina Geier, Andrea Mellitzer, Laura Näätsaari and Andi Braun and Post-Doc Andrea Camattari for showing me around and explaining the oePCR, RT-qPCR *etc*. protocols and bioreactor cultivations during the early days. Thanks to Martina and Rama Krishna Gudiminchi for continuous expertise on P450s.

Extra thanks to Martina for the collaboration on the monodirectional expression of the carotenoid pathway and your help on many other occasions. Also thanks to Anna-Maria Hatzl and Michaela Gerstmann for work on the monodirectional expression of the carotenoid pathway. Thanks to Julia Pitzer and Marlies Wagner for bidirectional expression of the carotenoid pathway.

Also big thanks to Katrin Weinhandl and Julia Pitzer for testing bidirectional histone promoters in CHO cells and baker's&fission yeast.

Huge thanks to super dedicated project lab students Armin Baumschlager, Pia Fauland, Eva-Maria Köhler and Richard Wasmayer. Your amazing work on BDPs, activators, BDPs & bidirectional

terminators (BDTs) and the MUT promoters really pushed the projects forward. Extra special thanks to Pia for modifying the conversion plasmids after hours and doing so extensive additional work.

Also big thanks to Astrid Weninger for the great project lab, the super independent work requiring minimal supervision and the collaboration during your master thesis. You managed to do the seemingly impossible successfully finishing an extra hard project. Also amazing work on the synthetic biology book chapter!

Huge thanks to Patrick Hyden for the hard work during the project lab on activators, BDPs and orthologous promoters and the following internship. Thanks a lot for performing all the parallel measurements, analyzing the data and doing the great figures.

Big thanks to Jasmin Elgin Fischer for the great project lab work co-supervised by Florian Krainer and the hard work during the following internship (taxadiene production, additional P_{CAT1}-ARS experiments).

Extra special thanks to my last three students Richard, Patrick and Jasmin for helping me to finish all the work of this thesis (and especially doing not-so-exciting work such as glucose measurements, parallel cultivations of hundreds of strains and preparing hundreds of vials for GC measurements). Without your extra dedicated participation, I would have never been able to finish all the work in time.

Brief summary on projects supervised, again extra big thanks to everybody involved:

Project labs

1. Lukas Sturmberger (MUT promoters)
2. Armin Baumschlager (synthetic hybrid BDPs)
3. Astrid Weninger (BDPs, CYPs)
4. Pia Carina Fauland (activators)
5. Eva-Maria Köhler (BDPs, BDTs)
6. Richard Wasmayer (MUT promoters, P_{CAT1}-ARS)
7. Patrick Hyden (activators, BDPs)

Co-supervised:

8. Marlies Wagner (BDPs)
9. Jasmin Fischer (BDPs, orthologous promoters.)

Master students

1. Lukas Sturmberger (P_{AOX1} circuits, microarrays)
2. Thomas Kickenweiz (BDP screening vectors, nBDPS, PPP, ROS promoters, P_{DAS} deletions)
3. Astrid Weninger (genome engineering)

Internships

1. Lukas Sturmberger (MUT promoters, orthologous promoters)
2. Alexander Korsunsky (BDPs, applications, P_{CAT1}-ARS)
3. Patrick Hyden (activators, BDPs)
4. Jasmin Elgin Fischer (P_{CAT1}-ARS follow up, taxadiene)

Thanks to Maria Freigassner for approaching me and allowing me to participate in the membrane protein work, thanks to Claudia Ruth for initiating the core promoter project. Extra big special thanks to Mudassar Ahmad for starting the plasmid project with me and being the major driving force behind it, never losing faith or motivation over the slow progress.

Big thanks to Gerhard Thallinger for his expertise on microarray experimental design and data analysis for the MUT pathway work and the activator circuits. Also big thanks to Karin Wagner at Medical University Graz for microarray hybridization and the great feedback on critical steps.

Thanks to Rui Portela, I really enjoyed participating in your core promoter projects on P_{AOX1} and the *de novo* core promoter design.

Last but not least also big thanks for scientific discussions to Flo W. Krainer, although we never really worked on a project together I got some of my best inputs form you!

However, possibly even more than the work related dedications, I also want to thank all of you on a personal level. I will miss the entertaining lunch conversations, the shared sweets and our beacon of unhealthy food: Kebap consumption on Dönerstag. Even if the scientific part was sometimes somewhat not working, there was always someone to talk to who would cheer you up. I especially remember meeting Mudassar at the institute on the weekends (and whenever I had to do something late at night or early in the morning, when do you sleep? ^^), thanks for the nice conversations and the company!
Big thanks to everyone for the amazing time on the second floor and the many nice evenings we spent. Special thanks to my office neighbors Andrea and Astrid and the whole room BK 02050. Also Amneris with her bench right behind me made the days more pleasant and even repetitive labwork enjoyable (and thanks for the nicknames Tommyboy and Tomassino x). Best thanks also to Sissy, filling out boring forms and getting pencils was always a highlight with you.
Extra special thanks to my homies Flo and Rui, it was a pleasure discussing science and life with you. I was really worried when Flo left for Manchester but luckily Rui had arrived ;) Especially the last half a year with the two of you was amazing (e.g. bamboo ^^). I really, really hope we stay in touch! Also special thanks to Flo and Ziska for the slack lining in summer and the evenings spent at your place (pizza from Gambaro Rosso and Casa Rustica!). Also Rui, you put that quite nice in essence “all the wing man fails (and the one really good win!!)” also much appreciated by me ^^

Also big thanks to my ‘private’ crew for forgetting work and chilling out together, especially Jakob&Chrissi, Michi&Maria, Berti&Laura (thanks a lot for all the nice dinners Laura), Alex&Serif [;p], ‘my good friend’ Thomas (especially the time between your girlfriends ;p) and Richie and Mathias.

Also huge thanks for the constant support from my family throughout my studies and my whole life!

So looking back at the last three and a half years and ahead into the great wide open: Time will pass, seasons will come and go, but friends and memories last! :D

1. Introduction and aims of this thesis

Single proteins such as lifesaving biopharmaceuticals and industrial biocatalysts, allowing more environmentally friendly production processes, are almost exclusively produced by recombinant gene expression. High titers of recombinant proteins have been produced in simple microbial expression hosts such as bacteria (e.g. *Escherichia coli, Bacillus sp.*), filamentous fungi (e.g. *Aspergillus sp., Trichoderma sp.*) or yeast (e.g. *Saccharomyces cerevisiae*) [1,2].

However, especially many plant derived pharmaceuticals are not protein based, but highly complex chemicals often derived from isoprenoid structures. Such terpenoids include the antimalarial artemisinin [3] or the anticancer drug taxol [4,5] (Fig. 1). Taxol for example is only found in the bark of the pacific yew tree and two to four trees had to be harvested to allow treatment of a single patient [6]. Chemical synthesis of this highly complex structure is possible, however requiring up to 51 steps resulting in maximal yields of less than one percent [5]. The natural taxol biosynthesis pathway in plants requires 19 enzymatic steps [5]. Reconstituting such a pathway in microorganisms requires not only the functional coexpression of all genes, but also optimization of the flux towards the desired product and the removal of kinetics bottlenecks [7,8]. Naturally, flux optimizations are achieved by varying transcription, translation levels and enzyme properties (*e.g.* feedback inhibition). In recombinant gene expression in eukaryotes, typically transcript levels are altered using different promoters.

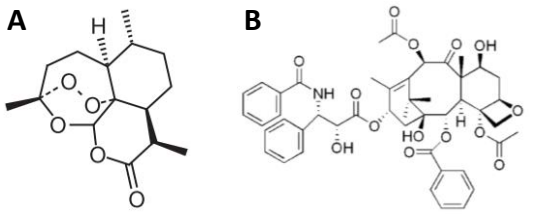


Fig. 1: Structure of A) artemisinin and B) Taxol. Structures taken from Paddon *et al.* [3] and Engels *et al*. [4].

Natural organisms and 'parts', such as promoters and terminators, are typically only in limited extents suitable for such endeavors. To this end, various synthetic biology approaches have been undertaken to engineer chassis organisms and to generate tailor made parts for metabolic engineering applications [9–13].

Concerning host platforms for metabolic engineering, most frequently *E. coli* and *S. cerevisiae* are used [14]. There is a lot of fundamental knowledge available on the 'classic' yeast and model organism *S. cerevisiae*. Therefore this yeast is especially for eukaryotic pathways the standard platform [3,15,16]. However, the number of tightly co-regulated promoters available in *S. cerevisiae* is limited: Typically only a few galactose (P_{GAL1}, P_{GAL3}, P_{GAL7}, P_{GAL10}) and copper (P_{CTR1}, P_{CTR3}, P_{CUP1}) regulated promoters are used [3,17]. While *S. cerevisiae* is a an excellent host for metabolite production relying on (micro-)anaerobic respiration, many pathways need oxidative growth for bioconversions involving cytochrome P450 enzymes for selective oxidations. In this respect the non-conventional yeast *Pichia pastoris* (*Komagataella phaffi*) is a highly favorable host for such processes, as demonstrated recently with cytochrome P450 expression [18]. A recently published approach of Wriesnegger *et al*. nicely showed a first efficient use the strengths of *P.* pastoris for sesquiterpenoid production [19].

For the heterologous expression of single genes the methylotrophic yeast *P. pastoris,* according to a recent literature survey, is even more frequently used than *S. cerevisiae* [20]. Advantages of *P. pastoris* include the suitability for high cell density cultivation and decent secretory capacities for heterologous proteins, while secreting little endogenous proteins [21]. *P. pastoris* has also been

assigned GRAS (generally regarded as safe) status by the U.S. Food and Drug Administration (FDA) and also the production of biopharmaceutical for use in humans has been approved (in 2009 Kalbitor by Dyax Corp., a Kallikrein inhibitor and in 2012 Jetrea by ThromboGenics NV, for the treatment of vitreomacular traction [2]).
The genome sequences of several *P. pastoris* strains [22–24] have been reported and provided the basis for the development of genome scale metabolic models (GSMMs) [25–27]. These GSMMs can be used for *in silico* predictions as a basis for engineering approaches, as recently demonstrated for improving production of cytosolic human superoxide dismutase (hSOD) [28]. However, in contrast to well-studied model organisms such as *E. coli* or *S. cerevisiae*, relatively little basic knowledge is available on *P. pastoris* [29] and widely available tools are still limited.

Nonetheless, several synthetic biology approaches have been undertaken in *P. pastoris* (Fig. 2) extending the scope and applicability of this organism (Chapter 2.1, [2]). Arguably the most prominent synthetic biology endeavor in *P. pastoris* was glycoengineering, generating strains with humanized glycosylation patterns. In a series of high profile publications the endogenous yeast type hyper-mannosylated glycosylation pattern was engineered into humanized forms [30–33].Until today, *P. pastoris* remains the only microbial expression host allowing the production of tailor made, fully humanized sialylated glycoproteins. Also novel molecular tools such as new strains facilitating the generation of knockouts and improved systems for marker recycling were reported. Moreover libraries of synthetic promoters were generated for *P. pastoris*, enabling transcriptional fine-tuning of heterologous gene expression.

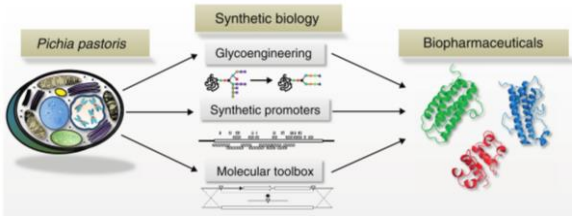


Fig. 2: State of the art synthetic biology approaches in *P. pastoris*. Figure taken from Vogl *et al*. [2]

The availability of strong, tightly regulated promoters has been a key advantage of *P. pastoris* and other methylotrophic yeasts such as *Hansenula polymorpha*, *Candida boidinii* and *Pichia methanolica* [34]. Genes involved in the metabolism of methanol (methanol utilization, MUT) are tightly repressed on repressing carbon sources and strongly upregulated upon addition of methanol (and depletion of the repressing carbon source). Natural peroxisomal MUT proteins such as alcohol oxidases and dihydroxyacetone synthases account on methanol for up to one third of total intracellular protein. If the promoters of these genes are cloned upstream of a heterologous gene, similarly high expression levels have been achieved. In *P. pastoris* up to 22 g/l of intracellular [35] and 15 g/l secreted protein [36] have been achieved using the wildtype promoter of the *alcohol oxidase 1* gene (P_{AOX1}). Using synthetic P_{AOX1} variants and codon optimized genes, even more than 18 g/l of secreted protein were achieved [37]. However, little is known on the molecular regulation of MUT genes (Chapter 2.2 and 2.3, [38]) and relatively few MUT promoters have so far been characterized [38]. While efficient tools for single gene expression have been widely available as simple kit systems for *P. pastoris*, the toolboxes for multi gene coexpression for pathway design still lagged behind *S. cerevisiae*.

The major aim of this thesis was to overcome these present limitations caused by shortage of parts in *P. pastoris.* Following synthetic biology approaches, a molecular toolbox should be created offering solutions for single gene expression and metabolic engineering applications. Although a library of elaborate synthetic promoters has been created for *P. pastoris* [39], these promoters are derived from a single parental sequence (*AOX1* promoter) by moderate changes. For metabolic engineering

applications, sequence diversified promoters are desirable, as identical or highly similar sequences might cause homologous recombination and thereby loss of entire expression cassettes or parts thereof by loop out recombination [40,41]. Synthetic P_{AOX1} variants had also shown different regulatory profiles, as some variants were derepressed *(i.e.* activation once the glucose in the medium was depleted, not requiring methanol). In related methylotrophic yeasts *H. polymorpha*, *C. boidinii* and *P. methanolica* even strong naturally derepressed promoters have been reported [34].
To this end, one of the main goals was to characterize a set of strong, methanol regulated promoters for *P. pastoris*, preferably also providing novel regulatory profiles such as derepression.

One of the strongest methanol regulated *P. pastoris* promoters previously reported [38] was a bidirectional promoter (BDP) driving the expression of the two *dihydroxyacetone synthase* genes *DAS1* and *DAS2* in opposite directions [42]. In higher eukaryotes a few synthetic BDPs have been reported [43,44]. We hypothesized that a library of BDPs with different expression strengths and regulatory profiles on both sides may constitute a valuable tool for optimizing dual and multi gene coexpression required for the production of multi subunit proteins or entire metabolic pathways. Therefore another aim of the thesis was to evaluate the applicability of such bidirectional promoters (BDPs) for gene coexpression and to develop strategies for generating BDPs in *P. pastoris*.

Aside of these projects focusing on *cis*-acting regulatory sequences, we also aimed to investigate the potential of creating simple transcriptional regulatory circuits by tinkering with *trans*-acting regulators. Acquiring control over the natural regulation of promoters and the design of synthetic artificial transcription factors have proven imperative for optimizing protein production, metabolic engineering and synthetic biology applications [10,45,46]. The goal was to alter the regulation of the *AOX1* promoter by transcription factor overexpression, ideally enabling methanol free activation by a different stimulus.

References

[1] Walsh G. Biopharmaceutical benchmarks 2010. Nat Biotechnol 2010;28:917–24.
[2] Vogl T, Hartner FS, Glieder A. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 2013;24:1094–101.
[3] Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 2013;496:528–32.
[4] Engels B, Dahm P, Jennewein S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 2008;10:201–6.
[5] Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 2010;330:70–4.
[6] Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 1971;93:2325–7.
[7] Stephanopoulos G. Synthetic biology and metabolic engineering. ACS Synth Biol 2012;1:514–25.
[8] Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 2012;40:e142.
[9] Mutalik VK, Guimaraes JC, Cambray G, Mai Q-A, Christoffersen MJ, Martin L, et al. Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods 2013;10:347–53.
[10] Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK, et al. A synthetic biology framework for programming eukaryotic transcription functions. Cell 2012;150:647–58.
[11] Chen Y-J, Liu P, Nielsen A a K, Brophy J a N, Clancy K, Peterson T, et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 2013;10:659–64.
[12] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009;27:946–50.
[13] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire M a, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010;329:52–6.
[14] Peralta-Yahya PP, Keasling JD. Advanced biofuel production in microbes. Biotechnol J 2010;5:147–62.
[15] Nielsen J, Larsson C, van Maris A, Pronk J. Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol 2013;24:398–404.

[16] Hong K-K, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cell Mol Life Sci 2012;69:2671–90.
[17] Da Silva NA, Srikrishnan S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res 2012;12:197–214.
[18] Geier M, Braun A, Emmerstorfer A, Pichler H, Glieder A. Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes - a comparative study. Biotechnol J 2012:1–13.
[19] Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, et al. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of Pichia pastoris. Metab Eng 2014.
[20] Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Front Microbiol 2014;5:85.
[21] Maccani A, Landes N, Stadlmayr G, Maresch D, Leitner C, Maurer M, et al. Pichia pastoris secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins. Biotechnol J 2014;9:526–37.
[22] Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb Cell Fact 2009;8:29.
[23] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6.
[24] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.
[25] Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, et al. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol J 2010;5:705–15.
[26] Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, et al. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 2010;9:50.
[27] Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol 2012;6:24.
[28] Nocon J, Steiger MG, Pfeffer M, Bum Sohn S, Yong Kim T, Maurer M, et al. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 2014;24:129–38.
[29] Dikicioglu D, Wood V, Rutherford KM, McDowall MD, Oliver SG. Improving functional annotation for industrial microbes: a case study with Pichia pastoris. Trends Biotechnol 2014:1–4.
[30] Choi B-K, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, et al. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris. Proc Natl Acad Sci U S A 2003;100:5022–7.
[31] Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, et al. Production of complex human glycoproteins in yeast. Science (80-) 2003;301:1244–6.
[32] Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, et al. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science (80-) 2006;313:1441–3.
[33] Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, et al. Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 2006;24:210–5.
[34] Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact 2006;5:39–59.
[35] Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Lückl J, et al. High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. Protein Expr Purif 1997;11:61–71.
[36] Werten MW, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA. High-yield secretion of recombinant gelatins by Pichia pastoris. Yeast 1999;15:1087–96.
[37] Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, et al. Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks. J Biotechnol 2014.
[38] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.
[39] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.
[40] Aw R, Polizzi KM. Can too many copies spoil the broth? Microb Cell Fact 2013;12:128.
[41] Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in P. pastoris. Chem Commun (Camb) 2014;62.
[42] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 2012;11:22.
[43] Xie M, He Y, Gan S. Bidirectionalization of polar promoters in plants. Nat Biotechnol 2001;19:677–9.
[44] Amendola M, Venneri MA, Biffi A, Vigna E, Naldini L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. Nat Biotechnol 2005;23:108–16.
[45] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature 2000;403:339–42.
[46] Blazeck J, Alper HS. Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol J 2013;8:46–58.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Biotechnology

New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*

Thomas Vogl[1], Franz S Hartner[2] and Anton Glieder[3]

Biopharmaceuticals are an integral part of modern medicine and pharmacy. Both, the development and the biotechnological production of biopharmaceuticals are highly cost-intensive and require suitable expression systems. In this review we discuss established and emerging tools for reengineering the methylotrophic yeast *Pichia pastoris* for biopharmaceutical production. Recent advancements of this industrial expression system through synthetic biology include synthetic promoters to avoid methanol induction and to fine-tune protein production. New platform strains and molecular cloning tools as well as *in vivo* glycoengineering to produce humanized glycoforms have made *P. pastoris* an important host for biopharmaceutical production.

Addresses
[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz A-8010, Austria
[2] Sandoz GmbH, Biochemiestrasse 10, A-6250 Kundl, Austria
[3] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, Graz A-8010, Austria

Corresponding author: Glieder, Anton (anton.glieder@acib.at, a.glieder@tugraz.at)

Current Opinion in Biotechnology 2013, **24**:1094–1101

This review comes from a themed issue on **Pharmaceutical biotechnology**

Edited by **Ajikumar Parayil** and **Federico Gago**

For a complete overview see the Issue and the Editorial

Available online 20th March 2013

0958-1669/$ – see front matter,

http://dx.doi.org/10.1016/j.copbio.2013.02.024

Introduction

Biopharmaceuticals are indispensable in modern medicine. The estimated market value is $70 to 80 billion (depending on the definition) and annual growth rates between 7 and 15% are expected [1–3]. This is another major reason for the worldwide focus of pharmacy and biotechnology on biopharmaceutical development and production. By definition, the term 'biopharmaceutical' refers to recombinant therapeutic proteins and nucleic acid based products and in the broader sense also to engineered cell or tissue-based products [2]. Vaccines, interferons and hormones like insulin, human growth hormone (hGH) and erythropoietin (EPO) are examples for protein biopharmaceuticals. Antibodies (including fragments like Fabs, scFvs and nanobodies) represent the biggest group of protein biopharmaceuticals [1–3].

Therapeutic proteins are typically produced in mammalian cell lines and *Escherichia coli*. While bacterial systems exhibit fast and robust growth in bioreactors using simple media, mammalian cells resemble their human counterparts more closely in terms of typical eukaryotic post translation modifications (PTMs) like glycosylation [2,4–6]. However, mammalian cell culture processes are relatively slow, require complex media, and are susceptible viral contaminations (Table 1).

Using yeasts enables to combine robust growth on simple media (in large scale bioreactors) with easily achievable genetic modifications and the introduction of the desired PTMs [7].

The 'classic' yeast *Saccharomyces cerevisiae* is one of the best studied eukaryotes and has been used as expression host for biopharmaceuticals since the early days of genetic engineering and recombinant protein production [8••]. Recently, the first biopharmaceutical produced in the methylotrophic yeast *Pichia pastoris* has been approved by the FDA (Kalbitor by Dyax Corp., a Kallikrein inhibitor) [1]. *P. pastoris* features all favorable traits of yeasts mentioned and has successfully been used to produce high titers of numerous heterologous proteins [7,9,10••]. Additionally, *P. pastoris* is suitable for high cell density cultivations, reaching more than 150 g dry cell weight per liter [11] and has high secretory capabilities for heterologous proteins, while secreting only low amounts of endogenous proteins (Table 1) [12].

In this review we focus on new opportunities for biopharmaceutical production by reengineered *P. pastoris* employing new tools, (semi-) synthetic parts and PTM pathways (see Figure 1). We also summarized already published approaches to identify regulatory elements and to reengineer promoters for bottom-up regulatory circuit design.

Recent developments in synthetic biology have extended the toolset of classical genetic engineering [13]. Tailormade expression systems have been created by modifying transcription, translation, PTMs and designing synthetic regulatory networks [14,15••].

Glycoengineering

The majority of therapeutic proteins contain post-translational modifications, with glycosylation being the most common and at the same time the most complex PTM [2].

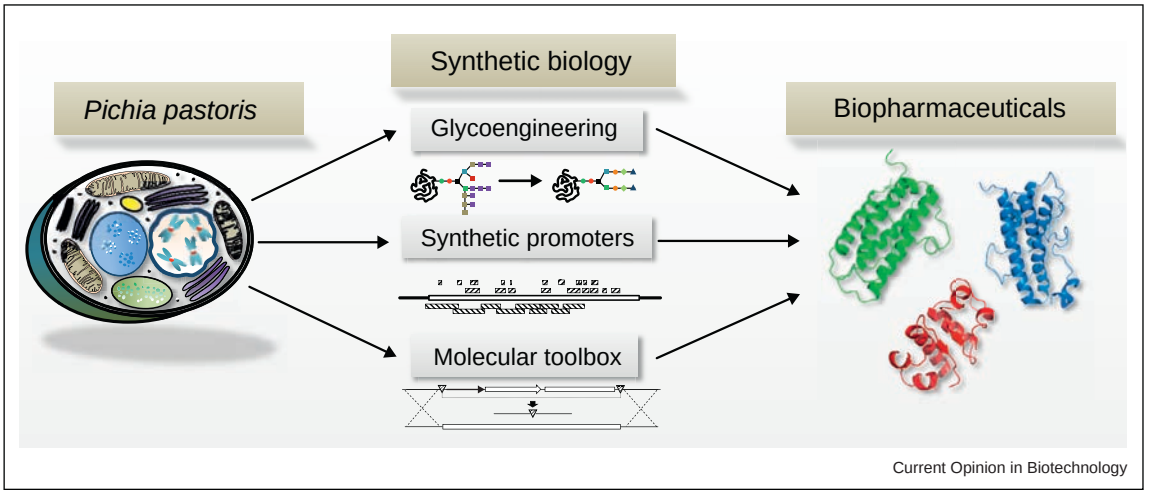
Table 1

Comparison of expression systems used for biopharmaceutical production [4,6,7]

	Higher eukaryotes		Yeast		*Escherichia coli*
Ease of genetic modifications	Moderate		Simple		Simple
Cultivation	Slow growth rates, expensive complex (or synthetic) media required		Fast and robust growth, defined minimal media		Fastest growth, defined minimal media
Contaminations	Risk of viral contaminations, viral clearance required		Little risks of endotoxins or viral DNAs		Endotoxins presence requires thorough purification, possible phage infections
Post translational modifications (PTMs)	Closely resembling human PTMs; usually mixtures of several glycoform variants		Most human PTMs achievable, but natural glycosylation patterns differ from humans, hypermannosylation, engineered strains can achieve human glycoforms and high uniformity		Limited set of PTMs, some human PTMs (e.g. glycosylation) difficult to achieve
Protein yields and secretory capacities	High yields, highly efficient secretion, high specific productivity		High yields, secretory capacities depending on the species		High expression capacities, secretion mostly inefficient, extensive purification and downstream processing required
Most commonly used species	Mammalian cells	Insect cells	*Pichia pastoris*	*Saccharomyces cerevisiae*	
Recently approved biopharmaceuticals[a]	32	2	2[b]	4	17
Additional information and specific differences between host species of the same class	Commonly used cell lines: CHO (Chinese Hamster Ovary), BHK (baby hamster kidney), murine-myeloma-derived NS0, SP2/0 cell lines [2] and HEK293	Baculo virus based systems most commonly used for transfection Easy scale up Contaminations less problematic Mammalianized glycosylation [5]	Efficient and selective secretion, often higher protein titers than *S. cerevisiae*, for example, [8••] Crabtree negative, high cell density cultivations GRAS status Hypermannosylation is less pronounced in *P. pastoris* and critical terminal α-1,3-mannose linkages were not observed [19], engineered strains providing fully humanized glycosylation not available for *S. cerevisiae*	Important eukaryotic model organism, high molecular- and cell biological knowledge Crabtree positive, leading to ethanol production	Fastest efficient expression system Inexpensive Well established processes suitable for mass production Folding problems may lead to the formation of inclusion bodies and require expensive refolding (yet, inclusion bodies provide a valuable strategy to achieve high protein yields and simple purification) Inefficient acetate metabolism may hamper high cell density cultivation of some strains

[a] Data from Walsh [1], time period: January 2006–June 2010, in total 58 biopharmaceuticals have been approved, two biopharmaceuticals produced in transgenic animals were not listed.
[b] In this number Jetrea by ThromboGenics is included (approved in 2012 and not listed by Walsh [1]).

Figure 1



Current synthetic biology approaches to improve biopharmaceutical yields and quality in *P. pastoris*. Glycoengineered strains provide humanized *N*-glycosylation patterns [14,15,16•], synthetic promoters allow the fine-tuning of expression levels [41,42,43•] and various tools for strain engineering [47–49,50•] and metabolic modeling [55•,56•,57•] are available.

Yeasts can perform typical eukaryotic PTMs, but final glycosylation patterns of yeasts and humans differ significantly. Hypermannosylation and terminal α-1,3-mannose linkages associated with glycoproteins from *S. cerevisiae*, can result in poor serum half-life or even immunogenic effects of therapeutic proteins [2,16•]. Thus, there have been efforts to humanize yeast glycosylation which has been accomplished in *P. pastoris* (see [16•–18•] for reviews). Also hypermannosylation is less pronounced in *P. pastoris* and terminal α-1,3-mannose linkages are not observed [19].

Here, we focus on recent developments of glycoengineering in *P. pastoris* and highlight the synthetic biology approaches and the heterologous and chimeric enzymes used for this purpose.

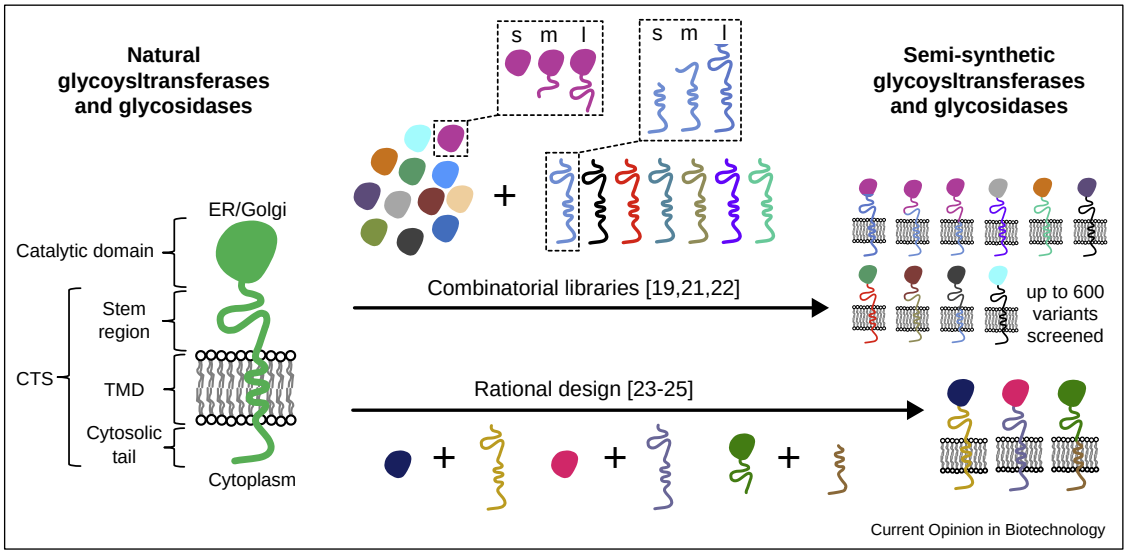
Achieving humanized glycosylation in yeast required on the one hand the elimination of hyperglycosylation by deleting the appropriate yeast genes, but on the other hand also the introduction of additional glycosidases and glycosyltransferases, including missing biosynthetic pathways and transporters for sugars not present in yeast, for example, sialic acid. In the case of galactose, UDP-glucose was converted to UDP-galactose in the Golgi by providing the respective epimerase activity [16•,17•].

In addition to simple expression of these genes, correct spatial positioning along the secretory pathway in the ER and Golgi is essential, as the sequential activity of one enzyme produces the substrate for the next. To achieve the suitable positioning of the required factors along this cellular assembly line in *P. pastoris*, synthetic glycobiology [20] approaches were used.

Tailor-made glycosyltransferases and glycosidases with the desired catalytic properties and localization characteristics were created [19,21–25]. The strategy was based on the knowledge, that eukaryotic glycosyltransferases and glycosidases are type II membrane proteins, consisting of an N-terminal cytoplasmic tail, a membrane anchor domain, a stem region and a C-terminal catalytic domain (see Figure 2) [20]. The C-terminal catalytic domain is active independently of the localization conferring N-terminal part, which is also termed 'CTS' (cytoplasmic, transmembrane, stem). Fusions of catalytic domains to CTS fragments allowed the creation of semi-synthetic glycosyltransferases and glycosidases. A combinatorial library approach paired with a high-throughput screening was used to create and evaluate these proteins [19,21,22]. Rational design led to similar results [23–25], but eventually input from combinatorial libraries was also used [25].

Notably, the initial publications of the combinatorial libraries [19,21] contained barely any information on their composition and how the chimeric glycosyltransferases were designed. More recently, a comprehensive report about the catalytic domains, the CTS fragments, and how they were fused was published [26••]. The authors had not only started from a large set of 33 catalytic domains from different eukaryotes (e.g. fungi, worm, fruit fly, mouse, rat, human) and 66 fungal leader sequences, but also tested fusions of various lengths of both the catalytic domain and the CTS. Up to 600 variants were screened for optimal desired activity and localization along the generated artificial glycosylation pathways in *P. pastoris* (see Figure 2).

Figure 2



Design strategies to create semi-synthetic glycosyltransferases and glycosidases for glycoengineering. On the left side, the general domain structure of glycosyltransferases and glycosidases is shown. These type II membrane proteins consist of an N-terminal cytosolic tail, a transmembrane domain (TMD), a stem region (these elements are referred to as CTS), and a C-terminal catalytic domain. In the middle and on the right side, design strategies for creating tailor-made enzymes with the desired catalytic activity and the proper localization in the sec pathway are shown. The combinatorial library approach involved the combination of large sets of catalytic domains with CTS fragments to fusion proteins, which were then screened for the desired activity [19,21,22]. Different lengths of the catalytic domains and the CTS fragments were tested (referred to as 's' for short, 'm' for medium, 'l' for long and shown exemplarily for one catalytic domain and one CTS). Rational approaches were also used to design these chimeric enzymes [23–25]. The schematic for the domain architecture and the combinatorial libraries is based on Czlapinski *et al*. [20] and Nett *et al*. [26••].

An essential milestone was achieved in 2006 by introduction of nine synthetic genes and deletion of six endogenous genes enabling the production of complex terminally sialylated glycoproteins in *P. pastoris* [22]. In the last five years, *N*-glycosylation site occupancy has been increased from 75–85% to 99% [27] and undesired β-linked mannose residues have been removed by creating a *P. pastoris* quadruple knock-out devoid of all four endogenous β-mannosyl transferases [28]. Furthermore, the production processes using glycoengineered *P. pastoris* strains have been optimized [29–31], antibody production in glycoengineered strains reached the g/l scale [32,33] and glycoengineered strains have also been established for surface display applications [34,35].

In addition to human like microbial glycosylation such heterologous synthetic pathways allow direct control of the intricate glycosylation process. Thereby, tailor-made glycoforms of a protein can be produced which can exhibit moderately differing pharmacodynamics. For example an antibody expressed in glycoengineered *P. pastoris* with a uniform, single glycoform showed improved antibody-mediated effector functions, compared to mammalian cell culture derived glycoforms with variable glycosylation patterns [36]. Therefore, better than nature glycoengineered *P. pastoris* strains pave the way for the creation of synthetic, supernatural glycoform preparations with altered properties compared to naturally occurring variants.

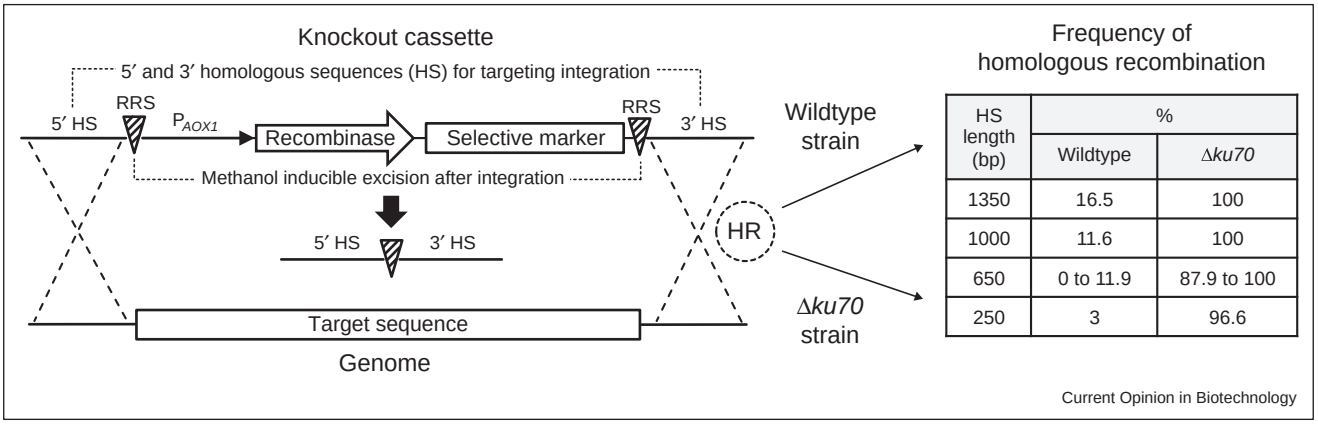
Synthetic promoters

Efficient transcription is a critical step in gene expression. Therefore strong and controllable promoters are an essential tool for high titers in recombinant protein production [7,37]. In addition to natural promoters there has been a growing interest in synthetic promoters driving enhanced expression, improving folding or showing tailor-made regulatory profiles [37–39]. In *P. pastoris*, up to 22 g/l intracellular protein and 15 g/l secreted protein have been obtained with the most frequently applied, tightly controlled, strong and methanol inducible *AOX1* promoter (P_{AOX1}) [40].

As result, this promoter was the starting point for creating synthetic variants with increased promoter strength and altered, methanol free regulation, as the use of toxic and flammable methanol can cause a considerable safety risk in industrial processes.

One semi-rational approach to create synthetic P_{AOX1} variants relied on an *in silico* analysis for putative conserved eukaryotic transcription factor binding sites (TFBS) in P_{AOX1}. Subsequently, the respective short

Figure 3



Recombinase based self-excisable knockout cassettes for marker regeneration (left side). Increased rates of homologous recombination in a *P. pastoris* Δ*ku70* strain (right side). The knockout cassettes consist of a recombinase (Cre or FLP [48,49,50•]) and a marker gene flanked by the respective recombinase recognition sites and are directed to the genome via the 5′ and 3′ homologous sequences to delete the desired target sequence. After integration via a double cross-over event, self-excision of the recombinase and the marker gene can be initiated by the expression of the recombinase from the methanol inducible *AOX1* promoter (P_{AOX1}), leaving only the recombinase recognition site in the genome (notably Marx *et al.* [49] provided the recombinase transiently on a CEN/ARS plasmid). The initial integration in the genome is dependent on homologous recombination (HR). Exemplary frequencies of homologous recombination (in %) of the wildtype compared to the Δ*ku70* strain are shown (right side). The length of the homologous sequence indicates the number of base pairs (bp) added on both sides of the cassette [50•]. For 650 bp two different integration loci were tested, therefore two % values are given.

sequence stretches were deleted [41]. These deletion variants showed both increased and decreased reporter gene expression levels spanning 6–160% of wildtype P_{AOX1} driven expression. Alternative approaches relied on the systematic deletion of larger adjacent fragments of almost the entire promoter [42]. Surprisingly, some small deletions and point mutations resulted in altered regulation as these variants were moderately active when glucose was depleted, without requiring the inductor methanol [41]. This derepression effect was further optimized by combinations of deletions and insertions of important sequence stretches. Such altered induction properties now enable the consecutive induction of coexpressed proteins such as chaperons and the therapeutic protein of interest. Putative TFBS of P_{AOX1} were also fused to natural core promoter fragments to create short semi-synthetic variants, which again showed altered regulation and surpassed the full-length wildtype promoter in certain applications especially when multiple copies of the expression cassettes were integrated [41,43•]. Also the constitutive promoter of the glyceraldehyde-3-phosphate dehydrogenase gene (P_{GAP}) of *P. pastoris* has been engineered by a random mutagenesis approach [44] showing the potential of additional promoters for expression fine tuning or the generation of new regulatory circuits. Bio-process strategies for biotechnologically relevant enzymes have been improved by employing these synthetic promoters [41,43•,44,45] and similar effects can be expected for biopharmaceuticals. Furthermore, multiple positive and negative factors involved in P_{AOX1} regulation have been identified since this initial semi-rational promoter engineering (see [40] for a recent review), opening the way for the design of novel synthetic regulatory circuits for gene expression and pathway design.

Molecular toolbox for synthetic biology in *P. pastoris*

Synthetic biology applications require efficient tools for strain engineering. For example, the creation of *P. pastoris* strains providing a fully humanized glycosylation pattern necessitated in the first place the development of suitable genetic strategies to knock out and introduce multiple genes [46]. Efficient strategies for gene replacements and marker recycling have now become available for *P. pastoris*. Namely, systems based on new counter selective markers [47], a Cre/loxP strategy [48,49] and an advanced flipper cassette application [50•] have been reported and applied. The recombinase based strategies [48,49,50•] allow active excision of the marker gene used in a deletion cassette and to thereby recycle markers and perform sequential rounds of deletions. This is achieved by designing a deletion cassette, in which the marker gene and the recombinase are flanked by two recombinase recognition sites and the recombinase is placed under the tight control of the methanol inducible *AOX1* promoter (see Figure 3). Näätsaari *et al.* [50•] applied such a strategy to generate a new platform of *P. pastoris* expression strains and Marx *et al.* [49] boosted riboflavin production in *P. pastoris* by subsequently overexpressing all six genes of the riboflavin biosynthetic pathway by inserting the strong constitutive *GAP* promoter upstream of

these genes. Coupling such approaches with synthetic promoter variants [41,42,43•,44] might support the transcriptional fine tuning of individual enzyme activities of biosynthetic pathways.

Site specific integration and knock-out strain generation rely on endogenous homologous recombination. While in *S. cerevisiae* HR is working highly efficiently, non-homologous end joining is the preferred pathway in most other filamentous fungi and yeasts, including *P. pastoris*. HR occurs at less than 1% and up to 30% of all integration events, depending on the length of the homologous targeting sequence [50•]. For example during glycoengineering of *P. pastoris* only 5 out of 460 clones showed the desired gene replacement [46]. Targeted integration and deletion should become more efficient in the future by employing a *P. pastoris ku70* deletion strain with increased rates of HR [50•]. By the deletion of a Ku70 homologue, a protein involved in NHEJ, HR rates of up to 100% were achieved (see Figure 3). The Δ*ku70* strain did not show genetic instability, but the growth rates were 10–30% lower than those of the wildtype (depending on the carbon source) and the strain showed a decreased survival rate under UV light. This hints an increased susceptibility to DNA damage and complementing the wildtype *KU70* gene after completion of strain engineering was recommended [50•].

In addition to precise deletions, site specific integration, and marker recycling, new cloning techniques facilitate the construction of the respective gene expression and deletion constructs. Efficient *in vitro* recombination methods such as Gibson assembly [51] enable flexible restriction free cloning and library generation allowing the simple testing of libraries of promoters, artificial or natural expression enhancers and signal-sequences or other targeting sequences. Although, bottom up approaches to design individual parts for *P. pastoris* strain reengineering and expression cassette constructions are ongoing, there is no systematic synthetic biology parts collection for this yeast so far.

Bioinformatics tools complete the toolbox for synthetic biology applied in *P. pastoris*. High-quality genome sequences [52–54] and metabolic models [55•,56•,57•] of *P. pastoris* have recently become available. This comprehensive new background knowledge enables research towards systems wide understanding of the *P. pastoris* expression system and provides the basis for reengineering this host using synthetic parts and pathways to improve biopharmaceutical production. For example, recent studies in *P. pastoris* have hinted an interconnection of both the carbon metabolism [58] and the cellular redox state [59] with protein production and secretion. Thus, similar to *S. cerevisiae* [60,61], a systems biology view on secretion coupled with a synergistic use of metabolic engineering and synthetic biology approaches [62,63] promise coming improvements for biopharmaceutical production by *P. pastoris*.

Conclusions

Over the last two decades, *P. pastoris* has been established as one of the most frequently used expression systems in both industry and academia. Beside a large number of various enzymes, many human proteins and biopharmaceuticals were also efficiently produced by *P. pastoris*. The adaptation of the yeast high-mannose type glycosylation to the complex humanized glycosylation was a major achievement and resulted in uniform glycoforms from microbial production. Synthetic promoter variants with altered regulatory profiles and expression levels surpassed their natural counterparts for enzyme production. Equally, these variants can be used to optimize and fine-tune the expression of therapeutic proteins. Also other new key methodologies for synthetic biology such as efficient gene deletion and assembly strategies, metabolic models and strains with altered recombination properties have become available. Together with milestones such as the FDA approval, these new tools and techniques have a high potential to boost the production of biopharmaceuticals and for efficient metabolic engineering (see Box 1).

Altered and new biosynthetic pathways for posttranslational modifications such as precise glycosylation are enabling techniques giving access to new therapeutics with uniform and excellent quality.

Synthetic biology will certainly not only further improve industrial enzyme production, but also stimulate and facilitate innovative approaches for biopharmaceutical production in *P. pastoris*.

Box 1 Milestones and recent accomplishments for biopharmaceutical production in *P. pastoris*

(1) FDA GRAS (generally regarded as safe) status in 2006 (Phospholipase C by Diversa Corp., for degumming vegetables oils for food use).

(2) FDA approved biopharmaceutical production processes in 2009 (Kalbitor by Dyax Corp., a Kallikrein inhibitor) and 2012 (Jetrea by ThromboGenics NV, for the treatment of vitreomacular traction).

(3) Glycoengineered strains providing humanized, uniform *N*-glycosylation patterns [22,25].

(4) Synthetic promoters for fine-tuning expression levels [41,42,43•].

(5) Efficient strategies for knockouts of multiple genes and overexpression of entire pathways [48,49,50•].

(6) High quality genome sequences [52–54].

(7) Establishment of *in silico* metabolic models for strain engineering [55•,56•,57•].

Acknowledgements

T.V. is financed by theEuropean Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 289646 (Kyrobio). We also gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) and the Austrian Centre of Industrial Biotechnology (ACIB) contribution was supported by FFG, bmvit, mvwfi, ZIT, Zukunftsstiftung Tirol and Land Steiermark within the Austrian COMET programme (FFG grant 824186). We would also like to thank Andrea Camattari and Claudia Ruth for valuable discussions.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Walsh G: **Biopharmaceutical benchmarks 2010**. *Nat Biotechnol* 2010, **28**:917-924.

2. Walsh G: **Post-translational modifications of protein biopharmaceuticals**. *Drug Discov Today* 2010, **15**:773-780.

3. Goodman M: **Market watch: sales of biologics to show robust growth through to 2013**. *Nat Rev Drug Discov* 2009, **8**:837.

4. Demain AL, Vaishnav P: **Production of recombinant proteins by microbes and higher organisms**. *Biotechnol Adv* 2009, **27**:297-306.

5. Palmberger D, Klausberger M, Berger I, Grabherr R: **MultiBac turns sweet**. *Bioengineered* 2012, **4**:1-6.

6. Berlec A, Strukelj B: **Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells**. *J Ind Microbiol Biotechnol* 2013 http://dx.doi.org/10.1007/s10295-013-1235-0.

7. Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D: **Recombinant protein production in yeasts**. *Methods Mol Biol* 2012, **824**:329-358.

8. •• Martínez JL, Liu L, Petranovic D, Nielsen J: **Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation**. *Curr Opin Biotechnol* 2012, **23**:965-971.
Concise overview of biopharmaceutical production in *S. cerevisiae*. Covers approaches for strain engineering and also shortly biopharmaceutical production in *P. pastoris*.

9. Daly R, Hearn MTW: **Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production**. *J Mol Recognit* 2005, **18**:119-138.

10. •• Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D: ***Pichia pastoris*: protein production host and model organism for biomedical research**. *Future Microbiol* 2013, **8**:191-208.
Detailed overview on protein production in *P. pastoris* covering key aspects for high level expression, strain engineering, secretion and novel methods. Protocols are also provided and the use of *P. pastoris* as a model for human cell biology is summarized.

11. Jahic M, Veide A, Charoenrat T, Teeri T, Enfors S-O: **Process technology for production and recovery of heterologous proteins with *Pichia pastoris***. *Biotechnol Prog* 2006, **22**:1465-1473.

12. Damasceno LM, Huang C-J, Batt CA: **Protein secretion in *Pichia pastoris* and advances in protein production**. *Appl Microbiol Biotechnol* 2012, **93**:31-39.

13. Liang J, Luo Y, Zhao H: **Synthetic biology: putting synthesis into biology**. *Wiley Interdiscip Rev Syst Biol Med* 2012, **3**:7-20.

14. Lynch SA, Gill RT: **Synthetic biology: new strategies for directing design**. *Metab Eng* 2012, **14**:205-211.

15. •• Krivoruchko A, Siewers V, Nielsen J: **Opportunities for yeast metabolic engineering: lessons from synthetic biology**. *Biotechnol J* 2011, **6**:262-276.
Excellent, comprehensive review highlighting the use of synthetic biology to engineer *S. cerevisiae*. Describes approaches ranging from synthetic promoters, pathway engineering, riboswitches, protein engineering to synthetic regulatory networks.

16. • De Pourcq K, De Schutter K, Callewaert N: **Engineering of glycosylation in yeast and other fungi: current state and perspectives**. *Appl Microbiol Biotechnol* 2010, **87**:1617-1631.
See annotation to Ref. [18•].

17. • Hamilton SR, Gerngross TU: **Glycosylation engineering in yeast: the advent of fully humanized yeast**. *Curr Opin Biotechnol* 2007, **18**:387-392.
See annotation to Ref. [18•].

18. • Bollok M, Resina D, Valero F, Ferrer P: **Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production**. *Recent Pat Biotechnol* 2009, **3**:192-201.
Refs. [16•,17•,18•] provide an in-depth coverage of the glycoengineering efforts in *P. pastoris* before the period under review. The most recent reviews from the two main groups involved (of Gerngross [17•] and Callewaert [16•]) and a review covering patent issues [18•] are listed.

19. Choi B-K, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU: **Use of combinatorial genetic libraries to humanize *N*-linked glycosylation in the yeast *Pichia pastoris***. *Proc Natl Acad Sci U S A* 2003, **100**:5022-5027.

20. Czlapinski JL, Bertozzi CR: **Synthetic glycobiology: exploits in the Golgi compartment**. *Curr Opin Chem Biol* 2006, **10**:645-651.

21. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H *et al.*: **Production of complex human glycoproteins in yeast**. *Science* 2003, **301**:1244-1246.

22. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K *et al.*: **Humanization of yeast to produce complex terminally sialylated glycoproteins**. *Science* 2006, **313**:1441-1443.

23. Callewaert N, Laroy W, Cadirgi H, Geysens S, Saelens X, Min Jou W, Contreras R: **Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1,2-alpha-D-mannosidase for *N*-glycan engineering in *Pichia pastoris***. *FEBS Lett* 2001, **503**:173-178.

24. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R: **In vivo synthesis of mammalian-like, hybrid-type *N*-glycans in *Pichia pastoris***. *Appl Environ Microbiol* 2004, **70**:2639-2646.

25. Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N: **Engineering complex-type *N*-glycosylation in *Pichia pastoris* using GlycoSwitch technology**. *Nat Protoc* 2009, **4**:58-70.

26. •• Nett JH, Stadheim TA, Li H, Bobrowicz P, Hamilton SR, Davidson RC, Choi B, Mitchell T, Bobrowicz B, Rittenhour A *et al.*: **A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of *Pichia pastoris***. *Yeast* 2011, **28**:237-252.
Revelation of the exact composition and design of the combinatorial genetic libraries (initially described in [19,21]) for the generation of the synthetic glycosyltransferases and glycosidases used to humanize *P. pastoris* *N*-glycosylation.

27. Choi B-K, Warburton S, Lin H, Patel R, Boldogh I, Meehl M, D'Anjou M, Pon L, Stadheim TA, Sethuraman N: **Improvement of *N*-glycan site occupancy of therapeutic glycoproteins produced in *Pichia pastoris***. *Appl Microbiol Biotechnol* 2012, **95**:671-682.

28. Hopkins D, Gomathinayagam S, Rittenhour AM, Du M, Hoyt E, Karaveg K, Mitchell T, Nett JH, Sharkey NJ, Stadheim TA *et al.*: **Elimination of β-mannose glycan structures in *Pichia pastoris***. *Glycobiology* 2011, **21**:1616-1626.

29. Jacobs PP, Inan M, Festjens N, Haustraete J, Van Hecke A, Contreras R, Meagher MM, Callewaert N: **Fed-batch fermentation of GM-CSF-producing glycoengineered *Pichia pastoris* under controlled specific growth rate**. *Microb Cell Fact* 2010, **9**:93.

30. Nett JH, Gomathinayagam S, Hamilton SR, Gong B, Davidson RC, Du M, Hopkins D, Mitchell T, Mallem MR, Nylen A *et al.*:

Optimization of erythropoietin production with controlled glycosylation-PEGylated erythropoietin produced in glycoengineered *Pichia pastoris*. *J Biotechnol* 2012, **157**:198-206.

31. Berdichevsky M, D'Anjou M, Mallem MR, Shaikh SS, Potgieter TI: **Improved production of monoclonal antibodies through oxygen-limited cultivation of glycoengineered yeast**. *J Biotechnol* 2011, **155**:217-224.

32. Potgieter TI, Cukan M, Drummond JE, Houston-Cummings NR, Jiang Y, Li F, Lynaugh H, Mallem M, McKelvey TW, Mitchell T *et al.*: **Production of monoclonal antibodies by glycoengineered *Pichia pastoris***. *J Biotechnol* 2009, **139**:318-325.

33. Ye J, Ly J, Watts K, Hsu A, Walker A, McLaughlin K, Berdichevsky M, Prinz B, Sean Kersey D, D'Anjou M *et al.*: **Optimization of a glycoengineered *Pichia pastoris* cultivation process for commercial antibody production**. *Biotechnol Prog* 2011, **27**:1744-1750.

34. Jacobs PP, Ryckaert S, Geysens S, De Vusser K, Callewaert N, Contreras R: **Pichia surface display: display of proteins on the surface of glycoengineered *Pichia pastoris* strains**. *Biotechnol Lett* 2008, **30**:2173-2181.

35. Lin S, Houston-Cummings NR, Prinz B, Moore R, Bobrowicz B, Davidson RC, Wildt S, Stadheim TA, Zha D: **A novel fragment of antigen binding (Fab) surface display platform using glycoengineered *Pichia pastoris***. *J Immunol Methods* 2012, **375**:159-165.

36. Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi B-K, Cook WJ, Cukan M *et al.*: **Optimization of humanized IgGs in glycoengineered *Pichia pastoris***. *Nat Biotechnol* 2006, **24**:210-215.

37. Juven-Gershon T, Cheng S, Kadonaga JT: **Rational design of a super core promoter that enhances gene expression**. *Nat Methods* 2006, **3**:917-922.

38. Ruth C, Glieder A: **Perspectives on synthetic promoters for biocatalysis and biotransformation**. *Chembiochem* 2010, **11**:761-765.

39. Blount BA, Weenink T, Vasylechko S, Ellis T: **Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology**. *PLoS ONE* 2012, **7**:e33279.

40. Vogl T, Glieder A: **Regulation of *Pichia pastoris* promoters and its consequences for protein production**. *New Biotechnol* 2012 http://dx.doi.org/10.1016/j.nbt.2012.11.010.

41. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A: **Promoter library designed for fine-tuned gene expression in *Pichia pastoris***. *Nucleic Acids Res* 2008, **36**:e76.

42. Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y: **An upstream activation sequence controls the expression of AOX1 gene in *Pichia pastoris***. *FEMS Yeast Res* 2009, **9**:1271-1282.

43. • Ruth C, Zuellig T, Mellitzer A, Weis R, Looser V, Kovar K, Glieder A: **Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris***. *Syst Synth Biol* 2010, **4**:181-191.
Describes the use of P_{AOX1} deletion variants and the design of small synthetic promoters (by the fusion of *cis*-acting elements to core promoter fragments) to improve protein production.

44. Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J: **GAP promoter library for fine-tuning of gene expression in *Pichia pastoris***. *Appl Environ Microbiol* 2011, **77**:3600-3608.

45. Mellitzer A, Weis R, Glieder A, Flicker K: **Expression of lignocellulolytic enzymes in *Pichia pastoris***. *Microb Cell Fact* 2012, **11**:61.

46. Nett JH, Gerngross TU: **Cloning and disruption of the PpURA5 gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris***. *Yeast* 2003, **20**:1279-1290.

47. Yang J, Jiang W, Yang S: **mazF as a counter-selectable marker for unmarked genetic modification of *Pichia pastoris***. *FEMS Yeast Res* 2009, **9**:600-609.

48. Pan R, Zhang J, Shen W-L, Tao Z-Q, Li S-P, Yan X: **Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette**. *FEMS Yeast Res* 2011, **11**:292-298.

49. Marx H, Mattanovich D, Sauer M: **Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris***. *Microb Cell Fact* 2008, **7**:23.

50. • Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A: **Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology**. *PLoS ONE* 2012, **7**:e39720.
Report on a *P. pastoris* Δ*ku70* strain with increased rates of homologous recombination, achieved by deleting a gene (*KU70*) coding for a protein involved in non-homologous end joining. Thereby access to deletions, which were otherwise difficult to obtain for *P. pastoris*, was facilitated. Contains also the description of self-excising knock out cassettes (similar to [47–49]) enabling subsequent rounds of strain engineering with only one selective marker.

51. Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO: **Enzymatic assembly of DNA molecules up to several hundred kilobases**. *Nat Methods* 2009, **6**:343-345.

52. De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N: **Genome sequence of the recombinant protein production host *Pichia pastoris***. *Nat Biotechnol* 2009, **27**:561-566.

53. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B: **Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris***. *Microb Cell Fact* 2009, **8**:29.

54. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R *et al.*: **High-quality genome sequence of *Pichia pastoris* CBS7435**. *J Biotechnol* 2011, **154**:312-320.

55. • Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee H, Lee D: **Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement**. *Microb Cell Fact* 2010, **9**:50.
See annotation to Ref. [57•].

56. • Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY: **Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production**. *Biotechnol J* 2010, **5**:705-715.
See annotation to Ref. [57•].

57. • Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J: **Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials**. *BMC Syst Biol* 2012, **6**:24.
Refs. [55•,56•,57•] describe metabolic models for two different *P. pastoris* strains GS115 [55•,57•] and DSMZ 70382 [56•] that can be used for *in silico* predictions as a basis for engineering approaches.

58. Heyland J, Fu J, Blank LM, Schmid A: **Carbon metabolism limits recombinant protein production in *Pichia pastoris***. *Biotechnol Bioeng* 2011, **108**:1942-1953.

59. Delic M, Rebnegger C, Wanka F, Puxbaum V, Haberhauer-Troyer C, Hann S, Köllensperger G, Mattanovich D, Gasser B: **Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast**. *Free Radic Biol Med* 2012, **52**:2000-2012.

60. Hou J, Tyo K, Liu Z, Petranovic D, Nielsen J: **Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae***. *Metab Eng* 2012, **14**:120-127.

61. Hou J, Tyo KEJ, Liu Z, Petranovic D, Nielsen J: **Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae***. *FEMS Yeast Res* 2012, **12**:491-510.

62. Nielsen J, Keasling JD: **Synergies between synthetic biology and metabolic engineering**. *Nat Biotechnol* 2011, **29**:693-695.

63. Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G: **The future of metabolic engineering and synthetic biology: towards a systematic practice**. *Metab Eng* 2012, **14**:233-241.

ELSEVIER

Regulation of *Pichia pastoris* promoters and its consequences for protein production

Review

Thomas Vogl[1] and Anton Glieder[1,2,]

[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz A-8010, Austria
[2] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, Graz A-8010, Austria

The methylotrophic yeast *Pichia pastoris* is a widely used host for heterologous protein production. Along with favorable properties such as growth to high cell density and high capacities for protein secretion, *P. pastoris* provides a strong, methanol inducible promoter of the alcohol oxidase 1 (*AOX1*) gene. The regulation of this promoter has been extensively studied in recent years by characterizing *cis*-acting sequence elements and *trans*-acting factors, revealing insights into underlying molecular mechanisms. However, new alternative promoters have also been identified and characterized by means of their transcriptional regulation and feasibility for protein production using *P. pastoris*. Besides the often applied GAP promoter, these include a variety of constitutive promoters from housekeeping genes (e.g. *TEF1*, *PGK1*, *TPI1*) and inducible promoters from particular biochemical pathways (e.g. *PHO89*, *THI11*, *AOD*). In addition to these promoter sequence/function based studies, transcriptional regulation has also been investigated by characterizing transcription factors (TFs) and their modes of controlling bioprocess relevant traits. TFs involved in such diverse cellular processes such as the unfolded protein response (UPR) (Hac1p), iron uptake (Fep1p) and oxidative stress response (Yap1p) have been studied. Understanding of these natural transcriptional regulatory networks is a helpful basis for synthetic biology and metabolic engineering approaches that enable the design of tailor-made production strains.

Contents

Corresponding author: Glieder, A. (anton.glieder@acib.at)

1871-6784/$ - see front matter http://dx.doi.org/10.1016/j.nbt.2012.11.010

Background

Methylotrophic yeasts like *Pichia pastoris*, *Candida boidinii*, *Hansenula polymorpha* (*Pichia angusta*) and *Pichia methanolica* are widely used host systems for heterologous protein production [1,2]. They unite the benefits of a eukaryotic expression system (efficient secretion and post-translational modifications like glycosylations or the introduction of disulfide bonds) with favorable growth properties of robust unicellular organisms (fast growth on minimal media, no production of intrinsic endotoxins or viral DNAs) [3].

Expression of the respective gene of interest (GOI) for recombinant protein production includes transcription, translation, folding and possible posttranslational modification as well as the correct targeting by the heterologous host organism. The initial transcription of the GOI is often a critical step in heterologous protein production, in both pro- and eukaryotes [4–7]. Therefore, strong and controllable promoters are a crucial tool for efficient heterologous protein production [7,8]. Promoters regulate transcription by providing specific DNA-binding sites for transcription factors (TFs) and the basal transcription machinery (e.g. the RNA polymerase and associated factors). TFs are *trans*-acting factors that can act either as activators or repressors and which bind to *cis*-acting regulatory DNA elements [9]. Even for the more complex eukaryotes, the understanding of the underlying mechanisms has drastically increased over the past decades and light has been shed on both *trans*-acting factors and *cis*-acting sequence elements including their effect on expression and involvement in various transcriptional regulatory circuits [10–12]. This knowledge and random approaches have been used to engineer synthetic promoter libraries. These promoter variants have also been applied in biotechnological protein production as they provide altered regulatory profiles and a range of different promoter strengths, often exceeding wild-type promoters available in a certain host [13,14]. Finding the optimal transcription level by comparing various natural promoters, synthetic variants or tunable promoters is an important aspect to optimize protein production in an expression system. This review is focused on recent findings on promoters and transcriptional regulatory circuits of TFs in the methylotrophic yeast *P. pastoris* and their application to improve heterologous protein production.

One of the most favorable properties of *P. pastoris* and the other methylotrophic yeasts are the strong methanol inducible promoters present in these organisms. Certain genes of the methanol utilization (MUT) pathway are completely repressed under growth on repressing carbon sources (e.g. glucose), but are strongly induced on methanol, constituting up to 20–30% of total intracellular protein [15,16]. This tight regulation is exerted on the transcriptional level, making MUT promoters ideal tools for heterologous protein production [17]. Therefore also most of the published knowledge about the regulation of *P. pastoris* promoters is focused on the most commonly used classic *P. pastoris* promoters, the promoters of the MUT pathway gene *alcohol oxidase 1* (P_{AOX1}), its rarely used homolog P_{AOX2}, and also the glycolysis gene *glyceraldehyde 3-phosphate dehydrogenase* (P_{GAP}), which was described as a constitutive promoter in the early days of development. However, although less characterized, alternative promoters offer advantages for some proteins and also first regulatory networks for important folding processes and stress response in *P. pastoris* have been reported.

Classic *P. pastoris* promoters

AOX1 *promoter*

Physiological role and regulation of AOX1 expression

Alcohol oxidase (Aox) catalyzes the first step in the metabolism of methanol, the conversion of methanol to formaldehyde. This reaction is taking place in the peroxisomes and involves the incorporation of oxygen and release of hydrogen peroxide as a side product [3]. The oligomeric flavoenzyme alcohol oxidase has however a low affinity for its substrates methanol and oxygen [18,19]. Therefore large amounts of the enzyme are needed to sustain growth on methanol and *AOX* accounts under these conditions for about 5% of total mRNA and 30% of total soluble protein. Cells grown on repressing carbon sources (e.g. glucose, glycerol, ethanol) [20] show however a complete lack of AOX activity. This tight regulation is exerted on the transcriptional level, as *AOX* mRNA was described as undetectable in cells grown on repressing carbon sources [21–23].

P. pastoris has two alcohol oxidase genes, *AOX1* and *AOX2*, of which *AOX1* is much more strongly transcribed than *AOX2* [22,23]. As P_{AOX1} is a remarkably strong promoter and at the same time directly controllable by simply changing the inexpensive carbon source [24], it is the most commonly used promoter for heterologous protein expression in *P. pastoris*. The highest expression levels of heterologous proteins in *P. pastoris* have been achieved with P_{AOX1} (up to 22 g/l intracellular and 15 g/l secreted protein) [25,26]. Concerning controlled bioreactor cultivation, the tight regulation of P_{AOX1} is used to enable cell growth to high densities on a repressing carbon source without heterologous protein expression hampering the cellular metabolism.

On the basis of the observation that P_{AOX1} induction requires the complete lack of a repressing carbon source and the presence of the inducer methanol, the current regulatory model involves three regulatory states. Catabolite repression suppresses expression as long as a repressing carbon source is present. Even if methanol is fed, the promoter is still repressed. As soon as the repressing carbon source is depleted or at low concentration in fed-batch processes, the promoter reaches the state of derepression. In *P. pastoris*, under these conditions the native *AOX1* promoter is still only slightly derepressed (2–4% of induced levels) and only methanol addition leads to full induction, which is the third state of regulation [20,27,28]. Fig. 1 shows a schematic representation of P_{AOX1}. The most commonly used sequence from the *Bgl*II restriction site downstream and features like the TATA box and the transcription start site are highlighted. Tschopp *et al*. [27] and Stroman *et al*. [29] had initially used a 1.1 kbp *Pst*I/*Bam*HI P_{AOX1} fragment (the respective sites are shown in Fig. 1).

Despite the favorable properties of P_{AOX1} and the basic understanding of its regulatory modes, only in recent years light has been shed on the molecular basis of P_{AOX1} transcriptional regulation, thereby enabling new tools and concepts for *P. pastoris* based gene expression. A series of *cis*-acting DNA sequences and also the first two transcriptional activators Mxr1p and Prm1p were identified. Furthermore there was also evidence for regulatory roles of a zeta crystalline and hexose transporters in P_{AOX1} regulation.

Cis-acting regulatory sequence elements

Cis-acting regulatory elements in the promoters of methanol regulated genes have been studied in different methylotrophic yeasts such as *H. polymorpha* [30,31], *C. boidinii* [32,33] and also extensively in *P. pastoris*. Research in *P. pastoris* has mainly focused on P_{AOX1} and P_{AOX2} by performing deletion and insertion studies or comparative sequence analysis. In addition an early study was performed using a previously uncharacterized *P. pastoris* strain [34]. An alcohol oxidase gene, named *ZZA1*, was identified by hybridization with an *AOX1* oligonucleotide and the promoter of *ZZA1* showed 66% identity to P_{AOX1}. This promoter was used to express an α-amylase gene in *S. cerevisiae* to produce ethanol from starch. The authors identified putative transcription factor binding sites (TFBS) in the P_{ZZA1} sequence and also highly conserved sequence motifs between P_{ZZA1} and the promoters of methanol inducible genes from various organisms such as *P. pastoris AOX1*, *AOX2*, *H. polymorpha MOX*, *DAS* and *C. boidinii* S2 *AOD1*. It was argued that these conserved motifs resemble the TFBS of *S. cerevisiae* Gcr1 and Rap1, both activating factors. Notable unique elements from P_{ZZA1} include a putative upstream repressing site and a putative binding site for a *S. cerevisiae* TF encoded by the *ADR1* (alcohol dehydrogenase regulator 1) gene. Adr1p activates in *S. cerevisiae* various glucose repressed genes such as alcohol dehydrogenase (*ADH2*) [35], fatty acid metabolism and peroxisomal biogenesis genes [36,37]. This indicates that the *P. pastoris* homolog of *Sc*Adr1p could have adapted to the regulation of methanol induced genes. Lin-Cereghino *et al*. identified later indeed a positive regulator of methanol inducible genes, Mxr1p (methanol expression regulator 1), which is a homolog of ScAdr1p (see also section '*Trans*-acting factors') [28].

In conclusion, the P_{ZZA1} study indicated first possible regulatory sequences which were however only identified by sequence

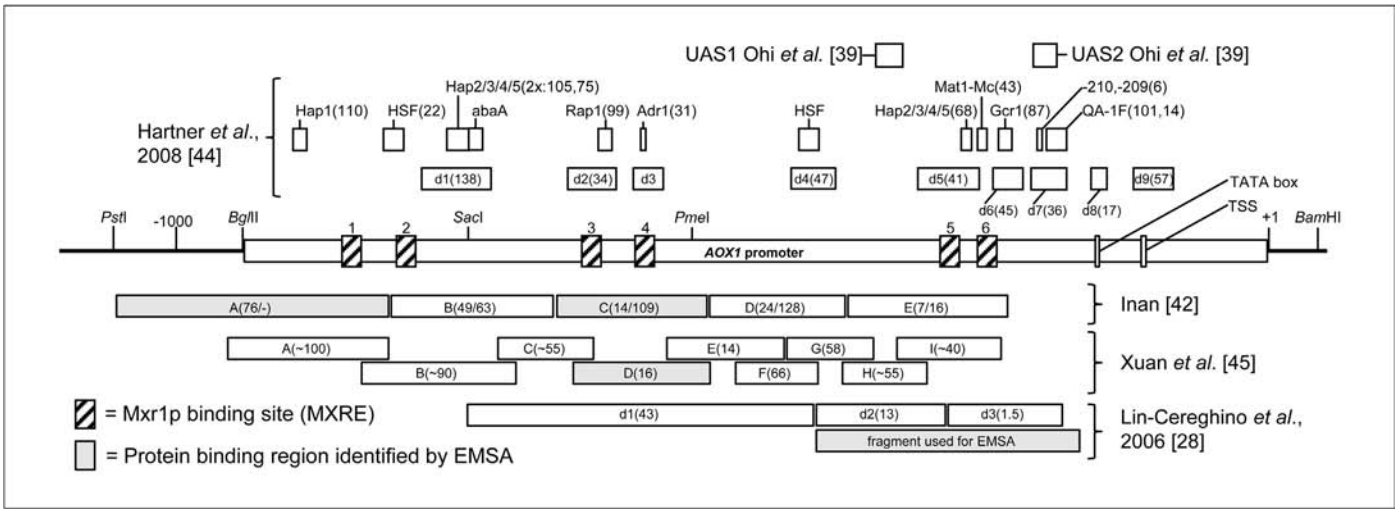


FIGURE 1

Cis-acting sequence elements of the *AOX1* promoter and their effect on expression. A schematic representation of the most commonly used *AOX1* promoter sequence (from *Bgl*II restriction site to start codon) is given. The *Pst*I, *Bgl*II, *Sac*I, *Pme*I, *Bam*HI restriction sites, TATA box and transcriptional start site (TSS) are shown. The translational start is indicated by +1. Sequence elements identified by systematic deletion studies by Inan [42], Xuan *et al.* [45] and Lin-Cereghino *et al.* [28] are shown below the strand ruler. Putative TFBS identified by Hartner *et al.* [44] using MatInspector software and related deletion studies are shown above the strand ruler. If reported, the effect of the deletions is given in parentheses in percent of wild-type promoter activity (under methanol induction). Inan performed sequential and drop out deletions, and both effects are given separated by a slash. Hartner *et al.* tested in total 46 deletion variants; here only deletions of the putative TFBS and longer stretches (d1 to d9) are shown. Hartner *et al.* did not provide percent of wild type activity values for all deletions, therefore the values shown here were calculated from Supplementary Table 3 of Hartner *et al.* (using the values at 72 hours induction). There are three putative Hap2/3/4/5 binding sites present, the first two are close together, and therefore just shown as one site. In case of the putative QA-1F site, the first value represents the deletion only of the putative TFBS, and the second value represents a deletion of additional six bp adjacent to the TFBS. Also the putative UAS suggested by Ohi *et al.* are depicted [39]. Elements shown in gray are bound by proteins, as they caused a band shift with cell extracts from cultures grown on methanol in the respective studies. Experimentally verified Mxr1p binding sites (MXRE, MXR response elements) are shown as dashed boxes and numbered according to Kranthi *et al.* [48].

comparison, but not verified by deletion or insertion studies [34]. Practically all other *P. pastoris* related research was performed using strains based on GS115/X33/KM71, CBS7435/NRRL Y-11430 or DSMZ 70382 and not the unique isolate bearing P_{ZZA1}. Furthermore P_{ZZA1} characterization was only described for the host *S. cerevisiae* [38], but not *P. pastoris* and due to the low sequence identity of 66% it is difficult to transfer the P_{ZZA1} derived results directly to P_{AOX1}.

The first deletion study in *P. pastoris* was performed by Ohi *et al.* using the promoter of the *AOX2* gene [39]. In early studies, Aox1p accounted under all tested growth conditions for most of the methanol oxidizing activity and no clear physiological role of Aox2p was found. Deletion of *AOX1* resulted in a strongly decreased growth rate on methanol, whereas the deletion of *AOX2* retained a wild type like growth. The two genes have 97% identity on the amino acid level but no obvious similarities in their promoter regions. It was furthermore shown that the difference in Aox activities is the effect of transcriptional regulation. This was done by placing the coding sequence (CDS) of *AOX2* under control of P_{AOX1}. The respective strain showed the same growth rate as the wild type, whereas the reverse exchange of the *AOX1* CDS under control of P_{AOX2} caused a strong growth defect [23].

Despite their different sequence and dissimilar expression levels, P_{AOX1} and P_{AOX2} show the same basic regulation pattern, as both require the lack of a repressing carbon source and presence of methanol for induction [23,39]. Therefore P_{AOX2} based studies also appear relevant for understanding the regulation of P_{AOX1}. Ohi *et al.* identified negative and positive regulatory elements of P_{AOX2} by deletion and addition of sequence stretches [39]. These studies indicated one upstream activating sequence (UAS) and two upstream repressing sequences (URS). The identified UAS showed similarities with P_{AOX1} and also the promoters from MUT genes of other methylotrophic yeasts such as *C. boidinii AOD1* and *H. polymorpha DAS*. Although it might be easier to improve a weak promoter like P_{AOX2} than the exceptionally strong P_{AOX1} promoter [17], remarkable progress was made in the understanding of methanol inducible promoters. Shortened promoter variants with added UAS showed a more than 100-fold increase in human serum albumin (HSA) expression (from 1 to 120 μg/ml). Whether such an improvement is transferable to the strong P_{AOX1}, where several mg/ml HSA can be obtained [40,41], remained unclear.

The first P_{AOX1} based systematic deletion studies were performed in frame of a PhD thesis [42] and a related patent [43]. Inan [42] truncated the P_{AOX1} sequence from the 5′ end by sequentially deleting five fragments named A to E (Fig. 1). The stepwise decrease of the promoter length concomitantly reduced the methanol induced expression of the reporter lacZ. This classical truncation approach for promoter studies resulted in generally lower lacZ activities, because large regions, probably containing several TFBS, were deleted in each individual truncation variant. Therefore internal 'dropout' deletions of fragments B to E were also performed, leaving the 5′ promoter part upstream of the deleted fragment intact. Deletions of fragments B and E resulted in decreased reporter activity (63 and 16% of wild-type P_{AOX1} activity were observed), whereas deletions of fragments C and D increased the activity to 109 and 128%. This seems to contradict results of the deletion of the d1 region by Hartner *et al.* [44]. The d1 region is part of region B deleted by Inan but a d1 deletion increased the promoter activity significantly. This indicates the location of several important binding sites in region B and a combined effect of lost positive and negative sites upon deletion of this region. All fragments were tested for protein–DNA interactions with cell extracts from methanol induced and ethanol repressed cultivations. In these electrophoretic mobility shift assay (EMSA) experiments only fragments A and C showed specific DNA–protein complexes, with cell extracts from methanol induction. These interactions were considered to arise from interactions with different proteins, as fragment A did not compete binding of fragment C and vice versa. Unfortunately, the bound proteins were not further characterized [42]. Three fairly long fragments of P_{AOX1} were sequentially deleted by Lin-Cereghino *et al.* [28] (Fig. 1) to identify binding regions of the TF Mxr1p (discussed in the section about *trans*-acting factors). However, while this study mapped a putative Mxr1p binding site, the deletion studies of Inan [42] and Xuan *et al.* [45] achieved a higher resolution by deleting shorter sequence stretches and also included internal drop out deletions. This allowed to map also none additive effects, which are solely obtained by sequential deletion studies.

An alternative design for deletion studies was followed by Hartner *et al.* [44,46]. In contrast to sequential deletions of large sequence stretches, putative regulatory *cis*-acting elements of the P_{AOX1} sequence were identified by sequence similarities to known TFBS from other organisms. Subsequently systematic single- and multisite deletions of those short sequence stretches were performed (Fig. 1). Thereby a first promoter library was created, that included 46 variants showing between ~6 and >160% of the wild-type promoter activity employing GFP as a reporter. For the most important promoter variants single copy integration of the expression cassette was verified by qPCR and the results were reevaluated with other reporter proteins. Transcript levels have not been determined but because GFP expression usually increases with copy numbers it might be speculated that strong GFP fluorescence also indicates high mRNA levels. In the same library variants with different regulatory patterns (e.g. derepression) were also identified. The deletions of varying short stretches (2–59 bp) of these TFBS and longer internal fragments (termed d1 to d9) showed different effects. For several putative TFBS (e.g. HSF, Adr1, Mat1-Mc), a clear reduction of P_{AOX1} driven expression was observed (22–43% fluorescence intensity compared to wild-type promoter driven GFP expression, Fig. 1). In one case the deletion of only two bp (at position −209 and −210) abolished GFP expression almost completely (only 6% of wild type fluorescence left) indicating the high importance of this putative TF binding site.

However, improvements of the strong *AOX1* promoter could also be achieved, for example, by deletion of a longer stretch at the 5′ end (fragment d1). Combination of multiple beneficial deletions also showed additive effects, and the best variant reached more than 160% of the wild-type promoter activity. The authors reasoned that deleted stretches leading to decreased expression might be binding sites for activators. Increased GFP fluorescence by contrast indicated possible repressor binding sites. Therefore they designed synthetic promoters, by fusing various putative TFBS of activators to a basal core promoter region. Compared to the core promoter alone, several synthetic promoters showed increased expression levels and altered regulatory patterns. Namely a tighter

glucose dependent repression, and enhanced expression under both derepressing and induced conditions have been achieved. Because all variants were made by deletions, structural effects in the DNA or disturbance of the interaction of different *trans*-acting proteins cannot be excluded.

Interestingly, the deletion variants d6 and d6* also showed a notable altered regulation. They caused increased expression after the carbon source was depleted (derepression). This variant still required methanol for full induction, but the protein production started earlier during the production process resulting for example in twice as much active horseradish peroxidase as the control employing the wild-type promoter [44].

On the basis of the findings of Hartner *et al.* [44], a similar strategy of fusing *cis*-acting elements to a core promoter was also applied by Ruth *et al.* and led to higher productivity and yields of porcine trypsinogen using P_{AOX1} derived variants [47]. Previously characterized sequence elements from Hartner *et al.* and novel regulatory elements were fused to the P_{AOX1} core promoter sequence. These new variants showed up to 218% activity of the wild-type P_{AOX1} and also altered regulation patterns. Most notably a derepressed variant reached without methanol induction 77% of the wild-type activity after 96 hours of methanol induction. However, it is notable that there were multiple copies of this variant present, but only one copy of the wild-type P_{AOX1} was present. Nevertheless this was the first proof that very short semisynthetic promoters built on the basis of a natural core promoter can be used to obtain higher titers of some target proteins. In addition this variant was also shown to be applicable for a methanol-independent fed-batch process.

Various regulatory sequences affecting P_{AOX1} expression were also identified by Xuan *et al.* They deleted nine adjacent fragments, named regions A to I (Fig. 1) [45]. In contrast to the studies of Inan and Lin-Cereghino *et al.*, these fragments were overlapping [28,42]. Most deletions led to a decrease of GFP reporter activity and decreased transcript levels as determined by RT-qPCR. Notably the deletion of regions D and E led to only 16% and 14% activity of the wild-type promoter under methanol induction. All fragments were analyzed for DNA–protein interactions by performing an EMSA. Cell extracts prepared from cultures grown on methanol or glucose and methanol displayed a specific band shift only with region D, but none of the other tested sequence elements. The authors noticed that a stretch of region D showed similarities to the TFBS of the transcriptional activator Adr1p from *S. cerevisiae* (also Hartner *et al.* had identified a putative Adr1p binding site in this region [Fig. 1] [44]). Additionally, insertion of three adjacent copies of region D to the region D deletion variant increased the reporter protein level to 157% of the wild-type promoter. These results indicated that region D is the binding site for a transcriptional activator involved in methanol induction. Lin-Cereghino *et al.* had shown that Mxr1p, the *P. pastoris* homolog of Adr1p, is a key activator of methanol induced genes [28]. Therefore Xuan *et al.* argued that the TFBS of Adr1p and Mxr1p could also be conserved and region D might be a binding region for Mxr1p. Indeed Kranthi *et al.* [48] showed independently by EMSA and DNase1 foot printing that region D contains even two Mxr1p binding sites (discussed in detail in the following section *trans*-acting factors, see also Fig. 1).

In summary, these sequence-based approaches identified *cis*-acting regulatory elements, improved the already strong P_{AOX1} and also provided new variants showing different regulatory patterns. Additionally, more simple short promoters composed of a basic core promoter sequence and additional positive regulatory sequences have been obtained. Generally, the results of the different studies match fairly well, but there are also a few notable discrepancies. Inan [42] and Xuan *et al.* [45] deleted approximately the same fragment (InanC and XuanD, Fig. 1). However, the dropout deletion of XuanD decreased the activity to 16% of the wild-type promoter, whereas the dropout deletion of InanC retained wild type like activity (109%). Consequences to the DNA structure influencing the interaction of DNA-binding proteins might be one possible explanation. Deletions by Hartner *et al.* [44] in this region also caused a loss of activity (e.g. fragment d2 only 34% of wild type activity). Notably, for both the InanC and XuanD fragment DNA–protein interactions were demonstrated in EMSA experiments with cell extracts from methanol grown cultures. Kranthi *et al.* [48] later identified two Mxr1p binding sites (MXRE3 and MXRE4, MXR reponse element) in the region covered by InanC and XuanD using the recombinantly produced DNA-binding domain of Mxr1p (Fig. 1 and also the next section about *trans*-acting factors). This suggests that the bandshifts arise from Mxr1p binding and that this region is crucial for methanol-dependent P_{AOX1} induction. Concerning the other Mxr1p binding sites MXRE1, 2, 5 and 6, the results from the EMSAs were somewhat inconsistent. Fragment A of Inan caused a band shift with methanol grown cells, but the experiments of Xuan *et al.* with fragment A did not confirm these results. InanA is longer than XuanA, which might explain the difference. However, MXRE1 identified by Kranthi *et al.* [48] lies within both regions (Fig. 1). As Inan showed that the binding events to region A and C are not competing each other, and InanC is most probably bound by Mxr1p, this would suggest that a different protein is binding to InanA, possibly in a region upstream of the *Bgl*II site which was not studied by others. Lin-Cereghino *et al.* [28] identified a Mxr1p binding fragment close to the 3′ end of the promoter, using a strain overexpressing Mxr1p. Fragment E of Inan also lies in this region, but did not show a band shift, possibly because Mxr1p was not overexpressed in this study. Kranthi *et al.* [48] identified two Mxr1p binding sites (MXRE5 and 6) lying exactly in InanE and the fragment used by Lin-Cereghino *et al.* [28]. Taking these results together suggest that Mxr1p binds MXRE3 and 4 with higher affinity than MXRE1,2, 5 and 6, because fragments containing MXRE3 and 4 formed a band shift directly with cell extracts of wild type strains, whereas a fragment containing MXRE5 and 6 could only cause a bandshift upon overexpression of Mxr1p. Determination of the Mxr1p dissociation constant (K_d) by Kranthi *et al.* showed indeed that MXRE4 has a lower K_d than MXRE6 (70 vs. 150 nM) [48]. MXRE1 even had a higher affinity than MXRE4 and 6 (K_d of 50 nM). It is also striking that the Mxr1p binding sites seem to be organized in tandem (Fig. 1). MXRE1&2, 3&4 and 5&6 are arranged in close proximity, however the significance (e.g. for cooperative binding) has not been investigated.

Concerning the outcome of the deletion studies, it is interesting to see, that the two different sequence-based approaches applied by Inan [42] and Xuan *et al.* [45] versus Hartner *et al.* [44] displayed similar results. Hartner *et al.* [44] based their analysis on literature research and computational analysis relying only on deletions of putative TFBS, whereas Inan and Xuan *et al.*

performed a systematic deletion analysis by dividing P_{AOX1} into adjacent overlapping fragments (Fig. 1). Still the best variants of both studies reached similar expression levels compared to the wild-type (Hartner *et al*. [44] >160%, Xuan *et al*. [45] 157%). Nevertheless the deletion of short predicted TFBS delivered different results for specific regions than deletions of larger regions, which resulted only in combined effects of several TFBS. In addition the knowledge of specific short positive and negative regulatory sequences as well as the known Mxr1p binding sites allow the design of short synthetic promoters.

All these studies aimed to identify UAS and URS relevant for transcriptional regulation, thereby focusing on sequences upstream the transcription start site and the TATA box core promoter region. Notably, the 5′ untranslated region (5′ UTR), which effects translation initiation and mRNA stability, was left mostly untouched (Fig. 1). Recently, Staley *et al*. investigated the effect of the *AOX1* 5′ UTR on expression by performing deletion and addition studies in a similar fashion as the promoter based studies [49]. This revealed that the *AOX1* 5′ UTR affects mostly translational efficiency, but not mRNA stability as determined by Northern blotting and that there are both positive and negative regulatory elements present.

Trans-acting regulatory factors

The sequence-based approaches of identifying *cis*-acting sequence elements, have to some extent given indications about *trans*-acting DNA-binding proteins and possibly involved TFs. Inan [42] and Xuan *et al*. [45] showed that certain sequence stretches of P_{AOX1} form a DNA–protein complex with cell extracts from methanol induced cells. However, the proteins involved had not been purified or characterized by mass spectroscopy, leaving their identity unknown.

The first TF involved in P_{AOX1} regulation in *P. pastoris* was identified by Lin-Cereghino *et al*. and was named methanol expression regulator 1 (Mxr1) [28]. *MXR1* was shown to be crucial for the activation of numerous genes involved in MUT and peroxisome biogenesis (*PEX* genes). Initially a *mxr1* mutant was obtained by random mutagenesis in a search for strains defective in peroxisome biogenesis [50]. In contrast to other *pex* mutants, the *mxr1* mutant could not induce P_{AOX1}. Deletion of *MXR1* in the wild type strain resulted in abolished growth on the peroxisomal substrates methanol and oleate. In conclusion, the *mxr1* deletion strain showed under methanol induction highly reduced protein and mRNA levels of *AOX1* and other genes involved in MUT, indicating that Mxr1p is a key regulator of methanol induced genes [28].

Mxr1p was found to show similarities to the zinc finger DNA-binding domain and other motifs of *S. cerevisiae* Adr1p, a TF involved in regulation of glucose repressed genes needed for growth on nonfermentable carbon sources [35–37]. Notably, already in the first P_{AOX1} related sequence based deletion study using P_{ZZA1} [34], the authors had identified putative Adr1p binding sites in this *P. pastoris* alcohol oxidase promoter and discussed its significance (see previous section about *cis*-acting sequences). Also later sequence-based studies of P_{AOX1} identified putative Adr1p binding sites [44,45]. Lin-Cereghino *et al*. reasoned that *P. pastoris* Mxr1p is a homolog of *S. cerevisiae* Adr1p and has evolutionary adopted to regulate genes involved in the methanol metabolism [28].

The binding of Mxr1p to P_{AOX1} was first evaluated by Lin-Cereghino *et al*. by investigating a putative regulatory region identified by deletion studies [28]. An EMSA experiment showed that cell extracts of an Mxr1p overexpressing strain grown on methanol caused a band shift with a 243 bp long region of P_{AOX1} (Fig. 1). However, as only three fairly long fragments (d1 to d3) of P_{AOX1} were tested by sequential deletion studies, the presence of multiple binding sites could not be ruled out.

Indeed, a systematic analysis for Mxr1p binding sites in P_{AOX1} by Kranthi *et al*. [48], revealed six binding sites (termed MXRE, Mxr1 response elements) by EMSA and DNaseI footprinting experiments (Fig. 1) [48]. These studies were performed *in vitro* using a 150 AA N-terminal fragment of Mxr1p, which contained the zinc finger DNA-binding domain and was produced recombinantly in *Escherichia coli*. The P_{AOX1} sequence was divided into 16 fragments, six of which showed a band shift with Mxr1p. The minimal Mxr1p biding sites in these six fragments were identified by DNase1 digestion and found to be approximately 20 bp long containing a 4 base pair 5′ CYCC/GGRG 3′ core element.

These results match previous research as the Mxr1p binding fragment identified earlier by Lin-Cereghino *et al*. [28] was shown to even contain two Mxr1p binding sites. Also the putative TFBS of ScAdr1p identified by Hartner *et al*. [44] was confirmed to be a Mxr1p binding site.

In a following study Mxr1p binding sites were also identified in the promoters of *dihydroxyacetone synthase* (*DAS*) and *peroxin* 8 (*PEX8*) [51]. As these two genes are involved in the MUT pathway and peroxisomal biogenesis, the results emphasize the key regulatory function of *MXR1* proposed earlier [28].

MXR1 seemed to be regulated post-translationally by the subcellular localization, as it was constitutively transcribed at low levels and found to localize to the cytosol in cells grown on glucose but to migrate to the nucleus upon shift to methanol or oleic acid. However, also when cells were grown on glycerol or ethanol Mxr1p was localized in the nucleus, while *AOX1* expression was still repressed on these carbon sources. Therefore regulation of *MXR1* cannot solely rely on the subcellular localization of the Mxr1p [28].

Very recently Parua *et al*. have shown that a *P. pastoris* 14-3-3 protein (GenBank accession no. CCA38880) is involved in the regulation of Mxr1p in a carbon source dependent way [52]. 14-3-3 proteins are highly conserved in all eukaryotes and regulate signaling proteins involved in diverse cellular processes. Typically 14-3-3 interaction depends on the phosphorylation of serine or threonine residues of the target protein. The name '14-3-3' dates back to the initial identification in frame of the systematic classification of brain proteins and refers to the purification fraction number in which these proteins were found [53,54]. 14-3-3 proteins have also been studied in yeast [55,56] and are notably involved in the regulation of *S. cerevisiae* Adr1p, the homolog of *P. pastoris* Mxr1p [57]. The *S. cerevisiae* 14-3-3 proteins Bmh1 and Bmh2 bind to the regulatory domain of phosphorylated Adr1p, thereby inhibiting it. A similar role was found for a 14-3-3 protein in Mxr1p regulation in *P. pastoris* [52]. A sequence comparison showed that Mxr1p contains a conserved 14-3-3 binding motif, that is highly similar to the respective region of Adr1p. An interaction between Mxr1p and the 14-3-3 protein was shown by pull down assays and found to depend on the phosphorylation of a

conserved serine residue of Mxr1p. 14-3-3 binding to Mxr1p did not affect DNA binding of Mxr1p, indicating a post-DNA binding regulatory mechanism (e.g. possibly affecting the recruitment of RNA polymerase II). Carbon source dependent binding of the 14-3-3 protein to Mxr1p seemed to inhibit the expression of Mxr1p regulated genes. A disruption of the interaction between 14-3-3 with Mxr1p by mutating the crucial serine residue of Mxr1p to alanine, led to an elevated expression of Mxr1p regulated genes on repressing carbon sources. The activating effect was clearly stronger on ethanol than on glucose and glycerol, which indicated that 14-3-3 inhibits Mxr1p regulated gene repression especially in the presence of ethanol [52]. This is an interesting finding, as glucose and ethanol were already considered before to repress MUT genes via different mechanisms in methylotrophic yeasts [58,59]. Therefore it seems probable that there exist additional regulatory factors, depending on the carbon source. Possible differences in transcript levels of 14-3-3 in methanol grown cells compared to cells grown on glucose or ethanol have not been discussed. Consequently, elucidating the exact physiological role of 14-3-3 and more notable also the factors involved in repression under other carbon sources than ethanol will require further studies.

Besides Mxr1p, another *trans*-acting factor was revealed in a patent by Takagi *et al*. [60]. Deletion of a gene referred to as *positive regulator of methanol* (*PRM1*, annotated in the *P. pastoris* CBS7435 genome sequence as uncharacterized transcriptional regulatory protein YIL130W, GenBank accession no. CCA40959.1) resulted in a growth defect on methanol but not glucose. This deletion also reduced P_{AOX1} and P_{DAS} (the *DAS* promoter is described in more detail in the section 'Alternative promoters') expression levels under methanol induction (~50-fold decrease of the activity levels of a phytase reporter gene for P_{DAS}). These findings indicated an activating role of Prm1p in regulation of MUT genes. Interestingly, constitutive expression of Prm1p and when grown on glucose resulted in an increase in activity of a phytase gene under control of P_{AOX1} or P_{DAS}. Thus a methanol-free expression from methanol inducible and natural glucose-repressed promoters was achieved by simply expressing the putative transcriptional activator Prm1p. In detail, the P_{AOX1} driven expression of the phytase was achieved by constitutively expressing *PRM1* under control of P_{GAP} and growth on glucose containing medium. A 1.86- to 3-fold increase of phytase gene expression compared to a not *PRM1* overexpressing strain was observed in individual transformants. However, the expression level was not compared to a wildtype strain expressing the reporter under control of P_{AOX1} under methanol induction, leaving the significance of these results unclear. In addition the growth behavior of strains constitutively overexpressing this TF was not described.

The corresponding experiment using P_{DAS} to control the phytase expression, led in different clones to a 7.6- to 44-fold increase of reporter gene expression compared to a not *PRM1* overexpressing strain. The expression compared to methanol induction was again not investigated.

The authors of this patent application did not provide experimental evidence for the mode of action, but argue that Prm1p is binding to the promoters of methanol inducible genes, so the role of Prm1p as a TF is not fully verified.

Additionally to the *PRM1* coexpression experiments, constitutive expression of *MXR1* under control of P_{GAP} was also investigated under growth on glucose. This resulted in a 7.6-fold increased activity of the phytase expressed from P_{AOX1} compared to the non *MXR1* coexpressing strain. Again no comparison to methanol induction or comparative biomass production was performed. The 7.6-fold increase might still account only for a low absolute expression level, as P_{AOX1} is increased ~1000-fold from glucose to methanol [27]. Furthermore, the results of the *MXR1* coexpression seem to oppose previous findings by Lin-Cereghino *et al.*, where overexpression of *MXR1* under control of P_{AOX1} caused a growth defect on methanol and inhibited expression of MUT genes [28]. Additionally, Mxr1p was shown to localize to the cytoplasm when grown on glucose and to migrate only upon shift to gluconeogenic carbon sources to the nucleus. This raises the question how constitutively overexpressed Mxr1p reaches the nucleus during growth on glucose. Therefore the findings by Takagi *et al.* and Lin-Cereghino *et al.* [28] are not entirely consistent and would require further investigation to clarify the activating mechanism of Mxr1p.

In addition to the activators Mxr1p and Prm1p, a hexose transporter responsible for glucose and fructose uptake has also been shown to be involved in P_{AOX1} regulation [61]. Two hexose transporter homologs, *HXT1* and *HXT2*, were identified in *P. pastoris* and shown to complement a *S. cerevisiae* mutant deficient in monosaccharide uptake. A *P. pastoris hxt1* deletion strain showed P_{AOX1} expression when grown on glucose, therefore indicating abolished catabolite repression. This deletion strain showed still a wild type like growth behavior on glucose, indicating that the derepression was not simply caused by lower intracellular glucose levels. Similar regulatory roles of hexose transporters have been studied extensively in other yeasts such as *S. cerevisiae* [62] and notably also the methylotrophic yeast *H. polymorpha* [63,64]. Deletion of the *H. polymorpha* hexose transporter homolog HpGCR1 led to the constitutive expression of the *H. polymorpha* alcohol oxidase (*HpAO*) gene and other peroxisomal proteins when grown on glucose [63]. Also a *H. polymorpha hxt1* deletion strain showed weak transient derepression of *HpAO* [64].

Another *trans*-acting factor possibly involved in P_{AOX1} regulation is zeta crystalline (Zta1), a single-stranded DNA (ssDNA)-binding protein [65]. The authors were looking for proteins binding to the P_{AOX1} sequence by EMSA and thereby identified a protein binding to the region of −288 to −115 of P_{AOX1}. Subsequent sequencing of a tryptic fragment identified the protein as a homolog of *E. coli*, *S. cerevisiae* and mammalian Zta1p. Zta1p from cell extracts or recombinant production showed an identical binding behavior to P_{AOX1} ssDNA. Interestingly, Zta1p binding is abolished in presence of NADPH but not NADH *in vitro*. As Zta1p is a NADPH-dependent quinone oxidoreductase and the homologs in other organisms are upregulated under oxidative stress [66], it could play a similar role in P_{AOX1} regulation [65]. Remarkably, cytochrome c, another redox protein and component of the mitochondrial electron transport chain, was shown to bind sequence specific to the *AOX2* promoter [67]. These findings could hint a possible regulatory involvement of oxidative stress and respiration in the regulation of *AOX* promoters.

In summary, the findings on both, *cis*- and *trans*-acting factors have revealed a complex regulation mechanism of P_{AOX1}, involving multiple factors. At least two activators, Mxr1p and Prm1p, were identified and one (Mxr1p) characterized in detail. Beside these direct activators, there was also evidence of a second level of

regulation. Mxr1p mediated activation was shown to be repressed by a 14-3-3 protein. However, because this effect was dependent on ethanol as carbon source, it seems possible that there are also alternate regulators taking over similar functions on other carbon sources. Also the exact molecular mechanisms of hexose transporter dependent repression of P_{AOX1} and also the function of Zta1p remain to be elucidated.

GAP *promoter*

The major advantages of P_{AOX1} driven expression are the strength and the tight regulation of this promoter. Especially for the production of toxic or detrimental heterologous proteins, the inducibility with methanol is favorable. Cultures can at first be grown to high cell densities under repressing conditions and in a second phase, heterologous protein production can be started by methanol induction [3]. However, this requires a two-step cultivation procedure, which increases the process time and handling efforts. For nontoxic proteins, which do not hamper cell growth, a strong constitutive expression system can help to minimize cultivation efforts and lead to a higher space/time yield.

Waterham *et al*. have characterized the *P. pastoris glyceraldehyde 3-phosphate dehydrogenase* gene (*GAP*) and showed that its promoter is strong and constitutive, reaching similar expression levels as P_{AOX1} [68]. Because of the constitutive expression, no induction step is needed which shortens and simplifies the process time for protein production and in some cases delivered similar and even higher product titers than P_{AOX1} driven expression. For these reasons, P_{GAP} has emerged as the most commonly used alternative to P_{AOX1}. Expression from P_{GAP} has been studied with respect to optimal vector design, the regulatory pattern and HCD bioreactor cultivations (reviewed by Zhang *et al*. [69]).

The *GAP* gene was initially isolated from a genomic library by colony hybridization with a probe of a *S. cerevisiae GAP* homolog. *GAP* is principally constitutively expressed, although there is a clear influence of the carbon source on the expression strength. Namely Northern blots showed, that the *GAP* mRNA levels were highest in cells grown on glucose and decreased approximately to two-thirds on glycerol, half on oleic acid and one-third on methanol [68]. Interestingly, the oxygen supply was also shown to affect P_{GAP} driven heterologous protein production [70]. Hypoxic conditions led to a three- to sixfold increased specific productivity in the production of three different but difficult to express heterologous proteins (a Fab fragment, human trypsinogen and porcine trypsinogen). Also the process times were decreased threefold and the volumetric productivity increased about twofold. In these studies *P. pastoris* was also shown to produce ethanol under hypoxic conditions (up to approximately 15 g/l), implying a change to oxidofermentative metabolism. The authors surmised that the hypoxic conditions could lead to a transcriptional upregulation of glycolysis genes like *GAP* [70]. However, these assumptions were not verified by the analysis of the transcript levels. Therefore it would also be possible, that the hypoxic conditions led to a weaker transcription by P_{GAP}, which could have been favorable for the expression of these difficult to express heterologous proteins.

Also in studies of *P. pastoris* alternative respiration, there were indications that P_{GAP} driven expression is not entirely constitutive but regulated under certain conditions [71]. The authors investigated the expression of a fusion protein of an alternative oxidase and GFP under control of P_{GAP}. They noticed that the fluorescence increased upon depletion of glucose strongly for approximately two doubling times where transiently also some ethanol was produced. Yet the mechanism is unknown, and this effect was also noticed in other cases before [71].

In contrast to P_{AOX1} there is hardly any information available on *cis*-acting sequence elements or *trans*-acting factors involved in P_{GAP} regulation. Still, there have been efforts to optimize this already very strong promoter. Qin *et al*. took a random approach and created a library of P_{GAP} variants spanning a wide range of activities, which helped to optimize the expression of a methionine adenosyltransferase [72]. The starting point was a random mutagenesis of the P_{GAP} sequence by error-prone PCR and the screening of ~30,000 clones with a GFP variant as reporter gene. This resulted in P_{GAP} variants ranging from ~0.006- to 19.6-fold wild-type promoter activity. Seven variants, spanning the whole range, were further studied by determination of transcript levels by RT-qPCR and the use of two additional reporter enzymes (lacZ and the methionine adenosyltransferase). These strains were also confirmed to contain a single copy of the expression cassettes by qPCR. These results matched the initial results, the transcript levels were in some cases even more than 50-fold increased compared to the wild-type promoter [72]. It should be noted that these results are in a somewhat unclear relation to P_{AOX1} based findings. P_{AOX1} and P_{GAP} were considered to provide similar expression levels under the respective optimal growth conditions [68]. Published reports about P_{AOX1} optimization resulted however in a maximal 1.6-fold increase compared to the wild-type promoter [44,45] and not an almost 20-fold increase as in the case of P_{GAP}.

Alternative promoters

Despite the advantages of tightly regulated P_{AOX1} driven expression, the induction with methanol (a toxic and flammable compound) was considered problematic for safety reasons, especially for large-scale production [69]. Furthermore the strong expression from P_{AOX1} and P_{GAP} is not always the most favorable condition for heterologous protein production. If the correct folding of the heterologous protein or processing in the secretory pathway is the limiting step for high amounts of active product, a weaker expression level can be more desirable [3,73,74]. For example in the studies of P_{AOX1} variants, weaker and differently regulated variants showed higher yields of porcine trypsinogen than the wild-type promoter [47].

For these reasons, and also for the purpose of simultaneous expression of multiple different genes, there have also been efforts to identify alternative promoters with different regulatory properties and expression levels (Table 1 for an overview). In general these promoters are by far less well-studied than P_{AOX1}, and hardly any published research has been reported on *cis*-acting sequence elements or *trans*-acting factors involved in their regulation. Some promoters have already been discussed in previous review articles [3,69,75], therefore this section will focus on novel promoters. Information already available will only be shortly summarized and supplemented with new findings.

Promoters, which were already reviewed include P_{FLD1} (formaldehyde dehydrogenase), P_{PEX8} (peroxin 8), P_{YPT1} (a GTPase involved in secretion) [3], P_{AOX2} (alcohol oxidase 2) and P_{ICL1} (isocitrate lyase, ICL) [75].

TABLE 1

Promoters of *P. pastoris* genes used for heterologous protein expression. The approximate expression levels of alternative promoters relative to the classic promoters of *AOX1* and *GAP* (listed at the top of the table) are given if they were provided in the respective reference. Promoters characterized by Stadlmayr *et al*. [93] were tested under different cultivation conditions with multiple reporter proteins, therefore the estimated expression range from weakest to highest is given

Gene name	Gene product	Regulation	Expression level	Reference
AOX1	Alcohol oxidase 1	Induced by methanol	Strong (naturally ~5% of mRNA and ~30% of total protein)	[3,21,27,29]
GAP	Glyceraldehyde 3-phosphate dehydrogenase	Constitutive	Strong (similar to P_{AOX1})	[68]
AOD	Alternative oxidase	Expression on glucose but not on methanol or upon glucose depletion if integrated in natural locus	~40% of P_{GAP}	[71]
AOX2	Alcohol oxidase 2	Induced by methanol	~5–10% of P_{AOX1}	[22,23]
DAS	Dihydroxyacetone synthase	Induced by methanol	Strong (similar to P_{AOX1})	[21,27,29,85]
ENO1	Enolase	Constitutive	~20–70% of P_{GAP}	[93]
FLD1	Formaldehyde dehydrogenase	Induced by methanol and methylamine	Strong (similar to P_{AOX1})	[76]
GPM1	Phosphoglycerate mutase	Constitutive	~15–40% of P_{GAP}	[93]
HSP82	Cytoplasmic chaperone (Hsp90 family)	Constitutive	~10–40% of P_{GAP}	[93]
ICL1	Isocitrate lyase	Derepression and ethanol induction	Not compared to P_{AOX1} or P_{GAP}	[83]
ILV5	Acetohydroxy acid isomeroreductase	Constitutive	~15% of P_{GAP}	[108]
KAR2	ER resident chaperone (also termed Bip)	Constitutive	~10–70% of P_{GAP}	[93]
KEX2	Endopeptidase involved in the processing of secreted proteins	Constitutive	~10% of P_{GAP}	[108]
PET9	ADP/ATP carrier of the inner mitochondrial membrane	Constitutive	~10–1700% of P_{GAP}	[93]
PEX8	Peroxisomal matrix protein	Induced by methanol or oleate	Weak	[3,79]
PGK1	Phosphoglycerate kinase	Constitutive	~10% of P_{GAP}	[92,93]
PHO89 or NSP	Sodium-coupled phosphate symporter	Induced by phosphate limitation	Strong (similar to P_{GAP})	[103,104]
SSA4	Heat shock protein	Constitutive	~10–25% of P_{GAP}	[93]
TEF1	Translation elongation factor 1 alpha	Constitutive and strong growth association	Strong (similar to P_{GAP})	[93,99,100]
THI11	Protein involved in thiamine biosynthesis	Completely repressed by thiamin	~70% of P_{GAP} on medium lacking thiamin	[93,111]
TPI1	Triose phosphate isomerase	Constitutive	~10–80% of P_{GAP}	[93]
YPT1	GTPase involved in secretion	Constitutive	Weak	[3,80]

P_{FLD1} was described to show similar strength as P_{AOX1} and also regulates the expression of a key enzyme of the MUT pathway [76]. However, in contrast to P_{AOX1}, P_{FLD1} is not only regulated by the carbon source but also the nitrogen source. This property was used to develop a methanol-free production process using methylamine for induction even in presence of glucose [77]. P_{FLD1} has been repeatedly used [78] and lately there were also commercial P_{FLD1} based vectors released (Life Technologies, Carlsbad, CA, USA).

P_{PEX8} and P_{YPT1} are weak promoters and considered alternatives in case P_{AOX1} or P_{GAP} driven expression is too strong, however they have only rarely been used [79,80]. Kranthi *et al*. identified five Mxr1p binding sites in the P_{PEX8} sequence [51]. Interestingly, the strong P_{AOX1} promoter contains only one site more, so the number of Mxr1p binding sites seems not to correlate with the promoter strength.

P_{AOX2} is providing about 10–20 times weaker expression than P_{AOX1}, but it is also methanol inducible and shows the same tight regulatory profile as AOX1 [22,23,75]. P_{AOX2} expression was also optimized by adjusting the growth medium [81] and identifying regulatory elements [39,82] (see also section about P_{AOX1} regulation).

P_{ICL1} was used to express a dextranase and shown to be regulated on the transcriptional level by the carbon source by derepression and ethanol induction [83]. However, no comparison between P_{ICL1} and the benchmark promoters P_{AOX1} and P_{GAP} was performed. Therefore the regulatory features, strength and applicability of P_{ICL1} remained unclear [75].

DAS promoter

P_{DAS} is a promoter from the MUT pathway showing a similar regulatory pattern and expression level as P_{AOX1} [27]. In an early patent [29] and a related research article [27], only a single *DAS* gene was identified. In principal, this gene had even been previously reported in the screening of a cDNA library as an

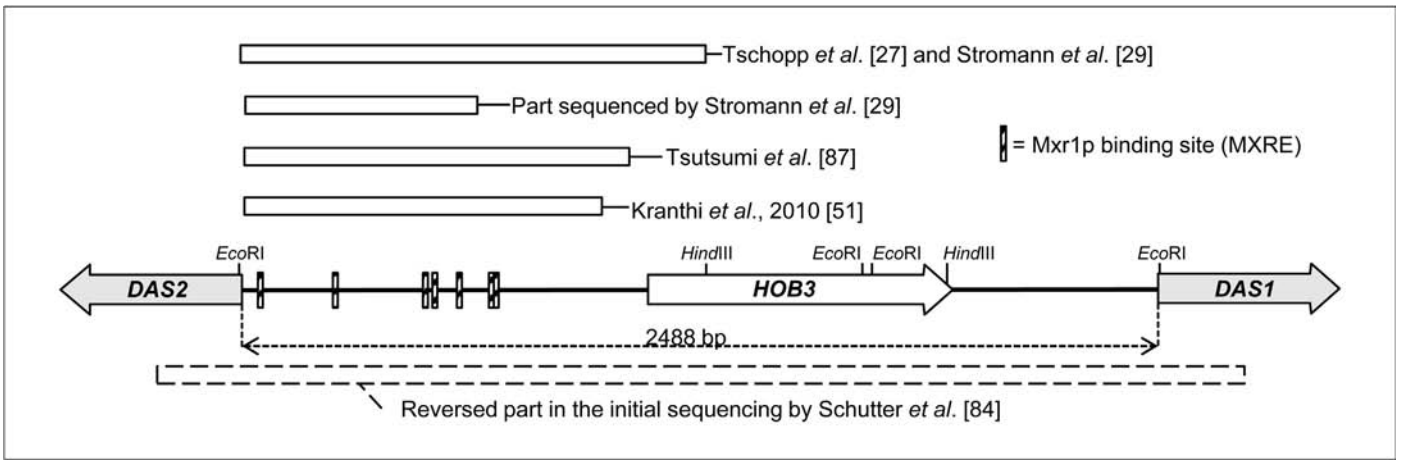


FIGURE 2

Putative promoter regions between the *P. pastoris DAS1* and *DAS2* genes. The annotation of the CBS7435 strain was used [85]. Only the first 500 bp of the 2124 bp long *DAS* genes are shown, and the other elements are drawn to scale. *HOB3*, annotated as a probable guanine nucleotide exchange factor FLJ41603 homolog is also shown. The part that was reversed in the sequencing of the GS115 strain by De Schutter *et al.* [84] is approximately shown, based on Fig. 4 of Küberl *et al.* [85]. The *Eco*RI/*Hind*III fragment described by Tschopp *et al.* [27] and Stroman *et al.* [29] is depicted. Also *Eco*RI and *Hind*III sites in the entire region are shown. The partial fragment sequenced by Stroman *et al.* is also shown. The P_{DAS} sequences used by Tsutsumi *et al.* [87] and Kranthi *et al.* [51] are also highlighted. Furthermore the Mxr1p binding sites identified by Kranthi *et al.* are also shown (the two most downstream binding sites are partially overlapping).

unidentified methanol regulated gene [21], but was not characterized as *DAS* until later [27]. Sequencing of the whole genome revealed, that there are two highly similar *DAS* genes in *P. pastoris*. The two genes are localized in the genome in close proximity to each other in reverse orientation, separated by only 2488 bp (Fig. 2) [84,85]. The two *DAS* genes show a high similarity of 91% [85]. In the first published sequence of the *P. pastoris* GS115 genome [84], the beginnings of the two *DAS* genes and the sequence between them were miss-annotated. Most probably due to the high sequence similarity of the two *DAS* genes the respective part was reversed. In a follow-up sequencing approach of the closely related wild type strain *P. pastoris* CBS7435, this issue was noticed and corrected (GenBank accession nos. CCA39320.1 and CCA39318) [85]. Notably the existence of the second *DAS1* gene was also described previously in a patent application and the related sequence is deposited in GenBank (accession no. FJ752551.1) [86]. Yet the very beginnings of the *DAS* genes and the upstream promoter sequence were also reversed in this sequence in a similar fashion as by De Schutter *et al.* [84]. Interestingly, there is also an additional gene annotated between the two *DAS* genes (*HOB3*, annotated as a probable guanine nucleotide exchange factor FLJ41603 homolog, GenBank accession no. CCA39319.1). The biological significance of this arrangement has however not been investigated. Research on the *DAS* promoter reported so far, has been performed using the promoter sequence of *DAS2*. Tschopp *et al.* used an approximately 1350 bp P_{DAS2} fragment containing an *Eco*RI/*Hind*III fragment of P_{DAS2}, to drive the expression of a lacZ reporter gene [27]. This promoter was also patented and thereby partially sequenced (the respective sequences are shown in Fig. 2) [29]. Kranthi *et al.* identified seven Mxr1p binding sites in P_{DAS2} (Fig. 2) [51]. This is even one site more than in the *AOX1* promoter [48]. However, in contrast to P_{AOX1}, where the Mxr1p binding sites appear to be organized in tandem (Fig. 1), the binding sites in P_{DAS2} are rather dispersed over the whole promoter. There have also been efforts on optimizing P_{DAS} by Tsutsumi *et al.* [87]. Similar to other studies on P_{AOX} they truncated the promoter length, performed internal deletions and added putative activating sequences. These variants were compared by using a phytase as reporter gene. For these studies also the sequence of P_{DAS2} (shown in Fig. 2) was used (GenBank accession no. GM709134). The authors did not mention the existence of two *DAS* genes and refer to the used promoter inconsistently as P_{DAS1}, indicating that they used the wrongly annotated genome sequence of the *P. pastoris* strain GS115. While the promoter sequence is the same in the different genome assemblies and annotations, the *DAS2* gene and not *DAS1* is transcribed from the reported promoter. The variants spanned 2–181% of the wild-type promoter activity and putative UASs were identified. These results were mainly achieved by focusing on an approximately 300 bp long internal fragment and the addition of putative UAS close to the 5′ end. Studies of the putative TF Prm1p also indicated that it is involved in P_{DAS2} regulation [60]. The constitutive overexpression of Prm1p led to P_{DAS2} induction when grown on glucose (see also the section about P_{AOX1} *trans*-acting factors).

In contrast to P_{DAS2}, P_{DAS1} has so far only been little characterized in frame of an engineering study of the *P. pastoris* MUT pathway. Krainer *et al.* noticed that both genes were induced to the same extent upon MeOH addition [88].

PGK1 *promoter*

Phosphoglycerate kinase (*PGK*) is a constitutively expressed house keeping gene involved in glycolysis and gluconeogenesis. The individual native promoters of *PGK* have been widely used for the construction of expression vectors in several yeasts such as *S. cerevisiae* [89], *Yarrowia lipolytica* [90], *Candida maltosa* [91] and others [92]. *P. pastoris PGK1* was isolated by a PCR based approach. Conserved amino acid regions of homologs from other yeasts were identified and subsequently respective degenerate primers were designed. *PGK1* was then analyzed by a sequence-based approach for homologies to other yeasts, conserved domains, its codon

usage and regulatory *cis*-acting sequences in its promoter. The promoter was shown to be regulated moderately by the carbon source, as mRNA levels were two times higher on glucose than glycerol. P_{PGK1} was used to express an alpha-amylase but expression levels were only measured semiquantitatively and only in rough relation to P_{AOX1} [92]. Stadlmayr *et al*. characterized multiple novel *P. pastoris* promoters with three different reporter proteins [93]. In these studies P_{PGK1} showed consistently only 10% of P_{GAP} driven expression, indicating that P_{PGK1} is rather weak. Also deletion studies were performed. However, possible regulatory sequences have not been discussed and surprisingly a promoter variant which was truncated to only 250 bp was long enough to show full promoter strength [94].

TEF1 *promoter*

Translation elongation factor 1 alpha (Tef1) is a crucial component of the eukaryotic translation machinery and mediates the delivery of aminoacyl tRNAs to the ribosomes to sustain the elongation of the peptide chain [95]. The *TEF1* promoters from various fungi *Ashbya gossypii* [96], *Aspergillus oryzae* [97] and *Y. lipolytica* [98] show a strong constitutive expression suitable for heterologous protein production. The *P. pastoris TEF1* homolog was isolated by the use of degenerate primers of conserved regions and its promoter was studied in respect to regulation and expression levels [99,100]. Northern blots revealed a growth associated expression pattern of P_{TEF1}, resulting in high P_{TEF1} mRNA levels during exponential growth and clearly reduced levels in the stationary phase. P_{TEF1} driven expression was compared to P_{GAP} by using a lipase fused to a cellulose binding domain as reporter gene. The total activities measured from batch cultivation were similar for both promoters, however P_{TEF1} expression started earlier than P_{GAP} and showed a tighter coupling to the growth rate. In carbon-limited fed-batch cultivations the activity levels of P_{TEF1} driven expression exceeded P_{GAP} almost twofold. In studies by Stadlmayr *et al*., P_{TEF1} was also found to be a strong promoter, but depending on the reporter protein and the cultivation time, the expression was varying between a similar strength as P_{GAP} or less [93]. Similar to the GAP promoter no specific *cis*- or *trans*-acting elements have been described for this *P.pastoris* promoter.

AOD *promoter*

Alternative respiration is a shortcut of the respiratory pathway that skips certain steps of the respiratory electron transfer chain [101,102]. Also in *P. pastoris*, an alternative oxidase (Aod) was identified and its significance for cell growth and viability was investigated [71]. In course of this study, the expression level of a fusion protein of the alternative oxidase and GFP was tested under control of the *AOD* promoter and the *GAP* promoter. Interestingly, P_{AOD} showed in the presence of glucose approximately the same fluorescence levels as P_{GAP}. However, expression from P_{AOD} started with a delay of about two to three hours. After reaching a peak at glucose depletion, the fluorescence levels of P_{AOD} driven expression decreased steadily to zero. By contrast, P_{GAP} fluorescence continuously increased and reached about twice the fluorescence of the P_{AOD} peak value. The authors argue that Aod is an unstable protein. This would suggest protein degradation as the cause for the fluorescence decrease in the P_{AOD} driven expression of the Aod-GFP fusion protein. A possible explanation would be that expression from P_{AOD} ceases after glucose depletion, so the previously produced Aod-GFP fusion proteins get degraded. The constitutive *GAP* promoter would by contrast continue expression even upon glucose depletion, leading to the further fluorescence increase seen in the experiments. However, this explanation would require further proof by testing P_{AOD} and P_{GAP} with a stable reporter gene such as GFP alone. Interestingly P_{AOD} expression was also not induced by methanol, hinting a carbon source dependent regulation. The information currently available suggests that P_{AOD} is a strong promoter, reaching similar expression levels as P_{GAP} on glucose, but facing a repression or missing induction upon glucose depletion or presence of methanol [71]. On the one hand, these properties rather limit the achievable expression levels, as P_{AOD} driven expression stopped at about half of the maximal expression level of P_{GAP}. On the other hand this alternative regulatory profile might also be beneficial for the expression of certain proteins, as also some synthetic P_{AOX1} variants with lower expression levels and altered regulatory profiles led to higher yields than the wild-type promoter [47].

PHO89 promoter

P. pastoris PHO89 (also described as NPS promoter) was identified during a search for genes upregulated under phosphate-limited conditions and codes for a putative sodium-coupled phosphate symporter [103,104]. Phosphate-responsive promoters have been studied in other organisms as *E. coli* [105] and *S. cerevisiae* [106] and shown to be highly induced under phosphate-limited growth conditions. In an unrelated study, the *P. pastoris* acid phosphatase (*PHO1*) showed ~100-fold increased activity under phosphate limitation [107]. However, the transcriptional regulation of *PHO1* or the applicability for protein expression was not tested.

The putative promoter of *PHO89* showed a strong regulation by the phosphate concentration in the growth medium [103] by applying the same lipase reporter enzyme used for the studies of P_{TEF1} [99]. Under high phosphate concentrations no lipase activity was measured. Under reduced phosphate concentrations, similar activity levels as with P_{TEF1} or P_{GAP} were achieved. Notably the specific productivity (lipase activity/biomass/time) of the P_{PHO89} based expression process was 7.4 or 14.8 times higher than P_{TEF1} or P_{GAP}. The main reason was that the P_{PHO89} reached the maximal activity earlier and there was also less biomass formed than in the not phosphate-limited cultures of P_{TEF1} or P_{GAP}. The volumetric lipase activity from P_{PHO89} was however the same as P_{TEF1} and approximately twice of P_{GAP}. However, this combination of the reporter enzyme and growth parameters might not be representative for all conditions. P_{TEF1} driven expression was in similar experiments also twice as high as P_{GAP} [99], but with different reporter enzymes and cultivation conditions only about the same or less than P_{GAP} [93]. Hence, the patented P_{PHO89} driven expression appears to be a promising methanol-free system showing similar expression levels as P_{GAP}, but the maximum achievable expression levels remain to be verified with different reporter enzymes. Concerning *cis*-acting regulatory elements, the 1044 bp long promoter was also truncated, showing that a basic fragment of approximately 300 bp is able to mediate phosphate-responsive expression [104]. Furthermore, the authors also identified two putative binding sites for the TF Pho4p, which indicates that *PHO89* is regulated by the *PHO* regulatory system [103].

Other promoters which are less well studied include the *ILV5* and *KEX2* promoters of *P. pastoris* [108]. An initial characterization of the *ILV5* promoter showed that significant constitutive activity can be obtained using an approximately 1 kbp long region upstream of the *ILV5* gene. Also a truncated version of approximately 0.5 kbp was at least strong enough to drive the expression of selection marker genes in a new series of *P. pastoris* expression plasmids [109].

Postgenomic promoter studies

So far, all mentioned promoter studies were based on the isolation of the promoters by classic methods of genetics and molecular biology. This involved for example the search for homologs in other yeasts, then the design of degenerate primers binding to conserved regions and eventually the PCR amplification of a part of the CDS using these primers. Subsequently, the flanking region containing the promoter must be identified by inverse PCR [92,99,110]. Alternatively, to identify inducible promoters, mRNAs upregulated under certain growth conditions can also act as the starting point for the promoter search. This approach requires reverse transcription to cDNA and also the identification of the flanking regions [21,103]. However, sequencing of the *P. pastoris* genome and the first microarray experiments enabled a much simpler approach, as the promoters did not have to be isolated tediously by classic methods. Stadylmayr *et al*. identified 24 promising promoters from micro array data or by a literature search [93]. These promoters have also been reported as part of a patent [111]. The microarray data have been obtained by probing *S. cerevisiae* DNA microarrays with *P. pastoris* cDNA [112]. These heterologous microarrays were initially performed because *P. pastoris* DNA microarrays were not available at the time of this study. On the basis of the hybridization data of multiple cultivation conditions, the promoters of the 15 genes showing the highest expression levels were used for comparative studies [93]. Additionally nine promoters were rationally selected by searching the literature for strong and/or tightly regulated promoters from other yeasts.

Subsequently, the promoter sequences were obtained from a genome database (Integrated Genomics), then suitable primers were designed and the promoters were PCR-amplified from genomic DNA. All promoters were combined with three different reporter proteins including two intracellular proteins EGFP (enhanced GFP) and beta-galactosidase (lacZ) and the secreted protein HSA. Strains bearing the respective expression constructs were grown on different carbon sources (glucose, glycerol and methanol) and reporter levels were determined at multiple timepoints. Notably also P_{AOX1} and P_{GAP} were included as references for strong methanol inducible and constitutive expression. In these experiments, 12 out of the 24 novel promoters showed constitutive expression under standard growth conditions in a full medium containing glucose. Few other promoters showed weak activity on glycerol or methanol, but the expression levels were less than 10% of P_{GAP} driven expression (promoters showing >10% activity of P_{GAP} are listed in Table 1). Notable constitutive promoters stem from glycolytic genes as *ENO1* or *TPI1* and genes involved in stress response like *HSP82*, *KAR2* and *SSA4*. Depending on growth conditions and the reporter used, these promoters were giving similar or weaker expression than P_{GAP}. None of the new promoters were convincingly under multiple tested conditions stronger than P_{GAP}. Some novel promoters showed different regulatory profiles, like a stronger induction in the early or late growth phase, but reached similar total expression levels. The promoter of *THI11*, a gene involved in thiamin biosynthesis, showed the most remarkable regulation pattern. It was completely repressed on medium containing thiamin, but reached 70% of P_{GAP} on medium lacking thiamin. Therefore P_{THI11} driven expression could be used to set up a methanol-free production process, in which the starting point for expression could be simply set by the initial amount of thiamin in the medium.

An important issue about the study by Stadlmayr *et al*. is the copy number of the strains bearing the different promoter constructs. Multiple copies of a weak promoter can lead to the same expression as a single copy of a strong promoter [47], so this effect can strongly bias the results.

Stadlmayr *et al*. determined the copy numbers for four promoters by checking multiple transformants made with the same promoter construct. This revealed that 52% of the clones were single copy strains, 19% contained two copies, 13% three copies and the rest more than three copies. These results indicate a relative broad diversity. However as each promoter was tested with three different reporters and ten clones per reporter were tested, possible copy number effects might be compensated by the multiple analyses.

Indeed, the results of the three different reporter proteins were in general in accordance to each other, but in one case there was a drastic difference. P_{PET9} showed with EGFP a 10- to 17-fold higher fluorescence than P_{GAP}. Also the transcript levels were drastically increased, as shown by RT-qPCR measurements. This high expression was however not reproducible with the two other reporter enzymes HSA and lacZ with reporter activities of about only 10% of P_{GAP}. The authors did not find an explanation, but this example shows that the combination of promoter and protein of interest can have a strong influence on the expression levels. This observation is further underlined by a comparison of previous studies on P_{PGK1} [92] and P_{TEF1} [99] to the work by Stadlmayr *et al*. [93]. As mentioned above, P_{PGK1} was initially only assessed semiquantitatively and in approximate comparison to P_{AOX1} [92]. The thorough study by Stadlmayr *et al*. showed however clearly that P_{PGK1} reaches only about 10% of P_{GAP} driven expression. [93]. Similarly, P_{TEF1} driven expression was at first found to match or even twofold exceed P_{GAP} driven expression [99]. In the work by Stadlmayr *et al*., P_{TEF1} driven expression never outperformed P_{GAP} and showed rather similar or lower expression levels [99].

Important transcriptional regulatory circuits

Introduction

Beside these sequence-based studies, there is a growing interest for regulatory networks and the TFs involved in the process of gene expression and protein folding. For example TFs involved in such diverse cellular processes as protein secretion (Hac1p), iron uptake (Fep1p), oxidative stress response (Yap1p) and biotin auxotrophy (Rop1p) have been studied in addition to the already discussed TF Mxr1p (Fig. 3 for an overview). The knowledge about the regulatory mechanisms enables TF-based engineering approaches and the design of new (synthetic) regulatory circuits to improve heterologous protein production processes.

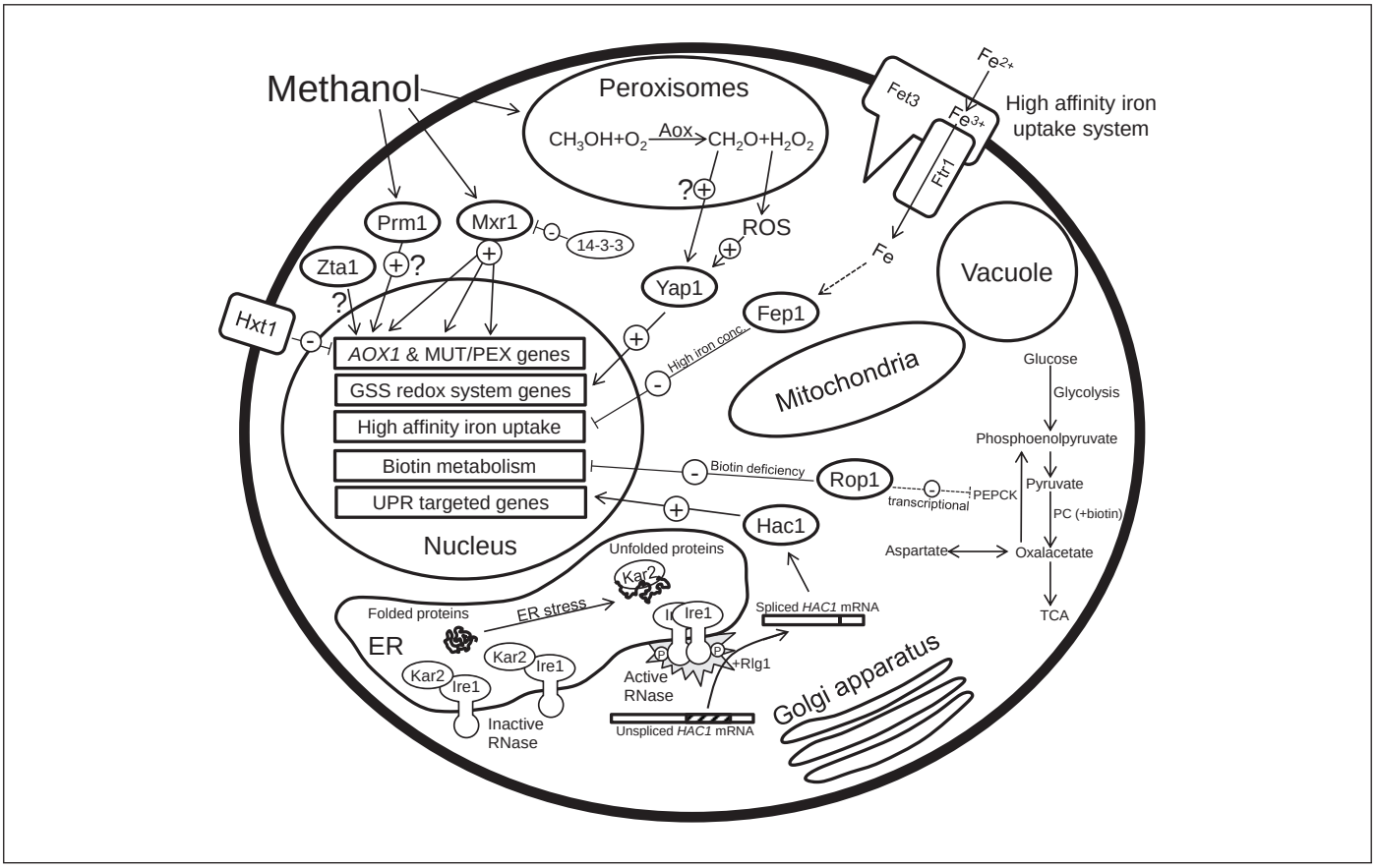


FIGURE 3

Transcription factors and regulatory circuits involved in various cellular processes relevant for heterologous protein production in *P. pastoris*. The regulation of the *AOX1* promoter by the *trans*-acting factors Mxr1p, Prm1p, Zta1p, Hxt1p and a 14-3-3 protein are shown. The oxidative stress response by Yap1p involved in ROS detoxification and methanol metabolism is depicted. Also regulation of the UPR by Hac1p (by splicing), regulation of the high-affinity iron uptake machinery by Fep1p and the role of Rop1p in biotin metabolism are shown. Positive regulatory effects are shown by a plus in a circle over the respective arrow. Negative regulatory effects are indicated by a minus in a circle and a truncated arrow. The schematic of the high-affinity iron uptake system is derived from Van Ho *et al.* [141], the schematic of the Rop1p regulated biochemical reactions are based on Kumar and Rangarajan [166] and the Hac1p regulation is based on Guerfal *et al.* [125] and Ma and Hendershot [132].

Regulation of the unfolded protein response by Hac1p

One of the main advantages of *P. pastoris* is that high titers of relatively pure secreted heterologous protein can be obtained on lab as well as on a large scale [113].

Gaining understanding of the secretory pathway is crucial for the identification of bottlenecks and improving the secretion of heterologous proteins further [114], for example, to obtain similar high yields as with filamentous fungi. Shortly summarized, secretory proteins need a signal sequence to enter the secretory pathway and are translated at the rough endoplasmic reticulum. Then they are transported from the ER lumen via the Golgi apparatus and secretory granules which fuse eventually with the plasma membrane [115]. The first step of translation at the rER, folding inside the ER and the introduction of posttranslational modifications are crucial steps in the standard secretory pathway. Unfolded or misfolded proteins in the ER cause cellular stress and activate the UPR pathway. The UPR upregulates chaperones to counteract the folding problems, but can also stimulate endoplasmic-reticulum-associated protein degradation (ERAD) or lead to programmed cell death (apoptosis) [116,117]. The overexpression of heterologous proteins is prone to activate the UPR, as the cellular machinery is not necessarily prepared to deal with the processing and folding of large amounts of foreign proteins [73]. Therefore the UPR is an important factor to tweak protein secretion and has been studied extensively in *P. pastoris*. Initially the effect of the copy number, promoter and carbon source on the UPR were tested [74]. Furthermore the transcriptional regulation of genes targeted by the UPR was tested by TRAC (transcript analysis with aid of affinity capture) [118], a magnetic bead-based sandwich hybridization system [119] and microarrays [120]. The UPR was also found to participate in providing osmotolerance [121,122], to depend on the stability of the secreted protein [123] and to affect the cellular redox state [124].

However, the exact transcriptional regulation of the UPR in *P. pastoris* and the factors involved has only recently been characterized [125,126]. The general model of the transcriptional regulation of the UPR in yeast was at first elucidated in *S. cerevisiae* and found to be mediated by the TF Hac1p. Hac1p activates genes of the UPR and underlies a distinctive post-transcriptional regulation mechanism [127]. The *HAC1* mRNA contains an intron that blocks translation and needs to be spliced to permit translation of the functional protein [128]. Splicing of the *HAC1* mRNA is not accomplished conventionally by the spliceosome, but is linked by a unique mechanism to folding stress in the ER [129]. Namely

Ire1p, an ER resident transmembrane kinase/endoribonuclease, Kar2p (also named Bip), an ER-chaperon, and Rlg1p, a tRNA ligase, participate in *HAC1* mRNA splicing. Kar2p is involved in the regulation of Ire1p by binding to it under nonstress conditions [130,131]. If un- or misfolded proteins accumulate in the ER, they are bound by Kar2p, which assists proper folding. Thereby Kar2p releases Ire1p, which forms oligomers and is thereby autophosphorylated. Phosphorylated Ire1p activates its endoribonuclease activity and cleaves the intron from the *HAC1* mRNA. The mRNA is then religated by Rlg1p and functional Hac1p can be translated, migrate to the nucleus and bind to UPR responsive elements (UPRE) in the promoters of UPR target genes (Fig. 3 for a schematic representation) [117,132]. Notably the *HAC1* mRNA must also be transported from the nucleus to the Ire1p oligomers at the ER, to permit splicing [133].

The *P. pastoris HAC1* homolog was characterized recently and also found to contain the conserved intron known from *S. cerevisiae* and other fungi [125,126]. Interestingly, the significance of the splicing mechanism was somewhat inconsistent between two research groups. Guerfal *et al*. identified potential splice sites by sequence similarities with *S. cerevisiae* [125]. However, under various tested growth conditions no unspliced *HAC1* mRNA was detected, indicating a constitutive splicing of the *HAC1* mRNA in *P. pastoris*. The *HAC1* mRNA level increased strongly under conditions of folding stress, which would hint rather a transcriptional regulation mechanism of *HAC1* in *P. pastoris* [125]. By contrast, Whyteside *et al*. clearly detected the unspliced *HAC1* mRNA under unstressed conditions [126]. Under conditions of folding stress, the spliced form showed up, but the unspliced form was still detectable. Notably, both groups used the same *P. pastoris* same strain (GS115) and also the same agent in the same concentration to cause the folding stress, namely 5 mM dithiothreitol (DTT). Therefore the differences could possibly arise from different cultivation conditions and would have to be further tested to clarify the splicing of *HAC1* mRNA in *P. pastoris*.

Guerfal *et al*. tested furthermore the effect of *HAC1* co-expression on heterologous protein production [125]. There have been earlier attempts to co-express *HAC1* and the homolog of *S. cerevisiae* (Sc*HAC1*) in *P. pastoris* [118,119,134]. Constitutive expression of Pp*HAC1* showed little to no effect compared to the not co-expressing strain [125], while inducible co-expression using P_{AOX1} resulted in an up to twofold increased expression of surface displayed proteins, but in one case also, but in one case also a complete loss of the initial expression. Also the amount of correctly processed and folded membrane protein was increased.

In conclusion, characterization of Pp*HAC1* has shed light on the UPR in *P. pastoris* and engineering of the UPR by *HAC1* co-expression has led to increased yields and improved quality of heterologous proteins. Still, the exact circumstances of the *HAC1* mRNA splicing reaction in *P. pastoris* remain to be clarified. Also a general model for predicting the effect of Pp*HAC1* overexpression on the secretion efficiency of a certain heterologous protein would be favorable.

Regulation of iron uptake by Fep1p

Iron is a crucial cofactor of many proteins, as part of prosthetic groups like heme, iron–sulfur clusters or in nonheme iron proteins. Many biotechnologically relevant enzymes like peroxidases and monooxygenases contain heme and the optimization of heme biosynthesis or addition of heme to the cultivations could significantly increase the yields of active enzyme [135,136]. Therefore the understanding of heme synthesis and iron uptake can be an important field for improvements in the production of heme or nonheme iron containing proteins and influence the general energy metabolism and productivity of the host. High-affinity iron uptake in *P. pastoris* has been found to involve similar machinery as in *S. cerevisiae* [137–139]. There are two main factors involved, the plasma membrane multicopper oxidase Fet3p and the iron permease Ftr1p [140,141]. Fet3p and Ftr1p interact and from a complex, in which Fet3p oxidizes Fe^{2+} to Fe^{3+}, which can subsequently enter the cell via Ftr1p (Fig. 3 for a schematic overview). Depending on the availability of iron, the amounts of proteins involved in iron uptake and iron-dependent enzymes have to be controlled. In *S. cerevisiae* under iron scarcity, the high-affinity iron uptake machinery is upregulated [142–144], whereas at the same time entire pathways involving iron-dependent enzymes are repressed [145]. This regulation is exerted mainly on the transcriptional level, although also posttranscriptional mechanisms have been described [146,147].

In *P. pastoris*, Fet3p and Ftr1p mRNA levels increase under iron scarcity indicating a transcriptional regulation [138,139]. Miele *et al*. identified a TF, named Fep1 (Fe protein 1) that acts under iron abundance as a repressor of genes involved in high-affinity iron uptake [148]. The promoter sequence of Fet3p was analyzed for consensus sequences for GATA-type TFs, as these TFs are frequently involved in iron-dependent transcriptional regulation in other fungi [149–151]. Indeed, a putative GATA sequence element was found and mutated to test its regulatory role [148]. This P_{FET3} mutant affected the iron-dependent expression of a lacZ reporter gene. In detail, the wild type strain showed an increase of lacZ activity levels under iron scarcity, whereas the P_{FET3} mutant showed independently of the iron availability the same low lacZ activity. This supported the involvement of a GATA type TF in iron regulation in *P. pastoris*. On the basis of conserved regions of iron-responsive GATA factors from other fungi, degenerate primers were designed and a part of the putative *FEP1* gene PCR amplified. A sequence comparison showed that Fep1p contains conserved Zn finger DNA-binding domains and a conserved motif of four cysteines that is typical of iron-responsive GATA TFs. A strain disrupted in *fep1* showed a constitutive expression of genes involved in high-affinity iron uptake independent of the iron concentrations in the medium. The wild type strain showed however strongly reduced expression levels of these genes under iron abundance. This indicates a repressing role of Fep1p, which is in accordance with the repressing function of Fep1p homologs from other fungi [149–151]. The authors showed furthermore that the DNA binding of Fep1p is dependent on the iron concentration [148]. In EMSA experiments, recombinantly produced Fep1p bound only in the presence of iron to the GATA sequence motif from the *FET3* promoter. This further supports a model in which Fep1p represses genes for high iron affinity iron uptake under conditions of iron abundance.

Interestingly, despite the conserved machinery for high-affinity iron uptake between *P. pastoris* and *S. cerevisiae*, the transcriptional regulation mechanisms are different between these two organisms. In *P. pastoris* the genes are repressed by Fep1p

under conditions of iron abundance, whereas the same genes are activated in *S. cerevisiae* under iron starvation. *P. pastoris* is a respiratory yeast, whereas *S. cerevisiae* is a fermentative yeast, therefore the different response to iron could be a consequence of the different metabolic setups. This consideration is supported by the fact that also the respiratory yeast *S. pombe*, which is evolutionary only distantly related to *P. pastoris*, exhibits repression of high-affinity iron uptake genes in a similar Fep1p dependent fashion [148].

Oxidative stress response by Yap1p

Oxidative stress caused by reactive oxygen species (ROS) is a major factor of cellular stress and damages all cellular components such as proteins, nucleic acids and lipids [152–154]. ROS are caused by the high reactivity of oxygen and the nonspecific electron transfer by redox enzymes to unintended substrates. Consequently respiratory metabolism and the mitochondrial electron transport chain are especially notable sources of ROS in eukaryotes [155]. ROS is a general term not only referring to molecules such as H_2O_2 but also free radicals like the superoxide radical ($O2^{\bullet-}$) or the hydroxyl radical ($OH^{\bullet}$) [156].

Cells contain a broad scope of defense mechanisms to counteract oxidative stress and ROS formation. This includes on the one hand antioxidants such as L-ascorbic acid (vitamin C), α-tocopherol (vitamin E), thioredoxin and glutathione (GSH). On the other hand this also includes enzymes like superoxide dismutase, catalase, peroxiredoxins and glutathione peroxidase protect cells against ROS [154].

Oxidative stress is also a relevant issue for recombinant protein production, as it decreases cell viability and triggers apoptosis [73]. Furthermore, ROS is involved in the UPR [124,157] and could therefore play a crucial role for stress associated with the industrially important protein secretion processes. ROS production is also an inevitable part of the methanol metabolism in methylotrophic yeasts, as the oxidation of methanol to formaldehyde by alcohol oxidase creates H_2O_2 as a side product (Fig. 3) [15]. Therefore the ROS detoxification plays a crucial role in the growth of methylotrophic yeasts and is tightly regulated.

In *P. pastoris*, the glutathione redox system was studied and found to be a crucial factor to permit growth on methanol [158,159]. Glutathione (γ-L-glutamyl-L-cysteinylglycine) is a nonribosomal peptide and protects via its low redox potential against ROS [160]. In the process the reduced form of glutathione (GSH) is oxidized and forms a self-dimer via a disulfide bond between thiol groups of cysteine residues (GSSG). In *P. pastoris* oxidized glutathione is regenerated by reduction by a NADPH dependent glutathione reductase encoded by *GLR1* [158]. Other crucial genes involved in the glutathione redox system include glutathione peroxidase (*GPX1*), peroxiredoxin (*PMP20*), that act as peroxisomal glutathione peroxidase, and the enzyme catalyzing the rate-limiting step of glutathione synthesis (*GSH1*.) Yano *et al*. showed that the glutathione redox system in *P. pastoris* is transcriptionally regulated by the TF Yap1p, which acts at the same time as a sensor for oxidative stress via the oxidation of conserved cysteine residues [158,159]. The *S. cerevisiae* Yap1p homolog has already been well characterized and shown to upregulate various genes involved in ROS defense [161,162]. ScYap1p contains a nuclear localization sequence, but is under normal growth conditions constantly exported to the cytoplasm by a nuclear exportin. Under conditions of oxidative stress cysteine residues of ScYap1p are oxidized and form a disulfide bond. This causes a conformational change and ScYap1p is no longer recognized by the export machinery [163]. *P. pastoris* Yap1p was found to play a similar role in the oxidative stress response during the methanol metabolism. A *yap1* deletion strain did not show any phenotypical changes when cultivated under standard conditions on glucose, but growth on methanol caused a prolonged lag phase similar to a *gpx1* deletion strain [158]. The *yap1* deletion strain was also very susceptible to various ROS generating agents when grown on glucose [159]. Furthermore, PpYap1p showed a similar localization pattern as ScYap1p and localized to the nucleus under growth on methanol [158] and conditions of oxidative stress [159]. Consistently, the transcription of genes involved in the glutathione redox system, most notably *GPX1*, were upregulated under growth on methanol in the wild type strain, but not in the *yap1* deletion strain [158]. Similar to ScYap1p, PpYap1p is sensing oxidative stress via the oxidation state of conserved cysteine residues, which was tested by mutating the respective amino acid positions [159].

In contrast to these similarities to ScYap1p, PpYap1p was shown to play a special role in the methanol metabolism, namely the detoxification of formaldehyde by the Yap1p regulated GSH redox system [158]. GSH detoxifies formaldehyde by the spontaneous reaction to S-hydroxymethylglutathione [15]. Consistently, it was shown that cellular GSH pools increased in cells grown on methanol and also regeneration of GSSG to GSH was crucial. Namely, deletion of *GLR1*, responsible for GSH regeneration, led to hypersensitivity to methanol but also formaldehyde underlining the importance of the GSH redox system for formaldehyde detoxification. So besides the PpYap1p activation via the oxidation of conserved cysteine residues mentioned above [159], an activation by formaldehyde or GSH depletion was also suggested [158]. In conclusion this work showed the crucial role of PpYap1p for the MUT pathway and revealed a new role of Yap1p, as Yap1p homologs in other yeasts regulate only a stress response, whereas it is in *P. pastoris* crucial to sustain the metabolism of the carbon source methanol. Furthermore, overexpression or engineering of Yap1p might be a promising target to counteract oxidative stress arising from heterologous protein production.

Involvement of Rop1p in biotin metabolism

Biotin (vitamin H) is required by many carboxylases as an essential prosthetic group, but not all organisms are capable of a *de novo* biotin synthesis [164]. Notably, *P. pastoris* and also certain laboratory strains of *S. cerevisiae* are among the biotin auxotroph organisms [165]. Gasser *et al*. have engineered a biotin-prototrophic *P. pastoris* strain by reconstituting the biotin biosynthesis pathway by the expression of the respective genes from *S. cerevisiae* [165]. Kumar and Rangarajan have recently characterized a TF involved in the regulation of biotin metabolism related genes [166]. Initially they searched for proteins similar to Mxr1p, a key TF of MUT genes (see section about P_{AOX1} *trans*-acting factors), and identified a protein, later termed Rop, that shared 58% identity to the zinc finger domain of Mxr1p (in this review Rop is referred to as Rop1 to meet standard yeast naming conventions and to avoid confusion when referring to the protein). *ROP1* mRNA was only detectable under growth on methanol, but not on glucose, glycerol or oleic

acid suggesting a possible involvement in the regulation of MUT genes. However, this role could not be experimentally verified, as the deletion of *ROP1* neither affected the levels of the mRNAs of MUT genes nor the growth rate compared to the wild type on methanol. The authors compared the transcriptomes of the *rop1* deletion strain with the wild type by microarray and noticed an upregulation of genes involved in the biotin metabolism (e.g. biotin synthase, a biotin transporter and a biotin–apoprotein ligase) in the *rop1* deletion strain. This hinted a repression of biotin metabolism genes by Rop1p. Additionally, the *rop1* deletion strain was in contrast to the wild type able to grow on biotin deficient medium. The authors tried then to further elucidate the regulatory function of Rop1p and to find a biochemical explanation how the *rop1* deletion strain could grow under biotin deficient conditions [166]. This requires understanding of the role of biotin in the cellular metabolism (Fig. 3). One of the major biotin dependent enzymes is pyruvate carboxylase (PC), which converts pyruvate, the final product of glycolysis, to oxalaecetate (OAA), a key intermediate of the tricarboxylic acid cycle (TCA). PC has been studied in *P. pastoris* by the identification of the gene and suppressor mutants [167], furthermore PC has also been shown to be involved in the assembly of Aox oligomers [168,169]. A PC deficient strain, caused by the lack of biotin or mutagenesis of PC, cannot replenish the OAA pool which stalls the TCA. However, the PC deficiency can be rescued by the addition of aspartate, which can be converted to OAA by transamination [170,171]. In case of *P. pastoris*, addition of aspartate to biotin deficient medium did restore the growth of the wild type strain, suggesting that the biotin dependent growth defect is caused by a lack of PC-dependent OAA synthesis [166]. The growth of the *rop1* deletion strain under biotin deficient conditions is therefore most probably permitted by a PC-independent pathway of OAA synthesis (PC bypass). These bypass reactions include ICL and malate synthase (MS) from the glyoxylate cycle and also phosphoenolpyruvate carboxykinase (PEPCK) [170,171]. In *P. pastoris* ICL and MS were expressed in the wild type and the *rop1* deletion strain grown under biotin deficient conditions. PEPCK was however only expressed in the *rop1* deletion strain and repressed in the wild type under biotin deficient conditions [166]. Rop1p is therefore involved in the repression of PEPCK under biotin- (and PC-) deficient growth conditions, therefore the name 'repressor of PEPCK' (ROP) was also chosen.

The authors also argued about putative binding sequences of Rop1p based on homologies of the zinc finger DNA-binding domain to Mxr1p. Furthermore, the transcriptional regulation exerted by Rop1p was additionally tested by comparing the *rop1* deletion strain and the wild type in a microarray experiment under biotin deficient conditions. The expression of several genes was clearly affected, but their exact physiological roles remain to be elucidated. It is also notable that the authors argue about the exact function of PEPCK sustaining growth on biotin deficient medium. PEPCK usually catalyzes the conversion of OAA to phosphoenolpyruvate (PEP), which is the first reaction of gluconeogenesis. However, the reaction can take place in the reverse direction and lead thereby to OAA and act as a PC bypass [170,171]. Yet, the glyoxylate cylcle genes ICL and MS can also provide OAA in *P. pastoris*. Therefore PEPCK could also be growth limiting because of lack of PEP formation for gluconeogenesis and not because of bypassing the biotin dependent PC reaction [166]. However, clarifying the exact role of PEPCK and thereby *ROP1* under biotin starvation will require further studies. Still, Kumar *et al*. found the first indication that a TF is involved in repression of PEPCK under biotin-deficiency and although not being the main focus of the study, they showed that a *rop1* deletion can grow under biotin deficient conditions. However, they did not compare the growth rate of the *rop1* deletion strain under biotin deficient conditions with the growth rate of the wild type on biotin. Therefore the biotechnological potential of using a *rop1* deletion strain for biotin free production processes remains unclear so far.

Conclusions

Recent research on promoters and the first partial knowledge about some transcriptional regulatory circuits in *P. pastoris* has shed light on the underling molecular mechanisms and on how to improve heterologous protein production. The regulation of P_{AOX1}, the most commonly used promoter in *P. pastoris*, was studied in respect to *cis*-acting sequence elements and also *trans*-acting factors. Although its exact regulation is still unknown, at least various independent sequence-based approaches have identified putative regulatory sequences and TFBS. In part, the predicted putative TFBS were verified by the characterization of *trans*-acting factors and their recognition sites. Namely, two activators Mxr1p and Prm1p were identified. Mxr1p has been well characterized and shown to regulate MUT and peroxisome biogenesis genes. Furthermore the DNA-binding sites of Mxr1p have been identified and also a 14-3-3 protein interacting with Mxr1p has been described and shown to repress Mxr1p activity especially when grown on ethanol. However these findings indicate that there exist also additional, so far uncharacterized regulators of Mxr1p, that act in addition to the 14-3-3 protein on other carbon sources. Also other aspects of P_{AOX1} regulation are only partially understood. Namely the regulatory mechanism of Prm1p or the exact regulatory functions of hexose transporters and Zta1p remain to be elucidated. This indicates a highly complex regulation pattern of P_{AOX1} and other MUT genes, that involves multiple activating and repressing factors. Still, especially the sequence-based approaches have helped to further improve the already exceptionally strong *AOX1* promoter and also variants with altered regulatory profiles for applications in heterologous protein production were identified.

Beside this progress in understanding and improving P_{AOX1}, alternative promoters with different modes of regulation have also been identified. P_{AOX1} is providing extremely strong, controllable expression, but requires methanol, which is toxic and flammable, for induction. As an alternative to derepressed P_{AOX1} variants which do not rely on methanol anymore and can be simply regulated by changed carbon source feeding, especially strong and inducible promoters like P_{PHO89} or P_{THI11} could also become significant alternatives to P_{AOX1}, as they are regulated by the phosphate and thiamin concentration, respectively.

Furthermore, this review also covered TFs and their regulatory modes on diverse cellular functions such as UPR (Hac1p), iron uptake (Fep1p), oxidative stress response (Yap1p) and biotin auxotrophy (Rop1p). Overexpression of *HAC1* led in some cases to increased levels of secreted proteins and also elevated levels of correctly folded processed proteins. Yet, this effect was protein

dependent and also the regulation of *HAC1*, namely the significance of the splicing reaction, in *P. pastoris* remains to be clarified. Also, the regulatory circuits of Fep1p, Yap1p and Rop1p bear the potential to engineer the expression system *P. pastoris*, but have not been practically tested to improve bioprocesses so far. Namely, engineering of Fep1p mediated iron uptake could help to tweak the production of iron containing proteins. The regulatory role of Rop1p has hinted a bypass of biotin auxotrophy, which could help to design biotin independent *P. pastoris* strains. Oxidative stress is a known problem associated with heterologous protein production and arises especially from growth on methanol. Therefore improving the oxidative stress response by Yap1p could be an approach to design more robust production strains.

Although *P. pastoris* has only been used since approximately two decades by a broad public, significant progress has been made on characterizing key transcriptional circuits. The majority of molecular mechanisms investigated up to now, are still far from being completely understood. New technologies in nucleic acid sequencing, bioinformatics and high throughput analysis will hopefully provide a profound basis and stimulate further research on transcriptional regulation in *P. pastoris*.

Note added in proof

Just after this review was written, two research articles describing additional *trans*-acting factors involved in *P. pastoris* P_{AOX1} regulation have been published:

- Polupanov *et al*. investigated the role of a glucose sensor (termed GSS1) in autophagic degradation of peroxisomes and catabolite repression (Polupanov AS, Nazarko VY, Sibirny AA. Gss1 protein of the methylotrophic yeast Pichia pastoris is involved in glucose sensing, pexophagy and catabolite repression. Int. J. Biochem. Cell Biol. 2012;44:1906–18). A strain with a deleted *gss1* gene exhibited Aox activity on glucose containing medium suggesting abolished glucose catabolite repression. Interestingly, ethanol catabolite repression was not affected in the *gss1* deletion strain. This supports previous findings that glucose and ethanol repression are exerted by different mechanisms (as discussed for the 14-3-3 protein [52] in the section ''*Trans*-acting regulatory factors'').
- Kumar *et al*. have shown that Rop1p, previously described as a transcriptional repressor of genes involved in the biotin metabolism and as a repressor of PEPCK (see the section ''Involvement of Rop1p in biotin metabolism''), is also repressing P_{AOX1} (Kumar NV, Rangarajan PN. The Zinc Finger Proteins Mxr1p and Repressor of Phosphoenolpyruvate Carboxykinase (ROP) Have the Same DNA Binding Specificity but Regulate Methanol Metabolism Antagonistically in Pichia pastoris. J. Biol. Chem. 2012;287:34465–73). Rop1p appears to act only on full media containing methanol as a repressor of P_{AOX1} and other MUT promoters by binding to the same binding sites as the main activator Mxr1p. Rop1p was shown to have a significantly higher DNA binding affinity than Mxr1p. This suggests a regulatory model, in which Rop1p and Mxr1p compete to regulate P_{AOX1} and other MUT promoters on full media in the presence of methanol. It is also notable, that Rop1p is regulating such diverse processes as biotin metabolism, methanol utilization and the expression of PEPCK.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 266025, BIONEXGEN. We also gratefully acknowledge the DK 'Molecular Enzymology' Graz. We would also like to thank Andrea Camattari and Claudia Ruth for fruitful discussions and critical reading of the manuscript.

References

[1] Gellissen G. Heterologous protein production in methylotrophic yeasts. Applied Microbiology and Biotechnology 2000;54:741–50.
[2] Houard S, Heinderyckx M, Bollen A. Engineering of non-conventional yeasts for efficient synthesis of macromolecules: the methylotrophic genera. Biochimie 2002;84:1089–93.
[3] Lin-Cereghino J, Cregg JM. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiology Review 2000;24:45–66.
[4] Desai PN, Shrivastava N, Padh H. Production of heterologous proteins in plants: strategies for optimal expression. Biotechnology Advances 2010;28:427–35.
[5] Goldstein MA, Doi RH. Prokaryotic promoters in biotechnology. Biotechnology Annual Review 1995;1:105–28.
[6] Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that enhances gene expression. Nature Methods 2006;3:917–22.
[7] Porro D, Sauer M, Branduardi P, Mattanovich D. Recombinant protein production in yeasts. Molecular Biotechnology 2005;31:245–59.
[8] Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiological Reviews 1996;60:512–38.
[9] Riethoven J-JM. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. Methods in Molecular Biology (Clifton NJ) 2010;674:33–42.
[10] Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits: predicting numbers from alphabets. Science 2009;325:429–32.
[11] Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology 2010;339:225–9.
[12] Heintzman ND, Ren B. The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. Cellular and Molecular Life Sciences 2007;64:386–400.
[13] Venter M. Synthetic promoters: genetic control through cis engineering. Trends in Plant Science 2007;12:118–24.
[14] Ruth C, Glieder A. Perspectives on synthetic promoters for biocatalysis and biotransformation. ChemBioChem 2010;11:761–5.
[15] Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. International Journal of Microbiology 2011;2011:101298.
[16] van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochimica et Biophysica Acta 2006;1763:1453–62.
[17] Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microbial Cell Factories 2006;5:39–59.
[18] Couderc R, Baratti J. Oxidation of methanol by the yeast, *Pichia pastoris*, purification and properties of the alcohol oxidase. Agricultural and Biological Chemistry 1980;44:2279–89.
[19] Ozimek P, Veenhuis M, van der Klei IJ. Alcohol oxidase: a complex peroxisomal, oligomeric flavoprotein. FEMS Yeast Research 2005;5:975–83.
[20] Inan M, Meagher MM. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*. Journal of Bioscience and Bioengineering 2001;92:585–9.
[21] Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. Molecular and Cellular Biology 1985;5:1111–21.
[22] Koutz P, Davis GR, Stillman C, Barringer K, Cregg J, Thill G. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast 1989;5:167–77.
[23] Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Molecular and Cellular Biology 1989;9:1316–23.
[24] Jahic M, Veide A, Charoenrat T, Teeri T, Enfors S-O. Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. Biotechnology Progress 2006;22:1465–73.
[25] Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Lückl J, et al. High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts. Protein Expression and Purification 1997;11:61–71.

[26] Werten MW, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA. High-yield secretion of recombinant gelatins by *Pichia pastoris*. Yeast 1999;15:1087–96.
[27] Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Research 1987;15:3859–76.
[28] Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, et al. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris*. Molecular and Cellular Biology 2006;26:883–97.
[29] Stroman DW, Brust PF, Ellis SB, Gingeras TR, Harpold MM, Tschopp JF, inventors; Phillips petroleum company 5th and Keeler Bartlesville Oklahoma 74004 (US), assignee. Regulatory region for heterologous gene expression in yeast. European patent EP/0183071/B1. 1992 Dec 9.
[30] Gödecke S, Eckart M, Janowicz ZA, Hollenberg CP. Identification of sequences responsible for transcriptional regulation of the strongly expressed methanol oxidase-encoding gene in *Hansenula polymorpha*. Gene 1994;139:35–42.
[31] Suckow M, inventor; Rhein Biotech Gesellschaft für neue biotechnologische Prozesse und Produkte Gmbh 40595 Düsseldorf (DE), assignee. Promoters having a modified transcription efficiency and derived from methylotrophic yeast. European patent EP 1 504 103 B1. 2003 May 08.
[32] Komeda T, Yurimoto H, Kato N, Sakai Y, Kondo K. Cis-acting elements sufficient for induction of FDH1 expression by formate in the methylotrophic yeast *Candida boidinii*. Molecular Genetics and Genomics 2003;270:273–80.
[33] Yurimoto H, Komeda T, Lim CR, Nakagawa T, Kondo K, Kato N, et al. Regulation and evaluation of five methanol-inducible promoters in the methylotrophic yeast *Candida boidinii*. Biochimica et Biophysica Acta 2000;1493:56–63.
[34] Kumagai MH, Sverlow GG, Della-Cioppa G, Grill LK. Conversion of starch to ethanol in a recombinant *Saccharomyces cerevisiae* strain expressing rice alpha-amylase from a novel *Pichia pastoris* alcohol oxidase promoter. Bio/Technology 1993;11:606–10.
[35] Ciriacy M. Genetics of alcohol dehydrogenase in *Saccharomyces cerevisiae*. II. Two loci controlling synthesis of the glucose-repressible ADH II. Molecular and General Genetics 1975;138:157–64.
[36] Tachibana C, Yoo JY, Tagne J, Kacherovsky N, Lee TI, Young ET. Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. Molecular and Cellular Biology 2005;25:2138–46.
[37] Simon M, Adam G, Rapatz W, Spevak W, Ruis H. The *Saccharomyces cerevisiae* ADR1 gene is a positive regulator of transcription of genes encoding peroxisomal proteins. Molecular and Cellular Biology 1991;11:699–704.
[38] Kumagai MH, Sverlow GG, inventors; Biosource Genetics Corporation; 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US), assignee. *Pichia pastoris* alcohol oxidase ZZA1 and ZZA2. World patent WO/1994/021802. 1994 Mar 24.
[39] Ohi H, Miura M, Hiramatsu R, Ohmura T. The positive and negative cis-acting elements for methanol regulation in the *Pichia pastoris* AOX2 gene. Molecular and General Genetics 1994;243:489–99.
[40] Barr KA, Hopkins SA, Sreekrishna K. Protocol for efficient secretion of HSA developed from *Pichia pastoris*. Pharmaceutical Engineering 1992;12:48–51.
[41] Romanos M. Advances in the use of *Pichia pastoris* for high-level gene expression. Current Opinion in Biotechnology 1995;6:527–33.
[42] Inan M. Studies on the alcohol oxidase (AOX1) promoter of *Pichia pastoris*. Dissertation. Lincoln, Nebraska: University of Nebraska; 2000. 101 pp.
[43] Inan M, Meagher M, Benson AK, inventors; The Board of Regents of the University of Nebraska (Lincoln, NE), assignee. Alcohol oxidase 1 regulatory nucleotide sequences for heterologous gene expression in yeast. United States patent US 6,699,691 B2. 2004 Mar 2.
[44] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Research 2008;36:e76.
[45] Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y. An upstream activation sequence controls the expression of AOX1 gene in *Pichia pastoris*. FEMS Yeast Research 2009;9:1271–82.
[46] Hartner F, Glieder A, inventors; Technische Universität Graz, assignee. Mutant AOX1 promoters. World patent WO2006/089329 A2. 2006 Aug 31.
[47] Ruth C, Zuellig T, Mellitzer A, Weis R, Looser V, Kovar K, et al. Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*. Systems and Synthetic Biology 2010;4:181–91.
[48] Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN. Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*. Biochimica et Biophysica Acta 2009;1789:460–8.
[49] Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, Mulye T, et al. Analysis of the 5′ untranslated region (5′UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in *Pichia pastoris*. Gene 2012;496:118–27.
[50] Johnson MA, Waterham HR, Ksheminska GP, Fayura LR, Cereghino JL, Stasyk OV, et al. Positive selection of novel peroxisome biogenesis-defective mutants of the yeast *Pichia pastoris*. Genetics 1999;151:1379–91.
[51] Kranthi BV, Kumar HRV, Rangarajan PN. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast *Pichia pastoris*. Yeast 2010;27:705–11.
[52] Parua PK, Ryan PM, Trang K, Young ET. *Pichia pastoris* 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction. Molecular Microbiology 2012;85:282–98.
[53] Aitken A. 14-3-3 proteins: a historic overview. Seminars in Cancer Biology 2006;16:162–72.
[54] van Heusden GPH. 14-3-3 proteins: regulators of numerous eukaryotic proteins. IUBMB Life 2005;57:623–9.
[55] van Heusden GPH, Steensma HY. Yeast 14-3-3 proteins. Yeast 2006;23:159–71.
[56] van Heusden GPH. 14-3-3 proteins: insights from genome-wide studies in yeast. Genomics 2009;94:287–93.
[57] Parua PK, Ratnakumar S, Braun KA, Dombek KM, Arms E, Ryan PM, et al. 14-3-3 (Bmh) proteins inhibit transcription activation by Adr1 through direct binding to its regulatory domain. Molecular and Cellular Biology 2010;30:5273–83.
[58] Sakai Y, Sawai T, Tani Y. Isolation and characterization of a catabolite repression-insensitive mutant of a methanol yeast, *Candida boidinii* A5, producing alcohol oxidase in glucose-containing medium. Applied and Environment Microbiology 1987;53:1812–8.
[59] Parpinello G, Berardi E, Strabbioli R. A regulatory mutant of *Hansenula polymorpha* exhibiting methanol utilization metabolism and peroxisome proliferation in glucose. Journal of Bacteriology 1998;180:2958–67.
[60] Takagi S, Tsutsumi N, Terui Y, Kong XY, inventors; Novozymes A/S, Bagsvaerd (DK), assignee. Method for methanol independent induction from methanol inducible promoters in Pichia. United States patent US 8,143,023. 2012 Jun 28.
[61] Zhang P, Zhang W, Zhou X, Bai P, Cregg JM, Zhang Y. Catabolite repression of Aox in *Pichia pastoris* is dependent on hexose transporter PpHxt1 and pexophagy. Applied and Environment Microbiology 2010;76:6108–18.
[62] Leandro MJ, Fonseca C, Gonçalves P. Hexose and pentose transport in ascomycetous yeasts: an overview. FEMS Yeast Research 2009;9:511–25.
[63] Stasyk OV, Stasyk OG, Komduur J, Veenhuis M, Cregg JM, Sibirny AA. A hexose transporter homologue controls glucose repression in the methylotrophic yeast *Hansenula polymorpha*. Journal of Biological Chemistry 2004;279:8116–25.
[64] Stasyk OG, Maidan MM, Stasyk OV, Van Dijck P, Thevelein JM, Sibirny AA. Identification of hexose transporter-like sensor HXS1 and functional hexose transporter HXT1 in the methylotrophic yeast *Hansenula polymorpha*. Eukaryotic Cell 2008;7:735–46.
[65] Kranthi BV, Balasubramanian N, Rangarajan PN. Isolation of a single-stranded DNA-binding protein from the methylotrophic yeast, *Pichia pastoris* and its identification as zeta crystallin. Nucleic Acids Research 2006;34:4060–8.
[66] Babiychuk E, Kushnir S, Belles-Boix E, Van Montagu M, Inzé D. *Arabidopsis thaliana* NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. Journal of Biological Chemistry 1995;270:26224–31.
[67] Bhatnagar A, Raghavendra PR, Kranthi BV, Rangarajan PN. Yeast cytochrome c is a sequence-specific DNA-binding protein. Biochemical and Biophysical Research Communications 2004;321:900–4.
[68] Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 1997;186:37–44.
[69] Zhang A-L, Luo J-X, Zhang T-Y, Pan Y-W, Tan Y-H, Fu C-Y, et al. Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. Molecular Biology Reports 2009;36:1611–9.
[70] Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D. Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. Biotechnology and Bioengineering 2008;100:177–83.
[71] Kern A, Hartner FS, Freigassner M, Spielhofer J, Rumpf C, Leitner L, et al. *Pichia pastoris* just in time alternative respiration. Microbiology (Reading UK) 2007;153:1250–60.
[72] Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J. GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Applied and Environment Microbiology 2011;77:3600–8.
[73] Mattanovich D, Gasser B, Hohenblum H, Sauer M. Stress in recombinant protein producing yeasts. Journal of Biotechnology 2004;113:121–35.
[74] Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotechnology and Bioengineering 2004;85:367–75.
[75] Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM. Heterologous protein production using the *Pichia pastoris* expression system. Yeast 2005;22:249–70.
[76] Shen S, Sulter G, Jeffries TW, Cregg JM. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. Gene 1998;216:93–102.
[77] Resina D, Cos O, Ferrer P, Valero F. Developing high cell density fed-batch cultivation strategies for heterologous protein production in *Pichia pastoris* using the nitrogen source-regulated FLD1 promoter. Biotechnology and Bioengineering 2005;91:760–7.
[78] Resina D, Maurer M, Cos O, Arnau C, Carnicer M, Marx H, et al. Engineering of bottlenecks in *Rhizopus oryzae* lipase production in *Pichia pastoris* using the nitrogen source-regulated FLD1 promoter. New Biotechnology 2009;25:396–403.
[79] Liu H, Tan X, Russell KA, Veenhuis M, Cregg JM. PER3a, gene required for peroxisome biogenesis in *Pichia pastoris*, encodes a peroxisomal membrane protein involved in protein import. Journal of Biological Chemistry 1995;270:10940–51.
[80] Sears IB, O'Connor J, Rossanese OW, Glick BS. A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast 1998;14:783–90.
[81] Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Tomomitsu K. Addition of oleic acid increases expression of recombinant human serum albumin by the AOX2 promoter in *Pichia pastoris*. Journal of Bioscience and Bioengineering 2000;89:479–84.

[82] Mochizuki S, Hamato N, Hirose M, Miyano K, Ohtani W, Kameyama S, et al. Expression and characterization of recombinant human antithrombin III in *Pichia pastoris*. Protein Expression and Purification 2001;23:55–65.
[83] Menendez J, Valdes I, Cabrera N. The ICL1 gene of *Pichia pastoris*, transcriptional regulation and use of its promoter. Yeast 2003;20:1097–108.
[84] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host *Pichia pastoris*. Nature Biotechnology 2009;27:561–6.
[85] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of *Pichia pastoris* CBS7435. Journal of Biotechnology 2011;154:312–20.
[86] Janesch B, Glieder A, Hartner F, Schroer K, Luef KP, Pscheidt B, inventors; Angewandte Biokatalyse Kompetenzzentrum GmbH; Petersgasse 14, A-8010 Graz (AT) Technische Universität Graz; Rechbauerstrasse 12, A-8010 Graz (AT), assignee. Recombinant *Pichia pastoris* cell. World patent application WO/2009/070822. 2009 Jun 11.
[87] Tsutsumi N, Takagi S, inventors; Novozymes A/S, Bagsvaerd (DK), assignee. *Pichia pastoris* DAS promoter variants. United States patent application publication US 2011/0129874 A1. 2011 Jun 2.
[88] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microbial Cell Factories 2012;11:22–35.
[89] Tuite MF, Dobson MJ, Roberts NA, King RM, Burke DC, Kingsman SM, et al. Regulated high efficiency expression of human interferon-alpha in *Saccharomyces cerevisiae*. EMBO Journal 1982;1:603–8.
[90] Le Dall M, Nicaud J, Tréton BY, Gaillardin CM. The 3-phosphoglycerate kinase gene of the yeast *Yarrowia lipolytica* de-represses on gluconeogenic substrates. Current Genetics 1996;29:446–56.
[91] Masuda Y, Park SM, Ohkuma M, Ohta A, Takagi M. Expression of an endogenous and a heterologous gene in *Candida maltosa* by using a promoter of a newly-isolated phosphoglycerate kinase (PGK) gene. Current Genetics 1994;25:412–7.
[92] de Almeida JRM, de Moraes LMP, Torres FAG. Molecular characterization of the 3-phosphoglycerate kinase gene (PGK1) from the methylotrophic yeast *Pichia pastoris*. Yeast 2005;22:725–37.
[93] Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D, et al. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. Journal of Biotechnology 2010;150:519–29.
[94] Sung-Jae L, Hong I-P, Baek S-Y, Choi S-G. Deletion analysis of *Pichia* PGK1 promoter and construction of an episomal vector for heterologous protein expression in *P. pastoris*. Korean Journal of Microbiology and Biotechnology 2007;35:184–90.
[95] Proud CG. Peptide-chain elongation in eukaryotes. Molecular Biology Reports 1994;19:161–70.
[96] Steiner S, Philippsen P. Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus *Ashbya gossypii*. Molecular and General Genetics 1994;242:263–71.
[97] Kitamoto N, Matsui J, Kawai Y, Kato A, Yoshino S, Ohmiya K, et al. Utilization of the TEF1-alpha gene (TEF1) promoter for expression of polygalacturonase genes, pgaA and pgaB, in *Aspergillus oryzae*. Applied Microbiology and Biotechnology 1998;50:85–92.
[98] Müller S, Sandal T, Kamp-Hansen P, Dalbøge H. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. Yeast 1998;14:1267–83.
[99] Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, et al. Translation elongation factor 1-alpha gene from *Pichia pastoris*: molecular cloning, sequence, and use of its promoter. Applied Microbiology and Biotechnology 2007;74:601–8.
[100] Lee HW, Ahn JO, Jung JK, Choi ES, Kim CS, Lee GL, et al., inventors; Korean Research Institute of Bioscience (Daejeon, KR), assignee. Translational elongation factor promoter from *Pichia pastoris* and method for producing recombinant protein using the same. World patent WO2007/058407. 2008 Aug 19.
[101] Veiga A, Arrabaça JD, Loureiro-Dias MC. Cyanide-resistant respiration is frequent, but confined to yeasts incapable of aerobic fermentation. FEMS Microbiology Letters 2000;190:93–7.
[102] Veiga A, Arrabaça JD, Loureiro-Dias MC. Cyanide-resistant respiration, a very frequent metabolic pathway in yeasts. FEMS Yeast Research 2003;3:239–45.
[103] Ahn J, Hong J, Park M, Lee H, Lee E, Kim C, et al. Phosphate-responsive promoter of a *Pichia pastoris* sodium phosphate symporter. Applied Microbiology and Biotechnology 2009;75:3528–34.
[104] Lee HW, Ahn JO, Jung JK, Choi ES, Kim CS, Lee GL, et al., inventors; Korea Research Institute of Bioscience and Biotechnology (Daejeon, KR), assignee. Auto-inducible sodium phosphate symporter promoter from *Pichia pastoris* and method for producing recombinant protein using it. World patent WO2007/117062. 2007 Oct 18.
[105] Boidol W, Simonis M, Töpert M, Siewert G. Recombinant plasmids with genes for the biosynthesis of alkaline phosphatase of *Escherichia coli*. Molecular and General Genetics 1982;185:510–2.
[106] Vogel K, Hinnen A. The yeast phosphatase system. Molecular Microbiology 1990;4:2013–7.
[107] Payne WE, Gannon PM, Kaiser CA. An inducible acid phosphatase from the yeast *Pichia pastoris*: characterization of the gene and its product. Gene 1995;163:19–26.
[108] Hubmann U. New constitutive promoters for recombinant protein expression. Diploma thesis. Graz, Austria: Graz University of Technology; 2005. 75 pp.
[109] Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 2012;7:e39720.
[110] Triglia T. Inverse PCR (IPCR) for obtaining promoter sequence. Methods in Molecular Biology (Clifton NJ) 2000;130:79–83.
[111] Gasser B, Mattanovich D, Sauer M, Stadlmayr G, inventors; Polymun Scientific Immunbiologische Forschung GmbH; Nussdorfer Lände 11, A-1190 Vienna (AT). Yeast expression systems. World patent WO/2008/128701. 2008 Oct 30.
[112] Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, et al. Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. Microbial Cell Factories 2004;3:17–30.
[113] Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in *Pichia pastoris*. Molecular Biotechnology 2000;16:23–52.
[114] Idiris A, Tohda H, Kumagai H, Takegawa K. Engineering of protein secretion in yeast: strategies and impact on protein production. Applied Microbiology and Biotechnology 2010;86:403–17.
[115] Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. Cell 2004;116:153–66.
[116] Patil C, Walter P. Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Current Opinion in Cell Biology 2001;13:349–55.
[117] Chakrabarti A, Chen AW, Varner JD. A review of the mammalian unfolded protein response. Biotechnology and Bioengineering 2011;108:2777–93.
[118] Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, et al. Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. BMC Genomics 2007;8:179–96.
[119] Resina D, Bollók M, Khatri NK, Valero F, Neubauer P, Ferrer P. Transcriptional response of *P. pastoris* in fed-batch cultivations to *Rhizopus oryzae* lipase production reveals UPR induction. Microbial Cell Factories 2007;6:21–31.
[120] Graf A, Gasser B, Dragosits M, Sauer M, Leparc GG, Tüchler T, et al. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. BMC Genomics 2008;9:390–402.
[121] Dragosits M, Stadlmann J, Graf A, Gasser B, Maurer M, Sauer M, et al. The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*. BMC Genomics 2010;11:207–22.
[122] Dragosits M, Mattanovich D, Gasser B. Induction and measurement of UPR and osmotic stress in the yeast *Pichia pastoris*. Methods in Enzymology 2011;489:165–88.
[123] Whyteside G, Alcocer MJC, Kumita JR, Dobson CM, Lazarou M, Pleass RJ, et al. Native-state stability determines the extent of degradation relative to secretion of protein variants from *Pichia pastoris*. PLoS One 2011;6:e22692.
[124] Delic M, Rebnegger C, Wanka F, Puxbaum V, Haberhauer-Troyer C, Hann S, et al. Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. Free Radical Biology and Medicine 2012;52:2000–12.
[125] Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, et al. The HAC1 gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microbial Cell Factories 2010;9:49–60.
[126] Whyteside G, Nor RM, Alcocer MJC, Archer DB. Activation of the unfolded protein response in *Pichia pastoris* requires splicing of a HAC1 mRNA intron and retention of the C-terminal tail of Hac1p. FEBS Letters 2011;585:1037–41.
[127] Cox JS, Walter P. A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell 1996;87:391–404.
[128] Rüegsegger U, Leber JH, Walter P. Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell 2001;107:103–14.
[129] Sidrauski C, Walter P. The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell 1997;90:1031–9.
[130] Kimata Y, Kimata YI, Shimizu Y, Abe H, Farcasanu IC, Takeuchi M, et al. Genetic evidence for a role of BiP/Kar2 that regulates Ire1 in response to accumulation of unfolded proteins. Molecular Biology of the Cell 2003;14:2559–69.
[131] Credle JJ, Finer-Moore JS, Papa FR, Stroud RM, Walter P. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. Proceedings of the National Academy of Sciences of the United States of America 2005;102:18773–84.
[132] Ma Y, Hendershot LM. The unfolding tale of the unfolded protein response. Cell 2001;107:827–30.
[133] Aragón T, van Anken E, Pincus D, Serafimova IM, Korennykh AV, Rubio CA, et al. Messenger RNA targeting to endoplasmic reticulum stress signalling sites. Nature 2009;457:736–40.
[134] Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D. Engineering of *Pichia pastoris* for improved production of antibody fragments. Biotechnology and Bioengineering 2006;94:353–61.
[135] Frankenberg N, Moser J, Jahn D. Bacterial heme biosynthesis and its biotechnological application. Applied Microbiology and Biotechnology 2003;63:115–27.
[136] Jiang F, Kongsaeree P, Charron R, Lajoie C, Xu H, Scott G, et al. Production and separation of manganese peroxidase from heme amended yeast cultures. Biotechnology and Bioengineering 2008;99:540–9.
[137] Bonaccorsi di Patti MC, Bellenchi GC, Bielli P, Calabrese L. Release of highly active Fet3 from membranes of the yeast *Pichia pastoris* by limited proteolysis. Archives of Biochemistry and Biophysics 1999;372:295–9.

[138] Paronetto MP, Miele R, Maugliani A, Borro M, Bonaccorsi di Patti MC. Cloning of *Pichia pastoris* Fet3: insights into the high affinity iron uptake system. Archives of Biochemistry and Biophysics 2001;392:162–7.
[139] Bonaccorsi di Patti MC, Miele R, Eugenia Schininà M, Barra D. The yeast multicopper oxidase Fet3p and the iron permease Ftr1p physically interact. Biochemical and Biophysical Research Communications 2005;333:432–7.
[140] Kosman DJ. Molecular mechanisms of iron uptake in fungi. Molecular Microbiology 2003;47:1185–97.
[141] Van Ho A, Ward DM, Kaplan J. Transition metal transport in yeast. Annual Review of Microbiology 2002;56:237–61.
[142] Blaiseau PL, Lesuisse E, Camadro JM. Aft2p, a novel iron-regulated transcription activator that modulates, with Aft1p, intracellular iron use and resistance to oxidative stress in yeast. Journal of Biological Chemistry 2001;276:34221–26.
[143] Rutherford JC, Jaron S, Ray E, Brown PO, Winge DR. A second iron-regulatory system in yeast independent of Aft1p. Proceedings of the National Academy of Sciences of the United States of America 2001;98:14322–27.
[144] Yamaguchi-Iwai Y, Dancis A, Klausner RD. AFT1: a mediator of iron regulated transcriptional control in *Saccharomyces cerevisiae*. EMBO Journal 1995;14:1231–9.
[145] Shakoury-Elizeh M, Tiedeman J, Rashford J, Ferea T, Demeter J, Garcia E, et al. Transcriptional remodeling in response to iron deprivation in *Saccharomyces cerevisiae*. Molecular Biology of the Cell 2004;15:1233–43.
[146] Yamaguchi-Iwai Y, Stearman R, Dancis A, Klausner RD. Iron-regulated DNA binding by the AFT1 protein controls the iron regulon in yeast. EMBO Journal 1996;15:3377–84.
[147] Puig S, Askeland E, Thiele DJ. Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation. Cell 2005;120:99–110.
[148] Miele R, Barra D, Bonaccorsi di Patti MC. A GATA-type transcription factor regulates expression of the high-affinity iron uptake system in the methylotrophic yeast *Pichia pastoris*. Archives of Biochemistry and Biophysics 2007;465:172–9.
[149] Pelletier B, Beaudoin J, Mukai Y, Labbé S. Fep1, an iron sensor regulating iron transporter gene expression in *Schizosaccharomyces pombe*. Journal of Biological Chemistry 2002;277:22950–58.
[150] Harrison KA, Marzluf GA. Characterization of DNA binding and the cysteine rich region of SRE, a GATA factor in *Neurospora crassa* involved in siderophore synthesis. Biochemistry 2002;41:15288–95.
[151] Lan C-Y, Rodarte G, Murillo LA, Jones T, Davis RW, Dungan J, et al. Regulatory networks affected by iron availability in *Candida albicans*. Molecular Microbiology 2004;53:1451–69.
[152] Imlay JA. Pathways of oxidative damage. Annual Review of Microbiology 2003;57:395–418.
[153] Evans MD, Cooke MS. Factors contributing to the outcome of oxidative damage to nucleic acids. Bioessays 2004;26:533–42.
[154] Poljsak B. Strategies for reducing or preventing the generation of oxidative stress. Oxidative Medicine and Cellular Longevity 2011;2011:194586.
[155] Rigoulet M, Yoboue ED, Devin A. Mitochondrial ROS generation and its regulation: mechanisms involved in H(2)O(2) signaling. Antioxidants & Redox Signaling 2011;14:459–68.
[156] Blokhina O, Virolainen E, Fagerstedt KV. Antioxidants, oxidative damage and oxygen deprivation stress: a review. Annals of Botany 2003;91:179–94.
[157] Santos CXC, Tanaka LY, Wosniak J, Laurindo FRM. Mechanisms and implications of reactive oxygen species generation during the unfolded protein response: roles of endoplasmic reticulum oxidoreductases, mitochondrial electron transport, and NADPH oxidase. Antioxidants & Redox Signaling 2009;11:2409–27.
[158] Yano T, Takigami E, Yurimoto H, Sakai Y. Yap1-regulated glutathione redox system curtails accumulation of formaldehyde and reactive oxygen species in methanol metabolism of *Pichia pastoris*. Eukaryotic Cell 2009;8:540–9.
[159] Yano T, Yurimoto H, Sakai Y. Activation of the oxidative stress regulator PpYap1 through conserved cysteine residues during methanol metabolism in the yeast *Pichia pastoris*. Bioscience Biotechnology and Biochemistry 2009;73:1404–11.
[160] Penninckx MJ. An overview on glutathione in *Saccharomyces* versus non-conventional yeasts. FEMS Yeast Research 2002;2:295–305.
[161] Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Molecular Biology of the Cell 2000;11:4241–57.
[162] Moye-Rowley WS. Transcription factors regulating the response to oxidative stress in yeast. Antioxidants & Redox Signaling 2002;4:123–40.
[163] Temple MD, Perrone GG, Dawes IW. Complex cellular responses to reactive oxygen species. Trends in Cell Biology 2005;15:319–26.
[164] Streit WR, Entcheva P. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. Applied Microbiology and Biotechnology 2003;61:21–31.
[165] Gasser B, Dragosits M, Mattanovich D. Engineering of biotin-prototrophy in *Pichia pastoris* for robust production processes. Metabolic Engineering 2010;12:573–80.
[166] Kumar NV, Rangarajan PN. Catabolite repression of phosphoenolpyruvate carboxykinase by a zinc finger protein under biotin- and pyruvate carboxylase-deficient conditions in *Pichia pastoris*. Microbiology (Reading UK) 2011;157:3361–9.
[167] Menéndez J, Delgado J, Gancedo C. Isolation of the *Pichia pastoris* PYC1 gene encoding pyruvate carboxylase and identification of a suppressor of the pyc phenotype. Yeast 1998;14:647–54.
[168] Ozimek PZ, Klompmaker SH, Visser N, Veenhuis M, van der Klei IJ. The transcarboxylase domain of pyruvate carboxylase is essential for assembly of the peroxisomal flavoenzyme alcohol oxidase. FEMS Yeast Research 2007;7:1082–92.
[169] Ozimek P, van Dijk R, Latchev K, Gancedo C, Wang DY, van der Klei IJ, et al. Pyruvate carboxylase is an essential protein in the assembly of yeast peroxisomal oligomeric alcohol oxidase. Molecular Biology of the Cell 2003;14:786–97.
[170] Blázquez MA, Gamo FJ, Gancedo C. A mutation affecting carbon catabolite repression suppresses growth defects in pyruvate carboxylase mutants from *Saccharomyces cerevisiae*. FEBS Letters 1995;377:197–200.
[171] Zelle RM, Trueheart J, Harrison JC, Pronk JT, van Maris AJA. Phosphoenolpyruvate carboxykinase as the sole anaplerotic enzyme in *Saccharomyces cerevisiae*. Applied and Environment Microbiology 2010;76:5383–9.

Chapter 2.3 - Natural promoters and synthetic promoter engineering strategies in *Pichia pastoris*

Cover page

Title of the chapter

Natural promoters and synthetic promoter engineering strategies in *Pichia pastoris*

Running head

P. pastoris promoters

Authors

Thomas Vogl[1], Anton Glieder[2]

[1] Institute for Molecular Biotechnology, Graz University of Technology, Petersgasse 14/2, 8010 Graz, Austria
[2] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14/5, 8010 Graz, Austria

Corresponding author:
Anton Glieder (anton.glieder@acib.at)

Summary/Abstract

The availability of exceptionally strong and tightly regulated promoters is a key feature of *Pichia pastoris*, a widely applied yeast expression system for heterologous protein production. Most commonly, the methanol inducible promoter of the *alcohol oxidase 1* gene (P_{AOX1}) and the constitutive promoter of the *glyceraldehyde 3 phosphate dehydrogenase* gene (P_{GAP}) have been used. Recently also promising novel constitutive (P_{GCW14}) and regulated (P_{GTH1}) promoters have been reported.
As natural promoters showed so far limited tuneability of expression levels and regulatory profiles as well as unknown regulatory effects, various promoter engineering efforts have been undertaken for *P. pastoris*. P_{AOX1}, P_{DAS2}, P_{GAP} and P_{GCW14} have been engineered by systematic deletion studies or random mutagenesis of upstream regulatory sequences. New engineering strategies have focused on P_{AOX1} core promoter and 5' UTR (untranslated region) modifications by random or semi-rational approaches. These promoter engineering efforts in *P. pastoris* have resulted in improved, sequence diversified synthetic promoter variants allowing coordinated fine-tuning of gene expression.

Key Words

Pichia pastoris, heterologous protein production, promoter engineering, synthetic promoters, core promoters

Background/Introduction

Transcription is an important step in gene expression and various natural and synthetic promoters are commonly used tools to regulate heterologous protein production [1]. A key feature of methylotrophic yeasts such as *Pichia pastoris*, *Candida boidinii*, *Hansenula polymorpha* (*Ogataea angusta*, *Pichia angusta*) and *Pichia methanolica* are the strong and tightly regulated promoters of genes involved in the metabolism of methanol [2–4]. These yeasts in combination with their respective methanol inducible promoters have been widely applied on an industrial scale to produce a large number of heterologous proteins ranging from biocatalysts to biopharmaceuticals [2, 5–7].

Typical methanol regulated promoters are tightly repressed on glucose and strongly induced on methanol, which allows the design of biphasic production processes. Due to its exceptional strength and tight regulation, most commonly the promoter of the *alcohol oxidase* gene (abbreviated P_{AOX}, P_{MOX}, P_{AOD}, P_{MOD} depending on the host organism) is used to drive the expression of the gene of interest (GOI) [3]. In the first phase, cells are grown on glucose to high cell density. During this phase P_{AOX} expression is repressed, thereby ruling out a negative effect of the heterologous GOI on cell growth. At high cell density, cells are induced with methanol, starting gene expression. If required, also a transition phase is applied to adapt the metabolism to the new carbon source methanol. While this system allows even the expression of toxic proteins, the induction with hazardous and flammable methanol may pose a problem, especially in larger scale industrial processes. In addition, the extremely strong expression of heterologous proteins by *P. pastoris* P_{AOX1} (approximately 1000 fold induction [8]) may overburden the cellular machinery [9]. If transcription is not the limiting step, strong overexpression may lead to folding problems and degradation. Also post translational modifications or trafficking (e.g. secretion) may become the limiting factor [10–12]. Therefore tunable promoter systems, offering a range of expression levels and regulatory profiles are desirable.

Similar problems of tuneability and lacking control have been noticed in other expression hosts and in recent years there has been a growing interest in engineering natural promoters or using synthetic promoters to counteract these issues [1, 13]. Synthetic promoters have been designed for prokaryotic organisms, lower eukaryotes as yeast and also higher eukaryotes [1, 14–17] and helped alongside other synthetic biology approaches to improve metabolic engineering and heterologous protein production [5, 18]. Concerning yeasts, most promoter engineering work has been performed in *Saccharomyces cerevisiae* [1], also a commonly used model system for studying fundamental principles of transcriptional regulation [19]. These approaches resulted in new parts for synthetic biology and metabolic engineering and the generation of improved production strains [15, 16, 20, 21]. Synthetic hybrid promoters have also been used in the nonconventional yeast *Yarrowia lipolytica* to counteract the lack of natural strong and easily controllable promoters [17, 22] and various promoter engineering strategies have been applied to improve and diversify promoter properties in *P. pasto*ris [9, 23–28].

The regulation of natural promoters and transcription factors (e.g. Mxr1, Hac1, Fep1, Yap1, Rop1) in *P. pastoris* has recently been comprehensively reviewed [29]. Here we focus on the application of different groups of natural promoters and engineering approaches to create synthetic promoters with tailor-made properties in *P. pastoris*.

Natural promoters

Methanol inducible P_{AOX1}

Most commonly the methanol inducible promoter of the *alcohol oxidase 1* gene (P_{AOX1}) [8, 30] or the constitutive promoter of the *glyceraldehyde 3 phosphate dehydrogenase* gene (P_{GAP}) [31] have been used to drive heterologous protein expression in *P. pastoris* [29]. *P. pastoris* harbors two *AOX* genes coding for highly similar alcohol oxidases catalyzing the oxidation of methanol to formaldehyde by molecular oxygen and producing H_2O_2 as a byproduct [32, 33]. However, P_{AOX1} is much stronger than P_{AOX2} and almost solely used to drive heterologous gene expression reaching titers up to 22 g/l intracellular [34] and 15 g/l secreted [35] heterologous protein. Upon methanol induction, the natural Aox proteins may reach up to 30 % of soluble cytoplasmic protein [2]. As Aox has relatively high Km values for methanol and oxygen (*i.e.* a low affinity for its substrates), it has been suggested that the natural high expression is compensating for the low catalytic efficiency [36]. P_{AOX1} has been studied in terms of *cis*-acting regulatory sequences and *trans*-acting transcription factors (recently reviewed in [29]). The most relevant and well-studied regulator is the methanol master regulator Mxr1 (methanol expression regulator 1) [37–40], which itself is regulated by a 14-3-3 protein [41]. Recently, P_{AOX1} driven expression has also been studied in detail in bioreactor cultivation processes [42] and additional details on molecular factors involved in P_{AOX1} regulation have been reported [40, 43, 44]. Namely, the zinc finger transcription factor Rop1 (repressor of phosphoenolpyruvate carboxykinase, PEPCK), previously reported to repress PEPCK under certain growth conditions [45], was reported to repress *P. pastoris* P_{AOX1} on full media containing methanol, but not on minimal media containing methanol. Additionally, a glucose sensor (*GSS1*) was shown to be involved in regulation of P_{AOX1} catabolite repression [43] and mechanistic differences thereof between *P. pastoris* and other yeasts were suggested [44]. These findings do not appear directly applicable to improve P_{AOX1} driven expression, but they demonstrate the intricate regulation of this promoter involving multiple, interwoven factors.

Alternative promoters

Besides P_{AOX1}, also other methanol inducible promoters have been characterized in *P. pastoris*: The promoter of the dihydroxyacetone synthase 2 gene (P_{DAS2}) was reported to reach higher expression levels than P_{AOX1} [8, 29, 30] while P_{PEX8} [46, 47] and P_{AOX2} [33] are much weaker promoters (Table 1). However, methanol inducible expression in *P. pastoris* has been almost exclusively achieved using P_{AOX1}; even P_{DAS2} has seldom been used to drive heterologous proteins expression.

Although methanol induction P_{AOX1} based processes are widely applied, the use of methanol, a toxic and flammable compound, is for industrial scale bioreactor cultivations a considerable concern due to safety issues. There have been several alternatives tested and previously reviewed [29], summarized in Table 1, including constitutive promoters (e.g. P_{GAP} [31], P_{TEF1} [48, 49], P_{PGK1} [49, 50], P_{TPI} [49]) and other means of regulation (e.g. methylamine induction of P_{FLD1} [51], phosphate regulated P_{PHO89}/P_{NSP} [52], thiamine regulated P_{THI11} [49, 53]. Also the heterologous *CUP1* promoter from *S. cerevisiae* (P_{ScCUP1}) has been demonstrated to be copper inducible in *P. pastoris* [54]. In *S. cerevisiae* a second copper regulated promoter (*CTR3*) has been reported, that is however not induced but repressed by the presence of copper. This promoter has proven to be a valuable part for metabolic engineering in *S. cerevisiae* [55], but has yet not been applied in *P. pastoris*.

Constitutive promoters

The most common alternative to P_{AOX1} is the constitutive *GAP* promoter [29], which in some cases under optimal growth conditions reached comparable titers to P_{AOX1} [31, 56]. P_{GAP} has been widely applied and delivered reliable results for the expression of many heterologous genes [56]. P_{GAP} shows some degree of regulation [29], namely lower expression levels on methanol and glycerol compared to the optimal substrate glucose [31], effects of the oxygen supply [57] and an upregualtion after glucose depletion [58].

Recently, P_{GCW14} has emerged as novel constitutive promoter showing similar strength as P_{GAP} [59]. The *GCW14* gene was identified in *P. pastoris* RNA-Seq (RNA sequencing) studies [60] as the highest expressed gene on glycerol showing also very high expression levels on methanol [59]. *GCW14* is coding for a putative glycosyl phosphatidyl inositol (GPI)-anchored protein. Using secreted eGFP (enhanced green fluorescent protein) as reporter, the expression levels of P_{GCW14} exceeded P_{GAP} and P_{TEF1} but were slightly lower than P_{AOX1} under the respective optimal growth conditions (comparing transformants containing a single copy of the expression vector). While for most recent promoter studies in *P. pastoris* intracellular expression of GFP (or variants thereof) has been used to assess expression levels (e.g. [23, 24, 49]), the comparisons of P_{GCW14} were performed using GFP fused to a signal sequence and measuring fluorescence levels in the supernatant. As GFP is naturally not a secreted but a cytosolic protein in *Aequorea victoria*, secretion may be inefficient and transcription not the limiting factor. Thereby the secreted levels in the supernatant may only reflect a part of the total accumulated levels. A recent study showed that there is indeed a saturation effect of eGFP secretion in *P. pastoris* [61]. By comparing different copy numbers of eGFP, Liu *et al*. showed that a plateau effect occurs at four to five copies of P_{AOX1} driven secretory eGFP expression in *P. pastoris* [61]. As Liang *et al*. were comparing single copy transformants of promoters with similar strength [59], secretion should not have been a limiting factor. In a follow-up publication, regulatory elements of P_{GCW14} were identified and it was also used to drive the expression of secreted *Candida antarctica* lipase B (CalB), however not compared to other promoters [28]. Furthermore, P_{GCW14} transcript levels were evaluated by relative RT-qPCR (reverse transcription quantitative real-time PCR) on different carbon sources (glucose, glycerol and methanol). The transcript levels of the *GAP* gene were used as a reference for normalization. The reference gene should show consistent transcript levels under all conditions tested to allow reliable normalization [62]. *P. pastoris GAP* transcript levels were shown to vary depending on the carbon source (reaching a maximum on glucose, approximately two thirds on glycerol and one third on methanol) [29, 31]. As P_{GCW14} driven transcript levels, relative to *GAP*, remained largely unchanged on the different carbon sources tested, this would imply that P_{GCW14} is also not entirely constitutive, but affected by the carbon source present.

Novel regulated promoters

Process design using constitutive promoters such as P_{GAP} or P_{GCW14} is simpler than biphasic (cell growth and induction) P_{AOX1} based protocols and space-time yields are often higher. However, expression using constitutive promoters is not feasible with every GOI. Toxic or difficult to fold proteins may hamper cell growth upon constitutive expression resulting in lower yields than regulated expression. Therefore a regulated system allowing methanol free expression by *P. pastoris* is desirable. Methanol free alternatively regulated promoters have been reported [29], but are limited as they require starvation for phosphate (P_{PHO89}/P_{NSP} [52]) or thiamin (P_{THI11} [49]) as inducing stimulus. P_{FLD1} [51] can be activated with the alternative inducer methylamine, which is however similarly to methanol toxic and flammable, making it not an actual alternative for safer process design.

Recently, promising new glucose-limited promoters have been reported that allow simple and safe induction relying on cell growth on glycerol followed by glucose addition as inducer. Prielhofer *et al*. used a transcriptomics approach to identify genes repressed on glycerol and upregulated upon glucose addition [63]. Six promoters were selected based on this regulatory profile and compared to P_{GAP} using intracellular expressed eGFP as reporter protein. The promoters reached between 20 and 150% of P_{GAP}. The two strongest promoters (P_{G1} and P_{G6}) were subsequently also tested with secretory expression of human serum albumin (HSA) reaching about 230% (P_{G1}) and 39% (P_{G6}) of the P_{GAP} control. P_{G1} and P_{G6} vary in the optimal glucose concentrations needed for induction. P_{G1} shows highest expression with less than 0.05 g/l glucose, whereas P_{G6} shows high expression up to 4 g/l. The gene being transcribed by P_{G1} was shown to be a high affinity glucose transporter (by deletion studies) and expression to be affected by the specific growth rate. The authors suggest the gene name *GTH1* (glucose transporter with high affinity) for the respective *P. pastoris* gene. Therefore P_{GTH1}/P_{G1} appears to be a valuable new natural promoter allowing tightly regulated methanol free expression.

Furthermore, novel repressible promoters have been identified, allowing tight downregulation of expression by addition of a repressor [53]. Delic *et al*. selected five putative repressible promoters (P_{SER1}, P_{MET3}, P_{THR1}, P_{PIS1} and P_{THI11}) based on a literature search and microarray data. Three of these promoters (P_{MET3}, P_{THI11} and P_{THR1}) showed tight repression upon addition of the respective compound triggering repression and varying expression levels compared to P_{GAP} (Table 1). The authors suggest the use of these promoters to achieve coordinated downregulation of essential genes, that cannot be deleted [53].

An interesting, yet unsuccessful approach to drive regulated GOI expression in *P. pastoris* has been recently tested by Hobl *et al*. [64]. The authors tried to use the prokaryotic T7 RNA polymerase (T7 RNAP) together with its respective promoter as a fully orthogonal system in *P. pastoris*. Similar approaches have been undertaken in mammalian cells, where transient expression and siRNA (short-interfering RNA) transcription was possible using T7 RNAP based systems [65, 66]. Hobl *et al*. were successful in achieving nuclear localization of T7 RNAP and transcribing a reporter gene in *P. pastoris* [64]. However, T7 RNAP RNA transcripts do not contain the 7-methyl guanosine caps at the 5' end that are required for canonical eukaryotic mRNA translational initiation. The authors tried to achieve translation initiation by providing an internal ribosomal entry site from hepatitis C virus (HCV-IRES), to achieve 5' cap independent translation. However, no translation of the reporter gene could be detected, suggesting that the HCV-IRES is not functional in *P. pastoris* and the limiting factor to use the T7 RNAP based system. In a recent RNA-Seq effort [60], *P. pastoris* endogenous IRES have been identified and shown to be functional in two tested cases (*GCN2* and *KOG1* 5'UTRs). These IRES parts may be used in combination with the established T7 RNAP system to achieve high level expression.

Promoter engineering

Despite their strength and/or tight regulation, the most commonly used natural *P. pastoris* promoters are not optimal for all applications. The high expression levels from P_{AOX1} and P_{GAP} may have negative effects on total yields if the GOI is coding for a toxic, secreted or difficult to fold protein. Therefore various promoter engineering approaches have been undertaken to allow fine tuning of expression levels by applying promoter libraries and using novel regulatory profiles offered by synthetic promoters. Various promoter engineering strategies have been applied in *S. cerevisiae*,

ranging from error prone PCR (epPCR) [16], over artificial transcriptional activators or repressors proteins [20, 67, 68] and hybrid promoter design [15] to core promoter engineering [20, 21].

In *P. pastoris*, most engineering approaches have focused on P_{AOX1}, where conventional deletion [24, 69], random mutagenesis [25] and hybrid promoter strategies [9] known from *S. cerevisiae* have been applied. However, also novel approaches based on putative transcription factor binding site (TFBS) engineering [23] and synthetic core promoters were successfully applied [26] (Figure 1). Deletion/hybrid promoter engineering strategies have also been applied to the P_{DAS2} promoter [70]. Recently also the constitutive *GAP* and *GCW14* promoters have been engineered by random mutagenesis and/or systematic deletion studies [27, 28] (Figure 2).

P_{AOX1} engineering

URS engineering

Eukaryotic promoters mainly consist of two major regulatory regions: the core promoter and upstream regulatory sequences (URSs). While yeast promoters are shorter and consist solely of these two elements, promoters of vertebrates are more complex and may contain in addition proximal elements, enhancers, long-range regulatory elements and insulators [19, 71]. Core promoters, are required for transcription initiation and bound by general transcription factors and by RNA polymerase II. The 5'UTR is an important determinant of mRNA stability and partially overlaps with the core promoter, as also the sequence after the transcription start site (TSS), i.e. the beginning of the 5' UTR, affects transcription initiation [72]. URSs contain specific TFBSs either bound by transcriptional activators or repressors conferring particular regulation. The URSs of *P. pastoris* methanol regulated promoters such as P_{AOX1}, P_{DAS2} and P_{PEX8} contain TFBSs for the methanol master activator Mxr1p [38, 39]. So far no regulatory RNAs as riboswitches or RNAi (RNA interference) have been reported in *P. pastoris*.

The first engineering efforts of P_{AOX1} relied on deletion studies of upstream promoter regions, leaving the core promoter region widely untouched [23, 24, 69]. A related study has also been performed by Ohi *et al*. on the much weaker *AOX2* promoter [73], leading also to the identification of putative regulatory regions in P_{AOX1} (Figure 1). Two systematic P_{AOX1} deletion studies were performed at different resolution; Inan divided the P_{AOX1} upstream sequence in five parts and concomitantly deleted them [69, 74] whereas Xuan *et al*. divided a similar P_{AOX1} region into nine fragments [24] (Figure 1). These deletions affected methanol inducible reporter gene expression in part strongly, decreasing expression to as little as 14 % of wildtype P_{AOX1} levels. Some deletion variants showed also up to 1.28 fold increased expression, others were bound by protein extracts from cells grown on methanol and under glucose/methanol-repressed conditions, suggesting regulatory regions (Figure 1). Xuan *et al*. subsequently introduced a putative regulatory region (shown to be bound by proteins by EMSA [electrophoretic mobility shift assay] in their study) in multiple copies into deletion variants, thereby further improving expression levels to 160 % of wildtype P_{AOX1} [24].

An alternative deletion approach by Hartner *et al*. relied on the prediction of putative TFBS from other eukaryotes in the P_{AOX1} upstream regions [23]. Subsequent deletions thereof led to variants spanning a range of approximately 6 to more than 160 % of wildtype P_{AOX1} methanol induced expression levels. Interestingly, some deletion variants showed also altered regulatory profiles. Deletions termed 'd6' and 'd6*' did not require methanol for induction anymore, but showed a strong derepression effect. Wildtype P_{AOX1} is tightly repressed on glucose or glycerol showing only very low

derepression after glucose or glycerol depletion (2 to 4% of methanol induction [8, 75]), requiring methanol for full induction. In the derepressed variants, expression was efficiently activated after glucose depletion, reaching about one third of methanol induced wildtype promoter levels. It was shown, that the derepressed conditions could be maintained in bioreactor cultivations by applying low feed rates immediately taken up and metabolized by the cells [23] and these variants have been demonstrated to outperform the wildtype *AOX1* promoter in certain applications [76]. In a follow-up study, putative regulatory regions identified by Hartner *et al*. were fused to the P_{AOX1} core promoter resulting in synthetic hybrid promoters [9]. Certain variants showed derepressed expression and when present in multiple copies, even outperformed the methanol induced wildtype P_{AOX1}.

Core promoter engineering

Engineering of URS has proven highly suitable to engineer variants spanning a range of expression levels that can be used to fine tune gene expression [23, 24, 69]. In addition, also synthetic variants with novel regulatory profiles were obtained [9, 23]. However, all these efforts relied on modifications of the approximately 700 bp long upstream sequence of P_{AOX1}. Introducing deletions in this region required assembly of the promoter variants by overlap extension PCR (oePCR) stitching [77, 78] and thereby considerable wet lab efforts.

Engineering the considerable shorter core promoter region was shown to reduce the work amount and also resulted in variants spanning a range of expression levels [25, 26]. This strategy is similar to an approach focused on modifications of the ribosome binding sites in prokaryotes [79], which has become a commonly used approach in recent years. The P_{AOX1} core promoter ($Pcore_{AOX1}$) is only 176 bp in length and variants thereof can be directly added on a long primer by standard PCR [26], not requiring tedious oePCR stitching. Recently, two different engineering strategies have been applied to engineer $Pcore_{AOX1}$. Berg *et al*. performed a random mutagenesis of $Pcore_{AOX1}$ [25] whereas Vogl *et al*. designed at first synthetic core promoters and used successively a sequence grafting strategy to modify $Pcore_{AOX1}$ [26].

The random mutagenesis by Berg *et al*. focused on two different regions highlighted in Figure 1. A core promoter region termed ‘LC’ and an adjacent upstream region termed ‘LU’ were mutagenized using degenerate primers with varying doping percentages for nucleotides (e.g. 79% C, 7% A, 7% T, 7%G or 79% T, 7% A, 7% C, 7% etc. [25]) at different positions. Berg *et al*. defined the LC region “as the region from about 15 base pairs upstream of the putative TATA-box to about 35 base pairs downstream of the transcriptional start site” whereas the LU region was defined as the “90 base pairs region directly upstream of the P_{AOX1} promoter core region” [25], see Figure 1. However, the transition between core promoter and upstream region is somewhat arbitrary and in *S. cerevisiae* up to 200 bp upstream of the TSS were considered to be part of the core promoter [72]. For $Pcore_{AOX1}$ this would also include the LU region of Berg *et al*. [25]. Yet, the core promoter length of P_{AOX1} was previously determined experimentally and applied as a core promoter by Hartner *et al*. [23]. The commonly used length of 176 bp was just one out of different lengths showing similar properties [23]. For these conflicting arguments, we are discussing both LC and LU libraries in this section, although the LU library may also be seen as an URS engineering approach associated with the previous section.

The LC and LU random mutagenesis libraries were screened for increased resistance to Zeocin by placing the Zeocin resistance gene under control of the promoter variants. More than 100000 (LC) and 5000 (LU) random transformants were screened for growth on Zeocin containing plates under

different conditions. On the one hand the variants were grown on methanol plates to identify improved inducible variants and on the other hand they were screened under derepressed conditions to identify abolished glucose repression. Some variants obtained in this plate screening showed up to 80 fold increased resistance levels under derepressed conditions and more than 3 fold increased resistance levels on methanol plates. The variants contained between two and 18 point mutations. Some variants were subsequently re-cloned and screened with a luciferase reporter gene. However, the authors did not provide data of the luciferase expression levels of the methanol induced wild type promoter, an important control. Therefore the expression levels of the derepressed variants relative to established methanol induction with the luciferase reporter gene are unclear. It is also notable that Berg *et al*. used an episomal (autonomously replicating) vector based on pBGP1 [80] to characterize the promoter variants. This is in contrast to other *P. pastoris* promoter studies and general gene expression in *P. pastoris* where expression cassettes are typically stably integrated in the genome [9, 23, 24, 26, 69, 81, 82].

An alternative core promoter engineering strategy was followed by Vogl *et al*. by designing synthetic core promoters as an engineering toolbox for $Pcore_{AOX1}$ modifications [26]. At first a synthetic core promoter sequence termed *Pcore11* was created by incorporating naturally occurring TFBSs into a consensus sequence derived from four natural core promoters. In this process not only the core promoter, but also the 5' UTR was designed *de novo*. Thereby the complex and largely unknown natural regulation of the *AOX1* 5'UTR [83] was removed. The synthetic *Pcore11* fused to the upstream P_{AOX1} sequence showed tight repression on glucose and reached upon methanol induction about 10 % of wildtype $Pcore_{AOX1}$, demonstrating that yeast core promoters can be in principle designed following similar design strategies as for prokaryotic promoters [1]. Short stretches of *Pcore11* were successively incorporated into $Pcore_{AOX1}$, resulting in hybrid core promoters spanning a range of 10 to 117 % of wildtype P_{AOX1} expression levels. These core promoters can easily be attached by PCR using long primers and allow thereby simple fine-tuning of gene expression e.g. required to regulate the flux through metabolic pathways. We expect that this approach can be transferred to other core promoters than $Pcore_{AOX1}$ by applying the same grafting strategy using *Pcore11*.

Comparing the fundamentally different core promoter engineering strategies applied by Berg *et al*. and Vogl *et al*., the total number of transformants screened and also the results were different. While Berg *et al*. screened more than 100000 random transformants, Vogl *et al*. characterized only 11 semi-rational constructs. Vogl *et al*. thereby obtained variants spanning a range of expression levels, including moderately improved variants, whereas Berg *et al*. (including parts of the URS) also achieved altered regulatory profiles (derepression to a so far unclear level of methanol induced wildtype P_{AOX1}) and a variant showing approximately two fold improved methanol induced levels of the wildtype promoter (LC-2 shown in Figure 1).

Engineering of alternative promoters

Beside the different engineering approaches focused on P_{AOX1}, also alternative promoters have been engineered in single studies (Figure 2). Tsustumi and Takagi engineered the strong methanol inducible P_{DAS2} promoter in a patent application [70]. In early promoter studies [8, 30], only a single *DAS* gene has been identified in *P. pastoris*. In a recent sequencing effort of the entire genome, De Schutter *et al*. [84] found a highly similar second *DAS* gene in reverse orientation next to the initially described one. A follow up sequencing effort by Küberl *et al*. [85] showed that due to the highly similar gene sequences, the promoter regions had been reversed by De Schutter *et al*. and they named

the corrected sequences *DAS1* and *DAS2*, whereas *DAS2* corresponds to the promoter sequence that had long been known [29]. In earlier publications, the single *DAS2* gene known had been referred to either as *DAS* or *DHAS* [8, 39] leading to no direct confusions. However in recent patent literature [70, 86] the P_{DAS2} sequence was used, but is referred to as P_{DAS1}.

Tsustumi and Takagi [70] performed truncations, internal deletions and added putative upstream activating sequences in multiple copies to P_{DAS2} similar to the studies of Xuan *et al*. with P_{AOX1} [24]. Using a secreted phytase as reporter gene, truncations led to a 155 % improved variant, whereas internal deletions led in some cases (e.g. pDd-14, 100 bp deleted) to a complete loss of function, suggesting the presence of an upstream activating sequence. Deleting an even smaller stretch of only about 20 bp in the pDd-14 region led also to a strong loss of reporter levels (24% of the P_{DAS2} wildtye). The putative regulatory region of pDd-14 was duplicated and fused in one to three copies to the truncated pDd2 variant (termed pDd-26 to pDd-28) leading to 108, 139 and 181 % improved expression compared to pDd2.

Qin *et al*. have performed a random mutagenesis over the whole length of P_{GAP} affecting simultaneously URSs and the core promote region [27]. After screening approximately 30000 transformants using eGFP as reporter, seven variants (G1 to G7) were further characterized with additional reporter genes (lacZ and a methionine adenosyltransferase) and RT-qPCR, obtaining a library spanning 0.6% to 19.6-fold of the wild-type promoter levels (Figure 2, B). However, in spite of the extraordinary strength of the improved P_{GAP} variants, no follow-up publications reproducing the almost 20 fold improvement have been published yet.

Similar to P_{AOX1}, engineering efforts of the constitutive P_{GCW14} have also only resulted in moderately improved variants (approximately 1.2 fold) [28]. Zhang *et al*. performed an analysis for promoter elements such as the TATA box and subsequently deleted these putative elements (Figure 2, C). In addition, the promoter sequence was truncated and random mutagenesis of the whole promoter length was performed. Approximately 3000 transformants were screened and 22 variants spanning a range of 20 to 120 % of the wild type promoter were further characterized showing between one and four mutations. The mutagenesis rates in this study were lower than in the P_{GAP} study and the authors noticed in several cases effects of single mutations (Figure 2, C). The best variant termed ‘M+20’ was also tested with secretory and cell surface display expression of CalB resulting in approximately 33 % higher activities.

Therefore the engineering efforts of P_{DAS2} and P_{GCW14} achieved comparable results to P_{AOX1}, studies, whereas the extremely high improvements obtained by Qin *et al*. for P_{GAP} stand apart.

Conclusion and outlook

Over more than two decades, most commonly the wildtype *AOX1* promoter and later on the *GAP* promoter have been applied to drive heterologous gene expression in *P. pastoris*. Recently, a series of promising new natural promoters and engineered variants of established promoters have emerged. These new tools enable fine-tuning of expression levels to optimize the expression of difficult to produce proteins and metabolic pathways. Although previously, methanol free, regulated systems have been suggested as alternatives to P_{AOX1}, none of these systems have been applied by a broader community. New natural promoters such as P_{GTH1} as well as synthetic P_{AOX1} variants offer novel

glucose induction or derepression based regulatory profiles that may succeed as a complementary system to the methanol based wildtype P_{AOX1}.

The basic understanding of *P. pastoris* promoters is still incomplete, yet synthetic biology and promoter engineering approaches have also aided to overcome the natural limitations of P_{AOX1}. Future efforts may rely on increasing the basic understanding of transcriptional regulation in *P. pastoris* and subsequently applying synthetic biology engineering approaches as demonstrated in *S. cerevisiae*.

Acknowledgement

This work was supported by the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 289646 (Kyrobio). We also gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz). The Austrian Centre of Industrial Biotechnology (ACIB) contribution was supported by FFG, bmvit, mvwfi, ZIT, Zukunftsstiftung Tirol and Land Steiermark within the Austrian COMET programme (FFG grant 824186). We would also like to thank Julia Pitzer for critical reading of the manuscript.

Figures

Figure 1

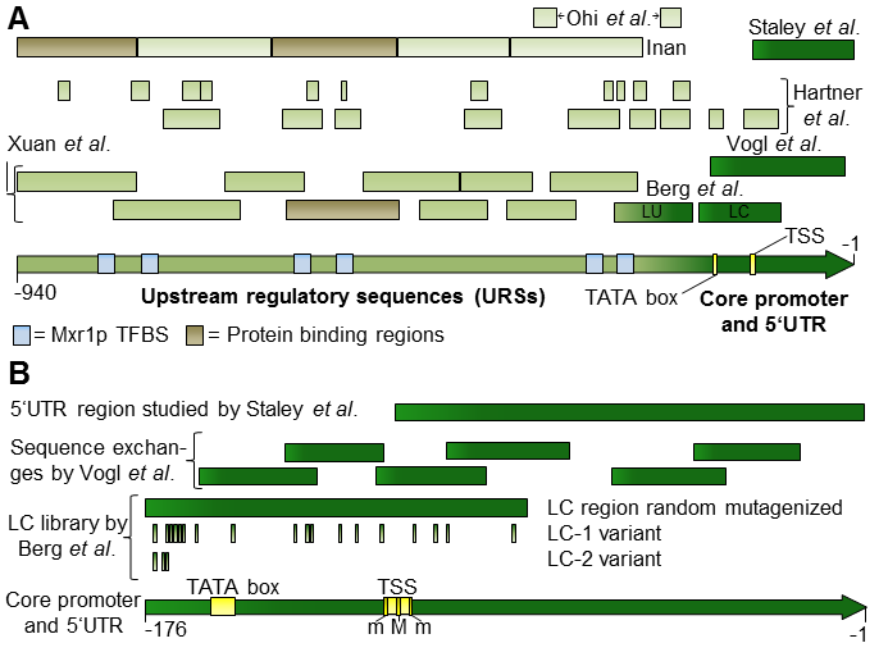


Figure 1: URS and core promoter engineering of the *AOX1* promoter.

(A) Overview on P_{AOX1} URS (light coloring) and core promoter/5'UTR (dark coloring) engineering approaches by Inan [69], Xuan *et al*. [24], Hartner *et al*. [23], Staley *et al*. [83], Berg *et al*. [25] and Vogl *et al*. [26]. Also two putative upstream activating sequences suggested by Ohi *et al*. from P_{AOX2} studies are shown [73]. The most commonly used length of P_{AOX1} is shown (940 bp), Inan [69] used a promoter that is 102 longer, therefore Inan fragment A is in this figure truncated. Protein binding regions identified by EMSA [24, 69] are indicated. Specific binding sites of the methanol master regulator Mxr1p as determined by Kranthi *et al*. [38] are shown. Numbering is relative to the translational start (+1).

(B) Detailed *AOX1* core promoter and 5' UTR engineering efforts. TATA box, major (M) and minor (m) transcriptional start sites (TSS) are shown [30]. Variants LC-1 and LC-2 obtained by random mutagenesis by Berg *et al*. showed initially 40 to 80 fold increased resistance to Zeocin, in a rescreening with an alternative reporter (firefly luciferase) expression reached an unclear level of methanol induced P_{AOX1} [25].

Figure 2

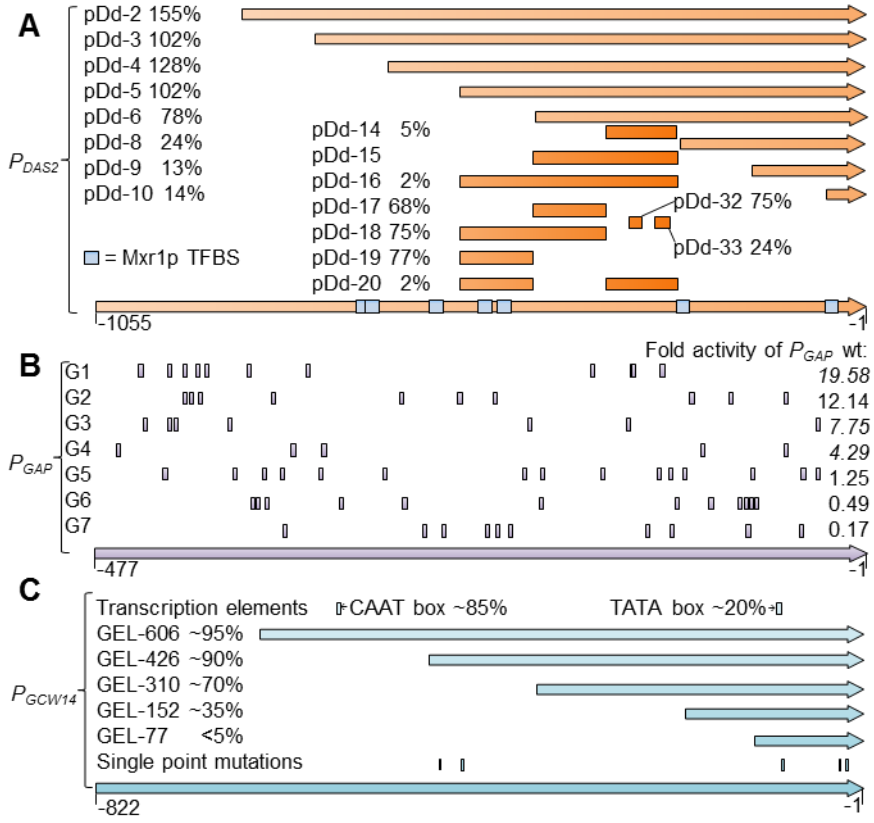


Figure 2: Promoter engineering of *DAS2*, *GAP* and *GCW14* promoters.

(A) P_{DAS2} engineering by Tsutsumi and Takagi [70]. Truncations pDd-2 to pDd-10 and internal deletions pDd14 to pDd-33 (darker shaded) are shown. The same naming as in the patent application [70] is used, therefore some constructs (e.g. pDd-7) are missing, as they were not provided. Percent activities of the wildtype promoter are given (for pDd-15 no information on the activity was provided). Mxr1p TFBS as determined by Kranthi *et al*. are also shown, the two most upstream binding regions are overlapping [39].

(B) P_{GAP} random mutagenesis variants characterized by Qin *et al*. [27]. Fold activity changes compared to the wildtype (wt) promoter using eGFP as reporter are given. The positions of point mutations are correct but not drawn to scale.

(C) P_{GCW14} engineering approaches by Zhang *et al*. [28]. Deletions of putative transcriptional elements, truncations (GEL-606 to GEL-77) and their effect on expression (percent of the wildtype promoter) are given. The single point mutations were characterized after screening a random library and led to expression levels between 21 and 128 % of the wildtype promoter. Numbering is relative to the translational start (+1).

Table 1

Table 1: Most commonly used and novel promoters in *P. pastoris*. Selected promoters are grouped by means of regulation. As the promoters were tested with different reporter genes and cultivation conditions, the indicated approximate expression levels are only a rough estimation. See [29] for a comprehensive list and detailed review of additional promoters. The terms 'inducible' and 'repressible' are somewhat analogous and here used to distinguish between conventional inducible promoters intended for high level protein production and the repressible promoters by Delic *et al*. [53] intended for down regulating the expression of essential genes for cell biological studies.

Regulation	Promoter	Gene name/function	Regulation	Strength	Reference
Inducible	P_{AOX1}	*Alcohol oxidase 1*	Methanol inducible, catabolite repressed	Strong (up to 22 g/l intracellular heterologous protein [34])	[8, 30]
	P_{AOX2}	*Alcohol oxidase 2*	Methanol inducible, catabolite repressed	~5–10 % of P_{AOX1}	[33]
	P_{DAS2}	*Dihydroxyacetone synthase 2*	Methanol inducible, catabolite repressed	Strong (even exceeding P_{AOX1})	[8, 29, 30]
	P_{FLD1}	*Formaldehyde dehydrogenase 1*	Methylamine and methanol inducible, catabolite repressed	Strong (similar to P_{AOX1})	[51]
	P_{GTH1} / P_{G1}*	*Glucose transporter with high affinity 1*	Repressed on glycerol, induced by low glucose concentrations	~150 to 230 % of P_{GAP}	[63]
	P_{PEX8}	Peroxisomal matrix protein	Oleate and methanol inducible, catabolite repressed	Weak	[46, 47]
	P_{PHO89} / P_{NSP}	Sodium-coupled phosphate symporter	Induced by low phosphate levels	Strong (similar to P_{GAP})	[52]
	P_{ScCUP1}	Metallothionein binding copper in *S. cerevisiae*, used as heterologous promoter in *P. pastoris*	Induced by copper, dependent on concentration	Not compared to P_{AOX1} or P_{GAP}	[54]
Constitutive	P_{GAP}	*Glyceraldehyde 3-phosphate dehydrogenase*	Constitutive, although affected by the carbon source	Strong, under optimal conditions similar to P_{AOX1}	[31]
	P_{GCW14}	Potential glycosyl phosphatidyl inositol (GPI)-anchored protein	Similar regulation pattern as P_{GAP}	Stronger than P_{GAP} (depending on the carbon source)	[28, 59]
	P_{TEF1}	*Translation elongation factor 1 alpha*	Constitutive, growth associated expression	Strong (similar to P_{GAP})	[48, 49]
	P_{TPI}	*Triose phosphate isomerase*	Constitutive	~10–80 % of P_{GAP}	[49]
	P_{PGK1}	*Phosphoglycerate kinase 1*	Constitutive	Weak (~10% of P_{GAP})	[49, 50]
Novel repressible promoters reported by Delic *et al*. [53]	P_{MET3}	ATP sulfurylase involved in methionine metabolism	Tightly repressed by methionine	~13 % of P_{GAP} (non-repressed)	[53]
	P_{PIS1}	*Phosphatidylinositol synthase 1*	Slightly repressed by inositol; moderately repressed by zinc sulfate (49 % of induced conditions)	~40 % of P_{GAP} (non-repressed)	[53]
	P_{SER1}	3-phosphoserine aminotransferase	Moderately repressed by L-serine (30 % of induced conditions)	~1 % of P_{GAP} (non-repressed)	[53]
	P_{THI11}	Involved in the synthesis of a thiamine precursor	Tightly repressed by thiamin	~63 % of P_{GAP} (non-repressed)	[49, 53]
	P_{THR1}	Homoserine kinase involved in threonine biosynthesis	Tight to slight repression depending on the combination of amino acids added (Thr, Ile, Val, Leu)	~13 % of P_{GAP} (non-repressed)	[53]

*The authors tested also additional, similarly regulated promoters with lower expression levels (P_{G3}, P_{G4}, P_{G6}, P_{G7} and P_{G8}) [63].

References

1. Blazeck J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* **8**, 46–58
2. Gellissen G (2000) Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* **54**,741–750
3. Hartner FS, Glieder A (2006) Regulation of methanol utilisation pathway genes in yeasts. *Microb Cell Fact* **5**, 39–59
4. Yurimoto H, Oku M, Sakai Y (2011) Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int J Microbiol*. doi:10.1155/2011/101298
5. Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. *Curr Opin Biotechnol* **24**, 1094–1101
6. Lin-Cereghino J, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* **24**, 45–66
7. Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in Pichia pastoris. *Mol Biotechnol* **16**, 23–52
8. Tschopp JF, Brust PF, Cregg JM, et al. (1987) Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. *Nucleic Acids Res* **15**, 3859–3876
9. Ruth C, Zuellig T, Mellitzer A, et al. (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. *Syst Biol Synth Biol* **4**, 181–191
10. Hohenblum H, Gasser B, Maurer M, et al. (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris. *Biotechnol Bioeng* **85**, 367–375
11. Gasser B, Saloheimo M, Rinas U, et al. (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact*. doi:10.1186/1475-2859-7-11
12. Mattanovich D, Gasser B, Hohenblum H, Sauer M (2004) Stress in recombinant protein producing yeasts. *J Biotechnol* **113**, 121–35
13. Ruth C, Glieder A (2010) Perspectives on synthetic promoters for biocatalysis and biotransformation. *Chembiochem* **11**, 761–765
14. Venter M (2007) Synthetic promoters: genetic control through cis engineering. *Trends Plant Sci* **12**, 118–124
15. Blazeck J, Garg R, Reed B, Alper HS (2012) Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnol Bioeng* **109**, 2884–2895
16. Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678–12683
17. Blazeck J, Liu L, Redden H, Alper H (2011) Tuning gene expression in Yarrowia lipolytica by a hybrid promoter approach. *Appl Environ Microbiol* **77**, 7905–7914
18. Krivoruchko A, Siewers V, Nielsen J (2011) Opportunities for yeast metabolic engineering: Lessons from synthetic biology. *Biotechnol J* **6**, 262–276
19. Hahn S, Young ET (2011) Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736
20. Blount BA, Weenink T, Vasylechko S, Ellis T (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*. doi:10.1371/journal.pone.0033279
21. Teo WS, Chang MW (2013) Development and characterization of AND-gate dynamic controllers with a modular synthetic GAL1 core promoter in Saccharomyces cerevisiae. *Biotechnol Bioeng*. doi:10.1002/bit.25001
22. Madzak C, Tréton B, Blanchin-Roland S (2000) Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast Yarrowia lipolytica. *J Mol Microbiol Biotechnol* **2**, 207–216

23. Hartner FS, Ruth C, Langenegger D, et al. (2008) Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Res.* doi:10.1093/nar/gkn369

24. Xuan Y, Zhou X, Zhang W, et al. (2009) An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. *FEMS Yeast Res* **9**, 1271–1282

25. Berg L, Strand TA, Valla S, Brautaset T (2013) Combinatorial mutagenesis and selection to understand and improve yeast promoters. *Biomed Res Int.* doi:10.1155/2013/926985

26. Vogl T, Ruth C, Pitzer J, et al. (2013) Synthetic Core Promoters for Pichia pastoris. *ACS Synth Biol.* doi:10.1021/sb400091p

27. Qin X, Qian J, Yao G, et al. (2011) GAP promoter library for fine-tuning of gene expression in Pichia pastoris. *Appl Env Microbiol* **77**, 3600–3608

28. Zhang X, Zhang X, Liang S, et al. (2013) Key regulatory elements of a strong constitutive promoter, P GCW14, from Pichia pastoris. *Biotechnol Lett* **35**, 2113–2119

29. Vogl T, Glieder A (2013) Regulation of Pichia pastoris promoters and its consequences for protein production. *New Biotechnol* **30**, 385–404

30. Ellis SB, Brust PF, Koutz PJ, et al. (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. *Mol Cell Biol* **5**, 1111–1121

31. Waterham HR, Digan ME, Koutz PJ, et al. (1997) Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186**, 37–44

32. Koutz P, Davis GR, Stillman C, et al. (1989) Structural comparison of the Pichia pastoris alcohol oxidase genes. *Yeast* **5**, 167–177

33. Cregg JM, Madden KR, Barringer KJ, et al. (1989) Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Mol Cell Biol* **9**, 1316–1323

34. Hasslacher M, Schall M, Hayn M, et al. (1997) High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. *Protein Expr Purif* **11**, 61–71

35. Werten MW, van den Bosch TJ, Wind RD, et al. (1999) High-yield secretion of recombinant gelatins by Pichia pastoris. *Yeast* **15**, 1087–1096

36. Couderc R, Baratti J (1980) Oxidation of Methanol by the Yeast, Pichia pastoris. Purification and Properties of the Alcohol Oxidase. *Agric Biol Chem* **44**, 2279–2289

37. Lin-Cereghino GP, Godfrey L, de la Cruz BJ, et al. (2006) Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. *Mol Cell Biol* **26**, 883–897

38. Kranthi BV, Kumar R, Kumar NV, et al. (2009) Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. *Biochim Biophys Acta* **1789**, 460–468

39. Kranthi BV, Kumar HRV, Rangarajan PN (2010) Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris. *Yeast* **27**, 705–711

40. Kumar NV, Rangarajan PN (2012) The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in Pichia pastoris. *J Biol Chem* **287**, 34465–34473

41. Parua PK, Ryan PM, Trang K, Young ET (2012) Pichia pastoris 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction. *Mol Microbiol* **85**, 282–298

42. Kim S, Warburton S, Boldogh I, et al. (2013) Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in Pichia pastoris. *J Biotechnol* **166**, 174–181

43. Polupanov AS, Nazarko VY, Sibirny AA (2012) Gss1 protein of the methylotrophic yeast Pichia pastoris is involved in glucose sensing, pexophagy and catabolite repression. *Int J Biochem Cell Biol* **44**, 1906–1918

44. Polupanov AS, Sibirny AA (2013) Cytoplasmic extension peptide of Pichia pastoris glucose sensor Gss1 is not compulsory for glucose signalling. *Cell Biol Int*. doi: 10.1002/cbin.10189

45. Kumar NV, Rangarajan PN (2011) Catabolite repression of phosphoenolpyruvate carboxykinase by a zinc finger protein under biotin- and pyruvate carboxylase-deficient conditions in Pichia pastoris. *Microbiol (Reading, U K)* **157**, 3361–3369

46. Liu H, Tan X, Russell KA, et al. (1995) PER3, a gene required for peroxisome biogenesis in Pichia pastoris, encodes a peroxisomal membrane protein involved in protein import. *J Biol Chem* **270**, 10940–10951

47. Sears IB, O'Connor J, Rossanese OW, Glick BS (1998) A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. *Yeast* **14**, 783–790

48. Ahn J, Hong J, Lee H, et al. (2007) Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. *Appl Microbiol Biotechnol* **74**, 601–608

49. Stadlmayr G, Mecklenbräuker A, Rothmüller M, et al. (2010) Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. *J Biotechnol* **150**, 519–529

50. De Almeida JRM, de Moraes LMP, Torres FAG (2005) Molecular characterization of the 3-phosphoglycerate kinase gene (PGK1) from the methylotrophic yeast Pichia pastoris. *Yeast* **22**, 725–737

51. Shen S, Sulter G, Jeffries TW, Cregg JM (1998) A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris. *Gene* **216**, 93–102

52. Ahn J, Hong J, Park M, et al. (2009) Phosphate-responsive promoter of a Pichia pastoris sodium phosphate symporter. *Appl Microbiol Biotechnol* **75**, 3528–3534

53. Delic M, Mattanovich D, Gasser B (2013) Repressible promoters -- A novel tool to generate conditional mutants in Pichia pastoris. *Microb Cell Fact.* doi:10.1186/1475-2859-12-6

54. Koller A, Valesco J, Subramani S (2000) The CUP1 promoter of Saccharomyces cerevisiae is inducible by copper in Pichia pastoris. *Yeast* **16**, 651–656

55. Paddon CJ, Westfall PJ, Pitera DJ, et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532

56. Zhang A-L, Luo J-X, Zhang T-Y, et al. (2009) Recent advances on the GAP promoter derived expression system of Pichia pastoris. *Mol Biol Rep* **36**, 1611–1619

57. Baumann K, Maurer M, Dragosits M, et al. (2008) Hypoxic fed-batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. *Biotechnol Bioeng* **100**, 177–183

58. Kern A, Hartner FS, Freigassner M, et al. (2007) Pichia pastoris "just in time" alternative respiration. *Microbiol (Reading, U K)* **153**, 1250–1260

59. Liang, S., Zou, C., Lin, Y., Zhang, X., & Ye, Y. (2013).Liang S, Zou C, Lin Y, et al. (2013) Identification and characterization of P GCW14: a novel, strong constitutive promoter of Pichia pastoris. *Biotechnol Lett* **35**, 1865–1871

60. Liang S, Wang B, Pan L, et al. (2012) Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. *BMC Genomics.* doi: 10.1186/1471-2164-13-738

61. Liu H, Qin Y, Huang Y, et al. (2013) Direct evaluating the effect of gene dosage on secretion of protein from yeast Pichia pastoris by expressing EGFP. *J Microbiol Biotechnol.* doi:10.4014/jmb.1308.08010

62. Bustin SA, Benes V, Garson JA, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611–622

63. Prielhofer R, Maurer M, Klein J, et al. (2013) Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris. *Microb Cell Fact.* doi:10.1186/1475-2859-12-5

64. Hobl B, Hock B, Schneck S, et al. (2013) Bacteriophage T7 RNA polymerase-based expression in Pichia pastoris. *Protein Expr Purif* **92**, 100–104

65. Fuerst TR, Niles EG, Studier FW, Moss B (1986) Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* **83**, 8122–8126

66. Yu J-Y, DeRuiter SL, Turner DL (2002) RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc Natl Acad Sci U S A* **99**, 6047–6052

67. Khalil AS, Lu TK, Bashor CJ, et al. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658

68. Alper H, Moxley J, Nevoigt E, et al. (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568

69. Inan M. Studies on the alcohol oxidase (AOX1) promoter of Pichia pastoris. Dissertation. Lincoln, Nebraska: University of Nebraska; 2000

70. Tsutsumi N, Takagi S, inventors; Novozymes A/S, Bagsvaerd (DK), assignee. Pichia pastoris DAS promoter variants. United States patent application publication US 2011/0129874 A1. 2011 Jun 2.

71. Allison LA (2007) Transcription in eukaryotes. In: Allison LA (ed) Fundamental Molecular Biology, Blackwell, London, pp. 312–391.

72. Lubliner S, Keren L, Segal E (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res* **41**, 5569–5581

73. Ohi H, Miura M, Hiramatsu R, Ohmura T (1994) The positive and negative cis-acting elements for methanol regulation in the Pichia pastoris AOX2 gene. *Mol Gen Genet* **243**, 489–499

74. Inan M, Meagher M, Benson AK, inventors; The Board of Regents of the University of Nebraska (Lincoln, NE), assignee. Alcohol oxidase 1 regulatory nucleotide sequences for heterologous gene expression in yeast. United States patent US 6,699,691 B2. 2004 Mar 2.

75. Inan M, Meagher MM (2001) Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. *J Biosci Bioeng* **92**, 585–589

76. Mellitzer A, Weis R, Glieder A, Flicker K (2012) Expression of lignocellulolytic enzymes in Pichia pastoris. *Microb Cell Fact.* doi:10.1186/1475-2859-11-61

77. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* **16**, 7351–7367

78. Bryksin A, Matsumura I (2013) Overlap extension PCR cloning. *Methods Mol Biol* **1073**, 31–42

79. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946–950

80. Lee CC, Williams TG, Wong DWS, Robertson GH (2005) An episomal expression vector for screening mutant gene libraries in Pichia pastoris. *Plasmid* **54**, 80–85

81. Daly R, Hearn MTW (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18**, 119–138

82. Cregg JM, Tolstorukov I, Kusari A, et al. (2009) Expression in the yeast Pichia pastoris. *Methods Enzymol* **463**, 169–189

83. Staley CA, Huang A, Nattestad M, et al. (2012) Analysis of the 5' untranslated region (5'UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in Pichia pastoris. *Gene* **496**, 118–127

84. De Schutter K, Lin Y-C, Tiels P, et al. (2009) Genome sequence of the recombinant protein production host Pichia pastoris. *Nat Biotechnol* **27**, 561–566

85. Küberl A, Schneider J, Thallinger GG, et al. (2011) High-quality genome sequence of Pichia pastoris CBS7435. *J Biotechnol* **154**, 312–320

86. Takagi S, Tsutsumi N, Terui Y, Kong XY, inventors; Novozymes A/S, Bagsvaerd (DK), assignee. Method for methanol independent induction from methanol inducible promoters in Pichia. United States patent US 8,143,023. 2012 Jun 28

Chapter 3.1
Pichia pastoris methanol utilization pathway promoters are different in sequence but share common technological characteristics

Title page

Title

***Pichia pastoris* methanol utilization pathway promoters are different in sequence but share common technological characteristics**

Author names

Thomas Vogl#, Lukas Sturmberger#, Thomas Kickenweiz#, Richard Wasmayer, Gerhard Thallinger, Martina Geier, Anton Glieder*

# These authors contributed equally.
* Corresponding author

Abstract

Background
The heterologous expression of metabolic pathways for pharmaceutical or fine chemical production requires a suitable expression host. The methylotrophic yeast *Pichia pastoris* is a commonly used protein production host and single genes have been expressed at high levels using the methanol inducible, strong and tightly regulated promoter of the *alcohol oxidase 1* gene (P_{AOX1}). In the heterologous expression of multi-enzyme pathways, the flux is typically balanced by stoichiometric fine-tuning of reaction steps by varying the transcript levels of the genes involved. Consequently a set of tightly regulated, yet sequence-wise distinct promoters is needed.
Results
Here, we have studied the complex regulation of the entire methanol utilization (MUT) pathway using transcriptome analyses and by assaying 45 promoters with reporter gene measurements. We noticed a pronounced involvement of the pentose phosphate pathway (PPP) and genes involved in the defense of reactive oxygen species (ROS), providing also strong promoters which in part even outperform P_{AOX1} and offer novel regulatory profiles. We identified 15 tightly methanol regulated promoters, spanning a range of 10 to 133 % of the P_{AOX1} benchmark. Half of the promoters reach more than 50 % of P_{AOX1}. Several promoters showed derepression (activation when the repressing carbon source is depleted) to varying extents and the promoter of the *catalase 1* gene could also be induced with oleic acid, offering an alternative to toxic and flammable methanol.
Conclusion
P. pastoris offers one of the largest sets of strong, co-regulated promoters, still showing pronounced differences in their particular regulation and high sequence diversity. Alongside recently developed genome-scale metabolic models, these promoters may propose *P. pastoris* as a powerful platform for metabolic engineering endeavors.

Keywords

Transcriptional fine-tuning; promoters; microarray; *Pichia pastoris*

1. Introduction

Metabolic pathways have been engineered and exploited for biofuel, pharmaceutical or fine chemical production and are commonly heterologously expressed in microbial host organisms. However, simple coexpression of the genes of a pathway is seldom sufficient to achieve high yields and productivity, typically requiring an optimization of the flux towards the desired product and the removal of kinetics bottlenecks [1,2].
Natural regulation of pathways is exerted at different levels ranging from transcription over translation to the protein level [3]. In recombinant protein expression in eukaryotes, most commonly transcript levels are varied by employing different promoters. Ideally, promoters for metabolic engineering should provide tight regulation by induction to enable a separation of cell growth from pathway expression and to avoid a constant additional metabolic burden [3]. Promoters covering a wide range of expression levels should be at hand to enable expression fine-tuning, ranging from tight down-regulation to high overexpression. To this end, either natural or synthetic promoters can be used. Synthetic promoters provide typically a wider range of expression levels (10 to 1000 fold) and finer increments of expression [3]. Commonly, promoter libraries are obtained from modifying a single natural sequence and the final variants vary only slightly between their sequences (e.g. [4,5]).
Fine-tuning the expression of a multi gene pathway using such libraries is troubled by the high identities of the sequences: On the one hand highly similar sequences complicate the *in vitro* DNA assembly of pathways when using overlap-directed DNA assembly methods such as Gibson assembly [6], as identical sequences can misalign. On the other hand similar sequences may lead to homologous recombination *in vivo* and loss of parts of the expression cassettes (by loop out recombination, [7] and references therein).

Concerning host platforms for metabolic engineering, most commonly *Escherichia coli* and *Saccharomyces cerevisiae* are used [8]. Due its long use as a basic eukaryotic model organism and the large fundamental knowledge, the 'classic' yeast *S. cerevisiae* is especially for eukaryotic pathways the standard platform [9–11]. However, *S. cerevisiae* provides a limited set of tightly co-regulated promoters, typically only few galactose (P_{GAL1}, P_{GAL3}, P_{GAL7}, P_{GAL10}) and copper (P_{CTR1}, P_{CTR3}, P_{CUP1}) regulated promoters are used [10,12].
Recently alternative, nonconventional yeasts have also attracted considerable attention having been successfully applied for various metabolic engineering endeavors [13]. The methylotrophic yeast *Pichia pastoris* is one of the most commonly used expression hosts for heterologous protein production, due to beneficial traits such as growth to exceptionally high cell densities (>150 g dry cell weight per liter [14]) and high yields of secreted proteins. A recent literature survey on recombinant gene expression suggests that *P. pastoris* is even more frequently applied for single protein production than *S. cerevisiae* [15]. Lately *P. pastoris* has also been used for an increasing numbers of metabolic engineering applications [16–24] supported by the development of genome scale metabolic models (GSMMs) [25–27]. Recently metabolic models have been revised on an improved annotation of the *P. pastoris* proteome [28]. Yet, despite recent synthetic biology approaches [29], the fundamental knowledge on *P. pastoris* is small compared to *S. cerevisiae* and also less molecular tools such as promoters and terminators are available [30].

In *P. pastoris* and related methylotrophic yeasts (*Hansenula polymorpha*, *Candida boidinii*, *Pichia methanolica*) typically methanol inducible promoters are used to drive heterologous gene expression [31–35]. The most commonly used *P. pastoris* promoter of the *alcohol oxidase 1* gene (P_{AOX1}) is tightly repressed on carbon sources such as glucose, glycerol and ethanol (*AOX1* mRNA undetectable), while reaching about 5 % of total mRNA and 30 % of total soluble protein upon methanol induction [34].

Despite the annotation of a several dozen genes putatively involved in the methanol metabolism [36–38] alongside transcriptomics [39,40] and proteomics studies [41,42] hinting targets of additional strong, methanol regulated promoters, so far only five methanol regulated promoters have been tested in *P. pastoris* (strong: P_{AOX1}, P_{DAS2}, P_{FLD1}; weak: P_{AOX2}, and P_{PEX8} [34]). Also in related methylotrophic yeasts transcriptomics studies have been performed [43,44], yet again only few promoters have been systematically characterized (*e.g.* five promoters in *C. boidinii* [45]).

Here we report a comprehensive characterization of the *P. pastoris* methanol utilization (MUT) pathway, providing a large set of tightly regulated, sequence-wise distinct promoters offering a wide range of expression levels (45 promoters tested). In frame of this work we noticed a distinct, carbon source dependent regulation of isoenzymes involved in steps of the central carbon metabolism (pentose phosphate pathway (PPP), glycolysis/gluconeogenesis), hinting an explanation for some issues of current genomic scale metabolic models (GSMMs).

2. Materials and methods

2.1 Strains, plasmids, chemicals and media

The *P. pastoris* CBS7435 wildtype strain was used for most expression studies. Only CalB and HRP were expressed in expression CBS7435 *mut*^{s} strain, as higher yields compared to the wildtype have previously been reported in the literature [20]. The reporter plasmids were based on the Zeocin selection based pPpT4_S vector reported by Näätsarri *et al.* [46]. See Supplementary materials S 1 for a detailed description of plasmid construction and cloning of the promoters (primer sequences provided in S 2).
Chemicals, enzymes, cloning kits and *E. coli* cultivations were used and performed as recently described [47]. *P. pastoris* media for standard deep well plate cultivations (glucose cell growth phase followed by methanol induction) were prepared as reported by Weis *et al.* [48]: buffered minimal dextrose (1% w/v) (BMD) and buffered minimal methanol medium (BMM) with 0.5% (v/v) methanol. For growth on alternative carbon sources 1% (w/v) glycerol (BMG), 1% (v/v) ethanol (BME), 1% (w/v) mannitol (BMMan), and 0.2% (w/v) oleic acid supplemented with 0.02% (v/v) Tween40 [49] (BMO) were used. For the cultivation of horse radish peroxidase (HRP) expression strains hemin was added to a final concentration of 25 µM to the starting medium and in the medium for the first induction step (Krainer *et al*., accepted manuscript).

2.2 *P. pastoris* transformations and screening

P. pastoris was transformed with *SwaI* linearized plasmids according to the condensed protocol of Lin-Cereghino *et al*. [50]. Low amounts of DNA were transformed to avoid multi copy integrations biasing comparisons. One µg of pPpT4_S is typically yielding only single copy transformants ([47] and unpublished results). Therefore equimolar amounts of the vectors equaling one µg of pPpT4_S were transformed. Transformant selection was performed as previously reported [47] to avoid clonal variation [51,52]. In short, 42 transformants were screened, typically this landscape showed uniform expression, except for a few transformants showing no expression or elevated expression. Three transformants from the linear range of the landscape were streaked for single colonies and confirmed by a rescreening for uniform expression. One representative transformant was used for the further work.

2.3 Cultivation conditions

DWP cultivations were performed following the protocol reported by Weis *et al*. [48]. In short, cell material from single colonies was inoculated into 96 well DWPs and cultivated for 60 h on BMD (250 µl). Subsequently an equal amount of BMM2 (1% methanol v/v to achieve a final concentration of 0.5%) was added. The cells were additionally induced with 50 µL BMM10 medium (5% methanol v/v) 12 h, 24 h and 48 h after the first induction. Samples were taken and measured at the indicated time points.

Shake flask cultivations were performed in 250 ml baffled flask (25 ml BMD starting volume) and inoculated to a starting OD_{600} of 0.05. The flasks were induced after 48 h with 25 ml BMM2 and after 12 h, 24 h after the first induction with BMM10. Glucose concentrations were measured using a hexokinase method based kit (Glucose UV kit, DIPROmed, Vienna, Austria). One ml of the reaction solution was mixed with 10 µl sample and incubated for 10 min at room temperature. The resulting NADH signal was measured at 340 nm and compared to a calibration curve.

Strains for the microarray experiments were cultivated in 1.5 l bioreactors. Due to precipitation occurring with standard *P. pastoris* bioreactor cultivation media (modified basal salt medium based on [53]), which might complicate RNA isolations and OD_{600} measurements, we used buffered minimal dextrose medium for all bioreactor cultivations (20 g/L glucose, 13.4 g/L BD Difco yeast nitrogen base (Franklin Lakes, NJ, USA) and 200 mM potassium phosphate buffer (pH 6.0)).

A *P. pastoris* strain bearing a pPpT4 based Zeocin resistance plasmid was used, to have a reference if the data should be compared to heterologous protein expressing strains. The 1.5-L fedbatch-pro® bioreactor system (DASGIP AG, Juelich, Germany) containing 600 mL of BMD medium were inoculated to an OD_{600} of 0.25. All cultivations were started in biological triplicates with a batch phase on glucose as sole carbon source at 28 °C and aeration at 0.7 L air/min. Agitation was set between 500 and 1200 rpm to keep oxygen saturation at 30 %. As a pH-control agent and nitrogen source 25 % ammonia solution was used.

After the entire glucose was consumed (shown with glucose detection strips (Combur Test strips, Rotkreuz, Switzerland)), methanol induction was started by addition of 0.5 % methanol. After methanol was consumed (deduced from differential nitrogen feed and oxygen consumption levels) reinduction at 1 % glucose was performed. The samples for RNA isolation were taken at the time points indicated in (Fig. 2A). All samples were centrifuged immediately after collection and stored in RNAse later solution (Life technologies, Carlsbad, CA, USA) at -80°C until further processing.

2.4 Microarray cultivations

Total RNA was isolated using a RiboPureTM Yeast RNA kit (Ambion, Austin, TX, United States). Purity and integrity were assessed with an Agilent Bioanalyzer 2100 using the RNA 6000 Nano Assay kit (Agilent Technologies, CA, United States). For reverse transcription and labelling an Affymetrix GeneChip 3'IVT Express Kit (Affymetrix, Santa Clara, United States) was used with an initial RNA amount of 450 ng. First strand and second strand synthesis were performed according to protocol at 40°C for 4 hours. Quantities and size distribution were again assayed with an Agilent Bioanalyzer 2100. Both for fragmentation and hybridization mix preparation 15 µg of DNA were used and incubated with custom Affymetrix microarray chips as previously reported in detail [55]. Analysis was performed with the new annotation of the CBS 7435 strain [38].

2.5 Reporter activity measurements

eGFP fluorescence measurements were performed in micro titer plates (Nunc MicroWell 96-Well Optical-Bottom Plates with Polymer Base, Black; Thermo Fisher Scientific) using a Synergy MX plate reader (Biotek, Winooski, VT, USA) [47]. Fluorescence was measured at 488/507 nm

(excitation/emission). Fluorescence measurements were normalized per OD_{600} measured to account for different dilution factors required to stay within the linear range of the plate reader.
HRP activity was measures using 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) as substrate as described in the literature [20,54]. CalB activity was measured using 4-nitrophenyl butyrate (Sigma-Aldrich, Vienna, Austria) as substrate as described in the literature [20].

3. Results and discussion

3.1. Genome wide transcriptional response towards different carbon sources

We hypothesized that the complex reactions of the MUT pathway of *P. pastoris* (Fig. 1) and other methylotrophic yeasts should encompass a large network of transcriptionally coregulated genes, offering tightly regulated promoters that can be harnessed for heterologous pathway expression. Typical considerations on the MUT pathway in *P. pastoris* have been limited to enzymes catalyzing direct reactions of methanol [20,37,55]. This `canonical´ MUT pathway is divided into a dissimilative branch (oxidation of methanol to CO_2 and the generation of NADH for respiratory ATP production) and an assimilative branch (to produce biomass by formaldehyde fixation) (Fig. 1).
However, recent transcriptome analyses suggest also non-canonical parts of associated processes (e.g. peroxisome biogenesis, stress response, respiratory function) [39,43,44]. Concerning stress response, the defense against reactive oxygen species (ROS) is crucial as the oxidation of methanol to formaldehyde creates equimolar amounts of reactive hydrogen peroxide [56,57].
Previous transcriptomics studies of methylotrophic yeasts were based on heterologous DNA microarray hybridization lacking MUT genes [40], different *P. pastoris* strains [39,40] or *H. polymorpha* [43,44] and only two conditions were compared (growth on glycerol vs. methanol). Yet typical MUT promoters show three levels of regulation [35]: 1) Repression in the presence of repressing carbon sources (in *P. pastoris* glucose, glycerol and ethanol). 2) Derepression once the repressing carbon source is depleted. 3) Induction by methanol.

Here, we used custom Affymetrix microarray chips [58] to compare the transcriptional response of *P. pastoris* grown in bioreactors under glucose repressed, derepressed, methanol induced and glucose readdition conditions (Fig 2A, S 3). The strong transcriptional changes between glucose and methanol resembled closely effects reported in previous studies between glycerol and methanol and have been comprehensively discussed [39,43,44]. However, gene regulation under previously uninvestigated derepressed conditions was more different from methanol induction and growth on glucose, than methanol and glucose from one another (Fig 2B,C), indicating vast transcriptional changes. Notably, already when preparing amplified RNA (aRNA) for the array hybridization we noticed lower yields of derepressed conditions and altered capillary electrophoresis migration patterns compared to the other samples (S 4). Compared to gene expression on glucose, 75 % of genes were significantly ($p<0.01$) differentially regulated under derepressed conditions (4413 of 5869 probe sets), with equal numbers of up- and down-regulated genes. Amongst biological processes (classified by COG terms [59]), especially genes coding for proteins involved in translation, RNA processing and modification, cytoskeleton, nucleotide transport and cell cycle control were downregulated (Fig 2D). Together with an upregulation of genes coding for proteins involved in defense mechanisms, cell wall and extracellular structures, this response is in line with the anticipated cellular reaction towards nutrient depletion and adaption to the stationary phase.

We investigated also the readdition of glucose to methanol grown cultures, to study mRNA turnover and create a basic data set relevant for pexophagy studies. Methylotrophic yeasts are model systems for peroxisome biogenesis and degradation [60] and similar shifts from methanol to glucose are commonly performed to trigger pexophagy (e.g. [61]). The transcriptional response of glucose

readdition appeared somewhat as an intermediate between the response towards glucose and methanol: Transcript levels of MUT genes were upregulated compared to glucose, yet not as high as on methanol, suggesting partial degradation of the mRNAs. Also sugar transporters were strongly upregualated (e.g. *HGT1* 557-fold), as reported in a study comparing transcript levels between growth on glycerol and glucose [62]. Downregulated genes compared to methanol included peroxisomal proteins (e.g. *PEX5*, *PEX6*, *PEX10*, *PER3*), which would be in line with anticipated pexophagy under glucose readdition conditions. However, the exact effects seen upon glucose readdition will depend on the time when RNA is isolated. We sampled after two hours, the same time interval applied for sampling after methanol induction (Fig 2 A). Therefore our experiment provides a large dataset, yet the most interesting gene candidates for further studies should however be verified by a time series *e.g*. by RT-qPCR.

Although MUT genes such as *AOX1* have been reported to be only slightly derepressed (2-4% of induced values) [34], they were amongst the most strongly upregulated genes in the derepressed set of the microarray experiments (Fig 2E, S 3). *AOX1* was for example 39-fold upregulated under derepressed conditions compared to presence of glucose. We assume that these surprisingly high values arise from the pairwise microarray comparison where the fold change (FC) between two conditions is calculated. If a gene is not expressed under a certain condition, the background noise will determine the FC value. In previous studies, the *AOX1* mRNA was undetectable on glucose and slightly derepressed upon glucose depletion (2-4% of induced values) [34,63,64]. Calibrating the moderate signal from derepression to the very low background expression is resulting in a high FC value. To this end, the FC values relative to repressed conditions have to be treated with care when analyzing putatively repressed genes. In these cases we based our interpretations on the reporter protein experiments (Fig. 3).

3.2. Reporter gene analysis of MUT promoters

Microarray results give only a relative fold change and no information on the actual absolute expression level. Therefore all promoters of canonical MUT pathway genes (following the most recent annotation/model [20,37]) and associated, non-canonical pathways (PPP, ROS defense, selected peroxisomal proteins and glycolysis/gluconeogenesis enzymes) were cloned upstream of an enhanced green fluorescent reporter protein (eGFP) and assayed for reporter protein fluorescence under various growth conditions (Fig. 3). The promoter of the *glyceraldehyde 3-phosphate dehydrogenase* gene P_{GAP} was included as reference of a strong constitutive *P. pastoris* promoter [34].

The length of yeast promoters can vary considerably (the median length in *S. cerevisiae* being 455 bp) and for promoter comparisons lengths of 1000, 800 and 600 bp have been used [65]. For assembling and expressing pathways, short promoters are desirable as they reduce the plasmid size (facilitating transformation) and mutations in PCR amplification are less likely to occur.

We have selected the promoter sequences as the distance from the start codon to the respective upstream gene (Supporting Table S 2 for gene names and promoter lengths tested and Supporting Figure S 5 for genomic organization). Promoter lengths reported in literature were also tested, in case of ambiguous sequence information (e.g. multiple ORFs, putative annotations) 1000 bp promoter length were used (S 6).

All promoters were seamlessly fused to the reporter gene (*i.e*. the transition from promoter to start codon is native, without additional restriction enzyme recognition sites in between). High throughput 96 deep well plate cultivations [48] of the 45 strains allowed to easily assay expression on different carbon sources and at different time points (Fig. 3, S 7).

We used protocols mimicking typical two-phase fed-batch bioreactor cultivations (Fig. 3A): The strains were at first grown on glucose until depletion and then induced with methanol over 72 h (full time series shown in S 7). Samples were taken during glucose repressed, derepressed and methanol

induced conditions (similar to the microarray cultivations shown in Fig. 2A). In methylotrophic yeasts species, the response towards alternative carbon sources is variable [35] and effects on *MUT* promoters have never been systemically assayed. Therefore we tested in addition growth on commonly used alternative carbon sources (glycerol, ethanol, oleic acid, mannitol) (Fig. 3 D). In cases where different lengths of the promoters were tested, no effects on reporter gene fluorescence were noticed (Supplementary figure S 6, only the shortest lengths are shown in Fig. 3).

Several canonical MUT promoters showed strong methanol inducible reporter fluorescence reaching at least half of the commonly used *AOX1* promoter (P_{FLD1}, P_{FDH1}, P_{DAS1}, P_{DAS2}). P_{DAS2} even outperformed P_{AOX1} as reported in early studies [66]. Some promoters showed weaker methanol inducible expression (P_{FGH1}, P_{DAK1}, P_{PEX5}, P_{FBA2}, P_{AOX2}). Further promoters showed constitutive expression (P_{ADH2}, P_{TPI1}, P_{FBP1}, P_{PGI1}) as in part previously reported [34], however clearly lower than the commonly used constitutive *GAP* promoter. Promoters of genes involved in ROS defense showed varying results: P_{CAT1} and P_{PMP20} showed strong expression on methanol, a few promoters showed low to intermediate expression (P_{SOD2}, P_{SOD3}, P_{MSR1c3}). However, several ROS promoters did not show any detectable reporter gene fluorescence, suggesting either no expression or low expression under the detection limit. In case of the five methionine sulfoxide reductase promoters tested, only a single promoter showed clear reporter gene fluorescence, suggesting that *MSR1c3* is the major methionine sulfoxide reductase in *P. pastoris*. PPP promoters showed a strong variability in regulatory profiles and expression levels, implications thereof are discussed below.

Cultivations on alternative carbon sources did not show remarkable general trends (Fig. 3B), but certain promoters were specifically regulated in correlation with the function of their natural gene product: P_{ADH2} (alcohol dehydrogenase) was clearly upregualted on ethanol, P_{CAT1} was strongly induced on oleic acid (presumably owing to the detoxification function of catalase of hydrogen peroxide arising from beta oxidation of fatty acids). Canonical MUT genes were only slightly upregulated on oleic acid compared to glucose, suggesting that the regulation thereof is not overlapping.
The trends observed in the microarray experiments (Fig. 2E) correlated with the output of the reporter gene constructs. A notable exception is the derepressed phase, where microarrays gave higher fold changes than could be deduced from the reporter fluorescence measurements. *AOX1* was for example in microarrays under derepressed conditions 39-fold upregulated compared to glucose, whereas reporter fluorescence did not increase over background level.
These differences may have arisen form the different cultivation conditions as bioreactor cultivations were used for the microarray experiment but small scale DWP cultivations for the reporter genes. In a recent study [67], P_{AOX1} driven expression of several heterologous genes reached in bioreactors substantial levels before methanol induction. In shake flasks this effect was not apparent. Bawa *et al*. did not discriminate explicitly between repressed and derepressed conditions, but the cultivations were left for several hours under carbon source depleted conditions prior to methanol induction. This would suggest a different extent of derepression between small scale and bioreactor cultivations. Bawa *et al*. discuss different oxygenation as possible explanation. Aeration is also in our experiments the major difference between the small scale cultivations and bioreactor cultivations. In the related methylotrophic yeast *C. boidinii* the expression of methanol inducible peroxisomal enzymes was dependent on mitochondrial respiratory function [68] and the expression of *P. methanolica* alcohol oxidases was also regulated by the oxygen levels [69]. It seems plausible that also expression in the derepressed phase could also be affected by oxygen levels. Also in another *P. pastoris* bioreactor study following GFP expression over time, an effect of the oxygen level on methanol induction was noticed [70]. However, no expression before methanol induction was apparent. Therefore understanding the exact effect of derepression in bioreactors will require further studies.

Alternatively, these differences may be attributable to the issues arising from pairwise microarray comparison of very tightly repressed genes, as suggested above. Therefore technical issues or cultivation conditions could be affecting the interpretation of the derepressed condition. Small scale cultivations appear to underestimate the derepression effect. So we rely for the interpretation more on the small scale reporter gene results, than the microarray data, to avoid overestimation of derepression effects.
We placed in the analysis special emphasis on derepression, regulation of isoenzymes and sequence features.

3.3 P_{CAT1} shows strong derepressed expression and matches P_{AOX1} also for some specific examples

In *P. pastoris*, so far no natural derepressed MUT promoters are known, but synthetic P_{AOX1} variants have been reported, enabling methanol free processes solely regulated by glucose/glycerol levels [5,71]. In related methylotrophic yeasts derepressed expression levels may reach up to 70 % of methanol induced expression [35], allowing simple, methanol free expression. Therefore we placed in the analysis special emphasis on the glucose repression and derepression phase. Derepression effects between MUT promoters varied considerably: Several promoters such as P_{AOX1}, P_{DAS1}, P_{DAS2}, P_{PMP20}, P_{TAL2}, P_{FBA2} showed very tight regulation with no derepressed reporter protein fluorescence detectable. P_{FGH1}, P_{DAK1}, P_{FLD1} and P_{FDH1} showed slight to intermediate derepression (Fig. 3).

The promoter of the *CAT1* gene showed tight repression on glucose and highest derepression (29 % of the strong constitutive P_{GAP}) and could be induced with methanol to similar levels as P_{AOX1} (Fig. 3A,E; Fig. 4). Therefore we focused on P_{CAT1} as representative promoter to evaluate the suitability of derepressed promoters for protein production in *P. pastoris*. P_{CAT1} is also the only *P. pastoris* MUT promoter, which can also be induced with oleic acid to similar levels as on methanol (Fig. 3E). Noteworthy, we experienced with the first tested promoter lengths $P_{CAT1\text{-}692}$ and $P_{CAT1\text{-}1000}$ constructs (S 6) problems of transformation background (colonies loosing expression when recultivated). The default length of P_{CAT1} for all standard applications should be 500 bp.

However, P_{CAT1} expression under derepressed conditions was here only inferred from the fluorescence levels of the eGFP reporter protein. Cytoplasmic eGFP is easily folded and well tolerated by *P. pastoris* [5,47,72]. Heterologous protein production in *P. pastoris* is preferably achieved by secretion, as downstream processing is facilitated (no cell breakage *etc*. required). Therefore we tested secretory expression of two industrially relevant enzymes *Candida antarctica* lipase B (CalB) and horseradish peroxidase (HRP) fused to the commonly used mating factor alpha secretion signal sequence. Under derepressed conditions we obtained 35 % (CalB) and 21 % (HRP) volumetric activity of the constitutive state of the art promoter P_{GAP} (Fig. 4B,C). Methanol induction of P_{CAT1} and P_{AOX1} showed a strong time dependent effect with activities considerably increasing over time. P_{CAT1} showed for both enzymes a faster response than the respective P_{AOX1} constructs. After 72 h of methanol induction, P_{CAT1} reached similar volumetric yields as in P_{AOX1} driven HRP production. For CalB production, P_{CAT1} even outperformed P_{AOX1} driven production 2.8 fold. For eGFP expression, reporter gene fluorescence changed only marginally after 24 h of methanol induction (S 7), suggesting a specific effect on secretory proteins. Judging from the results of eGFP, CalB and HRP the promoter used can have a strong effect on expression, suggesting to compare promoters with different regulatory profiles for optimal yields. These effects may be caused by effects on overall mRNA levels, degradation or even translation initiation (since also the 5'untranslated regions are different).
Considering space-time-yields, the shorter cultivation time of derepressed P_{CAT1} yielded after 60 h 44 % (CalB) or 32 % (HRP) of methanol induced P_{AOX1} after 132 h. These results suggest that

derepressed expression is also possible with secretory protein expression, yet reaching lower productivity than methanol induced processes in small scale cultivations. Secretion typically exerts more stress on the cell and may induce the UPR (unfolded protein response) or ERAD (Endoplasmic-reticulum-associated protein degradation) [73], resulting in higher metabolic demands. Carbon-source depleted growth conditions may affect secretory protein expression stronger than simple cytoplasmic expression resulting in lower relative yields.

In bioreactor cultivations, derepressed conditions have also been exerted by a limited carbon source feed (as demonstrated for synthetic derepressed P_{AOX1} variants [5,71]), providing the basis for more efficient expression, than in the small scale cultivations applied here. If desired, it should also be possible to extend the derepressed phase in small scale cultivations using a continuous enzymatic release of low levels of glucose from a glucose-based polymer, as reported for P_{AOX1} applications [74]. Alternatively, also oleic acid based feed strategies could be applied, since P_{CAT1} was induced to similar levels on oleic acid as on methanol (Fig. 3B).
However, for basic screening and comparison of transformants, derepressed P_{CAT1} driven expression, worked sufficiently without supplementary feeding, still shortening the screening protocol by 72 h (60 instead of 132 h) since no lengthy methanol induction phase is required.

In respect of the nutrient free derepressed phase, one major question is arising: Where does the energy come from? We presume that during exponential growth on glucose, *P. pastoris* is building energy depots (mostly lipids), that are subsequently liberated and used for adaption to carbon source depleted conditions/the stationary phase. Also autophagic processes may be involved, as also the adaption of *P. pastoris* to methanol is associated with autophagy [42,75].
Concerning its molecular regulation, P_{CAT1} could be induced by different stimuli and transcription factors: The peroxisomal *S. cerevisiae CAT1* homolog is regulated by carbon source-responsive transcription factors (Adr1p, Oaf1p) [76], correlating with *P. pastoris* P_{CAT1} activation after glucose depletion in our experiments (Fig. 4 A). However, also processes associated with the adaption to glucose depletion may be indirectly involved in P_{CAT1} regulation. The primary function of Cat1p after glucose depletions is detoxification of H_2O_2, arising from ROS stress due to adaption to the stationary phase (e.g. fatty acid beta-oxidation results in considerable amounts of H_2O_2). Therefore also ROS could act as activating stimulus of P_{CAT1}, similar to the cytosolic catalase (*CTT1*) of *S. cerevisiae* which is activated by various stress conditions and transcription factors such as Yap1p [77] (*P. pastoris* contains only a single catalase gene with a predicted peroxisomal targeting sequence). Also an involvement of nitrogen regulation has been suggested for *P. pastoris CAT1* [78].
For the basic handling in standard cultivations understanding the molecular regulation of P_{CAT1} is not required, yet elucidating the exact regulation of P_{CAT1} may enable the design tailor-made transcriptional regulatory circuits.

Ultimately we consider derepressed promoters such as P_{CAT1} important new tools enabling shorter, methanol free processes alongside glycerol-repressed/glucose-inducible promoters previously reported [62].

3.4. Regulatory implications and sequence features

Judging from reporter protein fluorescence and the microarray data, certain MUT associated pairs of isoenzymes are peculiarly regulated. While P_{DAS1} and P_{DAS2} are identically regulated and expressed to similar extents, P_{FBA1} is constitutively expressed on glucose at moderate levels, but downregulated on methanol. In contrast P_{FBA2} is repressed on glucose and induced on methanol. Similar behavior is also evident for PPP isoenzymes pairs: P_{TAL2} is tightly repressed by glucose and induced by methanol.

Vice versa P_{TAL1} is moderately expressed on glucose. P_{RPE2} is clearly induced by methanol, whereas P_{RPE1} appears to be at best weakly expressed.
This suggests a peculiar role of isoenzymes taking over different functions on glucose and methanol to channel metabolic flux towards alternative metabolites. We presume that the flux of the non-oxidative PPP is channeled towards Xu_5P required for formaldehyde fixation (Fig. 1). It is also tempting to speculate, that the oxidative PPP is upregulated to boost NADPH regeneration (required for reduction of GSSG to GSH by the glutathione redox system). However, the promoters of the respective genes show only low, inconclusive reporter protein fluorescence little above background values, which is in line with previous studies [79]. Therefore this assumption is not experimentally validated and in a recent *P. pastoris* metabolomics studies [80] alternative explanations such as a yet unidentified NADPH reoxidation mechanism (as described in the related methylotrophic yeast *H. polymorpha* (*P. angusta*) [64] have been discussed.
Jordà *et al*. [80] performed flux analysis of glucose-methanol co-utilization using a model previously developed [81]. Notably, they discussed possible inconsistencies of the model regarding PPP metabolites, that could only be resolved modifying the transketolase and transaldolase reactions [80]. It appears plausible that these effects are caused by the *TAL1* and *TAL2* isoenzymes that are antagonistically transcriptionally regulated in our study. In *S. cerevisiae* also different isoenzymes of PPP steps are known and yet their significance for accurate metabolic models has been highlighted [82], their exact roles are unknown.
A biochemical characterization of the antagonistically regulated isoenzymes of the MUT pathway (*FBA1*/*FBA2*, *TAL1*/*TAL2*) regarding their substrate specificity, feedback regulation *etc*. may considerably increase the understanding of the metabolic adaption of *P. pastoris* to methanol.

In several cases, isoenzymes or related genes were organized closely to each other in the genome (e.g. *DAS1*/*DAS2*, *TAL1*/*TAL2*, *RPE1*/*RPE2, RKI1*/*RKI2, SOD1*/*SOD3, MSR1c4*/*MSRAB,* see Supporting Figure S 5). Notably, also the antagonistically regulated PPP gene pair (*TAL1*/*TAL2*) is also adjacently organized separated by an intragenic region of less than 1 kbp. The annotations and sequences of these loci differ between sequencing efforts of the CBS7435 strain [37] (used in this study) and the GS115 strain [36] (S 5 and extended discussion therein). Most strikingly, almost the entire *RPE1/RPE2* locus is missing from the GS115 annotation and neither gene is annotated. Since the *P. pastoris* KEGG [83,84] pathway maps are based on the GS115 annotation, this key step of the PPP is missing in KEGG based analyses. As the coding sequences of the isoenzymes show high similarity, we presume that the sequencing/annotation differences in these loci have arisen because of difficulties associated with sequencing of repetitive sequences. In a recent GSMM based study [17], *TAL1* was even omitted as target for overexpression studies because of these uncertainties in annotation. Therefore resolving these annotation issues and understanding the function of the antagonistic regulation of MUT isoenzymes may help to improve existing GSMMs.

Interestingly, the core promoters of all MUT and associated promoters show a striking enrichment in TATA box motifs (S 2). Yeast core promoters follow two fundamental types of architecture [85,86]: 1) TATA box containing promoters are highly regulated and depend on the SAGA coactivator. 2) TATA-less promoters are rather constitutively active and depend on the TFIID coactivator (TFIID). While TFIID dependent promoters account for ~90% of promoters in *S. cerevisiae*, only ~19% of promoters contain a TATA box and approximately half of those are SAGA dependent [85].
Remarkably, 9/13 promoters of genes of the assimilative and dissimilative MUT pathway contained a clear TATA box consensus motif in close proximity (<200 bp) to the start codon. In associated pathways TATA boxes appeared also enriched (PPP: 5/11; ROS: 4/14 promoters containing a TATA box). Overall, ~47 % of promoters of MUT related genes in this study contain a TATA box which correlates with the tight regulation observed in the microarray and reporter gene measurements. In *S. cerevisiae* TATA box promoters are often associated with stress regulated genes [86], suggesting

that the tight transcriptional regulation of MUT genes has evolved to respond to the major physiological changes demanded from methanol metabolism.

Based on their strength and regulation we suggest that *P. pastoris* MUT promoters can be classified in groups of strong, intermediate and weak inducible promoters with different extents of derepression. All methanol inducible promoters are tightly repressed on glucose. But while some promoters remain repressed under glucose depleted conditions, others show clear derepressed expression (Tab. 1). These results suggest that *P. pastoris* offers a toolbox of about 15 methanol regulated promoters. About half show strong expression comparable to P_{AOX1} and the other half spans a range of lower expression levels. Due to the tight glucose repression, strength and their high sequence diversity, these promoters appear as a versatile toolbox for fine-tuning the expression of heterologous pathways.

4. Conclusion

Here we have shown, that *P. pastoris* provides to our knowledge the largest set of tightly co-regulated promoters known in biotechnological expression hosts. Testing 45 promoters of the MUT pathway and associated processes resulted in 15 methanol regulated promoters, about half showing very high expression levels (Tab. 1). Such a toolbox of strong promoters is valuable for overexpressing heterologous multi genes pathways, yet also weaker promoters appear useful for transcriptional fine-tuning of stoichiometric ratios between pathway steps. Up to seven genes can be coexpressed using only strong MUT promoters and up to eleven genes can be expressed using also promoters of intermediate strength. Our microarray data and related studies [39–44] suggested even more genes, that are similarly highly upregulated: About 35 additional genes are more than 20 fold upregulated and about 10 genes more than 70 fold (comparing methanol induction with glucose, S 3). Therefore these datasets can be mined for additional tightly methanol regulated promoters.

Despite their similar regulation, the promoters tested show little to none sequence similarities, possibly as short regulatory elements such as transcription factor binding sites may be variably dispersed over the whole promoter (as shown for P_{AOX1}, P_{DAS2} and P_{PEX8} [34,87,88]). In contrast to highly similar variants of a synthetic promoter library, which are typically derived by small changes of a single promoter, the low sequence similarity between natural MUT promoters appears favorable for *in vitro* overlap-directed DNA assembly and also *in vivo* stability. From a basic research perspective this low sequence similarity between highly co-regulated promoters is puzzling and may suggest MUT promoters as a model system for studying transcriptional regulation and as a repertoire of parts for synthetic biology.
Growth and metabolism of the cheap inducer methanol may bear also additional benefits for certain applications. In recent work on artemisinic acid production in *S. cerevisiae* [10], oxidative stress arising due to cytochrome P450 (CYP) expression was counteracted by increasing the levels of cytosolic catalase. *P. pastoris* has been shown to be a highly suitable expression system for CYPs [16,89,90] and growth on methanol necessitates a high tolerance to ROS stress (caused by oxidation of methanol to formaldehyde, Fig. 1).
Over the last two decades *P. pastoris* has become a popular expression host for high-level single protein production, even surpassing *S. cerevisiae* [15]. Considering the recent development of GSMMs and the availability of suitable promoters, *P. pastoris* appears also as a promising alternative to *S. cerevisiae* for metabolic engineering endeavors requiring tight transcriptional regulation of large heterologous pathways, CYP expression or oxidative stress tolerance.

Acknowledgments

We would like to thank Clemens Farnleitner for excellent technical assistance and Florian W. Krainer for fruitful discussions and critical reading of the manuscript. T.V. was supported by the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) while performing this research. The authors gratefully acknowledge support from NAWI Graz.

References

[1] Stephanopoulos G. Synthetic biology and metabolic engineering. ACS Synth Biol 2012;1:514–25.
[2] Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 2012;40:e142.
[3] Blazeck J, Alper HS. Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol J 2013;8:46–58.
[4] Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 2005;102:12678–83.
[5] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.
[6] Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.
[7] Aw R, Polizzi KM. Can too many copies spoil the broth? Microb Cell Fact 2013;12:128.
[8] Peralta-Yahya PP, Keasling JD. Advanced biofuel production in microbes. Biotechnol J 2010;5:147–62.
[9] Nielsen J, Larsson C, van Maris A, Pronk J. Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol 2013;24:398–404.
[10] Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 2013;496:528–32.
[11] Hong K-K, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cell Mol Life Sci 2012;69:2671–90.
[12] Da Silva NA, Srikrishnan S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res 2012;12:197–214.
[13] Liu L, Redden H, Alper HS. Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces. Curr Opin Biotechnol 2013:1–8.
[14] Jahic M, Veide A, Charoenrat T, Teeri T, Enfors S-O. Process technology for production and recovery of heterologous proteins with Pichia pastoris. Biotechnol Prog 2006;22:1465–73.
[15] Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Front Microbiol 2014;5:85.
[16] Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, et al. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of Pichia pastoris. Metab Eng 2014.
[17] Nocon J, Steiger MG, Pfeffer M, Bum Sohn S, Yong Kim T, Maurer M, et al. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 2014;24:129–38.
[18] Schroer K, Peter Luef K, Stefan Hartner F, Glieder A, Pscheidt B. Engineering the Pichia pastoris methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation. Metab Eng 2010;12:8–17.
[19] Gasser B, Dragosits M, Mattanovich D. Engineering of biotin-prototrophy in Pichia pastoris for robust production processes. Metab Eng 2010;12:573–80.
[20] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 2012;11:22.
[21] Jeong E, Shim WY, Kim JH. Metabolic Engineering of Pichia pastoris for Production of Hyaluronic Acid with High Molecular Weight. J Biotechnol 2014.
[22] Araya-Garay JM, Feijoo-Siota L, Rosa-dos-Santos F, Veiga-Crespo P, Villa TG. Construction of new Pichia pastoris X-33 strains for production of lycopene and β-carotene. Appl Microbiol Biotechnol 2012;93:2483–92.
[23] Araya-Garay JM, Ageitos JM, Vallejo JA, Veiga-Crespo P, Sánchez-Pérez A, Villa TG. Construction of a novel Pichia pastoris strain for production of xanthophylls. AMB Express 2012;2:24.
[24] Geier M, Fauland PC, Vogl T, Glieder A. Compact multi enzyme pathways in Pichia pastoris. Chem Commun 2014.

[25] Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, et al. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol J 2010;5:705–15.
[26] Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, et al. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 2010;9:50.
[27] Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol 2012;6:24.
[28] Renuse S, Madugundu AK, Kumar P, Nair BG, Gowda H, Prasad TS PA. Proteomic analysis and genome annotation of Pichia pastoris, a recombinant protein expression host. Proteomics 2014;14:2769–79.
[29] Vogl T, Hartner FS, Glieder A. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 2013;24:1094–101.
[30] Dikicioglu D, Wood V, Rutherford KM, McDowall MD, Oliver SG. Improving functional annotation for industrial microbes: a case study with Pichia pastoris. Trends Biotechnol 2014:1–4.
[31] Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. Int J Microbiol 2011;2011:101298.
[32] Van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochim Biophys Acta 2006;1763:1453–62.
[33] Yurimoto H. Molecular Basis of Methanol-Inducible Gene Expression and Its Application in the Methylotrophic Yeast Candida boidinii. Biosci Biotechnol Biochem 2009;73:793–800.
[34] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.
[35] Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact 2006;5:39–59.
[36] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6.
[37] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.
[38] Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb Cell Fact 2009;8:29.
[39] Liang S, Wang B, Pan L, Ye Y, He M, Han S, et al. Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. BMC Genomics 2012;13:738.
[40] Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, et al. Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation. Microb Cell Fact 2004;3:17–30.
[41] Vanz AL, Nimtz M, Rinas U. Decrease of UPR- and ERAD-related proteins in Pichia pastoris during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. Microb Cell Fact 2014;13:23.
[42] Vanz AL, Lünsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N, et al. Physiological response of Pichia pastoris GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. Microb Cell Fact 2012;11:103.
[43] Van Zutphen T, Baerends RJS, Susanna KA, de Jong A, Kuipers OP, Veenhuis M, et al. Adaptation of Hansenula polymorpha to methanol: a transcriptome analysis. BMC Genomics 2010;11:1.
[44] Ravin N V, Eldarov MA, Kadnikov V V, Beletsky A V, Schneider J, Mardanova ES, et al. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1 2013.
[45] Yurimoto H, Komeda T, Lim CR, Nakagawa T, Kondo K, Kato N, et al. Regulation and evaluation of five methanol-inducible promoters in the methylotrophic yeast Candida boidinii. Biochim Biophys Acta 2000;1493:56–63.
[46] Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 2012;7:e39720.
[47] Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic Core Promoters for Pichia pastoris. ACS Synth Biol 2014;3:188–91.
[48] Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A. Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 2004;5:179–89.
[49] Wriessnegger T, Gübitz G, Leitner E, Ingolic E, Cregg J, de la Cruz BJ, et al. Lipid composition of peroxisomes from the yeast Pichia pastoris grown on different carbon sources. Biochim Biophys Acta 2007;1771:455–61.

[50] Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, et al. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 2005;38:44, 46, 48.
[51] Cregg JM, Tolstorukov I, Kusari A, Sunga J, Madden K, Chappell T. Expression in the yeast Pichia pastoris. Methods Enzymol 2009;463:169–89.
[52] Viader-Salvadó JM, Cab-Barrera EL, Galán-Wong LJ, Guerrero-Olazarán M. Genotyping of recombinant Pichia pastoris strains. Cell Mol Biol Lett 2006;11:348–59.
[53] Hellwig S, Robin F, Drossard J, Raven NP, Vaquero-Martin C, Shively JE, et al. Production of carcinoembryonic antigen (CEA) N-A3 domain in Pichia pastoris by fermentation. Biotechnol Appl Biochem 1999;30 (Pt 3):267–75.
[54] Näätsaari L, Krainer FW, Schubert M, Glieder A, Thallinger GG. Peroxidase gene discovery from the horseradish transcriptome. BMC Genomics 2014;15:227.
[55] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. SUPPLEMENTARY MATERIALS Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6.
[56] Yano T, Takigami E, Yurimoto H, Sakai Y. Yap1-regulated glutathione redox system curtails accumulation of formaldehyde and reactive oxygen species in methanol metabolism of Pichia pastoris. Eukaryot Cell 2009;8:540–9.
[57] Yano T, Yurimoto H, Sakai Y. Activation of the oxidative stress regulator PpYap1 through conserved cysteine residues during methanol metabolism in the yeast Pichia pastoris. Biosci, Biotechnol, Biochem 2009;73:1404–11.
[58] Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, Glieder A, et al. Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression. N Biotechnol 2014;00.
[59] Tatusov RL, Galperin MY, Natale D a, Koonin E V. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 2000;28:33–6.
[60] Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, et al. Pichia pastoris: protein production host and model organism for biomedical research. Futur Microbiol 2013;8:191–208.
[61] Leao-Helder AN, Krikken AM, van der Klei IJ, Kiel J a KW, Veenhuis M. Transcriptional down-regulation of peroxisome numbers affects selective peroxisome degradation in Hansenula polymorpha. J Biol Chem 2003;278:40749–56.
[62] Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, et al. Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris. Microb Cell Fact 2013;12:5.
[63] Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. Mol Cell Biol 1985;5:1111–21.
[64] Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Mol Cell Biol 1989;9:1316–23.
[65] Kristiansson E, Thorsen M, Tamás MJ, Nerman O. Evolutionary forces act on promoter length: identification of enriched cis-regulatory elements. Mol Biol Evol 2009;26:1299–307.
[66] Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. Nucleic Acids Res 1987;15:3859–76.
[67] Bawa Z, Routledge SJ, Jamshad M, Clare M, Sarkar D, Dickerson I, et al. Functional recombinant protein is present in the pre-induction phases of. Microb Cell Fact 2014;13:127.
[68] Fujimura S, Yurimoto H, Kurimoto S, Matsufuji Y, Ito T, Hayakawa T, et al. Expression level of methanol-inducible peroxisomal proteins and peroxisome morphology are affected by oxygen conditions and mitochondrial respiratory pathway function in the methylotrophic yeast Candida boidinii. FEMS Yeast Res 2013;13:359–66.
[69] Fujimura S, Nakagawa T, Ito T, Matsufuji Y, Miyaji T. Peroxisomal metabolism is regulated by an oxygen-recognition system through organelle crosstalk between the mitochondria and peroxisomes 2007:491–8.
[70] Kim S, Warburton S, Boldogh I, Svensson C, Pon L, D'Anjou M, et al. Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in Pichia pastoris. J Biotechnol 2013;166:174–81.
[71] Ruth C, Zuellig T, Mellitzer A, Weis R, Looser V, Kovar K, et al. Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. Syst Biol Synth Biol 2010;4:181–91.
[72] Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y. An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. FEMS Yeast Res 2009;9:1271–82.
[73] Mattanovich D, Gasser B, Hohenblum H, Sauer M. Stress in recombinant protein producing yeasts. J Biotechnol 2004;113:121–35.

[74] Panula-Perälä J, Vasala A, Karhunen J, Ojamo H, Neubauer P, Mursula A. Small-scale slow glucose feed cultivation of Pichia pastoris without repression of AOX1 promoter: towards high throughput cultivations. Bioprocess Biosyst Eng 2013.
[75] Yamashita S, Yurimoto H, Murakami D, Yoshikawa M, Oku M, Sakai Y. Lag-phase autophagy in the methylotrophic yeast Pichia pastoris. Genes Cells 2009;14:861–70.
[76] Gurvitz A, Hiltunen JK, Erdmann R, Hamilton B, Hartig A, Ruis H, et al. Saccharomyces cerevisiae Adr1p governs fatty acid beta-oxidation and peroxisome proliferation by regulating POX1 and PEX11. J Biol Chem 2001;276:31825–30.
[77] Lee J, Godon C, Lagniel G, Spector D, Garin J, Labarre J, et al. Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. J Biol Chem 1999;274:16040–6.
[78] Rumjantsev a. M, Padkina M V., Sambuk E V. Effect of nitrogen source on gene expression of first steps of methanol utilization pathway in Pichia pastoris. Russ J Genet 2013;49:394–400.
[79] Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, et al. A multi-level study of recombinant Pichia pastoris in different oxygen conditions. BMC Syst Biol 2010;4:141.
[80] Jordà J, Suarez C, Carnicer M, ten Pierick A, Heijnen JJ, van Gulik W, et al. Glucose-methanol co-utilization in Pichia pastoris studied by metabolomics and instationary ^{13}C flux analysis. BMC Syst Biol 2013;7:17.
[81] Jordà J, Jouhten P, Cámara E, Maaheimo H, Albiol J, Ferrer P. Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures. Microb Cell Fact 2012;11:57.
[82] Kleijn RJ, van Winden W a, van Gulik WM, Heijnen JJ. Revisiting the 13C-label distribution of the non-oxidative branch of the pentose phosphate pathway based upon kinetic and genetic evidence. FEBS J 2005;272:4970–82.
[83] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42:D199–205.
[84] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
[85] Hahn S, Young ET. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics 2011;189:705–36.
[86] Basehoar AD, Zanton SJ, Pugh BF. Identification and distinct regulation of yeast TATA box-containing genes. Cell 2004;116:699–709.
[87] Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN. Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. Biochim Biophys Acta 2009;1789:460–8.
[88] Kranthi BV, Kumar HRV, Rangarajan PN. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris. Yeast 2010;27:705–11.
[89] Geier M, Braun A, Emmerstorfer A, Pichler H, Glieder A. Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes - a comparative study. Biotechnol J 2012:1–13.
[90] Gudiminchi RK, Geier M, Glieder A, Camattari A. Screening for cytochrome P450 expression in Pichia pastoris whole cells by P450-carbon monoxide complex determination. Biotechnol J 2013;8:146–52.
[91] Mead DA, Pey NK, Herrnstadt C, Marcil RA, Smith LM. A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology (N Y) 1991;9:657–63.
[92] Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, et al. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol Cell Biol 2006;26:883–97.
[93] Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 2000;11:4241–57.
[94] Aragon AD, Quiñones G a, Thomas E V, Roy S, Werner-Washburne M. Release of extraction-resistant mRNA in stationary phase Saccharomyces cerevisiae produces a massive increase in transcript abundance in response to stress. Genome Biol 2006;7:R9.
[95] Martinez MJ, Roy S, Archuletta AB, Wentzell PD, Anna-arriola SS, Rodriguez AL, et al. Genomic Analysis of Stationary-Phase and Exit in Saccharomyces cerevisiae : Gene Expression and Identification of Novel Essential Genes 2004;15:5295–305.

Figures

Fig. 1

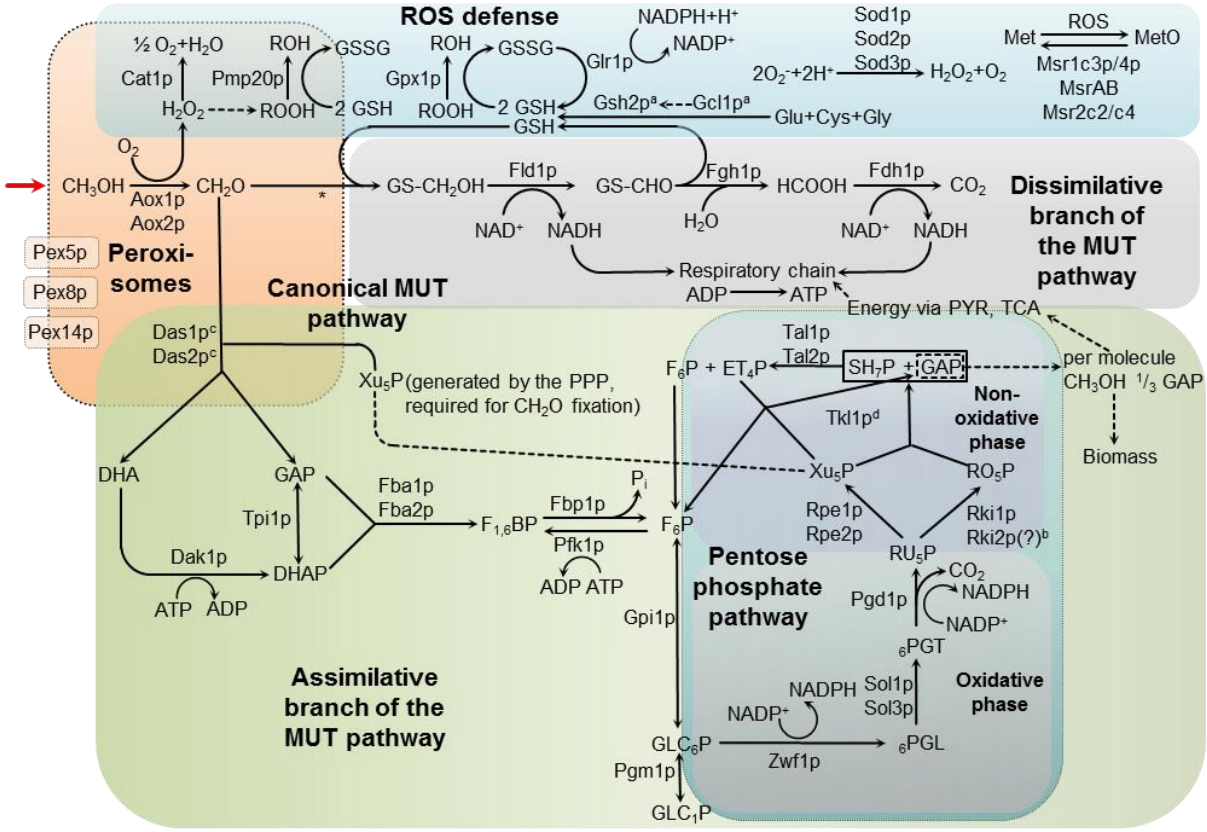
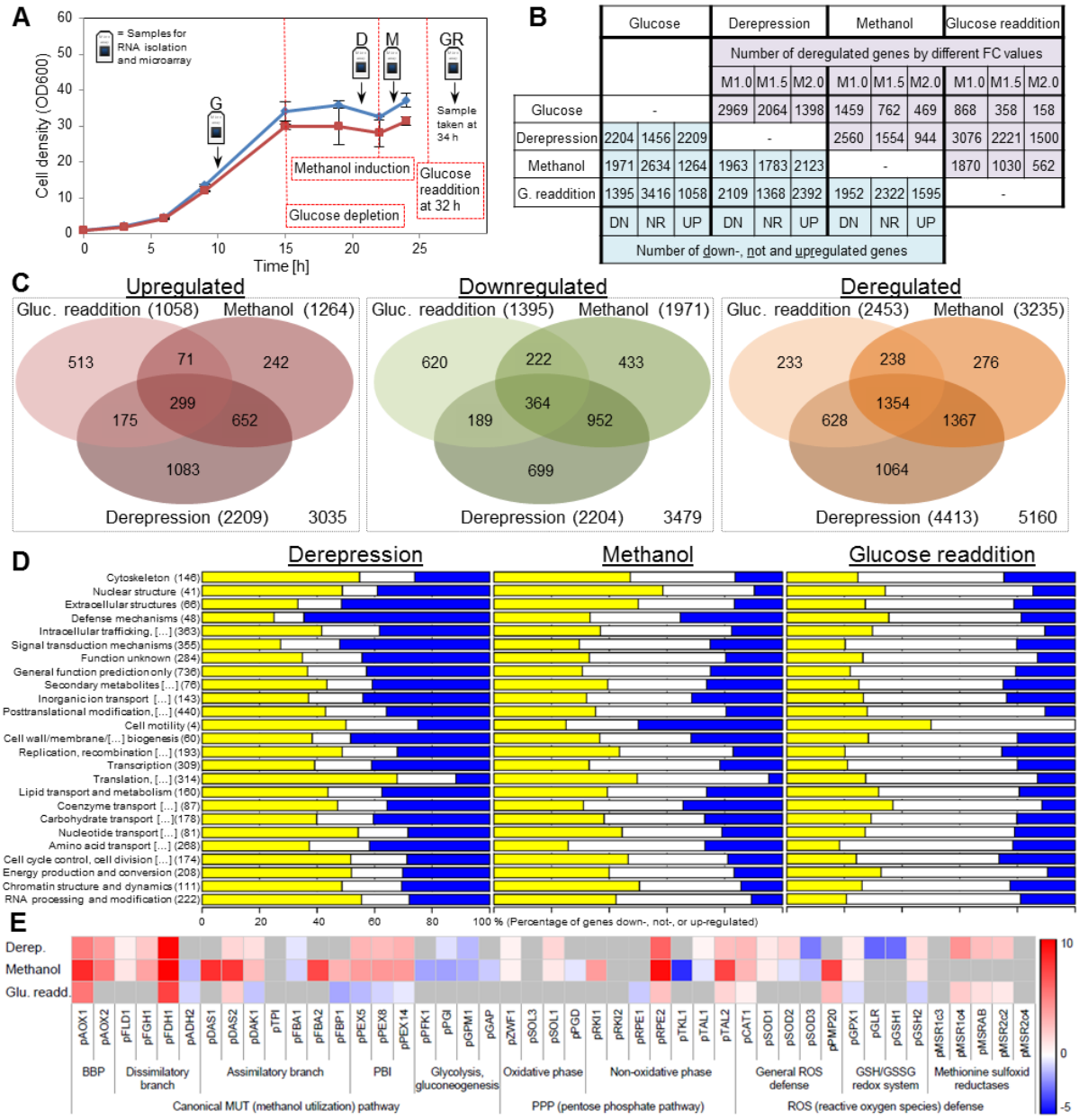


Fig. 1: Canonical and non-canonical parts of the *P. pastoris* methanol utilization pathway. The canonical MUT pathway is based on the most recent annotation [37] and physiological studies [20]. The part on ROS defense is based on [56]. The PPP is here shown as associated with the assimilative branch of the MUT pathway, the oxidative (for NADPH regeneration) and non-oxidative phase are highlighted. Enzymes are named according to the respective gene product names (see Table S 2 for the enzymatic functions). Peroxisomal membrane proteins Pex5, Pex8 and Pex14 involved in peroxisome biogenesis and peroxisomal signal sequence recognition and relevant reactions of glycolysis (catalyzed by Pfk1p, Gpi1p, Pfk1p) are also shown. Alternative steps of the dissimilative MUT pathway (methylformate formation [putatively by *ADH2*] and demethylation thereof by a yet unknown enzyme) as annotated by Küberl *et al*. [37] are not shown.

Abbreviations of metabolites: DHA: dihydroxyacetone; DHAP: dihydroxyacetone phosphate; ET_4P: erythrose 4-phosphate; $FRU_{1,6}P$: fructose- 1,6-bisphosphate; FRU_6P: fructose-6-phosphate; GAP: glyceraldehyde-3-phosphate; GLC_1P: glucose-1-phosphate; GLC_6P: glucose-6-phosphate; GSH: glutathione; GSSG: oxidized glutathione self-dimer; MetO: methionine sulfoxide; $_6PGL$: 6-phosphogluconolactone; $_6PGT$: 6-phosphogluconat; P_i: inorganic phosphorus; PYR: pyruvate; RO_5P: ribose 5-phosphate; RU_5P: ribulose 5-phosphate; SH_7P: sedoheptulose 7-phosphate; TCA: tricarboxylic acid cycle; Xu_5P: xylulose 5-phosphate.

Legend: ‚R' in chemical formulas denotes a hydrogen, aliphatic or aromatic organic group; [a] non-ribosomal peptide synthesis; [b] issues with annotation, see S 5; [c] Tkl1p has also been assigned putative dihydroxyacetone synthase activity [20,37]; [d] Transketolase activity is required for both the reaction of $Xu_5P+RI_5P \rightarrow SH_7P+GAP$ and $ET_5P+Xu_5P \rightarrow F_6P+GAP$. * The reaction of CH_2O and GSH is occurring non-enzymatically.

Fig. 2



	Glucose			Derepression			Methanol			Glucose readdition		
				Number of deregulated genes by different FC values								
				M1.0	M1.5	M2.0	M1.0	M1.5	M2.0	M1.0	M1.5	M2.0
Glucose	-			2969	2064	1398	1459	762	469	868	358	158
Derepression	2204	1456	2209	-			2560	1554	944	3076	2221	1500
Methanol	1971	2634	1264	1963	1783	2123	-			1870	1030	562
G. readdition	1395	3416	1058	2109	1368	2392	1952	2322	1595	-		
	DN	NR	UP	DN	NR	UP	DN	NR	UP			
	Number of down-, not and upregulated genes											

Fig. 2: Genome wide transcriptional response of *P. pastoris* towards growth on different carbon sources.

A) *P. pastoris* bioreactor cultivations and sampling points used for the microarray analyses. Growth curves for the same strain in biological duplicates are shown; standard deviations are technical replicates of the OD measurements. Glucose depletion (equaling the end of the exponential phase) was determined by the oxygen peak and confirmed by glucose measurements. Glucose readdition was performed one hour after methanol was depleted (31 h) and samples taken at the time points indicated. Abbreviations of sampling points for microarrays: 'G' –glucose, 'D' – derepression, 'M' – methanol, 'GR' – glucose readdition.

B) Comparison of the transcriptional response under all tested growth conditions. In the lower left corner the number of down- (DN), not regulated (NR) and upregulated (UP) genes is given ($p < 0.01$) [total number of probe sets: 5869]. In the upper right corner deregulated

(DN or UP) genes are listed by different fold change (FC) criteria [M1.0, 1.5 and 2.0 denoting the respective log2 values].

C) Comparative analysis of differential gene expression between growth under derepressed, methanol induced and glucose readdition conditions. These three conditions were each at first normalized to growth on glucose. The numbers of unique and overlapping genes showing up-, down- and deregulation are given (FC >log2(1.0), $p < 0.05$, the same criteria and normalization were applied for the analyses shown in panel D).

D) Functional grouping of differentially regulated genes to biologic processes by COG terms [59]. The relative numbers of down-, not- and upregulated genes assigned to the same COG terms are shown (each condition was compared to growth on glucose as in panel C). The total number of genes assigned to each COG term is given in parentheses. Full COG terms if abbreviated : “Intracellular trafficking, secretion, and vesicular transport”, “Secondary metabolites biosynthesis, transport and catabolism”, “Inorganic ion transport and metabolism”, “Posttranslational modification, protein turnover, chaperones”, “Translation, ribosomal structure and biogenesis”, “Replication, recombination and repair”, “Translation, ribosomal structure and biogenesis”, “Coenzyme transport and metabolism”, “Carbohydrate transport and metabolism”, “Nucleotide transport and metabolism”, “Amino acid transport and metabolism” and “Cell cycle control, cell division, chromosome partitioning”.

E) Regulation of MUT genes shown in Fig. 1 (relative to glucose). FC log2 values are shown; if the changes were not significant ($p < 0.01$) the values are not shown (grey). In case multiple probe sets were present on the array (e.g. *AOX1*), the mean FC value is shown. Abbreviations: BPP: before branch point (to assimilative and dissimilative branches), PBI: peroxisomal biogenesis and import.

Fig. 3

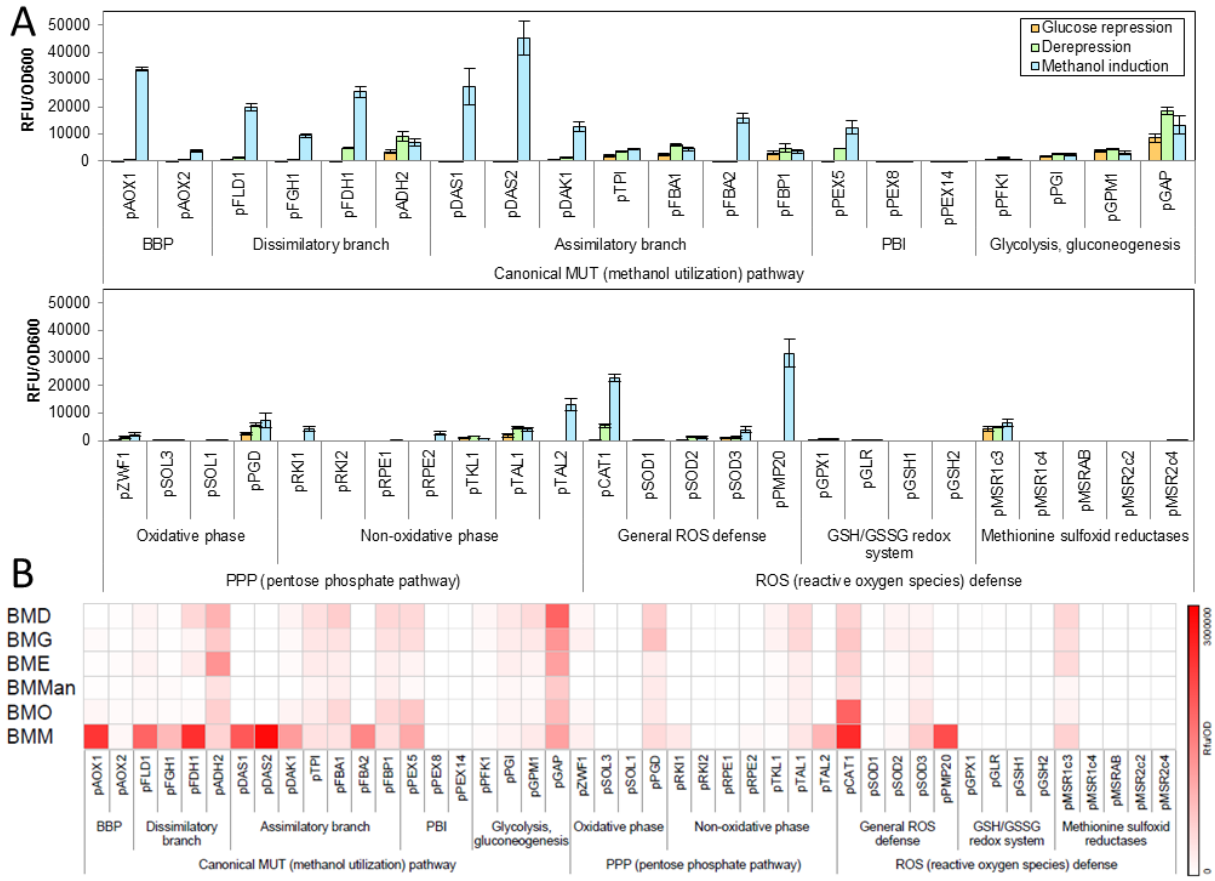


Fig. 3: MUT promoters show with an eGFP reporter gene a wide range of regulatory profiles and a range of expression levels.

(A) Reporter protein fluorescence of MUT promoters in a time series experiment mimicking typical biphasic bioreactor cultivations. The promoters of MUT and associated genes (S 2) were seamlessly cloned upstream of an eGFP reporter gene and transformed into *P. pastoris*. The strains were cultivated in a high throughput 96 well deep well plate format [48] in biological quadruplicates (mean value and standard deviation shown) on glucose containing medium (BMD: buffered minimal dextrose) for 60 h and subsequently induced with methanol (BMM). Samples were taken under glucose repressed conditions (15.5 h), derepressed conditions (60 h) and 24 h methanol induction. eGFP fluorescence was normalized per OD_{600}. Additional sampling points measured are shown in (S 7). Only the shortest promoter lengths tested are shown, alternative lengths tested did not affect reporter protein fluorescence (S 6). The same abbreviations as in Fig. 2 E are used.

(B) MUT promoter reporter protein fluorescence on different carbon sources. The same strains as in panel (A) were cultivated for 60 h on the media indicated (BMD: glucose, BMG: glycerol, BME: ethanol, BMMan: mannitol; BMO: oleic acid, BMM: methanol). Measurement procedures, replicate handling and data of glucose are identical to the 60 h data of panel A.

Fig. 4

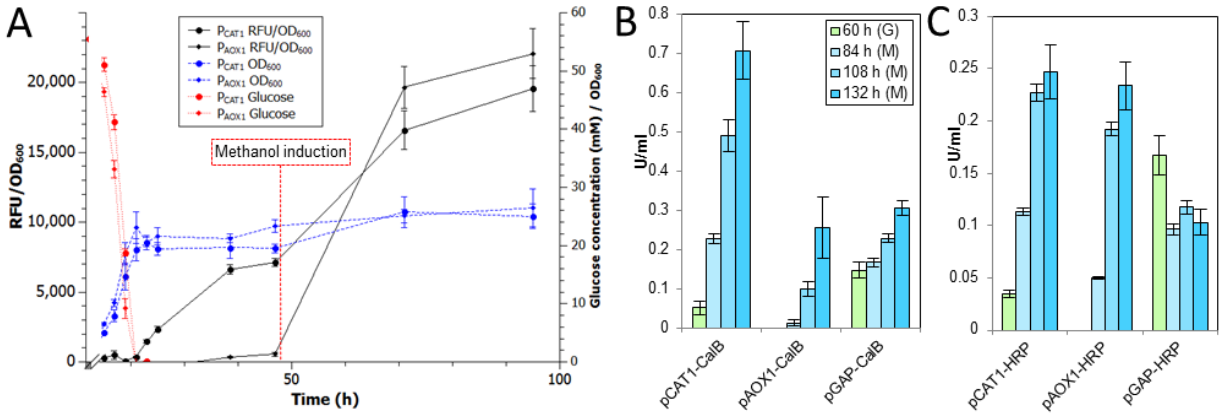


Fig. 4: P_{CAT1} expression starts after glucose depletion and this derepressed regulatory profile is also suitable for secretory protein production.

(A) Comparison of regulatory profiles of $P_{CAT1\text{-}500}$ to the methanol inducible state of the art promoter P_{AOX1}. The strains were grown in biological triplicates in shake flasks (mean value and SD shown). Reporter protein fluorescence of eGFP (under control of the respective promoters), glucose levels and OD_{600} (to assess growth) were measured at the time points indicated. Fluorescence measurements were normalized per OD_{600}. The cultures were induced with methanol after 48 h. Fluorescence/OD_{600} values at t=0 are not shown, as the starting OD_{600} (0.05) was outside of the linear range of the spectrometer used.
(B) Volumetric activities of $P_{CAT1\text{-}692}$, P_{AOX1} and P_{GAP} driven secretory production of *Candida antarctica* lipase B and (C) horseradish peroxidase (HRP). The strains were grown for 60 h on glucose (G) to achieve derepression and were subsequently induced with methanol (M) for 72 h. Volumetric activities in the supernatant were measured in 24 h intervals. Stable genomic integrated transformants of $P_{CAT1\text{-}692}$ were used (showing identical expression to $P_{CAT1\text{-}500}$; S 6). Mean value and standard deviations of biological triplicates are shown.

Tables

Tab. 1

Tab. 1: Grouping of *P. pastoris* MUT promoters.

Definitions of promoter strength (determined from reporter gene fluorescence of Fig. 3): strong >67%, intermediate 33-67%, weak <33 % of P_{AOX1}. [a] The *GAP* gene is not directly part of the MUT pathway but included as reference of the strongest constitutive *P. pastoris* promoter [34]. [b] P_{ADH2} is constitutively expressed, but also upregulated on ethanol.

	Methanol inducible		Constitutive
	Tightly repressed	Derepressed	
Strong	P_{AOX1}, P_{DAS1}, P_{DAS2}, P_{PMP20},	P_{CAT1}, P_{FDH1}, P_{FLD1}	P_{GAP} [a]
Intermediate	P_{TAL2}, P_{FBA2}	P_{PEX5}, P_{DAK1}	P_{ADH2} [b]
Weak	P_{AOX2}, P_{RKI1}, P_{RPE2}	P_{FGH1}	P_{TPI1}, P_{FBP1}, P_{PGI1}, P_{TAL1}, P_{MSR1c3}, P_{PFK1}, P_{GPM1}

Supporting information

S 1

S 1: Detailed construction of the vectors used in this study.

Promoter reporter gene plasmid

We based our reporter gene vector on the pPpT4_S vector [46] and enhanced green fluorescent protein (eGFP) sequence [47] previously established for promoter studies in *P. pastoris*. We adapted the vector for TA cloning [91] and simple insertion by Gibson assembly [6]. A stuffer fragment/place holder fragment was placed upstream of the eGFP reporter gene. Digestion with the type IIS restriction endonuclease *Bmr*I removes the stuffer, leaving thymidine overhangs suitable for cloning of adenine tailed PCR products. The stuffer is designed to enable seamless fusions of the promoter with the 5' end of the reporter gene, maintaining the natural sequence context of the start codon.
The *AOX1* promoter of the pPpT4_S vector was removed by *Pci*I and *Not*I digestion and replaced with an overlap extension PCR (olePCR) fragment consisting of an *ARG4* integration sequence (int*ARG4*), the stuffer fragment and the GFP gene. The stuffer fragment is a part of the *THI5* gene from *S. cerevisiae* and does not show any sequence homology to *P. pastoris* or *E. coli*. (blastN, standard settings). The following primers were used for PCR amplification (see S 2 for primer sequences): int.arg.fwd + int.arg.rev, stufferTHI5.fwd + stufferTHI5.rev and EGFPfwd.stufferTHI5 + EGFPrevNotI. *Bmr*I sites flanking the stuffer fragment were added on the PCR primers. The fragments were gel purified and mixed in equimolar ratios for olePCR. Primers int.arg.fwd and EGFPrevNotI were added after 20 cycles. The fragment of the correct size was gel purified, digested with *Pci*I and *Not*I and ligated with the vector backbone. A *Mly*I site present in the vector was removed by PCR amplifying the vector using primers ZeoCDS_mut_MlyI_fwd and ZeoCDS_mut_MlyI_rev Pfu Ultra polymerase followed by DpnI digestion. The whole vector was confirmed by Sanger sequencing. The vector is referred to as pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer.

The promoters were PCR amplified using the primers listed in S 2 and cloned by TA cloning or Gibson [6] assembly into the reporter vector. For TA cloning, orientation was confirmed by colony PCR and in all cases the promoter sequences were verified by Sanger sequencing (from both sides of the vector using primers seqintARG4fwd and seqEGFPrev).

CalB and HRP plasmids

CalB and HRP (isoenzyme A2A [[54]]) were seamlessly fused downstream of P_{AOX1}, P_{GAP} and $P_{CAT1\text{-}692}$. We started from pPpT4_S [46] based vectors containing the CalB and HRP sequences. The vector backbones were amplified using primers MFalpha-fwd and pILV5withoutBamHI-fwd. The promoters were amplified using primers pILV5-pAOX1-Gib + pAOX1-Gib-MFalpha-ins (for P_{AOX1}), pILV5-pCAT1-Gib + pCAT1-Gib-MFalpha-ins (for P_{CAT1}) and pILV5-pGAP-Gib + pGAP-Gib-MFalpha-ins (for P_{GAP}). The same primers could be used for the backbones and promoters for CalB and HRP since both are fused to the identical mating factor alpha signal sequence). The entire vectors were confirmed by sequencing.

S 2

S 2: List of promoters and primers used in this study.

- Sheet 1: Promoters
    - Nomenclature/annotation

      In several cases, the annotation used [37] did not follow standard abbreviations used in other yeasts (e.g. the gene commonly known as *PMP20* is in *P. pastoris* annotated as *PXR1*). In these cases we used the commonly used literature names. In some cases we suggest different gene abbreviations than provided in the annotation [37]: Methionine sulfoxide reductase genes are abbreviated *MXR1* and *MXR2* in *S. cerevisiae*, however in *P. pastoris* the methanol master regulation transcription factor is also termed *MXR1* [92]. To avoid confusions we suggest the abbreviation '*MSR*' for methionine sulfoxide reductase genes in *P. pastoris.* Note also that there are several *MSR* genes in *P. pastoris*; with some encoded proteins showing higher identity to *S. cerevisiae* Mxr1p and others to Mxr2p. We have named them according to these identities, in case of multiple paralogs we added the position on the genome for differentiation (e.g. *MSR1c3* [chromosome 3]). We assigned the names *SOD1* and *SOD3* to additional genes coding for proteins showing high identity to *S. cerevisiae* superoxide dismutase proteins.
    - TATA box motif

      The presence of the yeast TATA box consensus sequence (TATAWAWR [86]) is indicated (positioning on the reverse strand is indicated by 'rev'). The distance from the start codon is given in bp.
    - Primers

      Primers used for amplifying the promoters are given. Most promoters were cloned by TA cloning as outlined in the materials and methods section. In cases Gibson assembly was used, the overhangs to the vector are underlined.

- Sheet 2: Primers for assembly of the reporter vectors

S 3

S 3 Lists of genes differentially regulated in the microarray data.

Each sheet contains the data of the pairwise comparisons of two conditions (growth on glucose, under derepressed conditions, on methanol and after glucose readdition; see also Fig. 2 A). Significantly differentially regulated genes ($p < 0.01$) are shown.

Abbreviations:

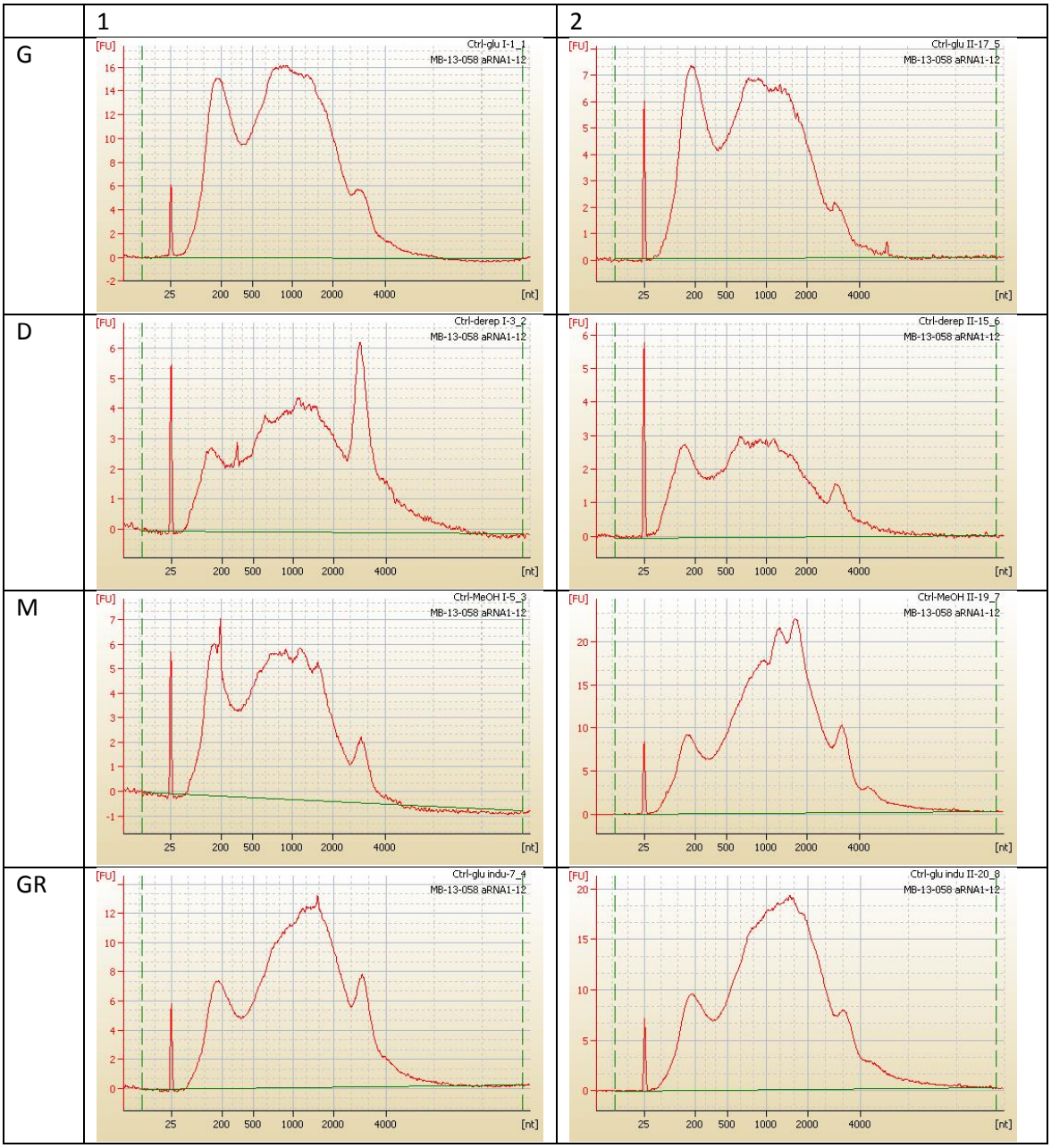
COG: Clusters of Orthologous Groups of proteins [59]; COGCat: COG category; COGCatDesc: COG category description; EC: Enzyme Commission number; GO: Gene Ontology term; log2FC: logarithm to the basis 2, calculated for fold changes; Abslog2FC: absolute log2FC; P.Value and adj.P.Val: probability values calculated from two biological replicates.

S 4

A

Condition	Glucose		Derepression		Methanol		Glucose readdition		Positive control
Replicate	1	2	1	2	1	2	1	2	
aRNA (µg)	128.00	114.00	44.75	36.75	91.50	98.00	105.25	127.25	54.50

B



S 4: Derepressed samples yield less aRNA for hybridizations.

A) Amplification yields of the amplified RNA (aRNA).

B) Capillary electrophoresis diagrams of aRNA, the different conditions glucose (G), derepression (D), methanol (M) and glucose readdition (GR) are shown. Note the different scales.

Derepressed conditions yielded lower aRNA concentrations and showed altered capillary electrophoresis migration patterns compared to the other samples The initial total RNA isolation had been normal regarding concentrations, purity and intactness (judged by capillary electrophoresis) and also after fragmentation no differences to the other samples were observed (data not shown). Derepression is equivalent to the stationary phase; therefore more mRNAs could have been degraded in this phase, possibly shifting the ratio between mRNA and rRNAs thereby effecting aRNA generation. The Affymetrix microarray platform includes multiple controls for normalization and quality controls and the derepressed samples did not show any deviations from the standard quality requirements. Even the lower aRNA yields of the derepressed samples were in the range of the positive control and more than sufficient to proceed with the protocol. Therefore we do not see a clear reason to challenge these results, especially as repetition with related technology (RNAseq, RT-qPCR) may run into similar limitations due to the reverse transcription and/or library generation step required. Also in *S. cerevisiae* massive transcriptional changes upon the entry into [93,94] and exit from [95] the stationary phase have been reported.
We also noticed the same effects on aRNA for additional three overexpression strains tested in duplicates under derepressed conditions (data to be published elsewhere). All three strains showed similarly low aRNA yields (strain 1: 44.00, 41.50µg; strain 2: 42.25, 44.0 µg; strain 3: 39.25, 25.50 µg).

S 5

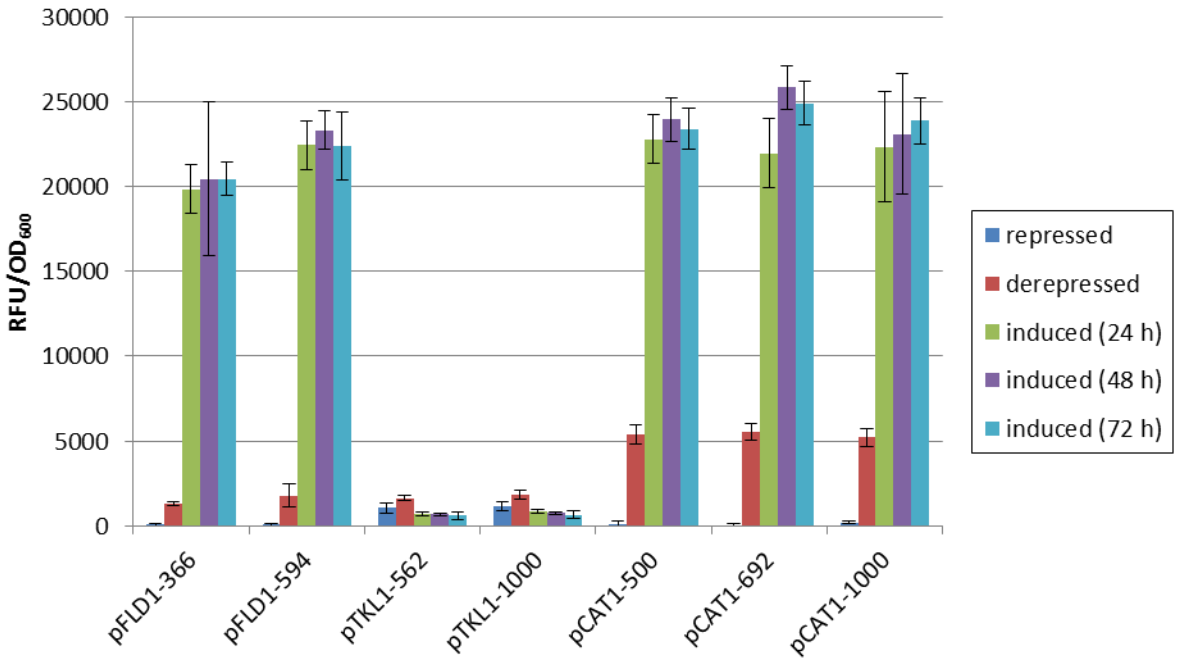
S 5: Genomic organization of the genes and promoter sequences used in the study and extended discussion of differences between *P. pastoris* CBS7435 and GS115 strains.

- Description
  10000 bp upstream and downstream of the start codon of the MUT genes from S 2 are shown. All genes are shown in forward orientation, even if they are naturally encoded on the reverse strand. The annotation of the *P. pastoris CBS7435* strain was used [37], in case there are notable differences also the annotation of the GS115 strain [36] is depicted.

- Extended discussion of differences between CBS7435 and GS115 strains
  In some cases we noticed strongly diverging annotations between the CBS7435 [37] and the GS115 [36] strain. The GS115 is a histidine auxotroph derived from the same strain as CBS7435 (deposited in a different strain collection as NRRL-Y 11430) by nitrosoguanidine mutagenesis. Therefore point mutations over the whole genome are to be expected, but complete rearrangements of entire loci are unlikely. Accordingly, most differences concerned annotations.
  Note that every single pair of adjacently organized genes coding for isoenzymes is differently annotated or sequence differences occur between CBS7435 and GS115: It has been previously noted that the *DAS1/DAS2* locus is flipped in the GS115 annotation [37]. In addition the *TAL1*/*TAL2* locus differs between the two strains: The *TAL2* gene is in the CBS7435 strain annotated longer than in the GS115 strain, with two introns in the 3' end. In CBS7435, *RKI1* is annotated as single long gene, whereas in GS115 two shorter genes are annotated (*RKI1* and *RKI2*). In case of *RPE1* and *RPE2*, large parts of the locus are missing in the GS115 strain, leading to a complete lack of *RPE* genes from the GS115 annotation (as discussed in the main manuscript).

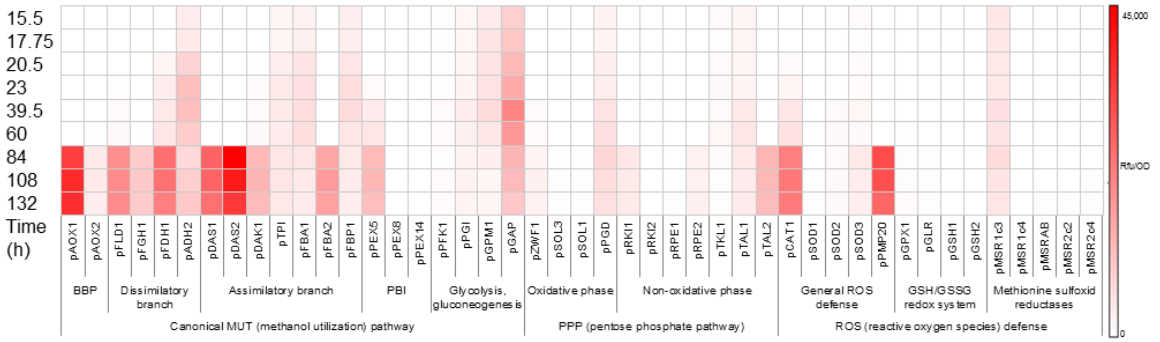
S 6

S 6: Different promoter lengths tested have no clear effect on reporter gene fluorescence.
The indicated promoters were tested following the same protocol as for the measurements in Fig. 3A. For $P_{CAT1\text{-}692}$ and $P_{CAT1\text{-}1000}$ stable genomically integrated transformants were used.



S 7

S 7: Full time series of DWP cultivations shown in Fig. 3 A. Additional time points measured are shown in heat map format. Methanol induction was performed from 60 h on, as in Fig. 3 A.



Chapter 3.2
Challenges and new opportunities by an autonomous replicating sequence coupled with the *CAT1* promoter of *Pichia pastoris*

Title page

Title

Challenges and new opportunities by an autonomous replicating sequence coupled with the *CAT1* promoter of *Pichia pastoris*

Authors

Thomas Vogl, Patrick Hyden, Richard Wasmayer, Astrid Weninger, Anton Glieder*

*Corresponding author

Keywords

Autonomously replicating sequence (ARS), strain stability, episomal expression, *Pichia pastoris*, *catalase 1* promoter

Abstract

The yeast *Pichia pastoris* is frequently used for heterologous protein production. Most commonly the methanol inducible promoter of the *alcohol oxidase 1* (P_{AOX1}) gene was used for gene expression. The promoter of the *catalase 1* (P_{CAT1}) gene has recently been reported to provide similarly tight regulation solely by varying glucose levels. Here we show that an AT-rich stretch in the previously used P_{CAT1} sequence (692 bp) contains an autonomously replicating sequence (ARS) promoting episomal plasmid propagation ambivalently affecting strain stability and protein yields.
For efficient heterologous gene expression in *P. pastoris*, expression cassettes are typically integrated into the *P. pastoris* genome resulting in strains stable even under non-selective conditions. Episomal plasmids were described to be lost upon growth under non-selective conditions. Removal of the ARS by shortening P_{CAT1} to 500 bp reduces background growth after transformation and increases strain stability under non-selective conditions. However, when maintaining selective pressure, the one-piece combination of an ARS and the strong *CAT1* promoter in a single sequence enabled up to seven-fold higher expression than genomic integration. Consequently, a truncated version of the *CAT1* promoter should be used for stable genomic integration, whereas the full length promoter including the ARS can be used for efficient episomal expression under selective conditions. Due to approximately 108-fold increased transformation rates and up to 3.5-fold less variability between transformants compared to genomic integration, episomal P_{CAT1} plasmids are ideal tools for screening large libraries for protein engineering or promoter studies.
Considering the frequency of ARSs in yeast (every 40-100 kbp), similar issues may also be faced in other yeast species, especially when evaluating a larger number of any genetic elements.

Introduction

Many industrially relevant proteins such as biocatalysts and also biopharmaceuticals are produced by heterologous gene expression. The yeast *Pichia pastoris* has emerged as one of the most commonly used microbial host systems for heterologous protein production due to its feasibility for high cell density bioreactor cultivations, high secretory capacities and strong promoters [1,2]. Most commonly the methanol inducible promoter of the *alcohol oxidase 1* gene (P_{AOX1}) is used to drive expression of heterologous genes [3]. P_{AOX1} is tightly repressed on carbon sources such as glucose and glycerol and approximately 1000-fold induced by methanol. This tight regulation allows to separate cell growth from heterologous protein production: At first *P. pastoris* is typically cultivated on glycerol to obtain a high cell density and subsequently induced with methanol to initiate expression of the gene of interest (GOI). Thereby even detrimental or toxic proteins can be produced. However, methanol is toxic and flammable making its use especially in large scale bioreactors undesirable.
Recently we have characterized 45 promoters of the *P. pastoris* methanol utilization pathway [4] and we have shown that the promoter of the *catalase 1* gene (P_{CAT1}) provides a distinct regulatory profile. P_{CAT1} is similar to P_{AOX1} tightly repressed on glucose or glycerol, but does not absolutely require methanol induction. Expression starts once the carbon source in the medium is used up ('derepression' [5]) reaching approximately 30 to 40 % of the space time yields (depending on the GOI) of methanol induced P_{AOX1} in small scale cultivations [4]. P_{CAT1} can also be induced with methanol and oleic acid reaching similar expression levels as P_{AOX1} at the end of a standard cultivation procedure. The derepressed regulatory profile allows methanol-free production, as the derepression phase can be maintained by feeding limiting amounts of glycerol or glucose in bioreactors (demonstrated with synthetic P_{AOX1} variants [5]).
However, we have encountered problems of background growth when transforming *P. pastoris* with vectors containing P_{CAT1} [4]: The majority of transformants were unstable resulting in loss of the selection marker and expression of the GOI, severely complicating the use of P_{CAT1}.
Here we show that removal of an AT-rich upstream sequence of P_{CAT1}, containing an autonomously replicating sequence (ARS), reduces background growth after transformation by increasing strain stability. Yet if used under selective conditions, the combination of P_{CAT1} and its endogenous ARS resulted also in up to seven-fold increased expression compared to genomic integration. Furthermore transformation rates were ~108 fold increased and the transformants showed up to 3.5-fold more uniform expression suggesting the application of the one-piece combination of P_{CAT1} and its ARS for screening large libraries required for example in protein or promoter engineering.

Materials & methods

Strains, materials, media and cultivation conditions

For cloning and plasmid propagation an *Escherichia coli* Top10 F' strain was used. *P. pastoris* transformations were mostly performed with the CBS7435 wildtype strain. The *GUT1* complementation plasmids [6] were transformed into a *gut1* knockout strain. Kits for plasmid isolation, gel purification and enzymes for cloning were used as recently described [7]. Additional Gibson assemblies were performed following standard procedures [8] using T5 exonuclease and *Taq* DNA ligase from New England Biolabs (Ipswich, MA, USA) and Phusion polymerase from Thermo Fisher Scientific (Waltham, MA, USA). Sanger sequencing was performed by LGC Genomics GmbH (Berlin, Germany) and Microsynth AG (Balgach, Switzerland). Media were prepared as outlined by Weis *et al*. [9], in short standard buffered minimal media with 1 % (w/v) glucose/dextrose (BMD) and full media (yeast extract, peptone, 2% glucose; YPD) were used. Additionally also buffered minimal medium with 1 % (w/v) glycerol (BMG) was used [6]. The following antibiotic concentrations were used: *E. coli*: LB-medium containing 25 µg/ml Zeocin, 50 µg/ml Kanamycin, 100 µg/ml Ampicillin; *P. pastoris*: 100 µg/ml Zeocin, 300 µg/ml Geneticin. Zeocin selection in liquid minimal BMD media was

attempted, but failed (presumably because of the pH or the high ionic strength). Therefore we used full media for Zeocin and Geneticin selection experiments. Deep well plate cultivations were performed as previously described [9], however no methanol induction was required for P_{CAT1} driven expression and the protocol therefore stopped after growth on glucose. Shake flask cultivations were performed in 250 ml baffled flask (25 ml BMD starting volume) with a starting OD_{600} of 0.05. The flasks were induced after 48 h with 25 ml BMM2 (1% methanol v/v to achieve a final concentration of 0.5%) and after 12 h, 24 h after the first induction with BMM10 (5% methanol v/v) [9]. Glucose concentrations were measured using a hexokinase method based kit (Glucose UV kit, DIPROmed (Vienna, Austria)).

Plasmid construction

eGFP reporter gene constructs with different selection markers are based on the shuttle vectors reported by Näätsaari *et al*. [6]. For Zeocin selection we used the restriction site free cloning (RSFC) [10] vector pPpT4mutZeoMlyI-intArg4-eGFP-BmrIstuffer previously reported [4] based on the pPpT4_S vector [6]. $P_{CAT1-1000}$, $P_{CAT1-692}$ and $P_{CAT1-500}$ vectors were available from a previous study [4].
putARS-P_{CAT1} was cloned into the vector by replacing the stuffer fragment by Gibson assembly [8] after PCR amplification using primers intARG4-pCAT1-764-Gib and eGFP-pCAT1-501rev-Gib (see Supporting information Table S 1) and verified by sequencing.
The one piece combination of the ARS and P_{CAT1} ($P_{CAT1-692}$) was also tested with alternative selection markers Geneticin and *gut1* complementation [6]. For Geneticin selection, the resistance cassette of the Zeocin vector was replaced with the Kanamycin/Geneticin cassette from pPpKan_S [6] (this cassette confers resistance to Kanamycin in *E. coli* and Geneticin in *P. pastoris*). The *GUT1* cassette was amplified from pPpGUT1 [6] (glycerol complementation for *P. pastoris*, Ampicillin for *E. coli*).
The Zeocin based reporter vector containing $P_{CAT1-692}$ was digested with *Bam*HI and *Pst*I and the backbones gel purified. The Kan/Gen resistance cassette was PCR amplified from pPpKan_S using primers AOX1TT-BamHI-pILV5-Gibson+pUC-Ori-PstI-AODTT-Gibson and incorporated into the vector backbones by Gibson assembly. The *GUT1* cassette was amplified from pPpGUT1 using primers AOX1TT-BamHI-pGUT1-Gibson and AmpR-GUT1TT-Gibson, the ampicillin cassette with primers GUT1TT-AmpR-Gibson and pUC-Ori-AmpR-Gibson. The two PCR fragments were assembled with the above mentioned *Bam*HI and *Pst*I backbone. In re-annotation efforts of the *P. pastoris* genome (personal communication Glieder A.), the maximal open reading frame of the *CAT1* gene is annotated with an additional DNA stretch in the 5'end containing an intron and 14 additional N-terminal amino acids. RNAseq data from two conditions (growth on glucose and methanol) do not provide evidence for this additional stretch and support the previously annotated translational start. The cassette for deleting the *GUT1* gene in *P. pastoris* was generated by amplifying homologous sequences 5' and 3' of the CDS and cloning them into a circular plasmid (*Pci*I and *Bgl*II digested backbone containing a Zeocin marker cassette, Ahmad M. personal communication/unpublished results). Roughly 1000 bp regions were selected to complement a *Swa*I restriction site (for linearization and targeting homologous recombination) and amplified with overhangs to the vector backbone (primers 3UTR-GUT1-F-Gibson+3UTRGUTR and 5UTRGUTF+5UTR-GUT1-R-Gibson). The PCR fragments were joined by OE-PCR and cloned into the vector backbone with Gibson Assembly. The PCR primers are listed in S 6.

Transformation, fluorescence and enzyme activity measurements and *gut1* strain

Competent *P. pastoris* cells were prepared and transformed using the condensed protocol of Lin-Cereghino *et al*. [11]. If applicable, plasmids were linearized with *Swa*I and equivalent amounts to one µg of the pPpT4_S vector transformed (as described in [4]), for circular plasmids 10 ng were

transformed. To avoid contaminations of linearized plasmids with uncut circular forms, the linearization reactions were loaded on agarose gels and the band corresponding to the linearized form was cut and purified.

For LuHNL activity measurements cells were lysed with Y-PER according to the manufacturer's recommendations. In case of activity exceeding the linear range of the plate reader used, the samples were diluted in 10-fold steps in citrate phosphate buffer (pH 5.0, 50 mM). Ten µl of the lysate supernatants or dilutions thereof were used for the activity assay and mixed with 140 µl citrate phosphate buffer (pH 5.0, 50 mM) and 10 µl of the substrate acetone cyanohydrin (300 mM) dissolved in citric acid (100 mM). The samples were incubated for 10 min and subsequently 10 µl of 100 mM N-chlorosuccinimid and 1 M succinimide were added to stop the reaction (incubated for 5 min). Thirty µl of barbituric acid (125 mM) and isonicotinic acid (65 mM) in 0.2 M NaOH were added for color development (measured at 600 nm for 10 min). A calibration curve with potassium cyanide (KCN) in a range of 0.025 to 0.2 mM was used to determine the absolute activities.

MeHNL activity was determined after aforementioned cell lysis using a mandelonitrile cyanogenesis assay described in literature [12] using a final mandelonitrile concentration of 15 mM.

eGFP fluorescence (ex./em. 488/507 nm) and absorption (600 nm, OD600) were measured and normalized using a Synergy MX plate reader (Biotek, Winooski, VT, USA) as outlined previously [7].

The *gut1* knock-out strain reported previously [6] was achieved in a *ku70* knockout strain. We aimed to use the wildtype strain background and created a *gut1* knockout following a similar strategy as Näätsaari *et al*. [6]. The strain was identified by screening transformants obtained on YPD+Zeo media for abolished growth on glycerol.

Results

Mapping the ARS region of P_{CAT1}

Expression cassettes are typically integrated into the *P. pastoris* genome resulting in strains stable even under non-selective conditions [1,2]. We hypothesized that the instability of P_{CAT1} bearing plasmids [4] may be related to problems with genomic integration. The P_{CAT1} length previously used was selected upstream from the start codon of the *CAT1* gene up to the end of the adjacent gene *LCP5*, resulting in a 692 bp fragment ($P_{CAT1\text{-}692}$; Fig. 1 A).

We analyzed the promoter sequence and noticed an AT-rich stretch in the 5' end of $P_{CAT1\text{-}692}$ probably serving as a terminator sequence of the *LCP5* gene (Fig. 1 A). Shortening the promoter to 500 bp length ($P_{CAT1\text{-}500}$) removes the AT-rich stretch and had increased strain stability [4]. AT-rich sequences are a common trait of transcription terminators and ARSs [13]. Recently ARSs of *P. pastoris* have been mapped by a high-throughput screen [14] based on deep sequencing (ARS-seq. [15]). Liachko *et al*. [14] thereby identified an ARS in P_{CAT1} and mapped the functional core to a 388 bp fragment (Fig. 1 A).

We cloned different fragments of P_{CAT1} into a vector containing an enhanced green fluorescent protein (eGFP) reporter gene to test if this ARS in P_{CAT1} is causing the strain instabilities and background growth (small colonies loosing expression over time). $P_{CAT1\text{-}1000}$, $P_{CAT1\text{-}692}$ and $P_{CAT1\text{-}500}$ provide different lengths of the promoter, a short AT-rich stretch of 264 bp was selected as putative ARS of P_{CAT1} (putARS-P_{CAT1}), 30 % shorter length than the functional core of Liachko *et al*. (Fig. 1 A).

P. pastoris cells were transformed with both circular and linearized forms of these vectors (Fig. 2 A). Linearization of plasmids results in highly recombinogenic DNA ends which drastically increase genomic integration rates compared to the circular form in yeast [16]. Standard *P. pastoris* vectors do not contain ARSs and cannot replicate episomally. Therefore transformation of cells with the circular form of the empty vector as control did not give any colonies (Fig. 2 A). Transformation with the circular forms of $P_{CAT1\text{-}1000}$, $P_{CAT1\text{-}692}$ and the short putARS-P_{CAT1} sequence did however show pronounced growth, whereas $P_{CAT1\text{-}500}$ did not show any growth. Transformations with linearized

forms of the plasmids resulted in transformants for all plasmids. These results confirm the function of the AT-rich stretch of P_{CAT1} as ARS.
Transformants with stable genomic integration of any promoter length tested showed identical reporter fluorescence (Fig. 1 B, Fig. 3 A), suggesting that the ARS is not affecting the strength of P_{CAT1}. Also the regulatory profile (repression/derepression/induction) was not affected, as demonstrated by comparing the three different promoter lengths in a time series (Fig. 1 B). As these sequences behaved identically, we conclude that the ARS is not required for the transcriptional regulation of P_{CAT1} by any means tested so far.

Vectors bearing the ARS of P_{CAT1} can replicate episomally even after linearization

For transformations with linearized ARS containing sequences ($P_{CAT1\text{-}1000}$, $P_{CAT1\text{-}692}$ and putARS-P_{CAT1}) two distinct types of colonies could be noticed: Big colonies (of similar size as the empty vector control and $P_{CAT1\text{-}500}$) and smaller colonies (Fig. 2 A). If the cells are incubated longer, the difference between the colonies gets less pronounced, suggesting different growth rates (data not sown). Initially we had not paid attention to the different colony sizes, and noticed strain instabilities when these transformants were further cultivated. Now we reasoned that the small colonies may be episomal, non-genomically integrated versions of the vectors, similar to transformants of the circular plasmids.
To prove this hypothesis, big and small colonies of the constructs were grown in liquid culture in 96 well deep well plates under selective (YPD+Zeocin) and non-selective conditions (YPD) and subsequently stamped to selective and non-selective media (Fig. 2 B). Big colonies of any construct showed uniform growth independent of the cultivation conditions with sizes comparable to colonies of the empty vector. Small colonies showed identical growth to big colonies on non-selective media. But if small colonies were transferred from non-selective media to selective media, they showed weak growth similar to circular plasmids. This would be the expected outcome for episomal plasmids: Under non-selective conditions the plasmids are not efficiently propagated and only maintained in a subset of the cell population, resulting in weaker growth. If the small colonies or circular plasmids are pre-grown under selective conditions, plasmid loss is depending on the experimental conditions: Either less severe than under non-selective conditions (Fig. 2 B) or completely rescued (S 2), fully restoring growth under selective conditions.
Since stamping from liquid culture involves a mixed population of cells, we streaked big and small colonies from linearized $P_{CAT1\text{-}692}$ and also colonies from a circular transformation on selective and non-selective agar plates. Subsequently single colonies were picked and streaked on selective media (Fig. 2 C). As expected, big colonies maintained growth under any condition (identical to $P_{CAT1\text{-}500}$) whereas small colonies and circular plasmids lost the ability to grow on selective media when precultivated under non selective conditions.
From these results we conclude, that big colonies contain stably integrated cassettes in the genome, whereas small colonies bear episomally replicating plasmids, providing an explanation for stability issues observed previously. This effect is not specific for selection with Zeocin but occurred also with Geneticin (S 3) and auxotrophy complementation of a glycerol auxotrophic *gut1* knockout strain.
Note that the empty vector and $P_{CAT1\text{-}500}$ are only showing tiny additional colonies (Fig. 2 A). Even if the plates are incubated for a longer time these colonies do not increase in size and also do not grow if streaked again on selective media (data not sown). Since transformation of circular forms of the empty vector and $P_{CAT1\text{-}500}$ are not showing any growth, we assume that the tiny colonies are not related to ARSs and caused by a different phenomenon.

The ARS of P_{CAT1} enables high episomal expression under selective pressure

Besides the stamping experiments of Fig. 2 B, also the fluorescence of the eGFP reporter was measured from the different lengths and colony sizes of P_{CAT1} (Fig. 3 A). Surprisingly, small and big colonies showed similar reporter fluorescence when cultivated under non-selective conditions,

suggesting that effect of the plasmid loss observed in Fig. 2 B, C is not severely affecting reporter protein fluorescence upon growth in deep well plates in full media. However, strains bearing episomal plasmids (linearized small colonies, circular) showed on selective media a five-fold higher reporter protein fluorescence than under non-selective conditions or compared to genomic integration (any big colonies, $P_{CAT1\text{-}500}$) (Fig. 3 A). This effect was even more pronounced with a different selection marker (Geneticin) leading to a more than seven-fold increase (Fig. 3 B).
These results suggest that the episomally replicating plasmids under selective pressure are simple tools for increasing expression.

Despite increased yields, it would be in most cases economically unfeasible to maintain selective pressure in larger scale cultivations using Zeocin or Geneticin, two relatively expensive antibiotics. Therefore we aimed to combine the ARS with selection by auxotrophy. Hereto, we used a *glycerol kinase 1* (*gut1*) knock out strain unable to metabolize glycerol and transformed this strain with complementation plasmids containing the wild-type *GUT1* gene [6] and $P_{CAT1\text{-}692}$ driving expression of the eGFP reporter gene. Since $P_{CAT1\text{-}692}$ and $P_{CAT1\text{-}1000}$ showed identical behavior with Zeocin, alternative selection markers were only tested with the shorter $P_{CAT1\text{-}692}$. Similarly to Zeocin or Geneticin driven expression, we obtained also a more than 4.4-fold increased reporter protein fluorescence under selective conditions (glycerol as sole carbon source) (Fig. 3 C), proving that also auxotrophy based selection is suitable to strongly increase expression from episomal plasmids. Therefore the combination with a carbon source auxotrophy *gut1* marker allows even cost efficient scale up.

We had previously especially noted instability issues when re-cultivating P_{CAT1} ARS containing plasmids from glycerol stocks. Therefore we used a glycerol stock of the cultivations shown in Fig. 2 B and Fig. 3 A to inoculate selective and non-selective media (S 2). Plasmid loss from these cultures inoculated from the glycerol stocks by stamping assays on selective media was even more severe than from direct inoculation (S 2 A vs. Fig. 2 B). Interestingly, in this case also the fluorescence of plasmid bearing constructs strongly decreased under non-selective conditions, suggesting almost complete plasmid loss (S 2 B vs. Fig. 3 A). These results imply that the ARS plasmids are more prone to loss under stress conditions such as freezing and re-cultivation.

The combination of P_{CAT1} and its endogenous ARS provide a screening system with improved transformation rates, increased yields and higher landscape uniformity

Transformation efficiencies of the circular ARS plasmid were on average 108 fold higher than using linearized expression cassettes needed for genomic integration (Tab. 1). High transformation efficiencies are needed when performing protein engineering and screening large random libraries of variants. However, such screening systems must not add additional bias to the results. Differences between variants should solely arise from mutations in the gene of interest and not because of different copy numbers or integration events. We tested the episomal P_{CAT1} plasmid for the expression of biocatalysts (hydroxynitrile lyases from *Manihot esculenta* [cassava; *MeHNL*] and *Linum usitatissimum* [flax; *LuHNL*]). We screened a larger number of transformants to compare the uniformity of the expression landscape of episomal replication ($P_{CAT1\text{-}692}$) and genomic integration ($P_{CAT1\text{-}500}$) (Fig. 4). As for the eGFP reporter gene (Fig. 3), also MeHNL and LuHNL expressed from episomal plasmids under selective pressure showed increased expression compared to genomic integration (3.5 and 4.9 fold comparing the mean values of the whole landscapes). Therefore the beneficial effects of the easy to fold and maintain eGFP could also be reproduced for more complex enzymes. Due to higher transformation efficiencies, considerably lower amounts of the plasmid (10

ng) could be used to achieve similar numbers of transformants of linearized cassettes. In addition no restriction endonuclease digestion and purification/desalting steps are needed for the ARS plasmids shortening experimental time and reducing costs.

The episomal P_{CAT1} plasmid resulted also in up to 3.5-fold more uniform expression than genomic integration (comparing the standard deviations in percent). For *Me*HNL, the transformant with the highest activity from genomic integration reached similar activity as average ARS transformants. For *Lu*HNL the best genomically integrated transformant reached only activity comparable to the worst episomal transformant. Some genomically integrated transformants did not show any detectable activity, while all episomal transformants were active. Clonal variability of genomic integration is known for *P. pastoris* [17] and may be attributable to differences in copy number or the locus of genomic integration. *P. pastoris* has lower rates of homologous recombination than *Saccharomyces cerevisiae* and linearized cassettes integrate at rates between less than 0.1 % up to 30% [6]. We used relatively high amounts of linearized DNA (3.5 µg) to obtain also multi copy strains, which may lead to a higher variability of the landscape. To this end we transformed also lower amounts of plasmid typically resulting only in single copy integration. However, the landscape uniformities were only marginally improved (S 5). In addition the use of larger amounts of DNA for transformation is preferable in order to obtain libraries with high numbers of individual transformants. Several transformations have to be done and transformants need to be pooled if low amounts of DNA are employed.
We tested the ARS system also for secreted proteins, that are difficult to fold: *Candida antarctica* lipase B (CalB) and horseradish peroxidase (HRP). The beneficial effects observed for episomal, intracellular *MeHNL* and *LuHNL* expression were not reproducible with CalB and HRP (S 6), possibly owing to secretion saturation [18] or unfolded protein response (UPR)/ endoplasmic-reticulum-associated protein degradation (ERAD) caused by the strong overexpression from the episomal plasmids under selective conditions.

Discussion

Here we have shown that removal of an ARS in P_{CAT1} reduces background growth after transformation and increases strain stability under non-selective conditions. If selection pressure is maintained, e.g. using auxotrophy complementation (*gut1*), the combination of P_{CAT1} and its ARS is a powerful tool enabling increased expression, higher transformation efficiencies and reduced clonal variability.
Therefore two distinct promoter sequences should be applied for different purposes:
On the one hand, the short $P_{CAT1\text{-}500}$ can be used to obtain solely genomic integration resulting in strains stable under non-selective conditions. On the other hand, plasmids containing also the P_{CAT1} ARS (e.g. $P_{CAT1\text{-}692}$) can be transformed in a circular form. These plasmids need not to be linearized and the transformation efficiencies are increased ~108-fold compared to linearized plasmids. Selective pressure is required for plasmid maintenance, resulting in up to seven-fold higher reporter protein fluorescence than from a cassette integrated in the genome. Therefore the full length ARS P_{CAT1} promoter can also be used as a fast screening system.

Notably, the effects of the ARS and background growth varied between selection markers used. Using Zeocin selection, big and small colonies could be easily discriminated (Fig. 2 A) whereas for Geneticin it was difficult to distinguish between big and small colonies (S 3), presumably due to less stringent selection than Zeocin. Since small colonies are on Zeocin vastly outnumbering big colonies (Fig. 2 A), we assume that they consist the majority of colonies picked from Geneticin. Even when trying to pick big colonies with genomic integration, we have apparently picked only one big colony and the rest episomally replicating colonies for $P_{CAT1\text{-}692}$. Therefore the high standard deviations of

Fig. 3 B are caused by the failure of properly selecting big and small colonies. In this case the values of the single measurements of each transformant are shown.

We would have initially expected that background growth and instability issues associated with ARSs can be resolved by linearization, as typical yeast ARS plasmids are circular. However, also for linearized ARS plasmids, the majority of transformants contained episomally replicating elements, lost under non-selective conditions. There are two possible explanations for this issue: On the one hand the plasmids could have been incompletely linearized prior to transformation, still leaving some circular molecules. We can reasonably rule out this explanation, as the linearized plasmids were run on an agarose gel and only the band of the correct size cut out and purified. On the other hand the plasmids could be replicating linearly, especially as *P. pastoris* is known to contain naturally a linear plasmid [19]. If and how telomere regions are added, may be a topic for future studies.

Interestingly, the ARS in P_{CAT1} is right after the adjacent gene *LCP5* (Fig. 1), suggesting that it also acts as transcriptional terminator region of this gene. In *S. cerevisiae* cases have been reported where terminators and ARSs overlap [13], presumably as similar sequence features may be needed for both tasks. This also highlights the tight organization of the *P. pastoris* genome, with regulatory features such as a terminator, an ARS and a promoter overlapping within less than 700 base pairs.

Typically genomic integration is the method of choice to express heterologous genes in *P. pastoris* [1,2]. Episomal plasmids have been, apart from very early efforts [20], seldom been used in *P. pastoris*. Recent efforts were based on Zeocin selection and separate ARS and promoter sequences [21–24]. Only one study [24] reported a beneficial effect of an episomal plasmid over genomic integration. The one piece combination of P_{CAT1} and its ARS led consistently for all intracellular proteins and markers tested to a 3.5 to 7-fold increased expression compared to genomic integration. We presume that this increase in expression is caused by copy number amplification of the ARS plasmid as previously observed in *S. cerevisiae* ([25]).
Expression from genomic integration can also be increased by integrating multiple copies of the vector, however typically only a small subset of transformants exhibits high copy numbers (Fig. 4 B,D) requiring tedious screening [18]. However all transformants of the ARS P_{CAT1} based system grown under selective conditions showed uniformly high expression.

In conclusion, the truncated $P_{CAT1\text{-}500}$ missing the ARS should be used for stable genomic integration, whereas the full length promoter (e.g. $P_{CAT1\text{-}692}$) can be used for extra-chromosomal expression under selective conditions. If similar instability problems should be faced with other sequences such as promoters and terminators, we suggest transforming the circular plasmid to rule out presence of an ARS as the cause.
Considering the frequency of ARSs in yeast (in *S. cerevisiae* every 40-100 kbp [26]), similar issues may also occur in other host systems, especially when evaluating a larger number any of genetic elements (promoters, terminators, random libraries). Therefore this work may not only be relevant for the application of P_{CAT1} in *P. pastoris*, but also for resolving similar issues with ARSs in other organisms.

Acknowledgement

The authors gratefully acknowledge support from NAWI Graz. We would like to thank Alexander Korsunsky for technical assistance.

Figure captions

Fig. 1

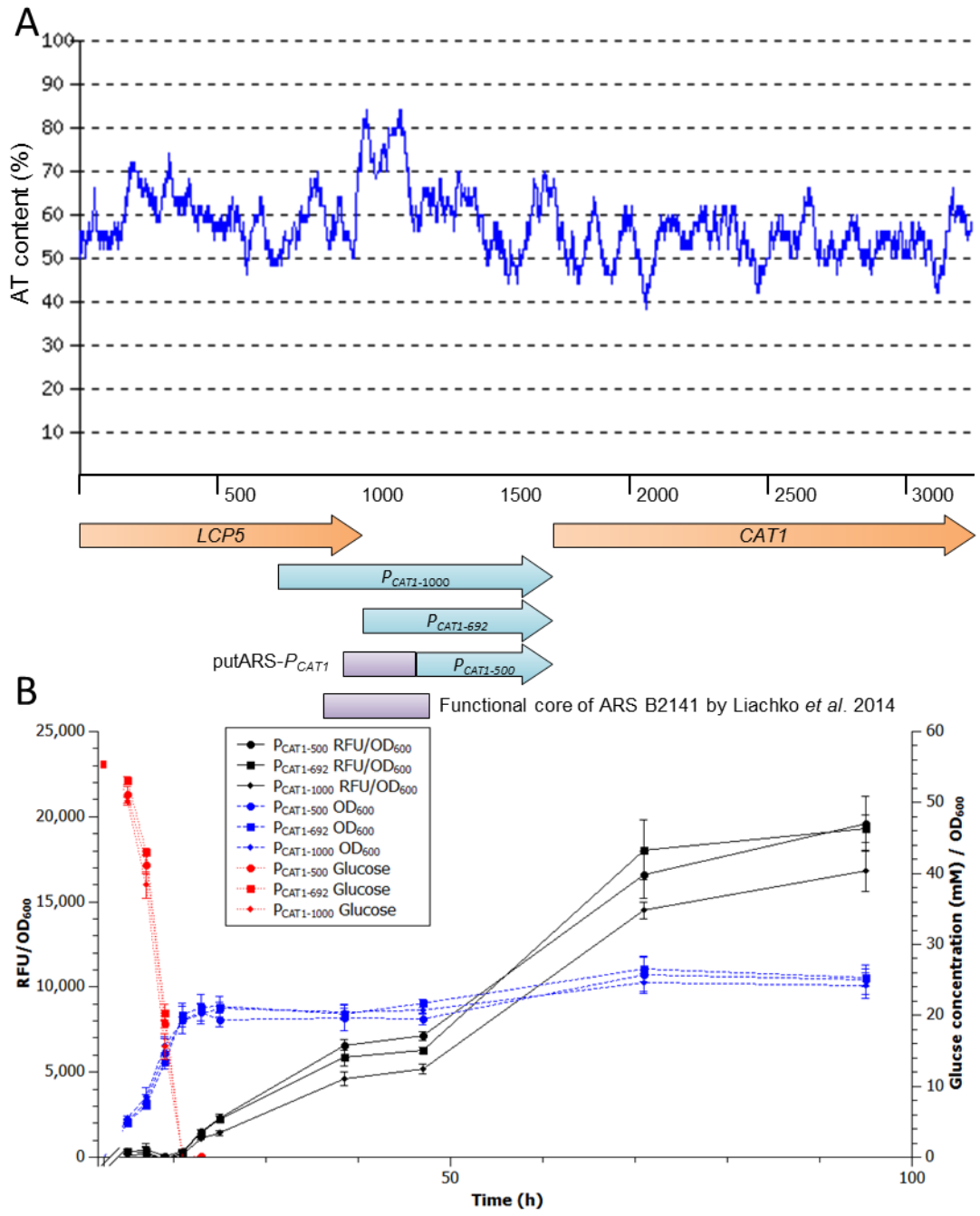


Fig. 1: The upstream region of the *CAT1* gene contains an AT-rich ARS which does not affect regulation or expression strength.

(A) The genomic locus of the *P. pastoris CAT1* gene is shown with gene annotations based on the sequencing of the CBS7435 strain [19]. An ARS identified by high throughput deep sequencing (ARS-seq) by Liachko *et al.* [14] is shown. The AT content was calculated with a sliding window of 50 bp using BitGene (http://www.bitgene.com/cgi/gene_analysis.cgi). The promoter sequences ($P_{CAT1-1000}$, $P_{CAT1-692}$ and $P_{CAT1-500}$) and the putative ARS of P_{CAT1} (putARS-P_{CAT1}, selected based on AT content) used in this study are indicated.

(B) The promoter lengths indicated were cloned upstream of an eGFP reporter gene and stable genomic *P. pastoris* transformants cultivated in shake flasks. Reporter protein fluorescence, OD_{600} and glucose concentrations were measured at the time points indicated. Mean values and standard deviations of biological triplicates shown. Cultures were induced with methanol after 48 h. At 0 h the flasks were inoculated to an OD_{600} of 0.05 and first measurements performed when the exponential growth phase was reached. The x-axis is broken between 1 and 14 h.

Fig. 2

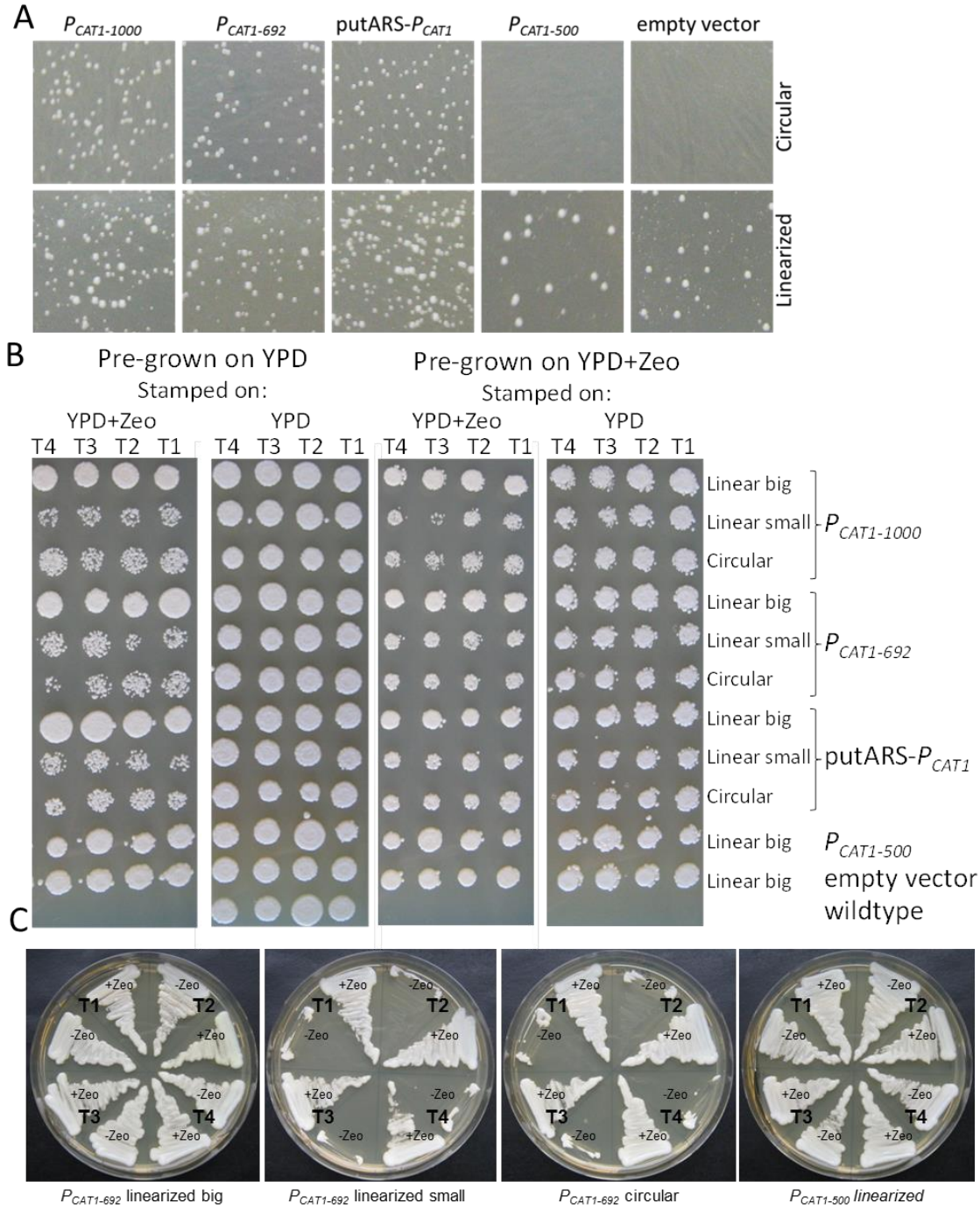


Fig. 2: The ARS of P_{CAT1} causes background growth and is unstable under non-selective conditions.

(**A**) Photos of agar plates after transformation of *P. pastoris* cells with circular or linearized plasmids containing the indicated lengths of P_{CAT1}. The empty vector control is the unmodified pPpT4_S vector [6] not containing P_{CAT1}. The circular plasmids showed higher transformation rates, therefore only 10 ng were transformed and the whole transformation reaction plated. For linear plasmids, one µg was transformed and one fifth of the transformation reaction plated.

(**B**) Evaluating plasmid stability by determining growth on selective (YPD+Zeo) and non-selective (YPD) media from liquid culture. Four colonies of transformants (T1-T4) of the indicated plasmids and colony sizes were inoculated in liquid culture under selective (YPD+Zeo) and non-selective (YPD) conditions for 60 h and subsequently stamped (diluted 1:1000) on selective and non-selective agar plates. The empty pPpT4_S vector is included as control for stable genomic integration, the wildtype strain to test Zeocin selection.

(**C**) Stability of single colony separated transformants under selective and non-selective conditions. Single colonies from four transformants (T1-T4) of $P_{CAT1-692}$ (as a representative ARS containing plasmid) and $P_{CAT1-500}$ (ARS free control) were re-streaked on selective (YPD+Zeocin) and non-selective conditions (YPD). Subsequently single colonies were streaked adjacently on selective media to monitor plasmid loss by growth (photos after 48 h incubation at 28°C). In case of $P_{CAT1-692}$ transformants of linearized vector (big and small colonies) and circular plasmid were used.

Fig. 3

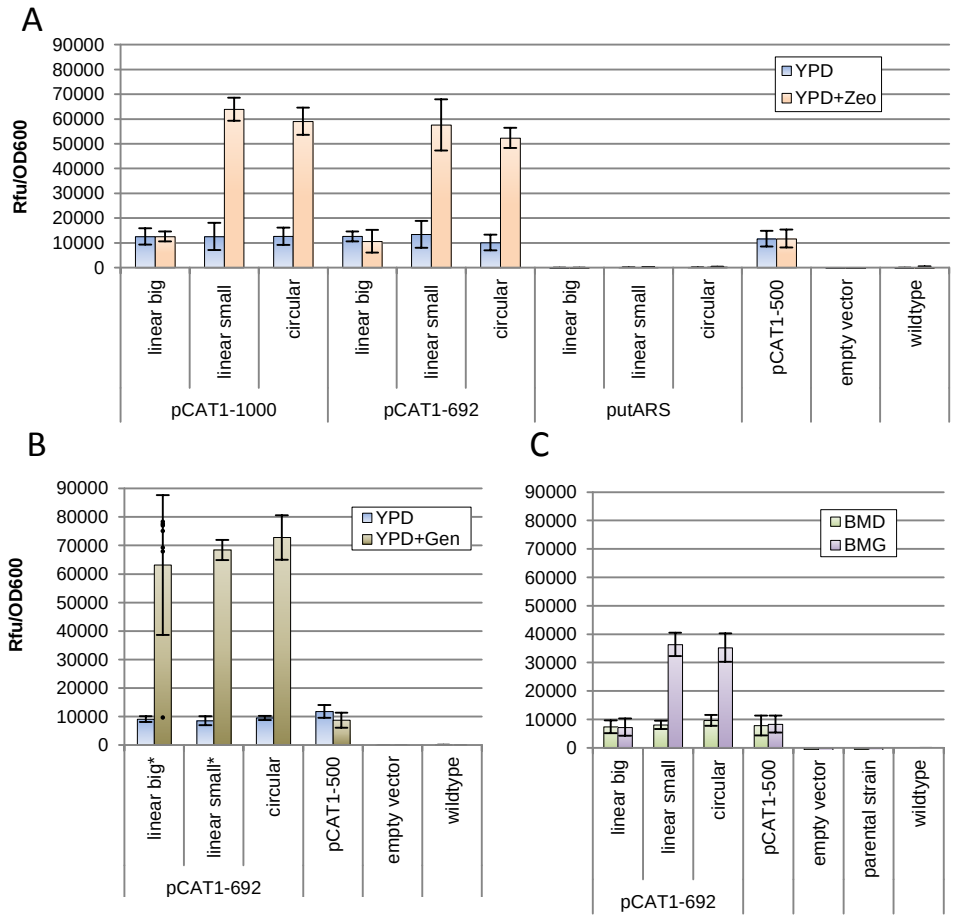


Fig. 3: Episomal replicating plasmids containing the ARS of P_{CAT1} show increased expression under selective pressure compared to genomic integration.

Reporter protein fluorescence of the indicated plasmids and colony sizes was measured with the selection markers (**A**) Zeocin (**B**) Geneticin and (**C**) *GUT1* complementation. The strains were grown in selective (YPD+Zeo/Gen; BMG [buffered minimal glycerol] and non-selective (YPD, BMD (buffered minimal dextrose)) media for 60 h (see also Fig. 2 B). The empty vector controls are: for Zeocin pPpT4_S, for Geneticin pPpKan_S, and for glycerol auxotrophy pPpGUT1 [6]. For *GUT1* complementation selection, also the auxotrophic parental strain was included. Mean values and standard deviations of seven different transformants are shown. '*': In case of Geneticin selection, barely any growth differences were noticeable between colonies (see photos in S 3), therefore the big and small colonies indicated are only putative. Due to the high standard deviation, for putative big colonies of $P_{CAT1\text{-}692}$ on Geneticin, also the fluorescence value of each single transformant is shown as an inlet.

Fig. 4

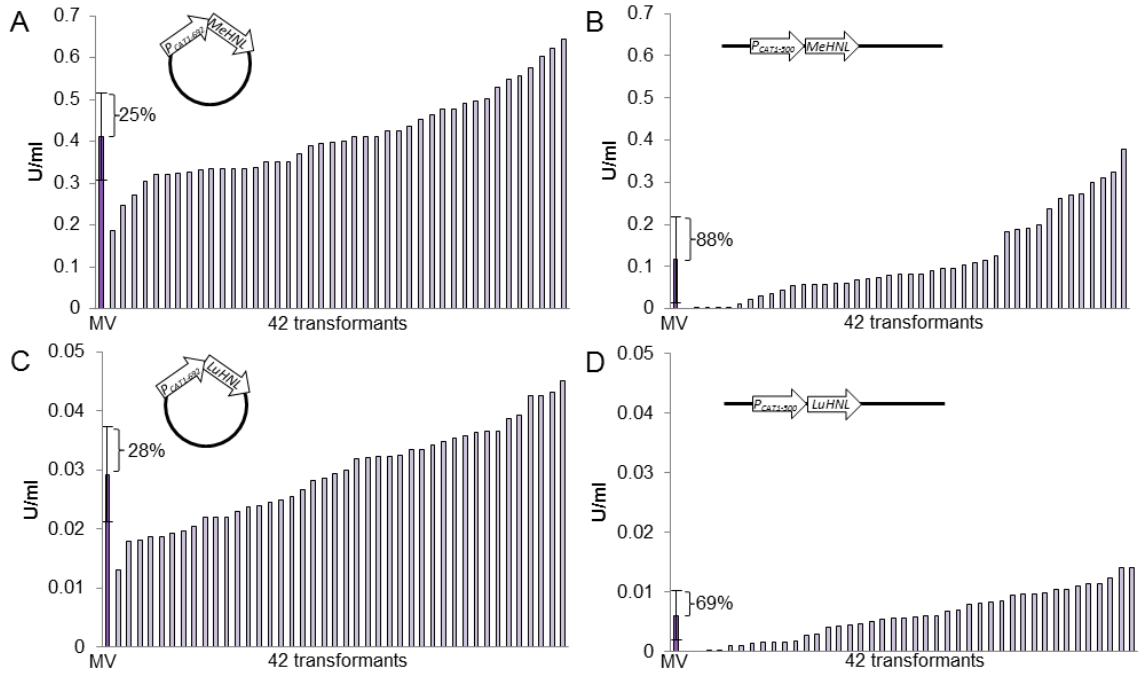


Fig. 4: The combination of the *CAT1* promoter and its endogenous ARS gives up to 4.9-fold higher yields for the biocatalysts *MeHNL* and *LuHNL* and the transformants show up 3.5-fold more uniform expression.

MeHNL **(A,B)** and LuHNL **(C,D)** were expressed from a circular plasmid bearing the $P_{CAT1\text{-}692\ +\ ARS}$ combination **(A,C)** or a linearized plasmid bearing $P_{CAT1\text{-}500}$ **(B,D)**. MeHNL and LuHNL activity were measured after growth under selective conditions (glycerol) for 60 h using. Forty-two transformants were compared per construct after growth in 96 well deep well plates on glycerol for 60h. The mean value (MV) and standard deviation (SD) of all transformants per construct are shown on the left side of each panel. The SD is also provided as percent of the MV. See S 4 for similar experiments using Zeocin as marker.

Tab. 1

Tab. 1: P_{CAT1}-ARS plasmids show in *P. pastoris* approximately 100 fold higher transformation efficiencies than linear cassettes targeting genomic integration. Ten ng of the circular ARS plasmids $P_{CAT1\text{-}692}$ and approximately 1 µg of the $P_{CAT1\text{-}500}$ plasmids (linearized to target genomic integration) were transformed. Transformation efficiencies were calculated as colony forming units (cfu) per µg DNA. Mean value and standard deviations of quadruplicates for ARS and genomic integration were calculated.

Type of plasmid	Circular ARS ($P_{CAT1\text{-}692}$)				Genomic integration ($P_{CAT1\text{-}500}$)			
Marker	GUT1		Zeocin		GUT1		Zeocin	
GOI	*MeHNL*	*LuHNL*	*MeHNL*	*LuHNL*	*MeHNL*	*LuHNL*	*MeHNL*	*LuHNL*
Transformation efficiency (cfu/µg DNA)	27500	40800	11900	88400	404	162	467	528
MV±SD	42150 ± 33018				390 ± 160			

References

[1] Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, et al. Pichia pastoris: protein production host and model organism for biomedical research. Futur Microbiol 2013;8:191–208.

[2] Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 2014.

[3] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.

[4] Chapter 3.1 of this thesis (Vogl *et al*., manuscript in prepration)

[5] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.

[6] Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 2012;7:e39720.

[7] Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic Core Promoters for Pichia pastoris. ACS Synth Biol 2014;3:188–91.

[8] Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.

[9] Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A. Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 2004;5:179–89.

[10] Chapter 3.6.3 of this thesis (Vogl *et al*., manuscript in prepration)

[11] Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, et al. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 2005;38:44, 46, 48.

[12] Wiedner R, Gruber-Khadjawi M, Schwab H, Steiner K. Discovery of a novel (R)-selective bacterial hydroxynitrile lyase from Acidobacterium capsulatum. Comput Struct Biotechnol J 2014;10:58–62.

[13] Chen S, Reger R, Miller C, Hyman LE. Transcriptional terminators of RNA polymerase II are associated with yeast replication origins. Nucleic Acids Res 1996;24:2885–93.

[14] Liachko I, Youngblood R a, Tsui K, Bubb KL, Queitsch C, Raghuraman MK, et al. GC-Rich DNA Elements Enable Replication Origin Activity in the Methylotrophic Yeast Pichia pastoris. PLoS Genet 2014;10:e1004169.

[15] Liachko I, Youngblood R a, Keich U, Dunham MJ. High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. Genome Res 2013;23:698–704.

[16] Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A 1981;78:6354–8.

[17] Cregg JM, Tolstorukov I, Kusari A, Sunga J, Madden K, Chappell T. Expression in the yeast Pichia pastoris. Methods Enzymol 2009;463:169–89.

[18] Aw R, Polizzi KM. Can too many copies spoil the broth? Microb Cell Fact 2013;12:128.

[19] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.

[20] Cregg JM, Barringer KJ, Hessler AY, Madden KR. Pichia pastoris as a host system for transformations. Mol Cell Biol 1985;5:3376–85.

[21] Lee CC, Williams TG, Wong DWS, Robertson GH. An episomal expression vector for screening mutant gene libraries in Pichia pastoris. Plasmid 2005;54:80–5.

[22] Uchima CA, Arioka M. Expression and One-Step Purification of Recombinant Proteins Using an Alternative Episomal Vector for the Expression of N-Tagged Heterologous Proteins in Pichia pastoris. Biosci Biotechnol Biochem 2014;76:368–71.

[23] Sandström AG, Engström K, Nyhlén J, Kasrayan A, Bäckvall J-E. Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid. Protein Eng Des Sel 2009;22:413–20.

[24] In-Pyo H, Anderson S, Choi S. Evaluation of a New Episomal Vector Based on the GAP Promoter for Structural Genomics in. J Microbiol Biotechnol 2006;16:1362–8.

[25] Bitoun R, Zamir A. Spontaneous amplification of yeast CEN ARS plasmids. Mol Gen Genet 1986;204:98–102.

[26] Poloumienko A, Dershowitz A, De J, Newlon CS. Completion of replication map of Saccharomyces cerevisiae chromosome III. Mol Biol Cell 2001;12:3317–27.

[27] Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS, Glieder A. Real-time PCR-based determination of gene copy numbers in Pichia pastoris. Biotechnol J 2010;5:413–20.

Supporting information

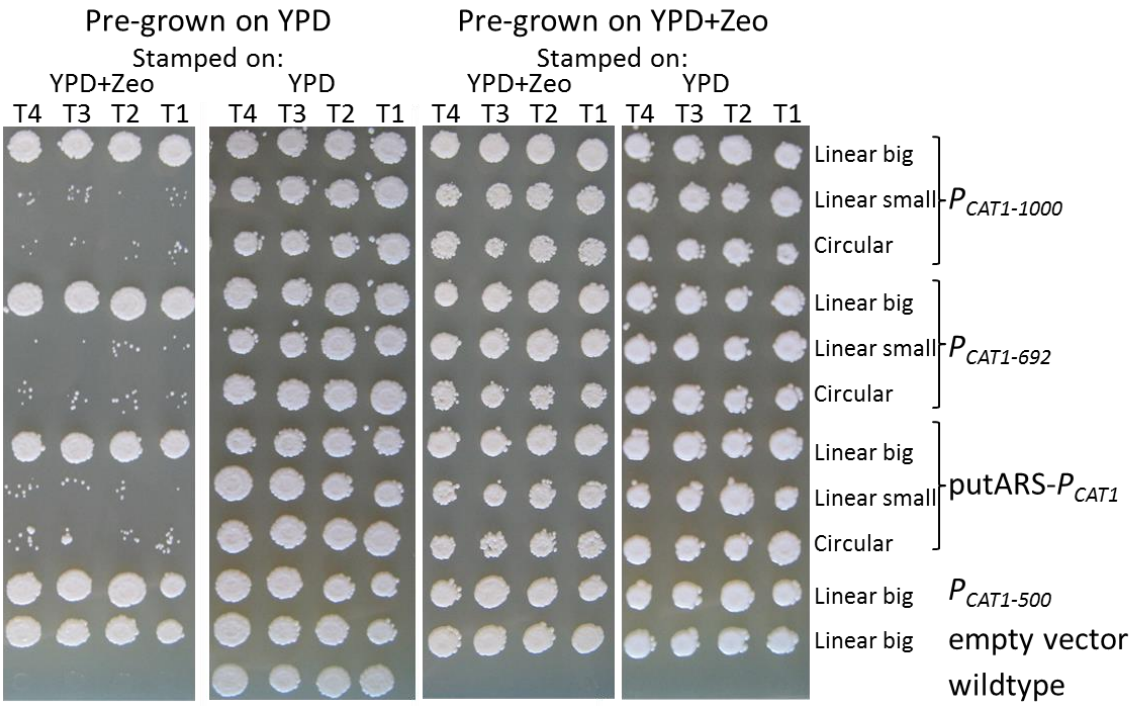
S 1

S 1: Primers used in this study.

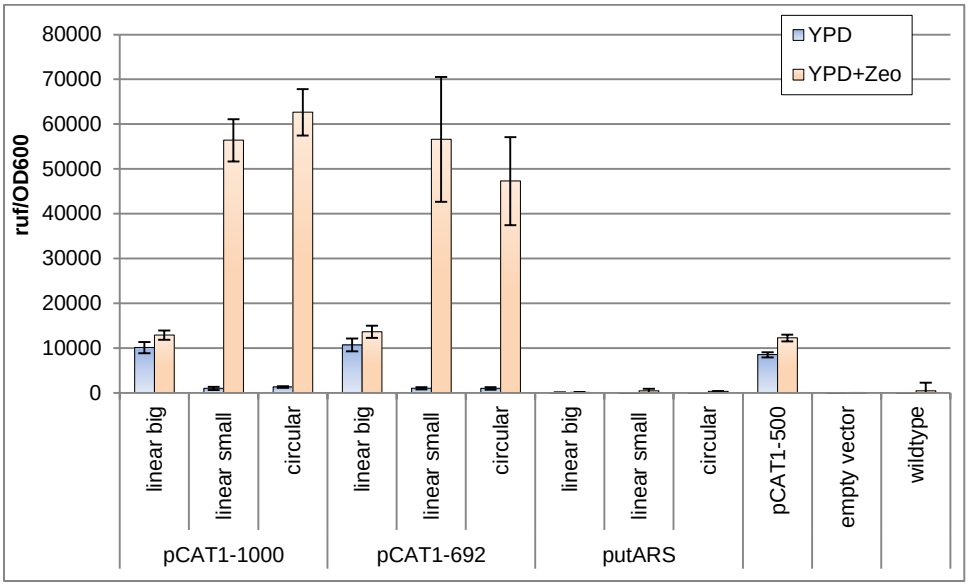
Name	Sequence
intARG4-pCAT1-764-Gib	gttagtagatatttataccattctgcgagaaggtcctaaaagtgcgaggaagaataaaaatactgcttc
eGFP-pCAT1-501rev-Gib	agtgaaaagttcttctcctttgctagccatcgtagaaaaaaatgtggtgaaacagtttcataagagttatatac
AOX1TT-BamHI-pILV5-Gibson	cagaagattaagtgagaccttcgtttgtgcggatccttcagtaatgtcttgtttctttgttgcag
pUC-Ori-PstI-AODTT-Gibson	ctacggggtctgacgctcagtggtacctgcagctaaggtaatcagatccaagtttccccaatc
AOX1TT-BamHI-pGUT1-Gibson	cagaagattaagtgagaccttcgtttgtgcggatccataccgaaaggtaaacaacttcggggaattg
AmpR-GUT1TT-Gibson	ggcgtatcacgaggccctttcgtctgccagagctgtcacatacttgaaatagggttg
GUT1TT-AmpR-Gibson	caaccctatttcaagtatgtgacagctctggcagacgaaagggcctcgtgatacgcc
pUC-Ori-AmpR-Gibson	gatcttttctacggggtctgacgctcagtaacgaaaactcacgttaagggattttggtc
3UTR-GUT1-F-Gibson	cctggccttttgctggccttttgctcagagcagctgtaattatattatcatgttaggtca
3UTRGUTR	gtgtttgctgtaggatgacctagatttaaatataagaggaaacaacgttcgtatcgtga
5UTRGUTF	cacgatacgaacgttgtttcctcttatatttaaatctaggtcatcctacagcaaacacc
5UTR-GUT1-R-Gibson	ctaagatacgttccgttcctatagtagatatatctgtggtatagtgtgaaaaagtagaag

S 2

A



B

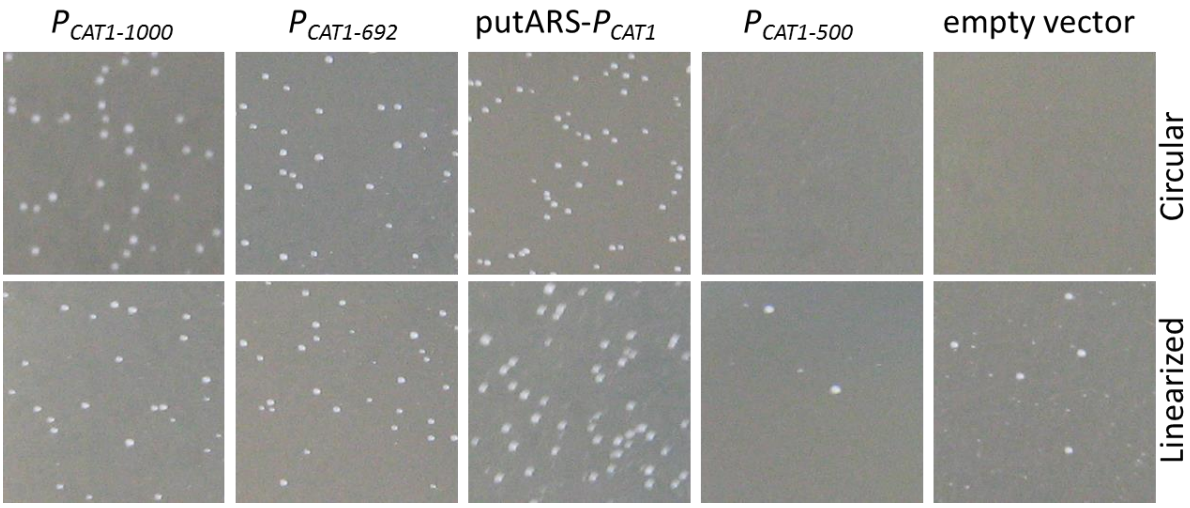


S 2: Effects of plasmid loss are even more severe when inoculated from glycerol stock.

(A) The same experiment as shown in (Fig. 2 B) was repeated with inoculation from 96 well glycerol stocks. Glycerol stocks in 96 well microtiter plates of the YPD cultivations shown in Fig. 2 B and Fig. 3 A were used to inoculate selective (YPD+Zeo) and non-selective (YPD) media. Therefore the same transformants (T1-T4) shown in Fig. 2 B of the indicated plasmids and colony sizes were used. After cultivation for 60 h the cultivations were diluted 1:1000 and stamped on selective and non-selective agar plates. The empty pPpT4_S vector is included as control for stable genomic integration, the wildtype strain to test Zeocin selection.

(B) Fluorescence measurements of the cultivations described in panel A, identical to Fig. 3 A except being inoculated from glycerol stocks.

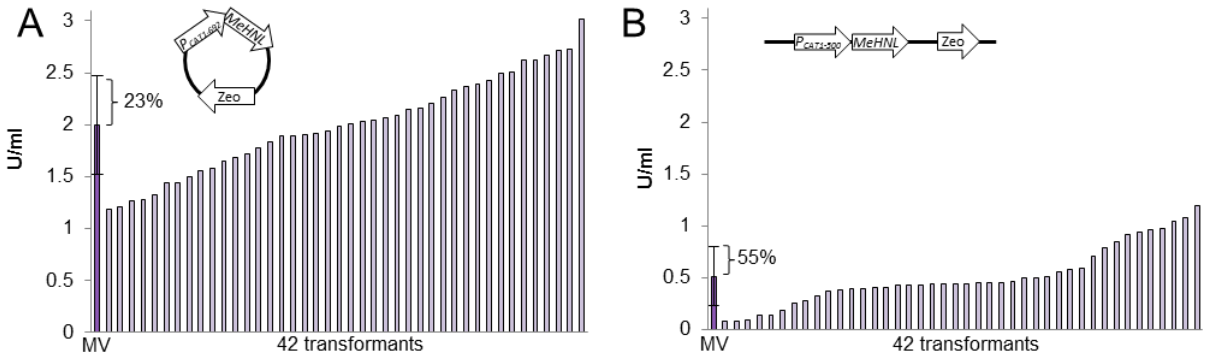
S 3



S 3: Geneticin as selectin marker shows similar results as Zeocin, yet it is difficult to discriminate between big and small colonies.

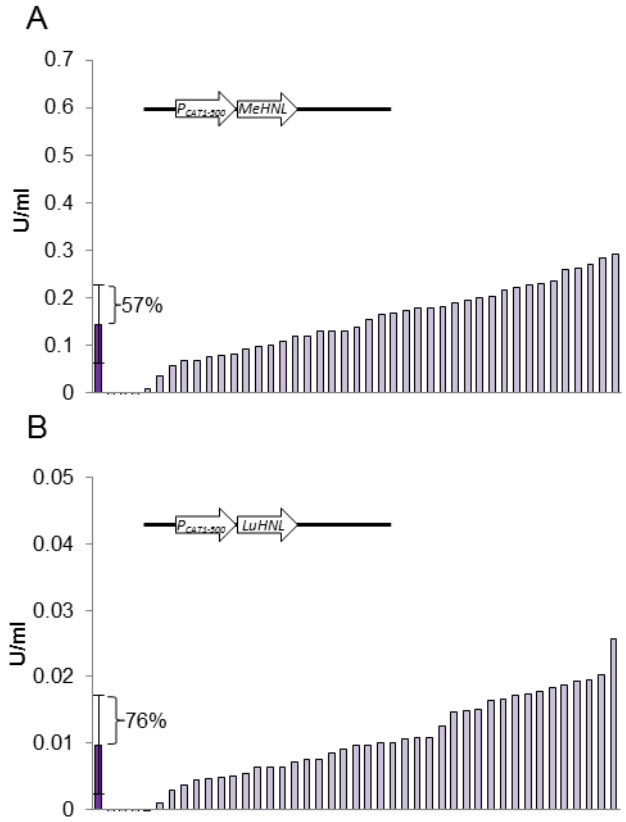
The same experiment as shown in (Fig. 2 A) was repeated with plasmids bearing a resistance cassette against Geneticin. Photos of agar plates after transformation of *P. pastoris* cells with circular or linearized plasmids containing the indicated lengths of P_{CAT1}. The empty vector control is the unmodified pPpKan_S vector [6] not containing P_{CAT1}. Non episomally replicating plasmids ($P_{CAT1-500}$, empty vector) showed lower transformation rates with Geneticin, than with Zeocin (compare to Fig. 2 A).

S 4



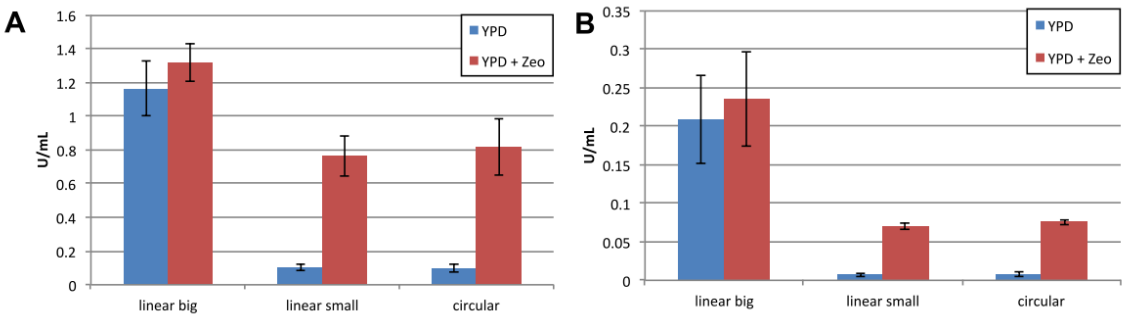
S 4: MeHNL tested with Zeocin instead of GUT1 selection shows similar results. Mean value of activity with ARS is about 3.9 times higher than genomic integration. Same experiment as Fig. 4, except the plasmids contained a Zeocin resistance gene and the cultivation was performed in YPD-Zeo full media. We also attempted selection in minimal BMD media with Zeocin, but also the untransformed wildtype strain showed growth (presumably due to the pH or high salt concentrations being unfavorable for Zeocin selection). We assume that the use of full media led also to higher activities compared to BMG minimal media used in Fig. 4. *LuHNL* was also tested with Zeocin selection on full media supplemented with zinc sulfate. However no clear activity was detected (data not shown). We presume that these negative effects were caused by the combination of zinc supplementation and zeocin selection. In similar experiments under derepressed conditions excess of heavy metals also abolished LuHNL activity (unpublished results). LuHNL expression on minimal media using the *GUT1* marker (Fig. 4C,D) was possibly successful as no additional Zeocin stress was present and the minimal media may contain less zinc.

S 5



S 5: Reducing the DNA amount of linearized $P_{CAT1-500}$ plasmids to approximately 1000 ng marginally improved landscape uniformity for *Me*HNL (A) and *Lu*HNL (B) activity (compare to Fig. 4 B,D). Same experiment as in Fig. 4 except only amounts equivalent to 1000 ng of the pPpT4S vector [6] were transformed. Landscape uniformities are changed for MeHNL from 88% percental standard deviation (Fig. 4 B) to 57% (A) and for LuHNL from 69% (Fig. 4 D) to 76 % (B). Further reducing the DNA amounts transformed may lead to uniformity improvements at the cost of further reduction of transformants numbers.

S 6



S 6: Increased yields observed with P_{CAT1}+ARS driven expression of intracellular proteins GFP, MeHNL and LuHNL were not reproducible for secreted proteins CalB (A) and HRP (B), possibly owing to secretion saturation UPR/ERAD upregulation or lack of nutrients in the derepressed phase. Mean value and standard deviations of the volumetric activity of 14 transformants measured after 60 h growth on the respective media are shown. It has previously been shown that multicopy strains of CalB [27] even show reduced activities compared to single copy if expressed without helper proteins. Similar effects were also noticed for HRP (Krainer F.W. *et al*. manuscript in preparation/personal communication). Since we used the derepressed expression from P_{CAT1} for production of CalB and HRP, another explanation would be that the cells, depleted of nutrients, could not provide energy to secrete high amounts of CalB and HRP.

Chapter 3.3
Orthologous promoters from related methylotrophic yeasts for protein expression in *Pichia pastoris*

Title

Orthologous promoters from related methylotrophic yeasts for protein expression in *Pichia pastoris*

Authors

Thomas Vogl, Patrick Hyden#, Jasmin Elgin Fischer#, Richard Wasmayer#, Lukas Sturmberger, Anton Glieder

# These authors contributed equally.

Abstract

Methylotrophic yeasts such as *Pichia pastoris* (*Pp*), *Hansenula polymorpha* (*Hp*), *Candida boidinii* (*Cb*) and *Pichia methanolica* (*Pm*) are widely used protein production platforms in biotechnology. Typically strong, tightly regulated promoters of their methanol utilization (MUT) pathways are used to drive heterologous gene expression. Despite the highly similar genes in the MUT pathways of the four yeasts, the regulation of the promoters varies strongly. While *Pp* MUT promoters remain tightly repressed on glycerol and after depletion of a repressing carbon, *Hp*, *Cb* and *Pm* MUT promoters are derepressed to up to 70 % of methanol induced levels, enabling methanol free production processes. Here, we have tested six orthologous promoters from *Hp*, *Cb* and *Pm* in *Pp*. Three promoters (P_{CbAOD1}, P_{PmMOD1}, P_{PmMOD2}) did not show any activity and P_{CbFLD1} reached weak expression. The promoter of the *HpMOX* gene reached about the same expression levels as the methanol state of the art promoter of the *Pp alcohol oxidase 1* gene (P_{PpAOX1}). The promoter of *HpFMD* reached methanol-free/derepressed up to 75 % of methanol induced P_{PpAOX1} and reached similar levels to the strong constitutive *Pp GAP* promoter. Upon methanol induction P_{HpFMD} in *Pp* even surpassed P_{PpAOX1} up to two-fold.
Our results demonstrate that orthologous promoters from related yeast species can give access to otherwise unobtainable regulatory profiles and may even considerably exceed endogenous promoters.

Introduction

Recombinant proteins such as biopharmaceuticals or industrially relevant biocatalysts are most commonly produced by heterologous gene expression in microorganisms. *Escherichia coli, Saccharomyces cerevisiae* and filamentous fungi have been widely used expression hosts since the advent of recombinant protein production. Over the last two decades, the methylotrophic yeasts *Pichia pastoris* (*Pp*), *Hansenula polymorpha* (*Hp*), *Candida boidinii* (*Cb*) and *Pichia methanolica* (*Pm*) have emerged as powerful alternatives, enabling high cell density fermentation and secretion of heterologous proteins [1–3]. Amongst these four yeast species, *P. pastoris* is most commonly applied for heterologous protein production, even surpassing *S. cerevisiae* according to a recent literature survey [4].
All methylotrophic yeasts offer tightly regulated, strong promoters that are naturally regulating the expression of genes involved in methanol utilization (MUT) [3]. Typically all promoters of MUT genes are firmly repressed on repressing carbon sources such as glucose and get strongly upregulated when shifted to methanol. However, derepression effects vary considerably between species [3] and even within the same organism [5]. Derepression leads to activation of the promoter when the repressing carbon source is depleted or when a non-repressing carbon source is present. The promoter of the *alcohol oxidase 1* gene in *P. pastoris* (P_{PpAOX1}) is under derepressed conditions only expressed to 2-4 %

of methanol induced levels [6]. In contrast, the promoter of the orthologous gene (named differently: *methanol oxidase, MOX*) in *H. polymorpha* (P_{HpMOX}) shows derepressed expression up to 70% of methanol induced levels. Also the promoters of the orthologous genes in *C. boidinii* (*alcohol oxidase 1*, abbreviated *AOD1*) and *P. methanolica* (*methanol oxidase 1/2*, abbreviated *MOD1/2*) show derepression up to 70 % of methanol induced levels [3]. Note that the *alcohol oxidase/methanol oxidase* genes fulfilling the same function were assigned different three letter abbreviations in all four yeasts. We keep these identifiers in addition to the prefixes *Pp*, *Hp*, *Cb* and *Pm* to differentiate between the organisms.

Induction with toxic and flammable methanol is especially in large scale production processes unwanted due to safety issues making strong derepressed promoters a favorable alternative. Lately, also P_{PpAOX1} variants [7], alternative promoters [8] and novel MUT promoters [5] showed derepression to varying extents in *P. pastoris* making them sought-after expression tools enabling methanol free processes.

Recent studies in metazoans [9] and yeast [10] showed that orthologous, highly divergent promoter sequences from different species can achieve similar expression. The promoters of the genes coding for orthologous ribosomal proteins in various yeast species, showed high expression conservation in *S. cerevisiae* [10]. We hypothesized that also different MUT promoters of related methylotrophic yeasts may show some extent of conservation as demonstrated previously by Raschke *et al*. [11] for the P_{PpAOX1} in *H. polymorpha*.

Here we have tested commonly used MUT promoters from *Hp*, *Cb* and *Pm* in *Pp* and some of these promoters performed surprisingly well even outperforming *Pp* endogenous promoters.

Results

We selected six heterologous promoters of the *HpFMD, HpMOX, CbFLD1, CbAOD1*, *PmMOD1* and *PmMOD2* genes for testing in *P. pastoris*. The promoters were compared to state of the art benchmark promoters in *P. pastoris*, the methanol inducible P_{AOX1}, constitutive P_{GAP} and derepressed/methanol inducible P_{CAT1}. The orthologous promoters were PCR amplified from genomic DNA and cloned into a reporter vector previously established for promoter comparisons in *P. pastoris* [5]. The promoters were seamlessly fused to the enhanced green fluorescent reporter gene (eGFP), (*i.e.* maintaining the natural transition of promoter to start codon without additional restriction sites or linker sequences in between). Sequencing showed that the promoter sequences contained minor differences compared to sequences previously reported (S 2). These differences are possibly arising from the use of genomic DNA from *Hp*, *Cb* and *Pm* strains from different strain collections than previously reported as PCR templates (see Materials and Methods section).

When transforming the plasmids into *P. pastoris*, we noticed seemingly higher transformation rates of the plasmid bearing P_{CbAOD1}. We had previously experiences similar problems with an autonomous replicating sequence (ARS) in *P. pastoris* P_{CAT1} (Vogl T. *et al*., unpublished results, [12]). Further experiments showed that also the highly AT-rich *CbAOD1* promoter sequence acts as an ARS in *P. pastoris* (S 3). Surprisingly the linearized plasmid bearing P_{CbAOD1} showed higher transformation rates than the most commonly used *P. pastoris* ARS (*ARS1* [13]). Further experiments may help to find an explanation for this effect and may show if P_{CbAOD1} bearing plasmids are also stably maintained and allow higher expression levels than *P. pastoris* endogenous ARSs [14].

P. pastoris transformants of plasmids bearing *CbAOD1*, *PmMOD1* and *PmMOD2* promoters did not show any reporter protein fluorescence (Fig. 1A). P_{CbFLD1} showed repression on glucose and weak methanol inducible expression of about 10 % of P_{PpAOX1}. Both *H. polymorpha* promoters tested maintained their natural regulation and showed repression, derepression and methanol induction profiles (Fig. 1A,B). The *HpMOX* promoter showed weak derepressed reporter protein fluorescence and reached similar reporter protein fluorescence on methanol as P_{PpAOX1}. The *HpFMD* promoter showed derepressed expression outperforming the constitutive P_{PpGAP} and reaching approximately 75 % of the methanol induced P_{PpAOX1}. Derepressed expression from P_{HpFMD} exceeded reporter protein fluorescence of the strongest derepressed endogenous MUT promoter from *P. pastoris* (P_{CAT1}) about 3.5 fold. These derepressed expression levels of P_{HpFMD} are similar to glucose regulated inducible promoters described by Prielhofer *et al*. [8]. Upon methanol induction P_{HpFMD} even outperformed P_{PpAOX1} about two-fold. However, in deep well plate cultivations (Fig. 1A) P_{HpFMD} seemed to give also a very weak signal under glucose repressed conditions, hinting constitutive activity. Expression from the P_{HpMOX} and *P. pastoris* P_{AOX1} and P_{CAT1} was undetectable. We repeated the experiments in shake flasks measuring also glucose levels (Fig. 1B). P_{HpFMD} showed again weak constitutive expression before glucose depletion. This result may suggest that the exceptional strength of P_{HpFMD} (clearly outperforming even *P. pastoris* endogenous promoters) is at the expense of less tight regulation. However, constitutive activity of P_{HpFMD} is less than 1 % of fully induced levels, showing still induction over two logs.

We were interested if the exceptionally strong expression of P_{HpFMD} could also be reproduced with other proteins than eGFP. Therefore the P_{HpFMD} promoter was cloned upstream of the coding sequence of secretory proteins horseradish peroxidase (HRP) and *Candida antarctica* lipase B (CalB) and the intracellularly expressed hydroxynitrile lyase from could *Manihot esculenta* (MeHNL) (Fig. 1C-E). Derepressed expression of all proteins matched constitutive expression from P_{PpGAP} and clearly outperformed derepressed expression from $P_{PpCAT1.}$ Methanol induced enzyme activities of P_{PpCAT1} and P_{PpAOX1} were similar, only for CalB expression P_{PpCAT1} outperformed all tested promoters, suggesting a specific beneficial effect. Methanol induced activities from P_{HpFMD} outperformed methanol induced P_{PpAOX1} up to 2.5-fold. However, the effect was stronger for the intracellular expression of *MeHNL* than the secretory expression of HRP and CalB. We assume that for the secretory proteins, not transcription but rather passage through the secretory pathway is the limiting factor. In line with this hypothesis, it has previously been shown that multicopy strains of CalB even show reduced activities compared to single copy if expressed without helper proteins [16]. Similar effects were also noticed for HRP (Krainer F.W. personal communication). Too strong overexpression of HRP and CalB by P_{HpFMD} may overburden the secretion machinery ('secretion saturation' [15]), whereas intracellular expression of MeHNL is well tolerated.

The strong expression from P_{HpFMD} was consistently reproducible using four reporter genes (eGFP, HRP, CalB, MeHNL), therefore we suggest that orthologues promoters from related organism can be valuable tools for protein production even exceeding endogenous promoters.

Discussion

Interestingly, none of the orthologous promoters tested show clear identify to the *P. pastoris* genome when performing a BLAST search (using standard parameters) and also alignments to their *P. pastoris* orthologs do not exhibit clear identities (data not shown). Similar results have been obtained

in in metazoans [9] and yeast [10]. We assume that the expression form the MUT promoters is governed by short, partially degenerative transcription factor binding sites (TFBS), that are apparently also conserved in some heterologous promoters. The *P. pastoris* methanol master regulator Mxr1p [17] binds for example a simple CYCCNY motif and this motif is dispersed over different positons in the *P. pastoris AOX1*, *DAS2* and *PEX8* promoter sequences [18,19].

Even more puzzling is the strong derepressed activity of the *HpFMD* promoter. It has previously been suggested, that regulation about derepression in methylotrophic yeasts is conferred primarily by the host regulatory machinery and not by the promoter sequences [3]. This assumption was taken, as the *P. pastoris AOX1* promoter (tightly repressed in its natural host) did not maintain its tight repression if transferred to *H. polymorpha.* P_{PpAOX1} showed in *Hp* derepression similar to endogenous *H. polymorpha* promoters [3,11,20]. A possible explanation would be that *P. pastoris* contains unique repressors to maintain tight repression under derepressed conditions. It appears that this machinery does not exist in *H. polymorpha* (or at least does not act on the *HpFMD* and *HpMOX* promoters, as these promoters are naturally derepressed). So it is unlikely that the *HpFMD* and *HpMOX* promoters contain binding sites for the *P. pastoris* machinery to maintain tight repression, which would explain their derepressed regulation in *P. pastoris*. Alternatively, the effect may also be explained by an activating model: *H. polymorpha* may contain activators that start expression under derepressed conditions. *P. pastoris* may contain similar derepressed activators, as the *PpCAT1* promoter is also moderately derepressed. The *HpFMD* promoter may contain more TFBS for these activators than P_{PpCAT1}, leading to stronger activation. However, these are just hypotheses and elucidating the exact mechanisms of the strong derepressed expression will require further studies.

There is a notable difference of our work on orthologous MUT promoters from related methylotrophic yeasts compared to a study on the promoters of genes coding for ribosomal proteins in various yeasts in *S. cerevisiae* [10]: The MUT promoters show different modes of regulation (repression, derepression, induction) whereas ribosomal proteins are constitutively expressed [10]. We have recently shown that even in *P. pastoris* 15 methanol inducible promoters of the MUT pathway are considerably differently regulated (regarding tightness of repression and strength) [5]. The strength and simple regulation of MUT promoters and the ease of cultivation and genetic modification of methylotrophic yeasts may suggest them also as promising model systems for studying eukaryotic promoter architecture and evolution (and not solely as prominent tools for heterologous protein production). Systematically studying the same set of MUT promoters in *Hp*, *Cb* and *Pm* and comparing these results to *Pp* and amongst each other may be a first starting step.

Surprisingly, orthologous promoters can be highly useful tools for single protein production, as demonstrated by up to 2.5-fold higher activities achieved form the P_{HpFMD} than P_{PpAOX1}. The orthologous promoters show also highly divergent sequences from *P. pastoris*. This is advantageous if multiple genes should be coexpressed. The repeated use of identical sequences can lead to 'loop out' recombination in yeast [15], leading to loss of copies or parts of expression cassettes [21,22]. To this end, orthologous promoters with similar regulation but divergent sequences may also become valuable tools for metabolic engineering endeavors, requiring the expression of multiple genes from similarly regulated promtoers [5]. To this end the promoters of the entire MUT pathways of *Hp*, *Cb* and *Pm* could be tested in *P. pastoris* to mine for additional strong, tightly methanol regulated promoters.

Acknowledgments

We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) for funding to T.V.

Materials and methods

Cloning of promoters

The orthologous promoters were PCR amplified and cloned upstream of an eGFP repoter gene into a previously established reporter plasmid for *P. pastoris* (pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer, [5]) based on the pPpT4 vector reported by Näätsaari *et al*. [23]. The promoters were cloned seemingly, *i.e.* maintaining the natural sequence context to the start codon without additional restriction endonuclease sites or linker sequences. Primers were designed according to the literature (*Hp FMD* and *MOX* promoters [24,25], *CbAOD1* [26] and *CbFLD1* [27], *Pm MOD1* and *MOD2* [28–30]) and the primer sequences are provided in S 1. Genomic DNA of the strains *Hp* DSM 70277, *Cb* DSM 70026 and *Pm* DSM 2147 was isolated and used as template for the PCR reactions. The PCRs were cloned into the reporter vector by TA cloning as outlined previously [5]. The resulting plasmids were sequenced, showing in part minor differences to previously reported sequences (S 2). The control vectors of the *P. pastoris* endogenous *AOX1*, *CAT1* and *GAP* promoters were available from previous studies [5].

The alternative reporter vectors bearing HRP (isoenzyme A2A [31]), CalB and MeHNL downstream of the respective promoters were in part available from previous studies [5] or generated by cutting out the eGFP reporter gene from the above mentioned vectors (via *Nhe*I and *Not*I restriction sites) and seamlessly inserting PCR products of the GOIs by Gibson assembly [32]. See again S 1 for the primer sequences. The HRP and CalB vectors previously reported [5] were used as PCR templates, the MeHNL sequence was codon optimized for *P. pastoris* and ordered with overhangs to the *AOX1* promoter and terminator for Gibson cloning (S 1). This vector was sequenced and used as template for amplification. Since the HRP and CalB genes were both fused to a mating factor alpha signal sequence, the same forward primer could be used for amplification (pHpFMD-MFalpha-Gib). The inserted genes were sequenced with primers binding to the *AOX1* terminator and the respective promoters, for P_{HpFMD} a Sanger sequencing primer (seq-pHpHMD-149..126fwd) was designed to allow sequencing of the downstream gene.

Strains, materials, fluorescence measurements and enzyme assays

Materials and strains were used as previously reported in detail [5]. Fluorescence measurements, HRP and CalB activity assays were also performed as previously reported [5]. For transformations of all basic promoter comparisons, the *P. pastoris* CBS7435 wildtype strain was used. HRP and CalB were transformed into a mutS strain, as higher yields have been reported [33] and the control plasmids bearing these genes of interest under the control of *P. pastoris* endogenous promoters had also been transformed into the muts strain [5]. MeHNL activity was measured as described in [12].

Figures

Fig. 1

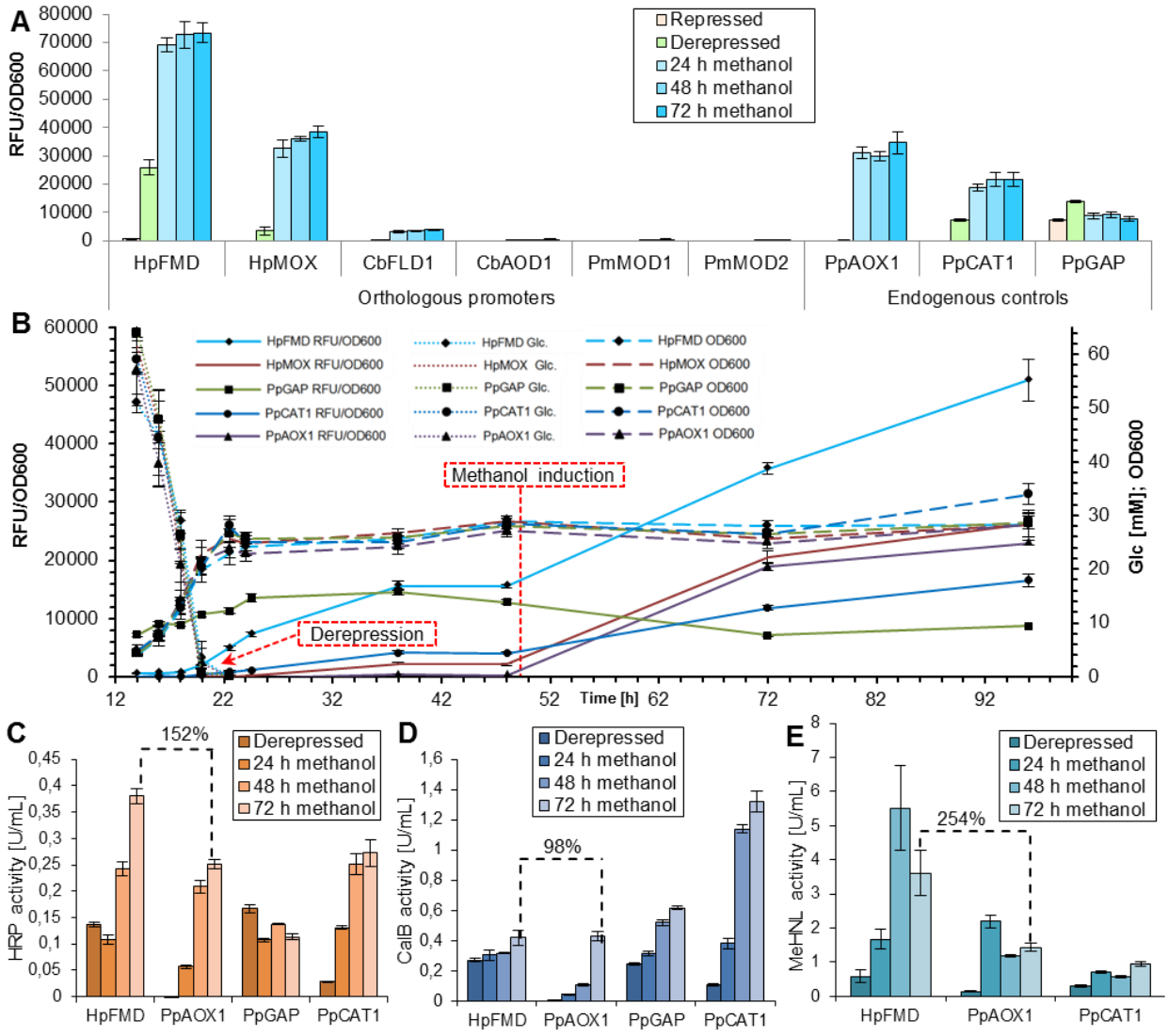


Fig. 1: The *HpFMD* promoter is the strongest orthologous promoter tested, even outperforming *P. pastoris* endogenous promoters.

A) Reporter protein fluorescence of all orthologous and endogenous promoters tested. The orthologous MUT promoters of different methylotrophic yeasts were cloned upstream of an enhanced green fluorescent protein (eGFP) and transformed into *P. pastoris*. The strains were cultivated in deep well plate (DWP) cultivation on BMD1 media and subsequently induced with methanol. Reporter protein fluorescence and OD_{600} were measured under glucose repressed (16 h) and derepressed (60h) conditions and different time points of methanol induction. Fluorescence measurements were normalized per OD_{600}. Mean values (MVs) and standard deviations (SDs) of biological quadruplicates are shown.

B) *HpFMD* is slightly constitutive active and shows strong derepression. Selected strains form panel A were cultivated in shake flasks and inoculated to a low starting OD_{600}. Reporter protein fluorescence, OD_{600} and glucose levels were measured. Fluorescence/OD_{600} values at t=0 are not shown, as the starting OD_{600} of 0.05 was outside the linear range of the spectrometer used. The initial glucose concentration of the media was 55.5 mM (10 g/l). MVs and SDs of biological triplicates are shown.

C-E) Similarly high yields as with eGFP can also be achieved with the enzymes HRP (C), CalB (D) and MeHNL (E). The strains were grown in DWPs on BMD1 media until glucose depletion for 60 h and were subsequently induced with methanol. HRP and CalB activities in the supernatants were measured and cells lysed to measure intracellular MeHNL activity. Mean values (MVs) and standard deviation (SDs) of biological quadruplicates are shown.

References

[1] Gellissen G. Heterologous protein production in methylotrophic yeasts. Appl Microbiol Biotechnol 2000;54:741–50.
[2] Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. Int J Microbiol 2011;2011:101298.
[3] Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact 2006;5:39–59.
[4] Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Front Microbiol 2014;5:85.
[5] Chapter 3.1 of this thesis (Vogl *et al.* manuscript in preparation)
[6] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.
[7] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.
[8] Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, et al. Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris. Microb Cell Fact 2013;12:5.
[9] Weirauch MT, Hughes TR. Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. Trends Genet 2010;26:66–74.
[10] Zeevi D, Lubliner S, Lotan-Pompan M, Hodis E, Vesterman R, Weinberger A, et al. Molecular dissection of the genetic mechanisms that underlie expression conservation in orthologous yeast ribosomal promoters. Genome Res 2014.
[11] Raschke WC, Neiditch BR, Hendricks M, Cregg JM. Inducible expression of a heterologous protein in Hansenula polymorpha using the alcohol oxidase 1 promoter of Pichia pastoris. Gene 1996;177:163–7.
[12] Chapter 3.2 of this thesis (Vogl *et al.* manuscript in preparation)
[13] Cregg JM, Barringer KJ, Hessler AY, Madden KR. Pichia pastoris as a host system for transformations. Mol Cell Biol 1985;5:3376–85.
[14] Liachko I, Youngblood R a, Tsui K, Bubb KL, Queitsch C, Raghuraman MK, et al. GC-Rich DNA Elements Enable Replication Origin Activity in the Methylotrophic Yeast Pichia pastoris. PLoS Genet 2014;10:e1004169.
[15] Aw R, Polizzi KM. Can too many copies spoil the broth? Microb Cell Fact 2013;12:128.
[16] Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS, Glieder A. Real-time PCR-based determination of gene copy numbers in Pichia pastoris. Biotechnol J 2010;5:413–20.
[17] Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, et al. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol Cell Biol 2006;26:883–97.
[18] Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN. Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. Biochim Biophys Acta 2009;1789:460–8.
[19] Kranthi BV, Kumar HRV, Rangarajan PN. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris. Yeast 2010;27:705–11.
[20] Rodriguez L, Narciandi RE, Roca H, Cremata J, Montesinos R, Rodriguez E, et al. Invertase secretion in Hansenula polymorpha under the AOX1 promoter from Pichia pastoris. Yeast 1996;12:815–22.
[21] Zhu T, Guo M, Sun C, Qian J, Zhuang Y, Chu J, et al. A systematical investigation on the genetic stability of multi-copy Pichia pastoris strains. Biotechnol Lett 2009;31:679–84.
[22] Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in P. pastoris. Chem Commun (Camb) 2014.
[23] Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 2012;7:e39720.
[24] Song H, Li Y, Fang W, Geng Y, Wang X, Wang M, et al. Development of a set of expression vectors in Hansenula polymorpha. Biotechnol Lett 2003;25:1999–2006.
[25] Ledeboer AM, Edens L, Maat J, Visser C, Bos JW, Verrips CT, et al. Molecular cloning and characterization of a gene coding for methanol oxidase in Hansenula polymorpha. Nucleic Acids Res 1985;13:3063–82.
[26] Yurimoto H, Komeda T, Lim CR, Nakagawa T, Kondo K, Kato N, et al. Regulation and evaluation of five methanol-inducible promoters in the methylotrophic yeast Candida boidinii. Biochim Biophys Acta 2000;1493:56–63.
[27] Lee B, Yurimoto H, Sakai Y, Kato N. Physiological role of the glutathione-dependent formaldehyde dehydrogenase in the methylotrophic yeast Candida boidinii. Microbiology 2002;148:2697–704.
[28] Raymond CK, Bukowski T, Holderman SD, Ching a F, Vanaja E, Stamm MR. Development of the methylotrophic yeast Pichia methanolica for the expression of the 65 kilodalton isoform of human glutamate decarboxylase. Yeast 1998;14:11–23.
[29] Nakagawa T, Sakai Y, Mukaiyama H, Mizumura T, Miyaji T, Yurimoto H, et al. Analysis of alcohol oxidase isozymes in gene-disrupted strains of methylotrophic yeast Pichia methanolica. J Biosci Bioeng 2001;91:225–7.
[30] Nakagawa T, Inagaki A, Ito T, Fujimura S, Miyaji T, Yurimoto H, et al. Regulation of two distinct alcohol oxidase promoters in the methylotrophic yeast Pichia methanolica. Yeast 2006;23:15–22.
[31] Näätsaari L, Krainer FW, Schubert M, Glieder A, Thallinger GG. Peroxidase gene discovery from the horseradish transcriptome. BMC Genomics 2014;15:227.
[32] Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.
[33] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 2012;11:22.
[34] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6.
[35] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.

Supplementary materials

S 1 (Primer sequences and synthetic *MeHNL* gene)

S 1: Primer sequences and synthetic MeHNL gene

Primers for amplification of the orthologous promoters:

HpFMDfwd	AATGTATCTAAACGCAAACTCCGAGCTG	LENGTH:28GC CONTENT:42.9 %MELT TEMP:59.0 ºC
HpFMDrev	GATTTGATTGATGAAGGCAGAGAGCGCAAG	LENGTH:30GC CONTENT:46.7 %MELT TEMP:61.3 ºC
HpMOXfwd	TCGACGCGGAGAACGATCTCCTCGAGCT	LENGTH:28GC CONTENT:60.7 %MELT TEMP:67.1 ºC
HpMOXrev	TTTGTTTTTGTACTTTAGATTGATGTCACCACCGTGCACTGGCAG	LENGTH:45GC CONTENT:42.2 %MELT TEMP:66.0 ºC
PmMOD1fwd	CGAGATGGTACATACTTAAAAGCTGCCATATTGAG	LENGTH:35GC CONTENT:40.0 %MELT TEMP:59.9 ºC
PmMOD1rev	TTTGAGAAATTAATAGTAAGATTTTTTTTTCGTAAAAGTTTTGATTGAGTTAATTC	LENGTH:56GC CONTENT:19.6 %MELT TEMP:59.9 ºC
PmMOD2fwd	GGATCCACTACAGTTTACCAATTGATTACGCCAATAG	LENGTH:37GC CONTENT:40.5 %MELT TEMP:61.1 ºC
PmMOD2rev	TTTGAATTTTAGTTTTAGATAGATAAATATAATTTTCAATCCTGTTATAAAATAGTATAT	LENGTH:60GC CONTENT:15.0 %MELT TEMP:58.4 ºC
CbAOD1fwd	GGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGAAAG	LENGTH:55GC CONTENT:18.2 %MELT TEMP:56.2 ºC
CbAOD1rev	TATTGAAAAATAATTTTGTTTTTTTTTTTTTGTTTTTTTAAAAGTTCGTTAAAATTCG	LENGTH:58GC CONTENT:13.8 %MELT TEMP:59.8 ºC
CbFLD1fwd	GGATCCCTTCAACAGCGGAGTCTCAAAC	LENGTH:28GC CONTENT:53.6 %MELT TEMP:62.5 ºC
CbFLD1rev	TTTTGTGGAATAAAAAATAGATAAATATGATTTAGTGTAGTTGATTCAATCAATTGAC	LENGTH:58GC CONTENT:22.4 %MELT TEMP:60.9 ºC

Primers for cloning the promoters upstream of different reporter genes:

pHpFMD-MFalpha-Gib	cttgcgctctctgccttcatcaatcaaatcATGAGATTCCCATCTATTTTCACCGCTGTC
AOX1TT-NotI-CalB	caaatggcattctgacatcctcttgaGCGGCCGCttatggggtcacgataccggaacaag
AOX1TT-NotI-HRPA2A	caaatggcattctgacatcctcttgaGCGGCCGCttaggatccgttaactttcttgcaatcaagtc
seq-pHpHMD-149..126fwd	actggtgtccgccaataagaggag
pHpFMD-MeHNL	cttgcgctctctgccttcatcaatcaaatcATGGTTACTGCTCACTTCGTCTTGATTCAC
AOX1TT-NotI-MeHNL	caaatggcattctgacatcctcttgaGCGGCCGCTTAAGCGTAAGCGTCGGCAACTTCCTG
pCAT1-MeHNL-Gib	cacttgctctagtcaagacttacaattaaaATGGTTACTGCTCACTTCGTCTTGATTCAC

Codon optimized MeHNL gene:

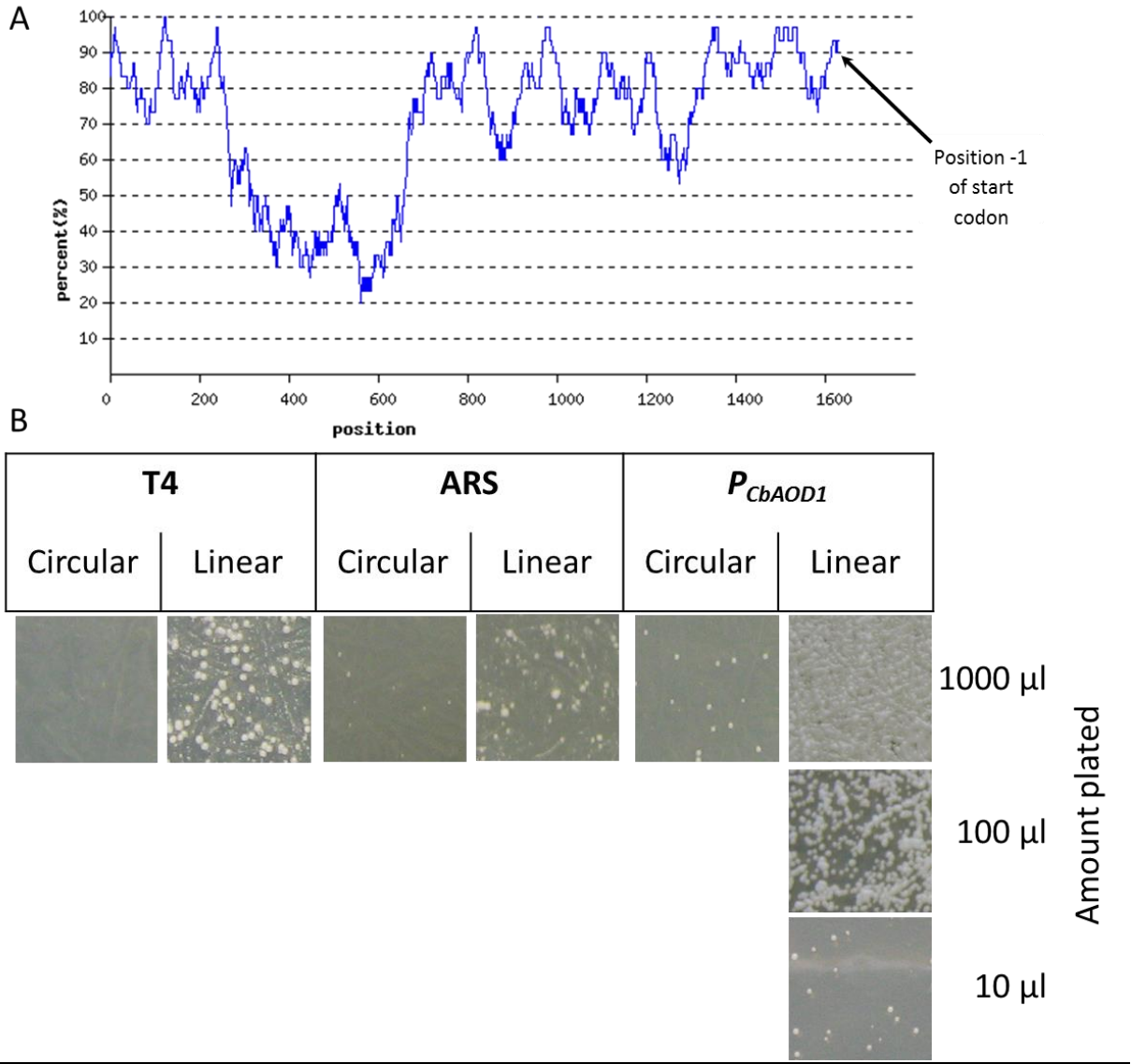
The CDS is written in upper case, start and stop codons written in bold, overhangs for Gibson assembly [32] to the entry vector are written in lower case, *EcoR*I and *Not*I restriction sites typically used for cloning in the pPpT4 vector family are underlined.

cgacaacttgagaagatcaaaaaacaactaattattgaaagaattccgaaacg**ATG**GTTACTGCTCACTTCGTCTTGATTCACACT
ATCTGTCATGGTGCTTGGATCTGGCACAAGTTGAAGCCAGCATTGGAGAGAGCTGGACATAAGGTTACCGCTC
TTGATATGGCTGCATCTGGTATTGATCCTCGTCAAATCGAACAAATCAATTCATTCGACGAGTACTCAGAGCCA
CTGCTGACCTTCTTGGAAAAGTTGCCTCAAGGTGAAAAGGTGATCATCGTTGGTGAATCCTGTGCTGGATTGA
ACATTGCCATTGCAGCTGATAGATATGTCGATAAGATCGCTGCTGGTGTCTTCCACAACTCTCTGTTACCAGAT
ACTGTTCACTCTCCATCTTACACTGTCGAGAAGTTGTTAGAATCATTCCCAGATTGGAGAGATACTGAATACTTT
ACTTTCACTAACATCACTGGAGAGACTATCACCACCATGAAACTTGGATTCGTTTTGTTGAGAGAAAACCTTTTC
ACCAAGTGTACTGATGGTGAATACGAATTGGCCAAGATGGTTATGAGAAAGGGTTCTTTGTTTCAGAATGTTC
TTGCACAAAGACCAAAGTTCACCGAAAAGGGTTACGGTTCTATCAAGAAGGTCTACATCTGGACTGATCAGGA
CAAGATCTTCCTGCCAGACTTCCAAAGATGGCAAATCGCAAACTACAAACCAGATAAGGTCTACCAAGTCCAA
GGTGGTGATCACAAGTTACAATTGACCAAGACCGAAGAGGTCGCTCACATCTTGCAGGAAGTTGCCGACGCTT
ACGCT**TAA**gcggccgctcaagaggatgtcagaatgccatttgcctg

S 2 (Alignments of promoter sequences to references)

S 2: The sequences of the orthologous promoters obtained in this study show minor differences to sequences reported in literature. Promoter sequences reported in the literature (denoted as 'ref', see materials and methods section) are compared to sequences obtained by PCR amplification and cloning in this study (denoted by an asterisk '*'). The lengths reported in the literature up to the start codon are shown. See separate file.

S 3 (ARS behavior of *CbAOD1* promoter)



S 3: P_{CbAOD1} containing plasmids result surprisingly in approximately 100-times higher transformation efficiency in linearized form and allow circular transformation of plasmids, suggesting the presence of an ARS.

A) AT content of the *CbAOD1* promoter. The average AT content of the *P. pastoris* genome is 58.9 % [34,35]

B) Transformation of circular and linearized forms of the empty vector (T4 = pPpT4_S [23]), a control vector containing the *P. pastoris ARS1* [13] and the *CbAOD1* promoter. Photos of the transformation plates after 48 h of incubation are shown. For linearized plasmids amounts equivalent to 1.0 µg of the empty T4 vector [23] were used and for circular plasmids 10 ng of plasmid was were transformed. Different volumes of transformed and regenerated cultures were plated (10 µL, 100 µL, 1000 µL).

ACS SyntheticBiology
Technical Note
pubs.acs.org/synthbio

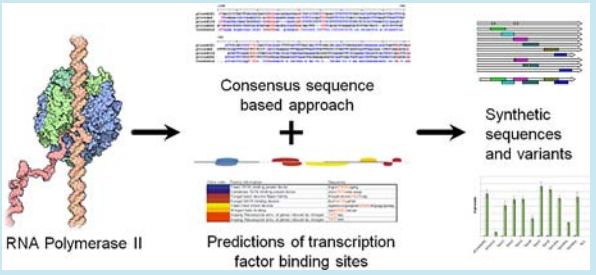
Synthetic Core Promoters for *Pichia pastoris*

Thomas Vogl,[†,#] Claudia Ruth,[†,‡,#] Julia Pitzer,[†] Thomas Kickenweiz,[†] and Anton Glieder*[,‡]

[†]Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz A-8010, Austria
[‡]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, Graz A-8010, Austria

Supporting Information

ABSTRACT: Synthetic promoters are commonly used tools for circuit design or high level protein production. Promoter engineering efforts in yeasts, such as *Saccharomyces cerevisiae* and *Pichia pastoris* have mostly been focused on altering upstream regulatory sequences such as transcription factor binding sites. In higher eukaryotes synthetic core promoters, directly needed for transcription initiation by RNA Polymerase II, have been successfully designed. Here we report the first synthetic yeast core promoter for *P. pastoris*, based on natural yeast core promoters. Furthermore we used this synthetic core promoter sequence to engineer the core promoter of the natural *AOX1* promoter, thereby creating a set of core promoters providing a range of different expression levels. As opposed to engineering strategies of the significantly longer entire promoter, such short core promoters can directly be added on a PCR primer facilitating library generation and are sufficient to obtain variable expression yields.



KEYWORDS: *synthetic promoters, core promoter construction, promoter library, transcriptional fine-tuning, Pichia pastoris*

Engineered promoters are commonly applied DNA parts for circuit design and used to increase titers of heterologous proteins and metabolites. In eukaryotes, where promoters are longer and more complex than in prokaryotes, the core or minimal promoter is the crucial region providing binding sites for general transcription factors and RNA polymerase II. In addition, transcript stability and efficient translation initiation depend on the 5′ untranslated region (UTR). In higher eukaryotes the design of synthetic core promoters has led to strongly improved synthetic variants.[1] In lower eukaryotes such as yeasts and fungi, promoter engineering has mostly concentrated on upstream regulatory sequences (URS),[2,3] their combination with natural core promoters[4,5] and/or random mutagenesis approaches of selected core promoter regions.[5,6] In the model yeast *Saccharomyces cerevisiae*, large scale studies have been used to analyze the sequence/function relationship of natural core promoters.[7]

Pichia pastoris is a commonly used expression system for the production of biocatalysts and biopharmaceuticals, as it allows secretion of pure heterologous proteins to the culture broth and grows to high cell densities.[8] The most commonly used, exceptionally strong, and tightly regulated methanol inducible promoter of the *alcohol oxidase 1* gene (p*AOX1*) has been studied in terms of regulatory sequences and factors.[9] It was also used for the design of synthetic promoter variants providing a range of expression levels and altered regulatory profiles. These engineering efforts have focused on URS,[2,3] random mutagenesis of core promoter sections[6] or the 5′ UTR.[10] Here we aimed for the first time to design a fully synthetic core promoter and 5′ UTR for *P. pastoris* and to employ such artificial sequences for p*AOX1* core promoter engineering to obtain a library of sequence-diversified promoters with different properties.

CONSTRUCTION OF SYNTHETIC CORE PROMOTER PCORE11

While promoters in bacteria can be rationally engineered considering conserved regions and spacing (e.g., conserved −35, −10 regions),[11] yeast core promoters have only been engineered by random mutagenesis methods.[5,6] We designed a synthetic yeast core promoter by using a consensus sequence of natural core promoters that was refined by incorporating common transcription factor binding site (TFBS) motifs (Figure 1A). The core promoter sequences of four natural, differently regulated promoters (p*AOX1*, p*GAP*, p*HIS4*, and p*ScADH2*) were aligned using MultAlin[12] to identify a general minimal consensus (Figure 1B). Detailed information on the promoter choices and exact sequence selection is provided in Supplementary Figures S 1, Supporting Information. This first core promoter sequence (pCore1) was successively re-engineered on the basis of an *in silico* analysis for putative TFBSs (Figure 1C, Supplementary Figures S 1, and Supporting File 1, Supporting Information). We aimed to identify common sequence motifs of the natural promoters and to integrate them into pCore1. Therefore the natural promoters used for the consensus design were analyzed for putative TFBSs using MatInspector.[13] TFBSs predicted in several natural promoters were incorporated into the pCore1 sequence, while superfluous

Special Issue: SB6.0
Received: July 28, 2013

   dx.doi.org/10.1021/sb400091p | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

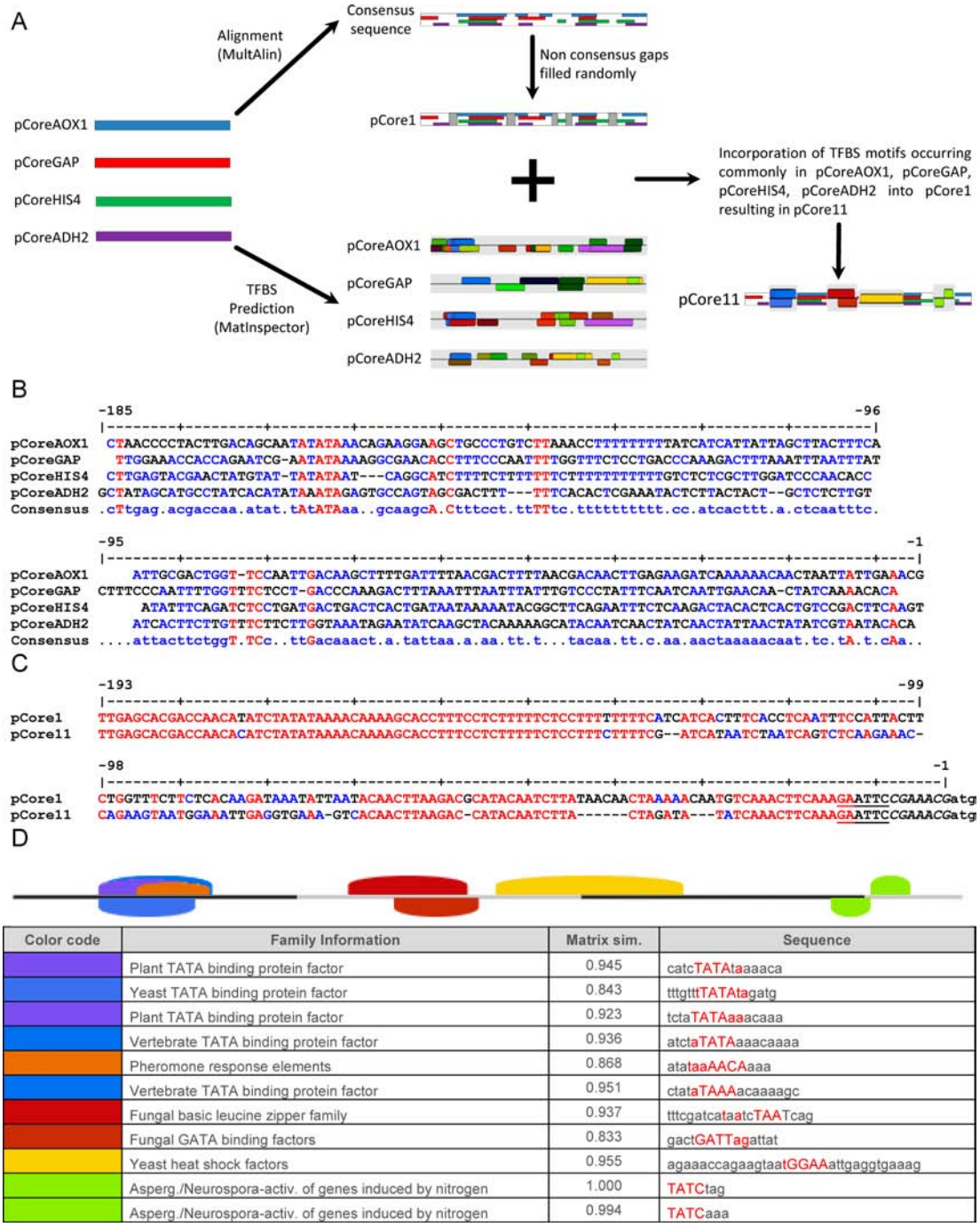


Figure 1. Design steps of the synthetic core promoter pCore11. (A) Schematic diagram of the design strategy. Four natural yeast core promoters were aligned resulting in a consensus sequence/pCore1. TFBSs predicted in at least two out of the four core promoters were incorporated into pCore1 resulting in pCore11 (Supplementary Figures S 1 and Supporting File 1, Supporting Information). Sequence elements and positions are not drawn to scale. (B) Sequence alignment of the core promoters used for the generation of a consensus sequence. Detailed information on the sequence selection is provided in Supplementary Figures S 1, Supporting Information. (C) The consensus sequence of the alignment was the basis for pCore1 generation. The four natural core promoters and pCore1 were analyzed for putative TFBSs using MatInspector (Supporting File 1, Supporting Information); common TFBSs were incorporated as described in Supplementary Figures S 1, Supporting Information, resulting in pCore11. The synthetic promoters fused to the *Eco*RI site (underlined), Kozak sequence (in italics) and the start codon (lower case) are shown (see also Supplementary Figures S 2, Supporting Information). (D) TFBSs prediction of the engineered pCore11 (for detailed information and analysis of the other core promoters, see Supplementary Figures S 1 and Supporting Figure 1, Supporting Information).

sites of pCore1 were removed, resulting in pCore11 (Figure 1C,D). Detailed information on the putative TFBSs found and the engineering thereof are provided in Supporting File 1 and Supplementary Figures S 1, Supporting Information. In *P. pastoris*, these predicted TFBSs are not necessarily bound by homologues of the factors identified by the MatInspector analysis, as these factors bind usually further upstream to URSs (except TATA binding protein, a crucial core promoter factor). As the MatInspector matrix-based model of TFBSs tolerates also variations in binding motifs, the motifs found may also

dx.doi.org/10.1021/sb400091p | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

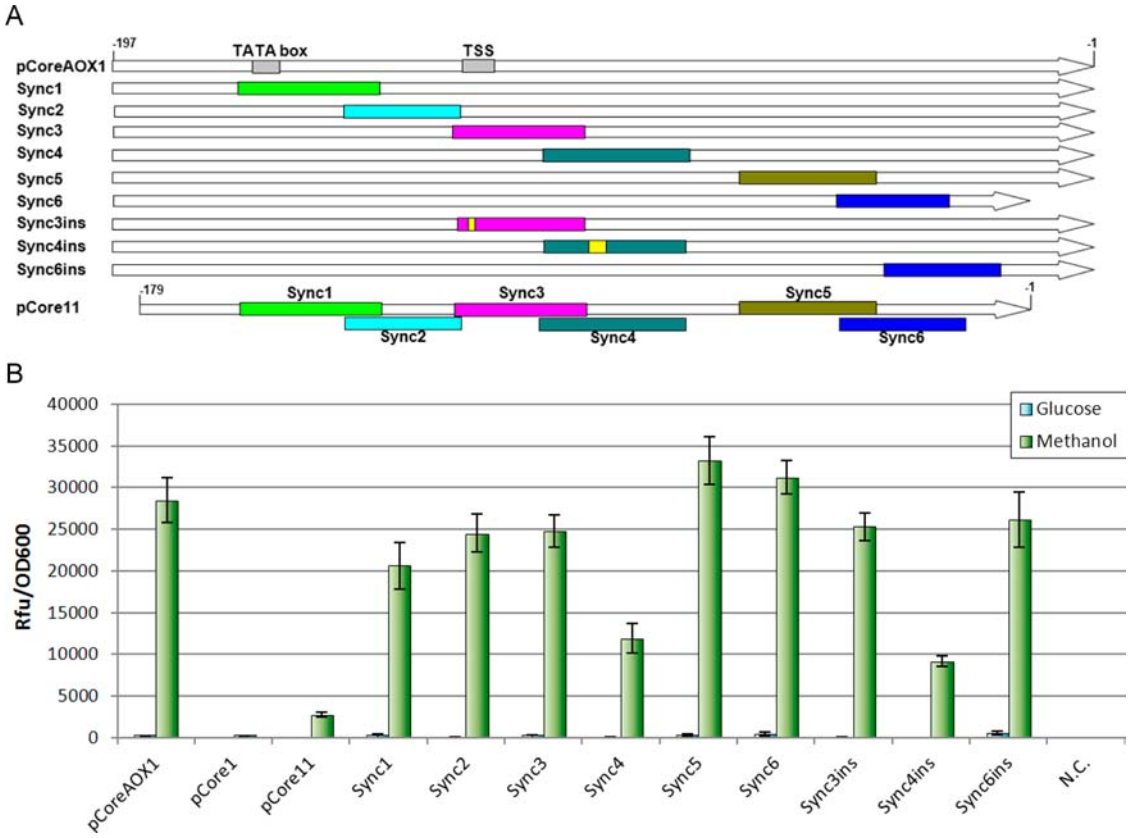


Figure 2. Engineered p*AOX1* core promoter variants exhibit a range of expression levels. (A) Design schematic of the variants. Regions from pCore11 were incorporated into pCore*AOX1*. Insertions for length corrections are indicated in yellow. Detailed information on the selection process and sequences is provided in Supplementary Figures S 2, Supporting Information 2. Numbering is relative to the translational start (+1). (B) Expression levels of the variants. The variants Sync1 to Sync6ins and controls (wildtype pCore*AOX1*, pCore1, pCore11) were cloned upstream of a GFP reporter (N.C.: negative control of untransformed wildtype strain). The strains were grown on glucose and induced with methanol, and fluorescence was measured as outlined in the methods section in the Supporting Information.

constitute degenerate determinants for the binding of general transcription factors or RNA polymerase II.

As core promoters alone provide either no or only low basal transcription,[14] we fused pCore1 and pCore11 to the upstream region of p*AOX1*. These fusions were subsequently assayed with a GFP reporter, established for yeast promoter studies.[2,3] While pCore1 fluorescence only marginally surpassed background fluorescence, the re-engineered pCore11 showed tight repression on glucose and upon methanol induction reached about 10% of the wildtype pCore*AOX1* (Figure 2B). These results show that functional synthetic yeast core promoters can in principle be obtained by complementation of a core sequence elements with additional nucleotides. This approach is similar to prokaryotic promoter engineering,[11] although conserved, tightly localized sequence motifs such as the −10 and −35 region of prokaryotic promoters are not obvious (except the TATA box).[14]

IDENTIFICATION OF VARIABLE SEQUENCE STRETCHES OF PCORE*AOX1*

Previous studies of the p*AOX1* URS focused either on systematic deletions,[3] putative TFBSs,[2] or the identification of TFBSs by DNA/protein interaction studies (see Supplementary Note, Supporting Information). However, while URS can be positioned variably over the promoter sequence (e.g., TFBSs of the methanol master regulator Mxr1 in p*AOX1* and p*DAS2*, reviewed in the literature[9]), core promoter recognition by RNA Polymerase II requires tight interaction with the DNA over more than 60 bp,[14] and deletions in the core promoter may interfere with the spacing of regulatory sequences. Therefore we replaced putative regulatory sequences in pCore*AOX1* with elements of the functional synthetic pCore11. As opposed to randomized sequences, such sequence transfer promised functional engineered *AOX1* core promoters variants. Since pCore11 is shorter than pCore*AOX1* also insertion variants have been made. The exact sequence selections for the synthetic core promoters (Sync1 to Sync6ins) are illustrated in Figure 2A and described in detail in Supplementary Figures S 2, Supporting Information. As the natural pCore*AOX1* sequence, all synthetic core promoters linked to the upstream region of p*AOX1* were repressed on glucose (Figure 2B). GFP fluorescence upon methanol induction of the synthetic variants, including pCore11, ranged from 10 to 117% percent of the wildtype promoter. Several variants showed only moderate changes in expression levels, even though key regions close to the TSS were changed (Sync2, Sync3, Sync3ins). Sync4 and Sync4ins show the strongest decrease of expression (30−40% of the wildtype promoter activity), suggesting an important region downstream the TSS in the beginning of the 5′UTR. In general the length correction by insertions did not cause significant differences.

Our results show that the construction of diversified synthetic core promoters and 5′ UTRs is also possible in yeast. A first generation synthetic core promoter sequence

dx.doi.org/10.1021/sb400091p | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

showing at least some functionality can be used to recruit suitable sequence stretches for the design and construction of a next generation library of fully functional core promoters with varying sequences and strength. By testing a moderate number of variants, we achieved a range of expression levels, and these variants can be applied to fine-tune gene expression. Compared to engineering by random mutagenesis (e.g., error prone PCR), much fewer constructs need to be tested. However, screening a higher number of randomly mutated sequences of pCore*AOX1* recently also led to considerably improved variants,[6] while our more rational approach resulted only in moderately improved variants so far (Sync5 and Sync6). Both kinds of core promoters might be combined with URS variants of other studies[2,3] to achieve synergistic effects. It remains to be shown whether the observed effects are a result of changed transcription or mRNA stability, or if the translation initiation was influenced. However, the replacement of parts of natural core promoter sequences also provides a tool for sequence diversification while maintaining natural expression levels and regulation. Engineering of URS may also interfere with regulation of the promoter.[2] For our synthetic core promoter/p*AOX1* URS fusions, the mode of regulation remained untouched. Therefore a similar strategy as in prokaryotes, where ribosome binding sites are modified to fine-tune strong natural promoters,[15] is feasible by engineering eukaryotic core promoters and UTRs.

■ ASSOCIATED CONTENT

Supporting Information

Complete methods, Table S1, supplementary figures S 1−S 3, supplementary note, supporting references, and supporting files 1 and 2. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +433168739300. E-mail: anton.glieder@acib.at.

Author Contributions

#T. Vogl and C. Ruth contributed equally to this work.

Conceived the concept of the study and sequence design: C.R., J.P., A.G. Planned and performed the experiments: T.V., T.K., C.R. Analyzed and interpreted the data: T.V., T.K., C.R. Wrote the paper: T.V., A.G.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the Austrian Science Fund (FWF) Project No. W901 (DK "Molecular Enzymology" Graz), and the Austrian Centre of Industrial Biotechnology (ACIB) contribution was supported by FFG, bmvit, mvwfi, ZIT, Zukunftsstiftung Tirol and Land Steiermark within the Austrian COMET programme (FFG Grant 824186). The image in the graphical abstract showing RNA Polymerase II was kindly provided by the RCSB PDB (taken from the Molecule of the Month feature by David S. Goodsell, doi: 10.2210/rcsb_pdb/mom_2003_4). We would like to thank Michaela Agnes Gerstmann for assistance with qPCR and Lukas Sturmberger for fruitful discussions and critical reading of the manuscript.

■ REFERENCES

(1) Juven-Gershon, T., Cheng, S., and Kadonaga, J. T. (2006) Rational design of a super core promoter that enhances gene expression. *Nat. Methods 3*, 917−922.

(2) Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., Lin-Cereghino, J., Kovar, K., Cregg, J. M., and Glieder, A. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res. 36*, e76.

(3) Xuan, Y., Zhou, X., Zhang, W., Zhang, X., Song, Z., and Zhang, Y. (2009) An upstream activation sequence controls the expression of AOX1 gene in *Pichia pastoris*. *FEMS Yeast Res. 9*, 1271−1282.

(4) Blazeck, J., Garg, R., Reed, B., and Alper, H. S. (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol. Bioeng. 109*, 2884−2895.

(5) Blount, B. A., Weenink, T., Vasylechko, S., and Ellis, T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One 7*, e33279.

(6) Berg, L., Strand, T. A., Valla, S., and Brautaset, T. (2013) Combinatorial mutagenesis and selection to understand and improve yeast promoters. *Biomed. Res. Int. 2013*, 926985.

(7) Lubliner, S., Keren, L., and Segal, E. (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res. 41*, 5569−5581.

(8) Vogl, T., Hartner, F. S., and Glieder, A. (2013) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. *Curr. Opin. Biotechnol.*, DOI: 10.1016/j.copbio.2013.02.024.

(9) Vogl, T., and Glieder, A. (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. *New Biotechnol. 30*, 385−404.

(10) Staley, C. A., Huang, A., Nattestad, M., Oshiro, K. T., Ray, L. E., Mulye, T., Li, Z. H., Le, T., Stephens, J. J., Gomez, S. R., Moy, A. D., Nguyen, J. C., Franz, A. H., Lin-Cereghino, J., and Lin-Cereghino, G. P. (2012) Analysis of the 5′ untranslated region (5′UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in *Pichia pastoris*. *Gene 496*, 118−127.

(11) Blazeck, J., and Alper, H. S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol. J. 8*, 46−58.

(12) Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res. 16*, 10881−10890.

(13) Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res. 23*, 4878−4884.

(14) Hahn, S., and Young, E. T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics 189*, 705−736.

(15) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946−950.

Chapter 3.4.2
A library of bidirectional promoters facilitates fine-tuning of gene coexpression

General information

Title

A library of bidirectional promoters facilitates fine-tuning of gene coexpression

Authors

Vogl T, Kickenweiz T, Pitzer J, Sturmberger L, Köhler E-M, Weninger A, Baumschlager A, Fischer JE, Hyden P, Geier M, Glieder A#

#Corresponding author

Abstract

The production of multimeric proteins or the expression of biocatalytic redox systems and enzyme cascades and metabolic pathways requires coexpression of several genes. Even the most simple heterologous expression experiments need the coexpression of a selection marker as a second protein. Each set of genes to be coexpressed needs optimization of cumulative expression levels, ratios and regulation. Optimizations using monodirectional promoters (MDPs) are complicated by plasmid size, cloning obstacles and a limited number of natural strong, tightly regulated promoters available.
We have generated a library of bidirectional promoters (BDPs) allowing fast screening of diverse expression profiles and ratios to optimize gene coexpression. Natural bidirectional histone promoters conserved between eukaryotes were used as a universal engineering framework applied in yeast. Synthetic BDPs were designed by promoter fusions, bidirectionalization of MDPs and assembly of two core promoters flanking *cis*-regulatory modules (CRMs) into bidirectional hybrid promoters. These libraries of short, tightly regulated sBDPs facilitate optimizing dual gene coexpression and ease the assembly of multi gene expression cassettes such as metabolic pathways.

Introduction

Coexpression of multiple genes is a common challenge in molecular biology and metabolic engineering to produce pharmaceuticals, fine chemicals or biofuels. Heterologous expression of an entire pathway or production of multimeric proteins requires coexpression of two or more genes. Each set of genes to be coexpressed needs optimization of cumulative expression levels, ratios and regulation: The cumulative expression especially of heterologous genes should not overburden the cellular capacities. Gene pairs such as enzymes and their redox partner (e.g. cytochrome P450 enzymes (CYP) and the associated reductase (CPR)) or any gene of interest (GOI) and a folding helper need to be coexpressed at a suitable ratio. Also in metabolic engineering the ratio of the genes coexpressed is critical for balancing the pathway flux [1]. In addition it may be necessary to time the coexpression of the GOIs: Production of metabolites or proteins that are difficult to fold may exert a

constant metabolic burden on the cell. In such cases it is desirable to control expression using regulated (inducible) promoters and separate cell growth form production.
The optimal cumulative expression level, ratio and regulatory profile are difficult to predict and typically optimized by screening different promoter combinations [2,3]. Mostly monodirectional promoters (MDPs), conferring expression in one direction, are used for fine-tuning gene coexpression [2–5]. However bidirectional promoters (BDPs), providing divergent expression in opposite directions, were shown to simplify coordinated gene coexpression [6–13].
We envisioned the application of a large set of BDPs for optimizing gene coexpression (Fig. 1a). Such a library of BDPs providing combinations of different expression levels and regulatory profiles on both sides can be cloned randomly between two genes or used for the assembly of multi-gene expression cassettes (S 1). Subsequent screening of the constructs yields the optimal coexpression strategy for the set of GOIs. However, this requires the availability of a library of diverse BDPs that could be selected from natural and synthetic BDPs (nBDPs and sBDPs).

Divergent transcription and nBDPs have been characterized in all model organisms [14–20]. RNAseq studies showed that eukaryotic promoters are intrinsically bidirectional and a source of low-level cryptic antisense transcripts [15,16]. nBDPs with non-cryptic expression in both orientations frequently coregulate functionally related genes [21,22] and are in part evolutionary conserved [23]. Yet only few nBDPs such as the *Saccharomyces cerevisiae GAL1-GAL10* promoter have been used to drive heterologous gene coexpression [24–26]. A few synthetic BDPs (sBDPs) have been created for *Escherichia coli* [13], *S. cerevisiae* [26–30], plants [6,7] and mammals [8–12] offering small sets of combinable regulatory profiles and expression levels. *S. cerevisiae* provides with less than 10 BDPs characterized to our knowledge the largest set of BDPs and thereby a rudimental resource for bidirectional gene coexpression. BDP engineering was so far focusing on single case studies and no generalizable strategies have been demonstrated.

Here, we have generated a library of 168 BDPs covering a 79-fold range of expression levels, variable ratios (equal expression up to 61-fold difference) and combinations of regulatory profiles for the methylotrophic yeast *Pichia pastoris*. Orthologs of the natural bidirectional histone promoters applied as engineering chassis in *P. pastoris* gave also high expression in *S. cerevisiae* and *Schizosaccharomyces pombe* (personal communication Pitzer J.) cells suggesting their suitability as general eukaryotic BDP engineering framework.

Results and discussion

nBDPs of *P. pastoris* show limited cumulative expression levels, ratios and combinations of regulatory profiles

A suitable library of BDPs for coexpression optimization must provide different cumulative expression levels, ratios and regulatory profiles. Therefore either nBDPs can be used or sBDPs can be generated. Assuming a random distribution of gene organization, one third of gene pairs should be arranged in head to head orientation forming BDPs. We adapted our BDP expression strategy (S 1) for vectors in the methylotrophic yeast *Pichia pastoris* and mined its genome for putative nBDPs (Fig. 1b). *P. pastoris* is even more frequently applied for heterologous protein production than *S. cerevisiae* [31] due to its growth to high cell densities (>150 g dry cell weight per liter [32]) and high

yields of secreted proteins [33]. Glycoengineered *P. pastoris* strains [33] are the only yeast species to allow the production of complex terminally sialylated glycoproteins [34] and humanized monoclonal antibodies [35]. *P. pastoris* offers also one of the largest sets of tightly coregulated promoters. Promoters of genes coding for enzymes involved in the methanol utilization (MUT) pathway are completely repressed by glucose and strongly induced by methanol [36]. Using a bioinformatics search (S 2), we found 1462 putative BDPs in the *P. pastoris* genome (Fig. 1b) and selected a subset of 40 promoters for detailed characterization (Fig. 1c, S 4 for a list of the promoters tested). We tested all putative nBDPs involving MUT genes [36] on one side to acquire regulated BDPs and we tested putative nBDPs of typical housekeeping genes to obtain constitutive promoters using green and red fluorescent proteins (FPs) as reporters (Fig. 1c). The relative fluorescence units (rfu) obtained with different FPs are not directly comparable, as they depend on the specific quantum yields of the FPs and the fluorescence spectrometer settings. We determined a normalization factor to allow direct comparison of the two promoter sides in our experimental setting (S 3). Amongst MUT promoters, only the *DAS1-DAS2* promoter ($P_{DAS1-DAS2}$) showed strong expression on both sides, as suggested from previous studies using 1000 bp from each side ([36] and S 5b). The other MUT promoters showed only strong monodirectional expression in one direction as observed in previous studies [36]. Various putative nBDPs of housekeeping genes showed detectable expression on both sides weaker than the monodirectional *GAP (glyceraldehyde-3-phosphate-dehydrogenase)* promoter (P_{GAP}, strongest constitutive promoter in *P. pastoris* [37]) (Fig. 1c). The vast majority of these nBDPs exhibited limited applicability for gene coexpression showing only on one side strong expression levels or relatively weak bidirectional expression.

Bidirectional histone promoters provide strong expression in yeast

Most strikingly, natural bidirectional histone promoters of *P. pastoris* (P_{PpHTX1}, P_{PpHHX1} and P_{PpHHX2}, see S 6 for nomenclature) provide equally strong expression on both sides, matching (Fig. 1c) or even outperforming (S 7) the P_{GAP} benchmark. Histones, required for DNA packaging, range amongst the most highly conserved eukaryotic proteins [38] and their genomic organization in gene pairs flanking a bidirectional promoter is also conserved (S 6). Also the bidirectional histone promoters of the yeasts *S. cerevisiae* (Fig. 1d) and *Schizosaccharomyces pombe* (Fig. 1e) showed strong bidirectional expression matching strong constitutive MDPs of these organisms (S 6 [preliminary data, personal communication Pitzer J.].

Therefore bidirectional histone promoters match or exceed strong MDPs in yeasts, suggesting their wide applicability for gene coexpression in any eukaryote. The histone promoters used in this study were combined with canonical poly(A) terminators, thereby removing histone typical 3'UTRs associated with cell cycle associated degradation (S 6).

However, natural histone promoters offer only a fixed ratio and cumulative expression strength, whereas a variety of expression profiles is desirable for the optimization of gene coexpression using our library approach. We engineered yeast (Fig. 1f) histone promoters using truncation and deletion strategies resulting in variants with diversified ratios and cumulative expression levels. Systematic deletion studies of these nBDPs were favored by their short length condensing the regulatory elements needed for strong bidirectional expression on minimal space (S 6). Removing the core promoter from one side of a bidirectional promoter from yeast (Fig. 1f, S 9) increases monodirectional expression on the other side up to 1.5-fold (S 9), hinting a regulatory model in which two core promoters are competing for transcription initiation by general TFs or RNAPII (extended discussion in S 9).

Bidirectionalization and fusions of MDPs yield BDPs with extended ratios and regulatory profiles

The natural histone promoters tested and engineered versions thereof provide a range of growth dependent (cell cycle regulated [38,39]) cumulative expression levels and ratios. However, inducible regulation is for many applications desirable, since allowing a separation of cell growth from recombinant protein/metabolite production. Amongst all methanol inducible promoters of the *P. pastoris* MUT pathway, we found only a single, very large (2488 bp) BDP ($P_{DAS1\text{-}DAS2}$, S 5). Detailed deletion studies (S 5) showed that $P_{DAS1\text{-}DAS2}$ can also be viewed as a fusion of two MDPs, since expression from both sides is not interdependent. In contrast expression from both sides of P_{HHX2} depends on a shared regulatory region (Fig. 1f). One may argue that 'true' BDPs rely on shared regulatory elements and other divergent transcription is just achieved by divergent fusions of MDPs. However, there is no clear definition in the literature and we see the concept of BDPs in this study rather as specific parts providing simple divergent transcription and use the term accordingly.

We aimed to generate additional inducible BDPs by bidirectionalizing MDPs. In plants [6] and mammalian cell lines [8] bidirectionalization was achieved by fusing a second core promoter in reverse orientation to a MDP. Core promoters are in eukaryotes the minimal region providing binding sites for general transcription factors and RNA polymerase II. Typically core promoters alone provide no or low basal expression, requiring additional regulatory sequences to activate transcription. In higher eukaryotes the term 'enhancer' is used, whereas in yeast such DNA stretches are called upstream activating sequences (UAS). We use the general term *cis*-regulatory modules (CRMs) [40], including also repressor binding sites. When analyzing the sequences of the bidirectional *P. pastoris* histone promoters, we noticed short core promoter/5' untranslated regions (5' UTRs), suggesting their convenient use for promoter engineering (S 6). We fused six histone core promoters to 12 monodirecitonal *P. pastoris* promoters, partly varying the lengths of the core promoters and the MDPs resulting in 30 variants. Two thirds of constructs were successfully bidirectionalized, showing detectable expression from the second core promoter added. Only in the case of three promoters (P_{AOX1}, P_{FLD1} and P_{DAS2}) bidirectionalized expression of more than 50% of the native monodirectional side was reached. The construct $Pcore_{HTA1\text{-}81}$+$P_{DAS2\text{-}699}$ reached on the bidirectionalized side exceptionally high expression outperforming even strong MDPs. Different core promoter lengths moderately effected expression, whereas the length of the MDP had a drastic effect (e.g. $Pcore_{HTA1\text{-}81}$+$P_{DAS2\text{-}699}$ vs. $Pcore_{HTA1\text{-}81}$+$P_{DAS2\text{-}1000}$: very high vs. no bidirectionalized expression). These results are in contrast to milestone bidirectionalization studies in higher eukaryotes [6,8] where testing few promoters in a single length led to suitable BDPs. These dissimilarities may be explained by a different function/distance relationship between CRMs from yeast and higher eukaryotes suggesting that it is in yeast more difficult to find the transcriptional sweet spot for bidirectionalization.

The nBDPs characterized and bidirectionalized sBDPs generated showed on both sides the same regulation and did not offer combinations of regulatory profiles on each side (e.g. constitutive and inducible expression). Different regulatory profiles are required to achieve expression cascades to temporarily separate gene expression. For example the expression of a chaperone should ideally precede the protein of interest to efficiently assist in folding. We generated fusions of constitutive, derepressed and inducible MDPs [36] resulting in 30 fusion BDPs with distinct regulation on each side (Fig. 2b,c; S 8). The BDPs maintained in all cases on each side the regulatory profile of the MDPs proving that fusions of MDPs are a suitable strategy to create differently regulated BDPs. Some variants showed synergistic or antagonistic effects (S 8). In previous studies of MDP fusions in *S.*

cerevisiae [26–30], no such interference effects were reported possibly owing to promoter specific effects and a smaller number of combinations tested. A subset of the fusion promoters (Fig. 2c) consisted of combinations of *DAS1* and *DAS2* deletion variants (S 5). These fusions demonstrated that separately engineered MDPs maintain their individual expression levels and can be used to generate regulated BDPs with varying ratios on both sides (ratios of 0.16 to 0.96).

Bidirectional hybrid promoters reach unprecedented expression efficiency

The nBDPs, bidirectionalization and fusions of MDPs offer a wide range of cumulative expression levels, ratios and regulatory profiles in *P. pastoris*. Yet, the bidirectionalization efforts showed that we have little rational understanding on how promoter length and core promoter properties affect divergent transcription. To this end we assembled short defined CRMs (30-175 bp, S 5, S 9) with histone core promoters (S 6) into compact bidirectional hybrid promoters (Fig. 2d). The CRMs were selected from methanol regulated promoters based on literature data available on P_{AOX1} ([37], S 5) and deletion studies on P_{DAS1} and P_{DAS2} (S 5). Each CRM was characterized with a single core promoter (S 9b), two core promoters, and combinations of CRMs were tested in different positions (upstream/downstream) and orientations (Fig. 2d). To create combinations of reg. profiles we fused a truncated histone promoter variant ($P_{HHT2\text{-}T3}$, Fig. 1f) to a single CRM and one core promoter. Inducible synthetic hybrid BDPs matched expression from the monodirecitonal *AOX1* promoter. In terms of expression efficiency (expression/bp promoter length; a central characteristic for the assembly of short expression cassettes), hybrid BDPs are up to 3.3-fold more efficient than typically used nMDPs and 2.1-fold more efficient than the most efficient nBDP (Fig. 3e). Amongst sMDP control constructs (S 9), also variants up to 2.4-fold more efficient than nMDPs were generated. The length of the core promoters used and the orientation of the CRMs generally effected expression only marginally. Orientation independency of yeasts CRMs has long been known [41], our results demonstrate that this property can also be harnessed to generate strong BDPs.

The library of BDPs facilitates optimization of dual gene coexpression and combinatorial pathway assembly

With a library of 168 *P. pastoris* BDPs spanning a 61-fold range of expression ratios between the two sides (Fig. 3f) at hand, we aimed to optimize dual gene coexpression and multi-gene pathway assembly (Fig. 1 A, S 1). We used BDP based strategies reducing the number of cloning junctions compared to conventional MDPs (S 1a,d) allowing seamless fusions of the BDPs (S 1b,c and extended discussion therein). We optimized dual gene coexpression of a human cytochrome P450 (CYP2D6) and its associated reductase (CPR, required for electron transfer from NADPH) using a subset of strong, differently regulated BDPs from the library (Fig. 3a). In addition the effect of the chaperone protein-disulfide-isomerase (PDI) on secretion of the disulfide-bond-rich biocatalysts *Candida antarctica* lipase B (CalB, Fig. 3b) and horseradish peroxidase (HRP, Fig. 3c) was evaluated. We also evaluated production of taxadiene, the first committed precursor of the potent anticancer drug taxol (paclitaxel), requiring coexpression of two enzymes: geranylgeranyl diphosphate synthase (GGPPS) and taxadiene synthase (TDS) [42]. Constitutive expression worked well for HRP, whereas CYP2D6 activity was below the detection limit, possibly owing to cellular stress by UPR/ERAD (extended discussion in S 10). For taxadiene production, constitutive expression of the GGPPS was detrimental resulting in approximately 100-fold decreased transformation rates (data not shown). Although the BDPs used had given similar reporter protein fluorescence, there was a 4.9 to 50-fold difference in activity/yields (depending on the gene pairs of interest, S 10). Most strikingly, for taxadiene production, the worst strain produced 0.1 mg/L whereas the best strain reached approximately 6

mg/L, in range with highly engineered *S. cerevisiae* strains [43]. For each gene pair, a different BDP gave the highest yield (CYP2D6+CPR: $P_{DAS1\text{-}DAS2}$, CalB+PDI: $P_{CAT1\text{-}AOX1}$, HRP+PDI: $P_{AOX1+GAP}$, GGPPS+TDS: $P_{GAP+CAT1}$), highlighting the importance of optimizing the expression strategy.

The assembly of multi gene coexpression cassettes is facilitated using BDPs and bidirectional transcriptional terminators (BDTs) by reducing the number of parts (promoters and terminators) needed and increasing the efficiency of commonly used overlap-directed DNA assembly methods (S 1 D, [65]). We generated BDTs (S 11) and used them together with the library of BDPs for fine-tuning the expression of the carotenoid pathway producing beta-Carotene. We tested combinations of differently regulated promoters (constitutive, inducible) and expression ratios. The cassettes were shorter than using repeatedly the P_{AOX1} and shorter than using different MUT promoters and less parts, easing assembly and sequencing. Bidirectional cassettes result in smaller vectors, increasing transformation/transfection efficiencies and less sequencing reactions are required for covering the entire cassette. MDPs and BDPs can principally achieve similar expression levels, also for dual gene coexpression (Fig. 3a-c) as mimicking the best BDPs by MDPs resulted in similar activities (S 11). However BDPs considerably facilitate assembly (S 1) and shorten expression cassettes due to higher expression efficiencies (Fig. 2e). Testing the same promoter combinations using MDPs would have approximately doubled the cloning effort for assembly (S 1).
For pathway assemblies, polycistronic strategies such as internal ribosome entry sites (IRESs) and 2A sequences provide alternatives to BDPs [8,44] that aid to even further reduce construct size. Yet, these post-transcriptional processes do not allow varying regulatory profiles. Ratios may be varied to an unclear extent by varying the position of the genes. Yet combining a library of BDPs with 2A sequences [44] may combine the favorable properties of each system.

Conclusion

According to our study in *P. pastoris*, only few nBDPs are versatile tools to drive bidirectional expression. However, a library of diversified sBDPs, covering a range of expression levels and regulatory profiles can be easily assembled following modular design strategies. sBDPs created by bidirectionalization, fusions and hybrid design are more efficient than nBDPs and MDPs. It appears that in nature only in rare occasions such as the requirement for exact equimolar expression and intricate regulation (e.g. core histones) strong, tightly regulated nBDPs have evolved. It seems that in *P. pastoris* the need for separated regulation and independent evolution of promoters outweighs the need for tighter organization. However, artificial constructs for dual and multi gene coexpression are not limited by these evolutionary constraints and may be assembled from sBDPs. The sBDPs designed using parts of natural histone promoters allowed to screen for the optimal cumulative expression strength, ratio and regulatory profile of various applications in *P. pastoris*. Given the high conservation of histone BDPs in eukaryotes and the design strategies for sBDPs outlined here, similar libraries of BDPs can be generated for other eukaryotes, facilitating metabolic engineering and synthetic biology applications.

Acknowledgements

The authors gratefully acknowledge support from NAWI Graz. We would like to thank Clemens Farnleitner, Alexander Korsunsky and Marlies Wagner for excellent technical assistance.

References

[1] Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. Microbial engineering for the production of advanced biofuels. Nature 2012;488:320–8.

[2] Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 2012;40:e142.

[3] Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 2013;496:528–32.

[4] De Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, et al. Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 2014;3:97–106.

[5] Casini A, Macdonald JT, Jonghe J De, Christodoulou G, Freemont PS, Baldwin GS, et al. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. Nucleic Acids Res 2013:1–13.

[6] Xie M, He Y, Gan S. Bidirectionalization of polar promoters in plants. Nat Biotechnol 2001;19:677–9.

[7] Patro S, Maiti IB, Dey N. Development of an efficient bi-directional promoter with tripartite enhancer employing three viral promoters. J Biotechnol 2013;163:311–7.

[8] Amendola M, Venneri MA, Biffi A, Vigna E, Naldini L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. Nat Biotechnol 2005;23:108–16.

[9] Golding MC, Mann MRW. A bidirectional promoter architecture enhances lentiviral transgenesis in embryonic and extraembryonic stem cells. Gene Ther 2011;18:817–26.

[10] Baron U, Freundlieb S, Gossen M, Bujard H. Co-regulation of two gene activities by tetracycline via a bidirectional promoter. Nucleic Acids Res 1995;23:3605–6.

[11] Fux C, Fussenegger M. Bidirectional expression units enable streptogramin-adjustable gene expression in mammalian cells. Biotechnol Bioeng 2003;83:618–25.

[12] Liu B, Paton JF, Kasparov S. Viral vectors based on bidirectional cell-specific mammalian promoters and transcriptional amplification strategy for use in vitro and in vivo. BMC Biotechnol 2008;8:49.

[13] Yang S, Sleight SC, Sauro HM. Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits. Nucleic Acids Res 2012:1–7.

[14] Wei W, Pelechano V, Järvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. Trends Genet 2011;27:267–76.

[15] Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature 2009;457:1033–7.

[16] Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 2009;457:1038–42.

[17] Park D, Morris a. R, Battenhouse a., Iyer VR. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res 2014:1–14.

[18] Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet 2013;14:880–93.

[19] Almada AE, Wu X, Kriz AJ, Burge CB, Sharp P a. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature 2013;499:360–3.

[20] Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010;465:182–7.

[21] Adachi, N. and Lieber MR. Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome. Cell 2002;109:807–9.

[22] Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. Genome Res 2004;14:62–6.

[23] Yang M, Elnitski L. Orthology-driven mapping of bidirectional promoters in human and mouse genomes. BMC Bioinformatics 2014;15 Suppl 1:S1.

[24] Andrianaki a, Siapati EK, Hirata RK, Russell DW, Vassilopoulos G. Dual transgene expression by foamy virus vectors carrying an endogenous bidirectional promoter. Gene Ther 2010;17:380–8.

[25] Liu X, Zhou X, Li Y, Tian J, Zhang Q, Li S, et al. Identification and functional characterization of bidirectional gene pairs and their intergenic regions in maize. BMC Genomics 2014;15:338.

[26] Miller CA, Martinat MA, Hyman LE. Assessment of aryl hydrocarbon receptor complex interactions using pBEVY plasmids: expressionvectors with bi-directional promoters for use in Saccharomyces cerevisiae. Nucleic Acids Res 1998;26:3577–83.

[27] Li A, Liu Z, Li Q, Yu L, Wang D, Deng X. Construction and characterization of bidirectional expression vectors in Saccharomyces cerevisiae. FEMS Yeast Res 2008;8:6–9.

[28] Partow S, Siewers V, Bjørn S, Nielsen J, Maury J. Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae. Yeast 2010;27:955–64.

[29] Vickers CE, Bydder SF, Zhou Y, Nielsen LK. Dual gene expression cassette vectors with antibiotic selection markers for engineering in Saccharomyces cerevisiae. Microb Cell Fact 2013;12:96.

[30] Da Silva NA, Srikrishnan S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res 2012;12:197–214.

[31] Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Front Microbiol 2014;5:85.

[32] Jahic M, Veide A, Charoenrat T, Teeri T, Enfors S-O. Process technology for production and recovery of heterologous proteins with Pichia pastoris. Biotechnol Prog 2006;22:1465–73.

[33] Vogl T, Hartner FS, Glieder A. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 2013;24:1094–101.

[34] Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, et al. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science (80-) 2006;313:1441–3.

[35] Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, et al. Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 2006;24:210–5.

[36] Chapter 3.1 of this thesis (Vogl *et al*. manuscript in preparation)

[37] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.

[38] Eriksson PR, Ganguli D, Nagarajavel V, Clark DJ. Regulation of histone gene expression in budding yeast. Genetics 2012;191:7–20.

[39] Artishevsky A, Wooden S, Sharma A, Resendez E, Lee AS. Cell-cycle regulatory sequences in a hamster histone promoter and their interactions with cellular factors. Nature 1987;328:823–7.

[40] Lelli KM, Slattery M, Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet 2012;46:43–68.

[41] Hahn S, Young ET. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics 2011;189:705–36.

[42] Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 2010;330:70–4.

[43] Engels B, Dahm P, Jennewein S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 2008;10:201–6.

[44] Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in P. pastoris. Chem Commun (Camb) 2014;62.

[45] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.

[46] Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 2012;7:e39720.

[47] Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol 2004;22:1567–72.

[48] Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.

[49] Chapter 3.5 of this thesis (Vogl *et al*. manuscript in preparation)

[50] Gudiminchi RK, Geier M, Glieder A, Camattari A. Screening for cytochrome P450 expression in Pichia pastoris whole cells by P450-carbon monoxide complex determination. Biotechnol J 2013;8:146–52.

[51] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 2012;11:22.

[52] Mead DA, Pey NK, Herrnstadt C, Marcil RA, Smith LM. A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology (N Y) 1991;9:657–63.

[53] Quan J, Tian J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 2011;6:242–51.

[54] Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 2007;4:251–6.

[55] Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 2012;40:e55.

[56] Crook NC, Freeman ES, Alper HS. Re-engineering multicloning sites for function and convenience. Nucleic Acids Res 2011;39:e92.

[57] Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, Mulye T, et al. Analysis of the 5’ untranslated region (5'UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in Pichia pastoris. Gene 2012;496:118–27.

[58] Liang S, Wang B, Pan L, Ye Y, He M, Han S, et al. Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. BMC Genomics 2012;13:738.

[59] Hesketh AR, Castrillo JI, Sawyer T, Archer DB, Oliver SG. Investigating the physiological response of Pichia (Komagataella) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. Appl Microbiol Biotechnol 2013.

[60] Chang DT-H, Wu C-Y, Fan C-Y. A study on promoter characteristics of head-to-head genes in Saccharomyces cerevisiae. BMC Genomics 2012;13 Suppl 1:S11.

[61] Kranthi BV, Kumar HRV, Rangarajan PN. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris. Yeast 2010;27:705–11.

[62] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.

[63] Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y. An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. FEMS Yeast Res 2009;9:1271–82.

[64] Blazeck J, Alper HS. Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol J 2013;8:46–58.

[65] Inan M. . Studies on the alcohol oxidase (AOX1) promoter of Pichia pastoris. University of Nebraska, 2000.

[66] Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN. Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. Biochim Biophys Acta 2009;1789:460–8.

[67] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539.

[68] Basehoar AD, Zanton SJ, Pugh BF. Identification and distinct regulation of yeast TATA box-containing genes. Cell 2004;116:699–709.

[69] Baldo AM, Les DH, Strausbaugh LD. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of Sophophora. Mol Biol Evol 1999;16:1511–20.

[70] Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet 2008;9:843–54.

[71] Dominski Z, Marzluff WF. Formation of the 3’ end of histone mRNA: getting closer to the end. Gene 2007;396:373–90.

[72] Mackenzie D a, Wongwathanarat P, Carter a T, Archer DB. Isolation and use of a homologous histone H4 promoter and a ribosomal DNA region in a transformation vector for the oil-producing fungus Mortierella alpina. Appl Environ Microbiol 2000;66:4655–61.

[73] Belshaw NJ, Haigh NP, Fish NM, Archer DB, Alcocer MJC. Use of a histone H4 promoter to drive the expression of homologous and heterologous proteins by Penicillium funiculosum. Appl Microbiol Biotechnol 2002;60:455–60.

[74] Kelemen Z, Mai A, Kapros T, Fehér A, Györgyey J, Waterborg JH, et al. Transformation vector based on promoter and intron sequences of a replacement histone H3 gene. A tool for high, constitutive gene expression in plants. Transgenic Res 2002;11:69–72.

[75] Gallie DR, Lewis NJ, Marzluff WF. The histone 3 ′ -terminal stem – loop is necessary for translation in Chinese hamster ovary cells 1996;24:1954–62.

[76] Gorgoni B, Andrews S, Schaller A, Schümperli D, Gray NK, Müller B. The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. RNA 2005;11:1030–42.

[77] Nevoigt E, Kohnke J, Fischer CR, Alper H, Stahl U, Stephanopoulos G. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. Appl Environ Microbiol 2006;72:5266–73.

[78] Matsuyama A, Shirai A, Yoshida M. A series of promoters for constitutive expression of heterologous genes in fission yeast. Yeast 2008;25:371–6.

[79] Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic Core Promoters for Pichia pastoris. ACS Synth Biol 2014;3:188–91.

[80] Waterham HR, Digan ME, Koutz PJ, Lair S V, Cregg JM. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 1997;186:37–44.

[81] Bawa Z, Routledge SJ, Jamshad M, Clare M, Sarkar D, Dickerson I, et al. Functional recombinant protein is present in the pre-induction phases of. Microb Cell Fact 2014;13:127.

[82] Inan M, Meagher MM. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. J Biosci Bioeng 2001;92:585–9.

[83] Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, et al. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol Cell Biol 2006;26:883–97.

[84] Geier M, Braun A, Emmerstorfer A, Pichler H, Glieder A. Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes - a comparative study. Biotechnol J 2012:1–13.

[85] Wang B, Yang L-P, Zhang X-Z, Huang S-Q, Bartlam M, Zhou S-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev 2009;41:573–643.

[86] Gajhede M, Schuller DJ, Henriksen A, Smith AT, Poulos TL. Crystal structure of horseradish peroxidase C at 2.15 A resolution. Nat Struct Biol 1997;4:1032–8.

[87] Uppenberg J, Hansen MT, Patkar S, Jones TA. The sequence, crystal structure determination and refinement of two crystal forms of lipase B from Candida antarctica. Structure 1994;2:293–308.

[88] Uppenberg J, Ohrner N, Norin M, Hult K, Kleywegt GJ, Patkar S, et al. Crystallographic and molecular-modeling studies of lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols. Biochemistry 1995;34:16838–51.

[89] Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 1971;93:2325–7.

[90] Prescott EM, Proudfoot NJ. Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A 2002;99:8796–801.

[91] Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. RNA polymerase II collision interrupts convergent transcription. Mol Cell 2012;48:365–74.

[92] Chen S, Reger R, Miller C, Hyman LE. Transcriptional terminators of RNA polymerase II are associated with yeast replication origins. Nucleic Acids Res 1996;24:2885–93.

[93] Cregg JM, Barringer KJ, Hessler AY, Madden KR. Pichia pastoris as a host system for transformations. Mol Cell Biol 1985;5:3376–85.

[94] Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A 1981;78:6354–8.

Materials and methods

Analysis of *P. pastoris* genome organization

The genome sequence of the *P. pastoris* CBS7435 strain [45] was analyzed chromosome by chromosome (GenBank IDs: FR839628.1, FR839629.1, FR839630.1 and FR839631.1) for genes in head to head, tail to head (head to tail) and tail to tail orientation similar to the analysis of Trinklein *et al.* [22] of the human genome. Lists of the gene pairs identified are provided in S 2. In rare cases genetic elements such as tRNAs, rRNAs, mobile elements or sequencing gaps were annotated between two genes transcribed by RNA polymerase II. The presence of genetic elements is denoted in S 2, gene pairs separated by gaps were omitted from the analysis.

Vector construction

P. pastoris promoter reporter vectors

For basic characterizations, a pPpT4_S [46] based expression vector bearing a single eGFP reporter gene previously reported was used (pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer [36]). With the single reporter vector, bidirectional promoters had to be cloned twice, once in forward and once in reverse orientation. The *P. pastoris* nBDPs were initially characterized by these means. To reduce the cloning effort and allow simultaneous detection of both sides, we designed a bidirectional screening vector. Based on the single reporter vector, we inserted a second reporter gene (a red fluorescent protein variant termed dTomato [47]), between the targeting sequence and the stuffer fragment of pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer. We also tested different fluorescent proteins and designed different vector variants of the RFP (data not shown). The vector was assembled by digesting the single reporter vector with *Asc*I and *Avr*II. Subsequently the RFP fused to a *P. pastoris* transcription terminator sequence was PCR amplified from a *P. pastoris* cloning vector using primers newTomatoAscIBmrIFWD and AOXTTSbfIAvrIIREV1. To add an additional SbfI restriction site, the obtained PCR fragment was used as template for a second PCR using primers newTomatoAscIBmrIFWD and AOXTTSbfIAvrIIREV2. The newly inserted part was confirmed by Sanger sequencing. Subsequently we cloned several natural bidirectional promoters and semi synthetic fusion promoters into this vector. The promoters were either inserted in random orientation by TA cloning or directional by Gibson assembly [48].
The bidirectional reporter vector described here was also used as entry vector for the coexpression of any gene pair. Therefore a cassette consisting of the two genes to be coexpressed with a stuffer fragment between them is assembled by olePCR, digested with *Not*I and cloned in the *Not*I digested bidirectional double reporter vector backbone. Alternatively also Gibson assembly can be used. *S. cerevisiae* and *S. pombe* reporter vectors were generated by Pitzer J. (personal communication).

Assays

Fluorescence measurements of eGFP and sTom were performed as previously reported [49] see S 3 for extended discussion on the normalization. CYP2D6 activity measurements were performed as outlined previously using 7-methoxy-4-(amino - methyl)-coumarin (MAMC) as substrates [50]. CalB and HRP activities in the supernatants were determined using p-Nitrophenyl butyrate and 2,2´-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS) as substrates as reported previously [51].

Figures

Fig. 1

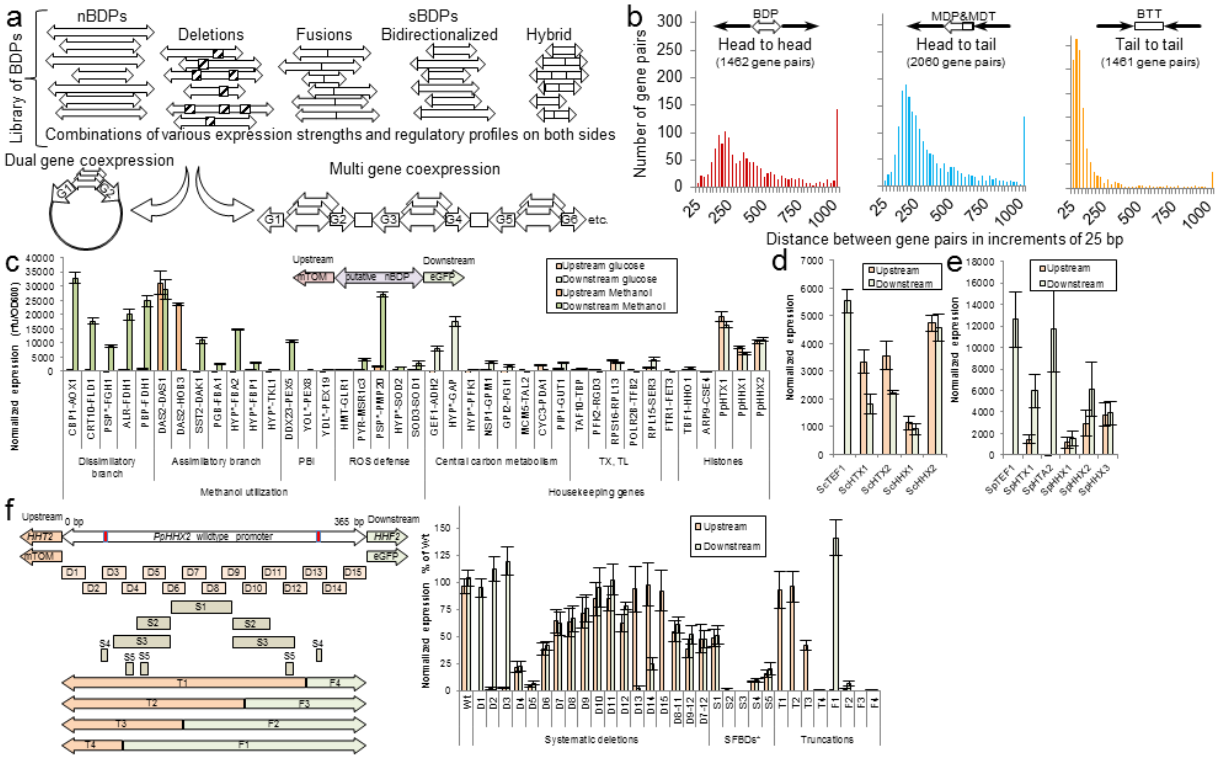


Fig. 1: Bidirectional histone promoters are amongst the few strong *P. pastoris* nBDPs and can be used as general eukaryotic engineering framework as demonstrated in *S. cerevisiae* and *S. pombe*.

A) A library of diversely regulated natural and synthetic bidirectional promoters (nBDPs and sBDPs) covering a wide range of regulatory profiles facilitates optimization of dual gene coexpression and the assembly of multi gene coexpression cassettes (S 1).

B) The *P. pastoris* genome harbors 1462 putative nBDPs (gene pairs in divergent head to head orientation, S 2). The distribution of distances between gene pairs is shown in 25 bp intervals. The last bar indicates gene pairs with an intragenic distance greater than 1000 bp. Also convergent tail to tail gene pairs (forming putative bidirectional transcription terminators, BDTs) and head to tail/tail to head gene pairs flanking a monodirectional promoter (MDP) and a monodirectional terminator (MTT) are shown. Genes are illustrated as bold single-line arrows, promoters as filled arrows, terminators as rectangles.

C) The natural bidirectional *DAS1-DAS2* promoter is the only *P. pastoris* methanol inducible promoter [36] showing strong reporter gene fluorescence on both sides and histone promoters are the strongest nBDPs of several housekeeping gene pairs tested in *P. pastoris*. All strains were grown on glucose media for 60 h and MUT promoters subsequently induced with methanol for 48 h (for MUT promoters measurements on methanol, for housekeeping genes on glucose are shown). The promoters were screened with a single reporter gene in both orientations and bidirectional expression confirmed using two FPs (normalization factor used as determined in S 3). Gene names denoted with an asterisk (*) were shortened and are provided in S 4. For nomenclature of histone promoters see S 6. Mean values and standard deviations of biological quadruplicates are shown. PBI: peroxisome biogenesis and import; ROS reactive oxygen species; TX,TL: transcription, translation.

D) and E) Histone promoters of *S. cerevisiae* (d) and *S. pombe* (e) match strong constitutive promoters described in these organisms (S 6) [preliminary data, personal communication Pitzer J.].

F) Systematic deletions and truncations of the *P. pastoris HHX2* promoter offer shortened variants with altered cumulative expression levels and ratios. On the left side of the panel a schematic on the sequence variants is shown (S 4 for exact positions). TATA boxes are denoted by red rectangles. On the right side expression levels after growth for 60 h on glucose are shown. *SFBDs: sequence feature based deletions (S 6).

Fig. 2

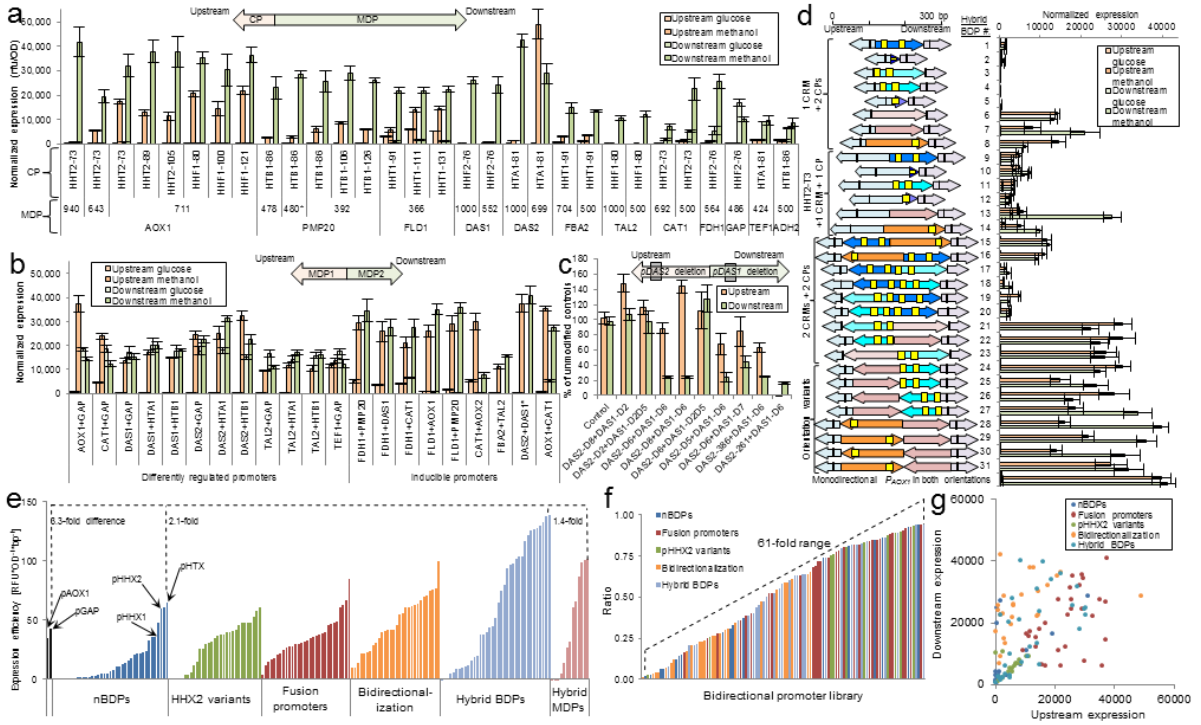


Fig. 2: Bidirectionalization and fusions of MDPs yield BDPs extending the repertoire of ratios and regulatory profiles in *P. pastoris*. Modularly designed bidirectional hybrid promoters achieve highest expression efficiency.

A) Bidirectionalization of MDPs by addition of core promoters (S 6) yielded functional BDPs in most cases, but few designs gave high expression. The core promoters (pCore) indicated were fused to the indicated MDPs. The length of the MDPs is given in bp, selection criteria are outlined in S 8. *: In case of the Pmp20 promoter slightly varying sequences from the CBS7435 and the GS115 strain were tested (S 8). In all panels of this figure mean values and standard deviations of normalized (S 3) reporter protein fluorescence measurements of biological quadruplicates grown on the respective carbon sources are shown.

B) Fusions of differently regulated MDPs yield BDPs with different regulatory profiles on each side. Fusions of methanol inducible MDPs provide a set of strong, tightly regulated, sequence diversified BDPs allowing coexpression of up to 10 genes without reusing any sequence. In case of P_{HTA1} and P_{HTB1} the truncated versions shown in S 6 and S 9 were used. *: Here only the fusion of $P_{DAS2\text{-}699}$+$P_{DAS1\text{-}552}$ is shown, for additional comparisons see S 5.

C) Fusing deletion variants of *DAS1* and *DAS2* promoters offers strong inducible BDPs with different expression ratios between the sides demonstrating that variants of MDPs can be combined into BDPs maintaining their properties on each side. The rationale for the selection of the deletions in P_{DAS1} and P_{DAS2} and the measurements of the separate promoters are shown in S 5. Fluorescence was measured after 48 h methanol induction and shown as percent of the unmodified fusion promoter ($P_{DAS2\text{-}1000}$+$P_{DAS1\text{-}1000}$).

D) Bidirectional hybrid promoters assembled from histone core promoters (S 6) and CRMs of methanol regulated promoters (S 9) enable modular design matching the strong monodirectional *AOX1* promoter. The color code for the regulatory elements used is provided in S 9, a list of the exact designs of sBDP1-32 is provided in S 4. Yellow boxes indicate experimentally confirmed Mxr1p (methanol master regulator) binding sites in P_{AOX1} and P_{DAS2} (S 5, S 9), red boxes: TATA boxes. Additional bidirectional variants, controls and extended discussion are provided in S 9. P_{AOX1} is a reference of a monodirectional, strong, methanol inducible promoter. P_{AOX1} was cloned in forward and reverse orientation in the bidirectional reporter vector, therefore the values shown are derived from separate constructs and not from bidirectional activity. Abbreviations: CP: core promoter, CRM: *cis*-regulatory module. 'HHT2-T3' is the truncated side of a bidirectional histone promoter (Fig. 1f) used to generate hybrid promoters with growth associated expression from one side.

E) Expression efficiencies of sBDPs surpass nBDPs up to 2.1 fold and nMDPs up to 3.3-fold. The expression efficiencies were calculated by adding up the normalized reporter protein fluorescence measurements of both sides (under optimal growth conditions) and dividing the sum by the length of the promoter. The monodirecitonal *AOX1* and *GAP* promoters are included as references for state of the art nMDPs. Fold differences between the most efficient hybrid promoters and the most efficient nBDPs, hybrid MDPs and the monodirectional reference promoters are shown.

F) The library of BDPs offers different ratios between the two sides of the promoters, ranging from equal expression to a 61-fold difference. The ratios were calculated from the normalized reporter protein fluorescences (of the optimal growth conditions), by dividing the lower value by the higher value. Different growth conditions of the strains with differently regulated promoters even extend the ratios achievable. Only promoters clearly exceeding the background signal of the measurements (>500 rfu for eGFP, >100 rfu for dTom) were included in the calculations.

G) The library of BDPs covers the whole expression space. Normalized upstream and downstream reporter fluorescence is shown (by the default orientation in which the BDPs were cloned in the reporter vector).

Fig. 3

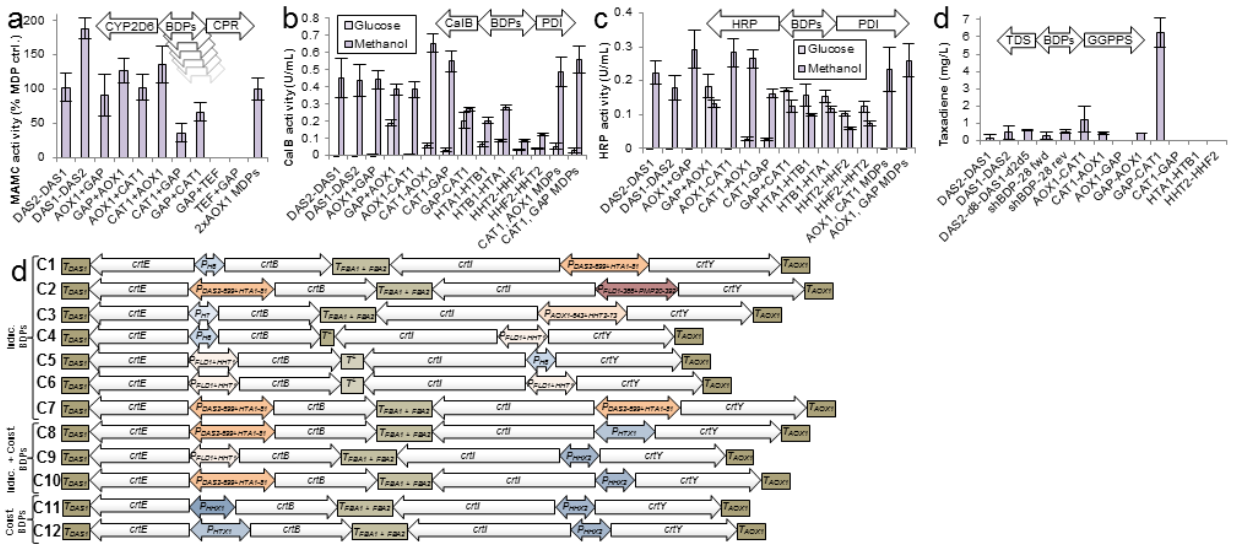


Fig. 3: Applying the library of BDPs helps to find the optimal expression condition for dual gene coexpression and facilitates the assembly and transcriptional fine-tuning of multi-gene pathways.

A) Highest activity for the coexpression of human CYP2D6 and its associated CPR was achieved using the natural $P_{DAS1\text{-}DAS2}$ promoter in reverse orientation. A subset of strong BDPs was cloned between the two genes and screened for highest activity compared to a control strain expressing the two genes using two monodirectional *AOX1* promoters. The strains were pregrown for 60 h on glucose and induced with methanol for 72 h. Activity was measured by a whole cell bioconversion assay using 7-methoxy-4-(aminomethyl)-coumarin (MAMC) as substrate. See S 10 for details on the controls used and extended discussions of all panels of this figure.

B) Bidirectional fusions promoters of P_{CAT1} to P_{AOX1} or P_{GAP} give highest volumetric activities in the coexpression of secreted CalB and the chaperone PDI. Similar approach as in panel A, however activities in the supernatant were also measured after growth for 60 h on glucose using a pNPB assay.

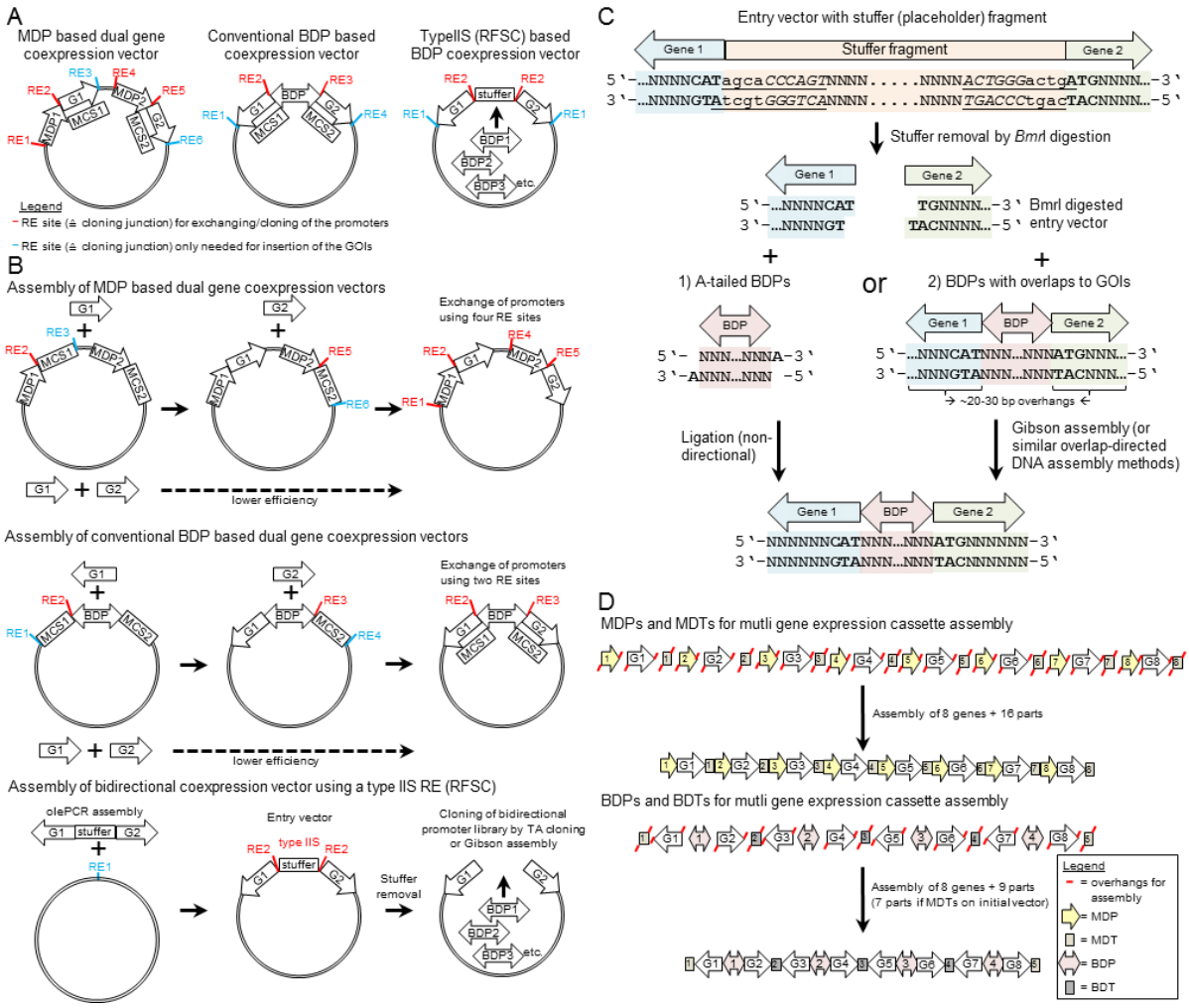
C) Similar activities for coexpression of secreted HRP and the chaperone PDI are achieved using inducible BDPs (fusions of P_{AOX1} or P_{CAT1}) or growth associated/constitutive BDPs (P_{GAP}, P_{HTX1}). Same experiment as in panel B, except use of an ABTS assay.

D) Highest taxadiene yields were achieved using a $P_{GAP\text{-}CAT1}$ fusion promoter for GGPPS and TDS coexpression. Constitutive expression of the GGPPS gene was detrimental (data not shown). Yields determined by GC-MS from shake flask cultivations with a dodecane overlay.

E) Using BDPs and BDTs for pathway assembly reduces construct length and the number of parts required, demonstrated for the four gene (*crtE*, *crtB*, *crtI*, *crtY*) model pathway of beta-carotene biosynthesis. Twelve bidirectional constructs were assembled by combining inducible or constitutive BDPs and combinations thereof (Induc. + const.) with a BDT and two MTTs. See S 1 for assembly strategy, S 11 for BDTs used and S 4 on detailed information on the constructs generated. For the BDPs, a coloring scheme similar to Fig. 2a-c was used. *T**: natural bidirectional terminator between the *S. cerevisiae IDP1* and *PEX19* genes; T^+: natural bidirectional terminator between the *P. pastoris TEF1* and *GDM1* genes. The bidirectionalized $P_{FLD1\text{-}366+HHT1\text{-}91}$ was used.

Supplementary materials

S 1



S 1: Molecular cloning of BDPs via TA cloning or Gibson assembly facilitates optimization of dual and multi gene coexpression compared to MDPs.

A) Dual gene coexpression vectors based on BDPs and type IIS restriction endonucleases (REs) require less restriction sites/cloning junctions than MDP based vectors or conventional bidirectional vectors (e.g. [26,27]). We use the term 'cloning junctions' to refer to identical sequences required by overlap-directed DNA assembly methods such as Gibson assembly. MCS elements depict multiple cloning sites required for cloning of the genes of interest (G1 and G2).

B) Comparison of vector assemblies using MDP based vectors, conventional bidirectional vectors and the stuffer/typeIIS RE strategy reported here. Removal of a stuffer (placeholder) fragment from an entry vector using a single typeIIS RE enables the testing of a library of seamlessly linked BDPs.

C) Applying type IIS restriction endonucleases for seamless, sequence independent cloning of BDPs by TA cloning [52] or providing junctions for Gibson assembly [48]. The start codons of the two genes are written bold, the entire BmrI site is underlined and the recognition sequence is written in uppercase in italics.

D) Using BDPs and BDTs (bidirectional terminators) cuts the number of parts (promoters and terminators) approximately in half compared to MDPs and MDTs (monodirectional terminators) facilitating the assembly of multi gene expression cassettes. The assembly of eight genes is shown as an example. The bidirectional cloning (entry) vectors used in this study for inserting bidirectional multi gene expression cassettes provide already two MDTs, therefore the number of parts is reduced from nine to seven.

Extended discussion

Optimization of dual gene coexpression

The coexpression of two genes can be fine-tuned using vectors based on MDPs or BDPs (S 1A). Using MDPs for optimization of dual gene coexpression requires at first cloning of the genes of interest (GOIs) into the reporter vector and subsequent insertion of differently regulated MDPs (S 1B). Therefore in total six unique restriction endonucleases (RE) sites are required and four RE sites are needed for exchanging the promoters. Alternatively to RE based cloning, overlap-directed DNA assembly methods such as Gibson assembly [48], CPEC [53] or SLIC/SLiCE [54,55] can be used requiring overlapping regions (here referred to as 'cloning junctions') with the vector. Thus for each gene pair a new set of primers is needed to amplify the MDPs to be tested for fine-tuning coexpression.

Currently used bidirectional vectors (e.g. [26,27] or a mammalian expression vector (pBI-CMV1) offered by Clontech (/Takara Bio Inc., Mountain View, CA, USA)) rely on a fixed bidirectional promoter and subsequent cloning steps using multiple cloning sites (MCSs) (S 1 A, B). This design requires a separate cloning vector and sequential cloning steps (requiring four RE sites) for each BDP to be tested or tedious replacement of the BDP using two RE sites. This strategy enables the basic use of BDPs for gene coexpression but is unfeasible for testing a library of BDPs required for optimization of cumulative expression levels, ratios and regulatory profiles.

Here, we use a cloning strategy based on the removal of a stuffer (place holder) fragment via a type IIS RE in combination with TA cloning [52] or Gibson assembly [48] (S 1 C) allowing RE site free, seamless cloning of a large number of BDPs. Thereby the promoter providing the most suitable cumulative expression strength, the most favorable ratio and most suitable regulatory profile for the coexpression of two genes can be identified.

An expression cassette of the two genes of interest separated by a stuffer fragment is cloned into a starting vector using a single RE (S 1 B). In a subsequent cloning step the stuffer fragment is entirely cleaved out using a single type IIS RE resulting in vector ends suitable for inserting PCR amplified BDPs (S 1 C). In total only two REs are needed for preparing the vector. This approach does not require RE digestion of the BDPs or the presence of MCSs in the vector and maintains the natural sequence context of the BDP up to the start codon. MCSs contain several RE sites adding non-natural sequences to the 5' untranslated region of the mRNA that can interfere with mRNA structure thereby causing translation inhibition [56]. In *P. pastoris*, it has been shown that an increased length of the 5' UTR decreases the expression of the commonly used *alcohol oxidase 1* promoter (P_{AOX1}) [57].

Conventional type II enzymes such as *Eco*RI and *Eco*RV cleave within their palindromic recognition sequences creating sticky or blunt ends. Type IIS REs such as *Bsa*I, *Mly*I and *Bmr*I recognize non palindromic sequences and cleave in a variable sequence outside of their recognition sequence. We have placed two *Bmr*I sites at the end of the stuffer fragment in opposite orientations resulting in complete removal of the stuffer fragment by digestion with a single enzyme (S 1 C). *Taq* DNA polymerase adds by default a 3' adenine overhang in PCR amplification that can ligate with a thymidine overhang created by digestion with a type IIS RE (TA cloning [52]) [PCR products of proof reading polymerases such as *Pfu* can be A-tailed with *Taq* polymerase after purification]. *Bmr*I creates a variable single nucleotide 3' overhang. We positioned this overhang in the beginning of the start codon of the two genes, resulting in 3' thymidine overhangs on both sides of the vector (S 1 C). Adenine-tailed PCR fragments of BDPs can be directly cloned into the vector by TA cloning complementing the start codons. The primers for BDP amplification can be designed up to the start codons. Using this strategy, a completely natural promoter and 5'UTR sequence is achieved, omitting any bias from MCS or RE sites. In addition the BDPs do not have to be digested with REs and no overhangs are needed. Therefore the same library of BDPs can be used for cloning between any gene pairs. Alternatively overlaps and Gibson assembly can be used. However, in this case it is necessary to add overlaps to the GOIs to all promoters and new primers are needed for each gene pair to be coexpressed.

For cloning of a larger number of constructs and the availability of suitable screening systems, TA cloning is favorable. Due to higher efficiencies and less errors in the cloning junctions, we recommend Gibson assembly/overlap based cloning if a small set BDPs is to be tested.

Random insertion of fragments by TA cloning is a major disadvantage for the cloning of MDPs or coding sequences as only the forward orientation is functional. For BDPs it is however a beneficial trait, since the same BDP can be tested in both orientations in a single cloning experiment, thereby facilitating coexpression optimization.

Also blunt end ligations based on stuffer removal by MlyI can be directly used to clone PCR fragments but they show lower efficiencies (TA cloning showed up to 50 fold higher ligation efficiencies than blunt end cloning [52]).

Assembly and optimization of multi gene coexpression cassettes

BDPs and BDTs (bidirectional terminators) facilitate also the assembly of multi gene expression cassettes in comparison to MDPs and MDTs (monodirecitonal terminators). Typically the efficiency of overlap-directed DNA assembly methods is decreasing with the number of fragments in the assembly [4]. The number of parts (promoters and terminators) needed is approximately cut in half using BDPs and BDTs over MDPs and MDTs, considerably increasing the efficiency of multi fragment assemblies (depending on the method used [4]). In addition cassettes based on the bidirectional elements reported here are shorter than using monodirectional elements. Smaller expression cassettes can be verified with less sequence reactions and show typically higher transformation efficiencies.

S 2

S 2: Lists of gene pairs in the *P. pastoris* genome and extended discussion.

The analysis was performed as described in the materials and methods section. Genes in head to head, head to tail and tail to tail orientation are provided in the separate sheets of the excel file.

Legend:

length: length of the intergenic region in bp; type: orientation of the two genes to each other ('<' and '>' characters indicate the orientation arrow-like); g1-from/g1-to: begin/end of the upstream gene of the gene pair on the respective chromosome; g2-from/g2-to: begin/end of the downstream gene of the gene pair on the respective chromosome; g1-orientation: orientation of the upstream gene of the gene pair on the reverse (complement) or forward (normal) strand; g1-CDS-range: coding sequence of the upstream gene ('join' and multiple numbers indicate splicing events); g1-locus_tag: gene identifier containing chromosome number; g1-product: gene product of the upstream gene; g1-protein_id: accession number of the protein sequence; g1-gene: gene name (if assigned); g1-inference: protein motifs (if assigned); g1-EC_number: Enzyme Commission number (if assigned); the same terms (-orientation to -EC_number) are also given for the downstream gene (g2); inbetween: tRNA, rRNA or mobile_elements present in the intragenic region

Extended discussion

Analysis of genome organization

This search was limited to directly adjacent genes. Missannotations (or hypothetical genes present) may bias the results. For example the natural bidirectional *DAS1*/*DAS2* promoter is missing from the list of putative nBDPs, as a gene termed "Probable guanine nucleotide exchange factor FLJ41603 homolog " is between the *DAS1* and *DAS2* genes (S 5). To rule out bias of the annotation, the genes of the MUT pathway were manually curated for putative nBDPs.

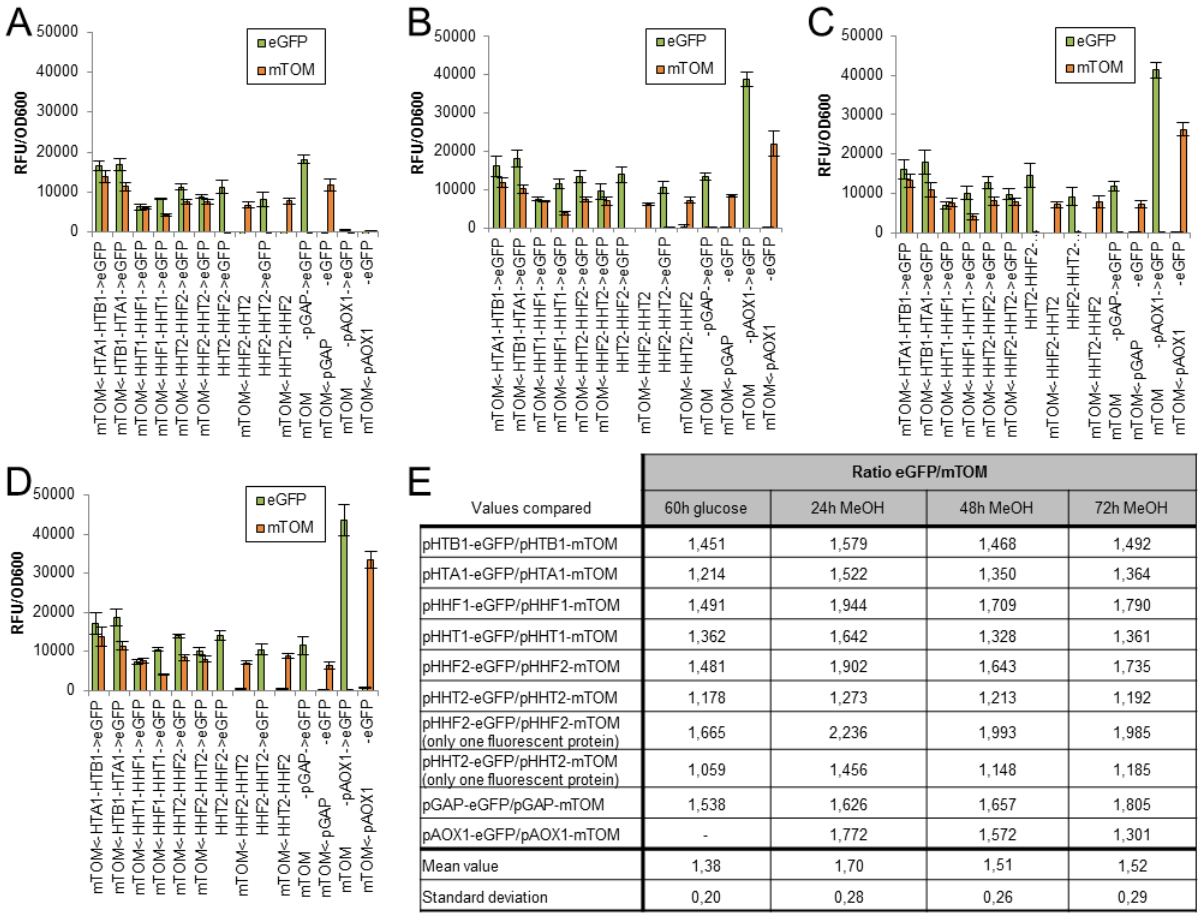
Selection and testing of putative nBDPs

The list of head to head genes was searched for putative nBDPs of typical housekeeping genes (Fig. 1c) to obtain constitutive promoters. Gene pairs containing annotations with the terms "putative", "hypothetical", "uncharacterized" or "probable" were omitted from the analysis. We focused on genes of the central carbon metabolism, general transcription machinery and ribosomal proteins.

Genome wide absolute quantification by RNA sequencing (RNAseq) may facilitate nBDP characterization, since promising nBDP targets can be directly selected from their expression strength. However, at the time we started this study, no RNAseq data for *P. pastoris* was available and RNAseq studies in *P. pastoris* remain scarce [58,59]. Yet, for widely studied model organisms such as *S. cerevisiae* with an abundance of RNAseq data studies at hand, pre-selection of putative nBDPs may considerably reduce screening efforts (aside studies on cryptic/pervasive bidirectional transcription [15,16], so far only a DNA sequence based study on BDPs has been performed in *S. cerevisiae* comparing sequence features such as the presence of TATA boxes and transcription factor binding sites [60]).

However even with RNAseq studies it may be impossible to find nBDPs with specific regulatory profiles, since they may not exist. For a library of BDPs to optimize gene coexpression, inducible nBDPs and combinations of inducible and constitutive promoters sides are desirable to fine tune expression in a time dependent manner. In *P. pastoris* only a limited set of methanol regulated promoters is known or anticipated [36] and we have tested all putative nBDP with MUT promoters on one side (Fig. 1c).

S 3



Values compared	Ratio eGFP/mTOM			
	60h glucose	24h MeOH	48h MeOH	72h MeOH
pHTB1-eGFP/pHTB1-mTOM	1,451	1,579	1,468	1,492
pHTA1-eGFP/pHTA1-mTOM	1,214	1,522	1,350	1,364
pHHF1-eGFP/pHHF1-mTOM	1,491	1,944	1,709	1,790
pHHT1-eGFP/pHHT1-mTOM	1,362	1,642	1,328	1,361
pHHF2-eGFP/pHHF2-mTOM	1,481	1,902	1,643	1,735
pHHT2-eGFP/pHHT2-mTOM	1,178	1,273	1,213	1,192
pHHF2-eGFP/pHHF2-mTOM (only one fluorescent protein)	1,665	2,236	1,993	1,985
pHHT2-eGFP/pHHT2-mTOM (only one fluorescent protein)	1,059	1,456	1,148	1,185
pGAP-eGFP/pGAP-mTOM	1,538	1,626	1,657	1,805
pAOX1-eGFP/pAOX1-mTOM	-	1,772	1,572	1,301
Mean value	1,38	1,70	1,51	1,52
Standard deviation	0,20	0,28	0,26	0,29

S 3: Normalization of the two fluorescent reporter proteins used for characterization of the BDPs in *P. pastoris*. Bidirectional histone promoters and the monodirectional *GAP* and *AOX1* promoters were cloned between the two reporter genes eGFP (enhanced GFP) and dTomato (a red fluorescent protein variant [47]). Reporter gene fluorescence was measured after **(a)** 60 h growth on glucose and subsequently after 24 h **(b)**, 48 h **(c)** and 72 h **(d)** methanol induction. For each construct and time point a normalization factor was calculated **(e)** by dividing the indicated eGFP value by the dTomato values.

Experimental outline and extended discussion

Due to different maturation times, quantum yields, stabilities and signal amplification by the fluorescence spectrometer used, the relative fluorescence measurements obtained from eGfp and dTom are not directly comparable. It would also be possible that FRET (Förster resonance energy transfer) is leading to biased results. We designed a set of controls and determined a normalization factor between the two FPs. Therefore a set of promoters was cloned in forward and reverse orientation between the two FPs. Subsequently the eGfp and the dTom signals of the same side were compared. We included the monodirectional state of the art *AOX1* and *GAP* promoters and three bidirectional histone promoters. We also included controls vectors with only a single FP present and cloned the *HHX2* promoter in both orientations into these vectors (the second FP was omitted and the promoter directly adjacent to the transcriptional terminator). These controls were performed to check for effects of coproduction of two FPs vs. production of a single FP and possible FRET effects.

Since we characterized constitutive and methanol inducible promoters, we compared the reporter fluorescence obtained from growth on glucose and different time points of growth on methanol (glycerol was also tested, but yielded similar results to glucose [data not shown]).

The normalization factors calculated from the different promoters (panel e) were in good agreement for each single time point measured. However, the mean value of the ratio/normalization factor for growth on glucose (a) was lower than for growth for 24 hours of methanol (b). When the cells were grown for a 48 h (c) and 72 h (d) on methanol, the normalization factors leveled off at similar values as on glucose. We assume that these effects are evoked by different maturation times of eGfp and dTom; as the eGfp variant was selected for improved folding: After 60 h growth on glucose both proteins have folded and accumulated, but after 24 h on methanol eGfp is folding faster than dTom resulting in a higher eGfp/dTom fluorescence ratio. After 48 h and 72 h enough time has passed to allow dTom folding, resulting in a similar ratio as on glucose. So for every measurement time point the respective normalization factor has to be used. For the normalizations shown in the main manuscript and the supplementary materials, the values of 60 h growth on glucose and 48 h methanol induction were used. The reporter fluorescence remained linear at increased levels up to at least three copies of a plasmid bearing the strong methanol inducible *AOX1* promoter (data not shown).

We had initially also tested alternative combinations of FPs and variants of Tomato (data not shown), and found the eGFP and dTom combination most suitable.

Flow cytometry measurements (*e.g*. FACS) provide more detailed information on the cell population measured, whereas the fluorescence plate reader measurements performed here give only a cumulative signal of the entire population. However, the FACS machines available to us did not provide the correct filter sets required for clearly discriminating the signals of eGFP and dTom. The high throughput characterization of the 168 *P. pastoris* BDPs was also rendered possible by the availability of monochromator based 96 well fluorescence microtiter plate format. Notably also several filter based fluorescence plate readers considered did not provide by default the suitable filter sets to discriminate eGfp and dTom fluorescence. We performed these extensive controls to ensure reliability of our plate reader measurements.

S 4

S 4: List of *P. pastoris* nBDPs tested, sBDPs generated, BTTs tested, *S. cerevisiae and S. pombe* promoters and detailed carotenoid pathway assemblies and primers used. The respective information is provided in different sheets of the excel file.

S 5

S 5: Characterization of the natural bidirectional *P. pastoris DAS1/DAS2* promoter, deletion variants of P_{DAS1} and P_{DAS2} and regulatory elements selected form literature studies on the *AOX1* promoter.

A) Genomic organization of the *P. pastoris DAS1* and *DAS2* locus (based on Figure 2 of Vogl and Glieder [37]) and the promoter lengths tested in this study. Most promoter lengths were tested with a single fluorescent protein (eGFP, indicated by single arrows), a subset also with two fluorescent proteins (dTom and eGFP, double arrows). The SbfI site in the 5' end of $P_{DAS1\text{-}1000}$ and $P_{DAS2\text{-}1000}$ was used for cloning the deletion variants outlined below and did not affect expression.

B) Reporter gene fluorescence measurements of the promoters shown in panel A. Fluorescence was measured at the respective wavelengths after 48 h of methanol induction, for dTom the normalization factor determined in S 3 was used.

C) Schematic overview on deletion studies on $P_{DAS1\text{-}1000}$ for the generation of variants with altered expression (panel E and Fig. 2c) and selected CRMs used for hybrid promoter design (Fig. 2d and S 5). The deleted regions termed D1 to D8 were selected based on sequence similarities to the promoters of methanol regulated *P. pastoris* genes (*DAS1/DAS2*, *AOX1*, *AOX2*, *FLD1*, *FGH1*, *FDH1* and *DAK1*). Similar stretches from pairwise alignments using ClustalO and LALIGN are shown. Stretches appearing multiple times were selected for the deletions.

D) Same as C) for $P_{DAS2\text{-}1000}$. Binding sites of the methanol master regulator Mxr1 reported by Kranthi *et al*. [61] are depicted.

E) Effect of single deletions (top panel *DAS1*, bottom panel *DAS2*)

F) *AOX1* CRMs used shown schematically, highlithging previously deleted regions by Xuan *et al*. and Harnter *et al*. [62,63]

Extended discussion

Selection of deleted regions

Deletion studies of promoters have been used in various organisms to identify regulatory regions and to generate variants with altered expression levels applicable as promoter library for fine tuning gene expression [62,64]. Either systematic deletions are performed (i.e. adjacent fragments [63,65]) or semi-rational considerations (such as the prediction of transcription factor binding sites [62]) applied, as exemplified by studies on the *P. pastoris AOX1* promoter [62,63,65].

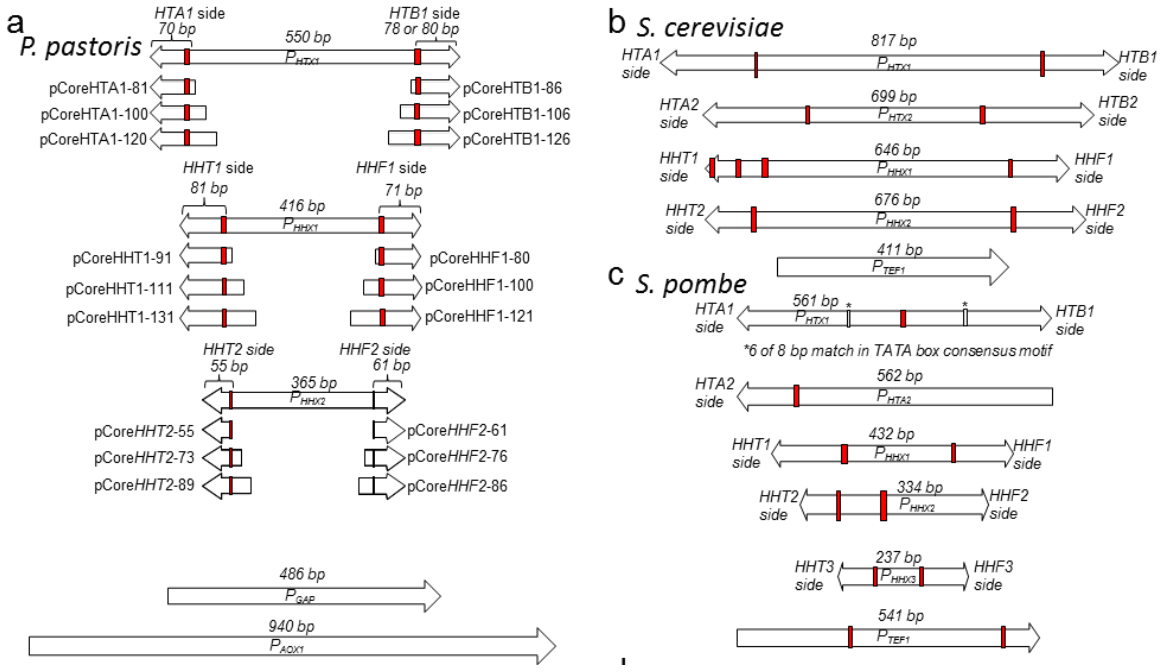
Here we used a different approach to select relevant regions for deletion in *DAS1* and *DAS2* promoters (although systematic deletion studies or TFBS predictions would likely yield similar results). Based on the recent finding that several promoters of the *P. pastoris* methanol utilization pathway are similarly regulated [36], we reasoned that this coregulation must be conferred by conserved sequence DNA stretches. Therefore we selected a set of eight methanol inducible and glucose repressed promoters including the *DAS* promoters to search for shared elements (*DAS1*, *DAS2*, *AOX1, AOX2, FLD1, FGH1, FDH1 and DAK1*). However, TFBS may be placed at different positions between promoters. Studies on the *P. pastoris* methanol master regulator Mxr1p [61,66] showed that its binding sites are arranged pairwise over the whole *AOX1* promoter [66], whereas they are generally closer together in the *DAS2* promoter [61] (reviewed in [37] and compare the Mxr1 binding sites in S 5D and F). In addition yeast TFBS are often short and degenerate exemplified by the Mxr1 consensus sequence CYCCNY (N = any base, Y = C or T) [61].

Performing a multiple sequence alignment of the eight MUT promoter sequences mentioned above using Clustal Omega [67] in standard settings (http://www.ebi.ac.uk/Tools/msa/clustalo/) did not show clearly conserved regions (data not shown). Therefore we performed pairwise comparisons of the *DAS1* and *DAS2* promoters with the other promoters, including LALIGN analysis in the standard settings (http://www.ch.embnet.org/software/LALIGN_form.html, LALIGN is suitable to identify local sequence similarities between two sequences). Sequences appearing multiple times were selected for deletion.

Effects of the single deletions

Several deletion variants showed up to 1.33 fold increased expression compared to the full length wildtype promoters (1000 bp length) suggesting either removal of repressor binding sites or beneficial effects from rearranging the spacing. Most strikingly, deletions D6 and D7 in the *DAS1* promoter led to strong decrease in expression (17 and 37 % of unmodified control), suggesting loss of a major activating region. Deletion of several regions in P_{DAS2} also had a negative impact on reporter fluorescence, however not as drastically as in P_{DAS1} (62 % of 1000 bp unmodified control).

S 6



S 6: *P. pastoris* (a), *S. cerevisiae* (b) and *S. pombe* (c) histone promoters used in this study and positions of regulatory elements including strong MDPs as reference. In case of the *P. pastoris* promoters also the core promoters sequences used for BDP engineering are shown. Extended background information on histone genes and nomenclature is provided below. Sequences matching the yeast TATA box consensus sequence (TATAWAWR [68]) are indicated as red boxes, in the *P. pastoris* promoters the distance of the TAT box to the start codon is indicated. Lengths of the promoters are indicated in bp and strong reference promoters of the respective organisms are shown.

Extended discussion

Background on histone genes

Histones are highly conserved eukaryotic proteins required for packaging DNA into chromatin. DNA is wrapped around a nucleosome consisting of four core histone: histone H2A (HTA), histone H2B (HTB), histone H3 (HHT) and histone H4 (HHF) [38]. HTA+HTB and HHT+HHF form dimers, which are assembled into an octamer [38]. Therefore HTA+HTB and HHT+HHF have to be produced in equal amounts which is typically achieved by their expression from a shared bidirectional promoter. This genomic organization is evolutionary conserved between eukaryotes and has been harnessed for phylogenetic studies of closely related taxa [69]. Additional histones such as the linker histone H1 (HHO) and various variants such as H2A.Z and CenH3 are also present in yeast, higher eukaryotes even contain more HTA and HHT variants [38]. Since these variants are typically not organized bidirectionally, we focused in this study on the four core histones (also referred to as 'canonical' [70] or 'replication dependent' [71] histones).

The function, structure and involvement in gene regulation of histones and their modifications were extensively investigated in various model organisms. Histone promoters were also studied, as their expression is tightly linked to the cell cycle [70,71] and been applied to drive heterologous gene expression in fungi [72,73] and plants [74]. However, the histone promoters were used as monodirectional promoters in these studies and not evaluated in terms of their bidirectional potential. As *P. pastoris* histone promoters showed strongest bidirectional expression of all housekeeping genes tested (Fig. 1c), we evaluated also *S. cerevisiae* and *S. pombe* histone promoters for bidirectional activity.

Histone stem loop (HSL) 3' ends in metazoans

Metazoan replication dependent histone mRNAs feature unusual 3' ends: Unlike most RNA polymerase II transcripts, they do not contain poly(A) tails, but a unique, conserved histone stem loop (HSL) structure [70]. The HSL is relevant for unique histone mRNA properties such as stability and cell-cycle associated degradation [71], but also involved in

translation initiation, fulfilling in this respect a similar function as the poly(A) tail [75,76]. In contrast to metazoans, yeast (and most other non-metazoans) histone mRNAs are poly-adenylated ([75] and references therein).
We used canonical transcriptional terminators containing poly(A) signals for the characterizations of all the bidirecitonal histone promoters reported in this work.

Nomenclature used in this study

We based the nomenclature used in this study on the gene names in *S. cerevisiae* where eight core histone genes are known, organized in pairs in four genomic loci (*HTA1+HTB1*, *HTA2+HTB2*, *HHT1+HHF1*, *HHT2+HHF2*) [38]. We applied the same systematic nomenclature to *P. pastoris* and *S. pombe* histone genes, where in part no three letter gene identifiers have been assigned (organisms discriminated by the prefixes *Pp*, *Sc* and *Sp*). When referring to the locus consisting of both genes we suggest the abbreviations *HTX* for *HTA+HTB* and *HHX* for *HHT-HHF*.

P. pastoris contains only a single *HTA+HTB* locus (*PpHTX1*) pair and two *HHT+HHF* loci (*PpHHX1*, *PpHHX2*) [Concerning the two histone *HHT+HHF* loci, we named the locus on chromosome one *HHX1* and the locus on chromosome two *HHX2*]. *S. pombe* contains three *HHT+HHF* loci (*ScHHX1*, *ScHHX2*, *ScHHX3*), one *HTA+HTB* locus (*ScHTX1*) and an additional single *HTA2* gene.

Reference promoters used and histone promoter lengths

Controls of strong constitutive promoters in the respective organisms were included to assess the strength of the bidirectional histone promoters. In *P. pastoris* the *GAP* promoter is the most commonly used promoter for constitutive expression and the *AOX1* promoter used for inducible expression [37]. In *S. cerevisiae* we used the promoter of the *translation elongation factor 1* gene (P_{ScTEF1}) previously characterized as a strong constitutive promoter [28,77]. Also in *S. pombe* the promoter of the *TEF1* gene (P_{SpTEF1}) previously reported as a strong constitutive promoter [78] was used. All promoters were seamlessly fused to the reporter gene, providing the natural context up to the start codon without any additional RE sites.

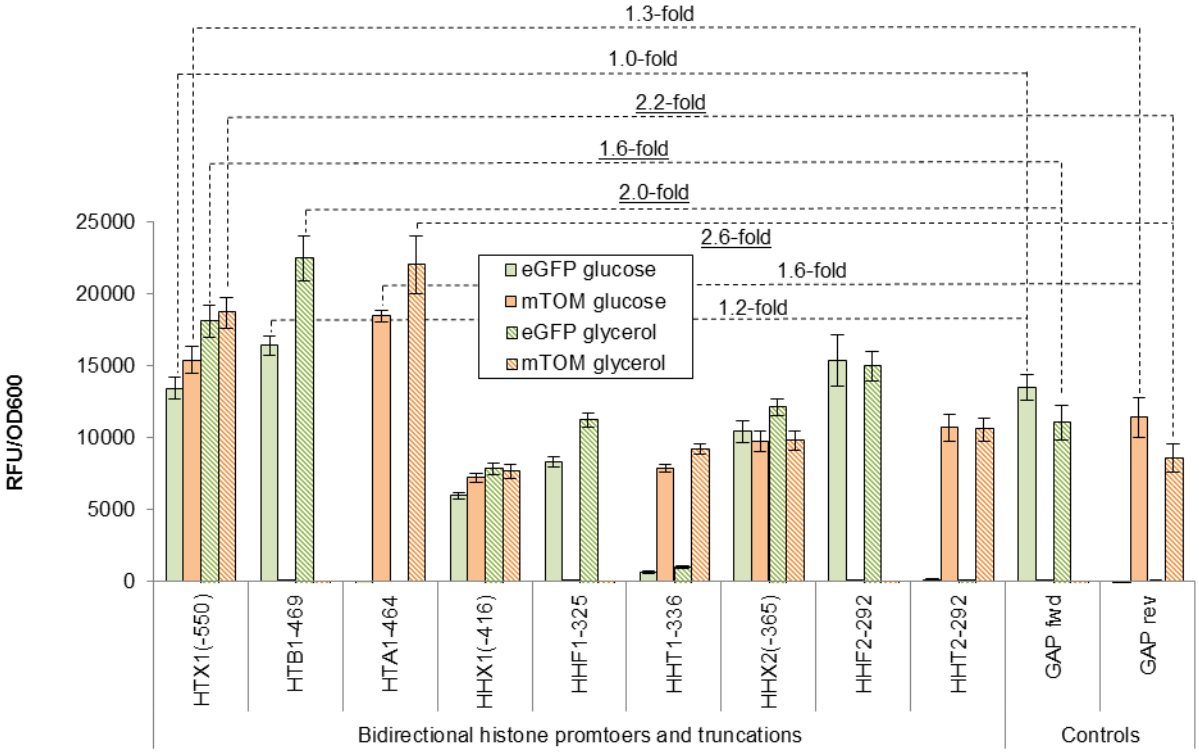
Bidirectional histone promoters condense the regulatory elements needed for the strong bidirectional expression on short sequences compared to the monodirectional promoters. In most cases the monodirecitonal controls were as long or at similar same length as the bidirectional promoters (Panel a to d). Still both bidirectional sides were giving on both sides expression levels comparable to the MDPs (Fig. 1). This is reflected in higher expression efficiency (expression strength per promoter length [in this study normalized fluorescence per bp]), calculated for *P. pastoris* promoters in Fig. 2e.

TATA boxes and short core promoters/5'UTRs in yeast histone promoters

There are two groups of core promoters in yeast: 1) TATA box containing promoters and 2) TATA-less promoters [41]. The TATA box is bound by TATA-binding protein (TBP) required for initiating transcription, in TATA-less promoters alternative factors are needed for transcriptional initiation. TATA box containing promoters are typically tightly regulated, e.g. for stress response. TATA less promoters are rather constitutively active [41]. Strikingly all histone promoters from the three yeasts contain clear TATA box motifs (red boxes in figure), which is in line with their tight cell cycle associated expression [38].

The distance of the TATA boxes relative to the start codon is in most promoters remarkably short. In *P. pastoris* the TATA boxes are 55 to 81 bp away from the start codon, compared to 160 bp in case of the well-studied P_{AOX1} [79] (panel a). We used the TATA boxes as a hallmark for determining the core promoter length. Short core promoters are desirable tools for promoter engineering as they can be simply provided on a PCR primer. The methanol regulated *P. pastoris* promoters used as parts repository for synthetic BDP design contain generally also TATA boxes [36]. To this end we used the *P. pastoris* histone core promoters for the following design of sBDPs.
Note that the core promoter sequences we are using contain also the 5' untranslated region of the natural histone mRNAs, as it is difficult to functionally separate the core promoter and 5'UTR, as also sequences downstream of the transcription start site (TSS) may affect its efficiency [79].
We used the TATA box motif as hallmark for core promoter/5'UTR selection and tested in part different lengths of the promoters (panel a and extended discussion in S 8 and S 9).

S 7



S 7: Bidirectional *P. pastoris* histone promoters show increased expression on glycerol compared to glucose, outperforming the strongest known constitutive promoter (P_{GAP} [37]**) up to 1.6-fold. This effect is even more pronounced for truncated variants showing only expression from one side (up to 2.6-fold increased expression compared to P_{GAP} on glycerol).** The strains were grown for 60 h in quadruplicates (mean value and standard deviations shown) in 96-well plates for 60 h in parallel in media containing 1% (w/v) glucose or glycerol. Reporter protein fluorescence was measured and dTomato signals normalized as outlined in S 3. Fold differences are shown for selected comparisons.

Most commonly the promoter of the glyceraldehyde 3-phosphate dehydrogenase ($_{PGAP}$) is used to drive constitutive gene expression in *P. pastoris* [37]. P_{GAP} requires cultivation on glucose as carbon source to reach highest expression, mRNA levels on glycerol decreased to approximately two thirds [37,80]. However, it was shown that *P. pastoris* reaches higher specific growth rates and higher amounts of biomass on glycerol than on glucose [81,82]. Histone promoters are known to be cell-cycle regulated [38], which would results in our experimental setup in growth associated expression. We expected that *P. pastoris* histone promoters should also show strong expression on glycerol. To this end we compared reporter gene fluorescence of the histone promoters and P_{GAP} grown on glucose and glycerol. The strongest histone promoter (P_{HTX1}) reaches on glucose similar expression levels as P_{GAP}. On glycerol, P_{GAP} expression decreases, while some histone promoters show elevated reporter protein fluorescence compared to glucose. P_{HTX1} even outperforms P_{GAP} 1.6-fold. The reporter protein fluorescence is normalized per biomass (determined by OD_{600} measurements) to rule out effects of different biomass yields between the carbon sources.

Note that deletion/truncation of the core promoter from one side of the BDP increases in some cases the expression from the other side. General implications thereof and systematic characterizations are shown and discussed in S 9c. These truncation variants were also included here, to demonstrate that the beneficial effect of growth on glycerol and the truncations are additive resulting in up to 2.6-fold higher expression than the P_{GAP} control.

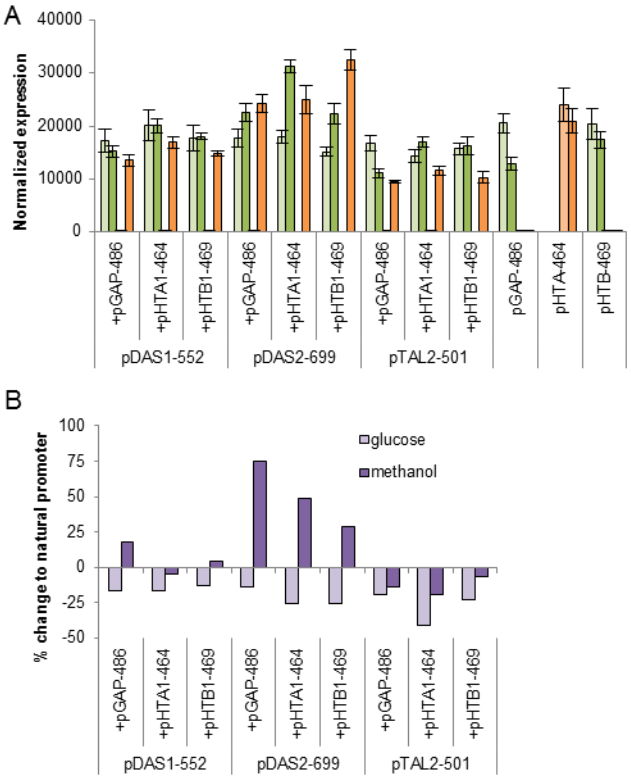
S 8

S 8: Design considerations for bidirectionalizations and fusions promoters and synergistic and antagonistic effects observed for fusions of inducible and constitutive promoters.

Combinations of P_{DAS1} and P_{DAS2} deletions

We generated methanol inducible BDPs with varying expression ratios by combing different monodirectional variants (Fig. 2c). We combined deletions showing increased reporter gene fluorescence in the monodirectional context (S 5E) into improved BDPs (*e.g. DAS2-D8+DAS1-D2*, *DAS2-D6+DAS1-D2D5*) BDPs with decreased expression (*DAS2-D5+DAS1-D6*, *DAS2-386+Das1-D6*) or altered ratios between both sides (*DAS2-D8+DAS1-D6*) were generated by fusing weaker monodirectional variants. $P_{DAS2\text{-}386}$ and $P_{DAS\text{-}261}$ are additional truncated variants to reduce expression from the *DAS2* side (since monodirectional P_{DAS2} deletions had only shown a decrease to 62 % of the unmodified control).

Synergistic and antagonistic effects observed for fusions of inducible and constitutive promoters



A) Comparison of bidirectional P_{GAP}, $P_{HTA1\text{-}464}$ and $P_{HTB1\text{-}469}$ fusions to methanol regulated promoters ($P_{DAS1\text{-}552}$, $P_{DAS2\text{-}699}$ and $P_{TAL2\text{-}501}$) with the single MDPs The same data shown in Fig. 2b and S 7 was rearranged to facilitate comparisons.

B) Changes in normalized reporter gene fluorescence of the fusions promoters compared to the single MDPs from panel a were calculated.

Fusions of growth-associated/constitutive *HTA1*, *HTB1* and *GAP* promoters to P_{DAS2} reached on methanol 1.3 to 1.8 fold increased expression compared to the single promoters (S 8). Notably, P_{GAP} is typically downregulated on methanol ([37], S 3), whereas fusions to P_{DAS2} showed increased expression suggesting a transcriptional 'spillover' from the methanol inducible promoters. In consistency with these results, the $P_{DAS2\text{-}699}$ fragment had also given high expression when fused to a core promoter Fig. 2b, underlining the strongly activating effect on upstream fusions.

Fusions of the same growth-associated/constitutive promoters to $P_{DAS1\text{-}552}$ showed less pronounces effects.

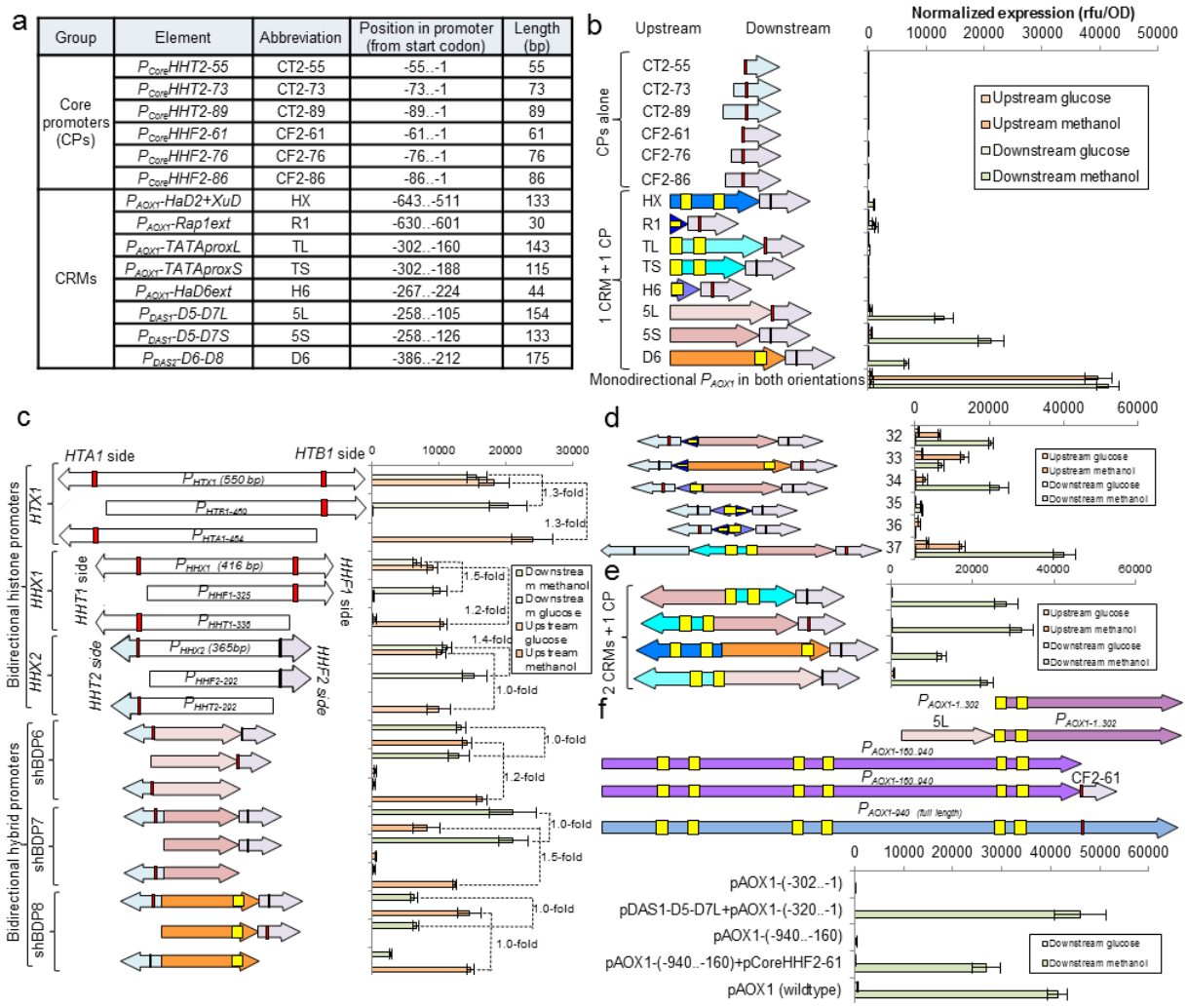
However all promoters fused to $P_{TAL2\text{-}501}$ show decreased expression, on both carbon sources tested. Most strikingly the fusion of $P_{TAL2\text{-}501}$ to $P_{HTA1\text{-}464}$ shows on methanol a 41 % decrease compared to the $P_{HTA1\text{-}464}$ promoter alone, suggesting a moderate repressing effect of the $P_{TAL2\text{-}501}$ sequence.

These results show that fusions of two differently regulated MDPs may interfere, affecting expression strength. Synergistic and antagonistic effects vary even between similarly regulated (methanol inducible) promoters. To this end, the properties of fusion promoters cannot entirely be foreseen and should be tested with reporter genes. However, the synergistic effects can be harnessed to design shorter, more efficient promoters and we expanded this principle for the design of hybrid promoters (Fig. 2d).

S 9

a

Group	Element	Abbreviation	Position in promoter (from start codon)	Length (bp)
Core promoters (CPs)	P_{Core}HHT2-55	CT2-55	-55..-1	55
	P_{Core}HHT2-73	CT2-73	-73..-1	73
	P_{Core}HHT2-89	CT2-89	-89..-1	89
	P_{Core}HHF2-61	CF2-61	-61..-1	61
	P_{Core}HHF2-76	CF2-76	-76..-1	76
	P_{Core}HHF2-86	CF2-86	-86..-1	86
CRMs	P_{AOX1}-HaD2+XuD	HX	-643..-511	133
	P_{AOX1}-Rap1ext	R1	-630..-601	30
	P_{AOX1}-TATAproxL	TL	-302..-160	143
	P_{AOX1}-TATAproxS	TS	-302..-188	115
	P_{AOX1}-HaD6ext	H6	-267..-224	44
	P_{DAS1}-D5-D7L	5L	-258..-105	154
	P_{DAS1}-D5-D7S	5S	-258..-126	133
	P_{DAS2}-D6-D8	D6	-386..-212	175



S 9: Detailed design considerations, supplementary control constructs for bidirectional hybrid promoters in *P. pastoris* and extended discussion.

A) Table on regulatory elements used for the bidirectional hybrid promoter design (see S 6a for histone core promoters, S 5c for P_{DAS1}, S 5d P_{DAS2} and S 5f for P_{AOX1} for illustrations of the elements in the natural promoter context.

B) Reporter protein fluorescence of Histone core promoters alone and combinations of the CRMs with a single core promoter. The *HHX2* core promoter lengths tested alone do not show any expression. Normalized fluorescence measurements after 60 h growth on glucose and 48 h of subsequent methanol induction are shown. The monodirecitonal *AOX1* promoter is include as a control. The experimental cultivation conditions and the P_{AOX1} control apply to all panels.

C) Truncation of nBDPs (P_{HTX1}, P_{HHX1}, P_{HHX2}) and hybrid sBDPs (#6 to 8) on one side leads in 7/12 cases to increased expression on the other side. The nBDPs show the effect more pronounced (5/6) than the sBDPs tested (2/6). The data on the histone promoters is also shown in S 7 in comparison to P_{GAP} and growth on glycerol. Values from growth on glucose are shown for the nBDPs, growth on methanol for the sBDPs. Fold changes of the truncated variant compared to the full length bidirectional promoter are shown.

D) Additional bidirectional hybrid promoter variants not included in Fig. 2d.

E) Additional monodirectional hybrid promoters (combinations of 2 CRMs with 1 core promoter).

F) The first 302 bp of P_{AOX1} do not show any detectable reporter protein fluorescence despite containing two Mxr1p binding sites. Control constructs include the fusion of a P_{DAS1}-*D5-D7L* activating sequence to the 302 core region, the *AOX1* promoter upstream sequence without a core promoter, a fusion of the upstream sequence to the *HHF2-61* core promoter and the full length wild type promoter.

Extended discussion

Selection of CRMs

Various synthetic monodirectional hybrid promoter have been engineered by fusing CRMs to core promoters [64]. We extended this strategy to BDPs, by flanking CRMs with two core promoters in opposite orientation. We used the short *PpHHX2* histone core promoters (S 6 A) successfully applied for bidirectionalization of MDPs (Fig. 2 A, S 8). Six short CRMs (30 to 175 bp) from methanol regulated promoters (P_{AOX1}, P_{DAS1}, P_{DAS2}) were used. Namely, four elements form P_{AOX1} (P_{AOX1}-*HaD2+XuD*, P_{AOX1}-*Rap1ext*, P_{AOX1}-*TATAproxL/S,* P_{AOX1}-*HaD6ext*) and a single element from each P_{DAS1} (P_{DAS1}-*D5-D7L/S*) and P_{DAS2} (P_{DAS2}-*D6-D8*) were used.

CRMs from the *AOX1* promoter were selected based on deletion studies from literature ([62,63,65,83] reviewed in [37]) and binding sites reported for the methanol master regulator Mxr1p [66]. The CRMs of the *AOX1* promoter contain Mxr1 (zinc finger transcription factor and master regulator of MUT genes in *P. pastoris*) binding sites and deletions within these regions strongly effected expression [62,63,66]. P_{AOX1}-*HaD2+XuD* is a fusion of D2 region of Hartner *et al*. [62] and region D of Xuan *et al*. [63] containing two Mxr1p binding sites [66]. P_{AOX1}-*Rap1ext* is a putative TFBS reported by Hartner *et al*. extended to contain an Mxr1p binding site. P_{AOX1}-*TATAproxL/S* contains two Mxr1p binding sites and several deletions in this region drastically effected expression. Due to its proximity to the TATA box we refer to this CRM as 'TATAprox'. P_{AOX1}-*HaD6ext* is the region D6 characterized by Hartner *et al*. extended to comprise the adjacent Mxr1p binding site.

CRMs from P_{DAS1} and P_{DAS2} were selected based on deletion studies performed in frame of this work (S 5) and Mxr1 binding sites reported in P_{DAS2} [37,61]. Variants with deletions of the regions D6 to D7 in the *DAS1* promoter showed strongly decreased expression, suggesting the presence of a major activating region. We extended this region to include the D5 region and tested it due to its close proximity to the core promoter/TATA box in two lengths (termed P_{DAS1}-*D5-D7L* and P_{DAS1}-*D5-D7S*). Deletions in the *DAS2* promoter had not shown as drastic effects as in the case of P_{DAS1}, however deletion of region D7 had notably reduced expression. We extended this sequence stretch to the adjacent elements resulting in P_{DAS2}-*D6-D8.*

CRMs adjacent to the core promoter/TATA box were in part tested in different lengths (P_{AOX1}-*TATAproxL/S* and P_{DAS1}-*D5-D7L/S;* 'L' for long, 'S' for short) to probe for carryover effects of the core promoter. The long variants of these CRMs were extended up to the TATA box (fusions of these CRMs with core promoters reconstitute the natural position of the TATA box in both core promoter and CRM).

Truncation of BDPs on one side leads in 7/12 cases to increased expression on the other side.

We had noticed in the deletion and truncations studies of P_{HHX2} (Fig. 1f) that removal of the core promoter from one sides increases expression from the opposite side. To confirm this effect we also truncated the core promoters from the histone promoters P_{HTX1} and P_{HHX1} and synthetic bidirectional constructs shBDP6 to shBDP8 (panel C). For P_{HTX1} we removed the 86 bp long *HTB1*-core promoter ($Pcore_{HTB1-86}$, S 6) resulting in a truncated P_{HTA1} promoter of 464 bp ($P_{HTA1-464}$). *Vice versa* the core promoter removal/truncated promoter pair on the other side of P_{HTX1} is $Pcore_{HTA1-81}$/$P_{HTB1-469}$. The pairs for P_{HHX1} are $Pcore_{HHT1-91}$/$P_{HHF1-325}$ and $Pcore_{HHF1-80}$/$P_{HHT1-336}$. For P_{HHX2} the truncations F1 and T1 already shown in Fig. 1f were used. For shBDP6 to shBDP8 we tested the CRMs flanked by two core promoters simultaneously and also a single core promoter on each side (panel C).

In 7 of 12 cases removal of a core promoter increased expression from the other side (up to 1.5-fold). This may be caused by transcriptional or translational effects: The two core promoters could be competing for RNA polymerase II (RNAPII) and general transcription factors. Alternatively transcription could be unaffected and solely the protein level affected. Producing two FPs at the same time may require more energy from the cell than expressing a single FP. If the burden of a second protein is removed, translation of the single one may be stronger.

For P_{HHX2} we assume that the effect is transcriptional and not translation: In frame of the normalization work to compare the two FPs used (S 3), we created constructs of the full length *HHX2* promoter flanked by two FPs and one FP and one transcriptional terminator (in both orientations: transcription terminator on the 5' end [directly next to the promoter] and a FP gene on the 3' end or *vice versa* [terminator on the 3' end and FP gene on the 5' end]). Expression of these constructs was not increased compared to coexpression of both genes simultaneously (S 3). Therefore we conclude that *P. pastoris* cells have sufficient energy to produce two FPs at high levels and translation is not the limiting factor. Truncating the *HHX2* core promoters (thereby lacking binding sites for general TFs and RNAPII) however increases expression on the other side (panel C), hinting that the effect is on the transcriptional level.

Transcriptional 'spillover' in hybrid promoter

In hybrid BDPs created by fusions of the growth-associated $P_{HHT2\text{-}T146}$ to glucose repressed/methanol regulated CRMs, similar antagonistic/synergistic 'spillover' effects as seen with some fusion promoters (S 8) were observed (hybrid promoters #9-14 in Fig. 2d). If the methanol regulated CRMs were not fused to any additional sequences (S 9a) or other methanol regulated CRMs (Fig. 2d), they were tightly repressed on glucose. However, if they were fused to the growth associated active $P_{HHT2\text{-}T146}$, they showed already on glucose clear reporter protein fluorescence. This effect depended on the CRM, but suggests that the growth associated expression of the $P_{HHT2\text{-}T146}$ to glucose repressed/methanol regulated CRMs partially alleviates the repression.
We did not observe so strong effects with fusion promoter (Fig. 2b), presumably as these promoters were considerably longer and regulatory regions not directly adjacent. The use of insulator sequences may also abolish the spill-over in hybrid promoters consisting of $P_{HHT2\text{-}T146}$ fusions to glucose repressed/methanol regulated CRMs.

The hybrid promoter assemblies and additional controls suggest different promoter architectures for P_{AOX1} and P_{DAS1}

In all studies of the hybrid promoters, the CRMs close to the TATA box and the 3' end of the *AOX1* promoter did not show any activity (P_{AOX1}-*TATAprox*). This is surprising as a CRM from the *DAS1* promoter (P_{DAS1}-*D5-D7*) stemming from a similar 3' region close to the TATA box, does show strong activation in all contexts tested (Fig. 2d and S 9).
These effects could be caused by an incompatibility of the *AOX1* CRM with the histone core promoters or indeed a lack of activating sequences.
We performed additional controls in S 9f, by testing the natural context of the *AOX1* CRM fused to the *AOX1* core promoter ($P_{AOX1\text{-}1..302}$). This part alone did not show any detectable reporter fluorescence. Fusion of the entire sequence upstream of the TATA box of the *AOX1* promoter ($P_{AOX1\text{-}160..940}$) to the histone core promoter of the *HHF2* gene, showed expression matching the wild type *AOX1* promoter. The negative control of the $P_{AOX1\text{-}160..940}$ sequence alone does not show any expression. These experiments rule out that the problem is arising from the fusion to the histone core promoter CRM. Fusion of the P_{DAS1}-*D5-D7* CRM to the $P_{AOX1\text{-}1..302}$ sequence restores P_{AOX1} wild type like expression levels.
It indeed seems that the TATA proximal region of P_{AOX1} does not have any activating function whereas the similar region in P_{DAS1}-*D5-D7* leads to strong activation. It is also puzzling, that the TATA proximal region of P_{AOX1} contains two experimentally confirmed binding sites for Mxr1p, a master activator for methanol inducible genes in P. pastoris [66]. The full length *AOX1* and *DAS1* promoters are however similarly regulated (tight glucose repression, strong methanol induction Fig. 1c). These results highlight the variability and flexibility of yeast promoters, achieving similar regulation by vastly different promoter architecture.

S 10

S 10: Extended discussion on applications of the BDP library (Fig. 3).

Extended discussion

Dual gene coexpression

Selection of gene pairs and BDPs

Cytochrome P450 enzymes are monooxygenases catalyzing typically hydroxylation reactions. Usually a cytochrome P450 reductase (CPR) is required for electron transfer from the cofactor NADPH to the CYP. The two genes coding for CYP and CPR need to coexpressed at a suitable ratio and were

therefore used as a model system for applying our library of BDPs. *P. pastoris* was shown to be a favorable expression host for CYP production [84]. We used human CYP2D6, one of several human CYPs involved in drug metabolism in liver microsomes [85], and its associated reductase (codon optimized for expression in *P. pastoris*).
Previously only the monodirectional *AOX1* promoter either on separate vectors [84] or on the same plasmid [50] has been used to drive both CYP and CPR expression in *P. pastoris*. Here we tested a subset of BDPs from the library to optimize CYP and CPR coexpression. We applied strong, differently regulated BDPs and omitted weaker ones since previous experiments had shown that CYP high CYP activities were favored by P_{AOX1} driven expression, even in multiple copies integrated in the genome. The natural bidirectional methanol inducible *DAS1-DAS2* promoter, a fusion of P_{AOX1} and P_{GAP} (combination of inducible and constitutive expression), a fusion of P_{AOX1} and P_{CAT1} (inducible and derepressed/inducible expression), a fusion of P_{CAT1} and P_{GAP} (derepressed/inducible and constitutive) were tested in both orientations.

We also applied BDPs to fine-tune chaperone/foldase coexpression, a common strategy to increase yields of proteins that are difficult to fold. We aimed to increase the secretion of the enzymes CalB (*Candida antarctica* lipase B, Fig. 3b) and HRP (horseradish peroxidase, Fig. 3c) by coexpressing protein-disulfide-isomerase (PDI). PDI helps in the formation of correct disulfide bonds and has also chaperone activity assisting in correct folding in the endoplasmic reticulum (ER). We reasoned that HRP and CalB may benefit from PDI coexpression since they contain four [86] and three [87,88] disulfide bonds respectively.

We also included a terpenoid production application. Many plant derived pharmaceuticals such as the anticancer drug taxol (paclitaxel) are not protein based, but highly complex chemicals derived from isoprenoid/terpenoid structures. Taxol is only found in the bark of the pacific yew tree and two to four trees have to be harvested to allow treatment of a single patient [89]. Taxadiene is the first committed precursor of taxol, requiring two enzymes for synthesis: geranylgeranyl diphosphate synthase (GGPPS) and taxadiene synthase (TDS) [42]. In previous studies in *E. coli* [42], it had been shown to be critical to find the optimal ratio of expression. We aimed to balance expression of GGPPS and TDS by using strong and differently regulated promoters.

Discussion of results and controls

Although all BDPs used showed strong expression with FPs, activity levels obtained with different gene pairs varied considerably. Comparing the lowest and highest activity for CYP2D6 and HRP coexpression gives a 5.2 fold difference ($P_{DAS1-DAS2}$ vs. $P_{CAT1+GAP}$). For CalB there is a 7.8 fold difference ($P_{CAT1+AOX1}$ vs. $P_{HHT2-HHF2}$ [comparing the methanol induced values]), for HRP there is a 4.9 fold difference ($P_{AOX1+GAP}$ vs. $P_{HHT2-HHF2}$). Yields in taxadiene production varied even approximately 50-fold ($P_{DAS2-DAS1}$ vs. $P_{GAP-CAT1}$). Also the best and worst promoters varied between the constructs, demonstrating that testing various promoters for optimal coexpression is a feasible strategy.

CYP2D6 and CPR activity was measured using a whole cell bioconversion assay with the substrate 7-methoxy-4-(aminomethyl)-coumarin (MAMC). We used a strain expressing CYP2D6 and the CPR from two monodirectional *AOX1* promoters as control. Constitutive promoters (fusion of *GAP+TEF1* promoters in both orientations) did not show any detectable expression, neither after growth on glucose (data not shown) nor on methanol (Fig. 3a). We assume that constitutive expression of these two ER localized proteins may exert too much stress on the cells, leading possibly to UPR (unfolded protein response upregulation) and degradation by ERAD (ER-associated protein degradation). Since

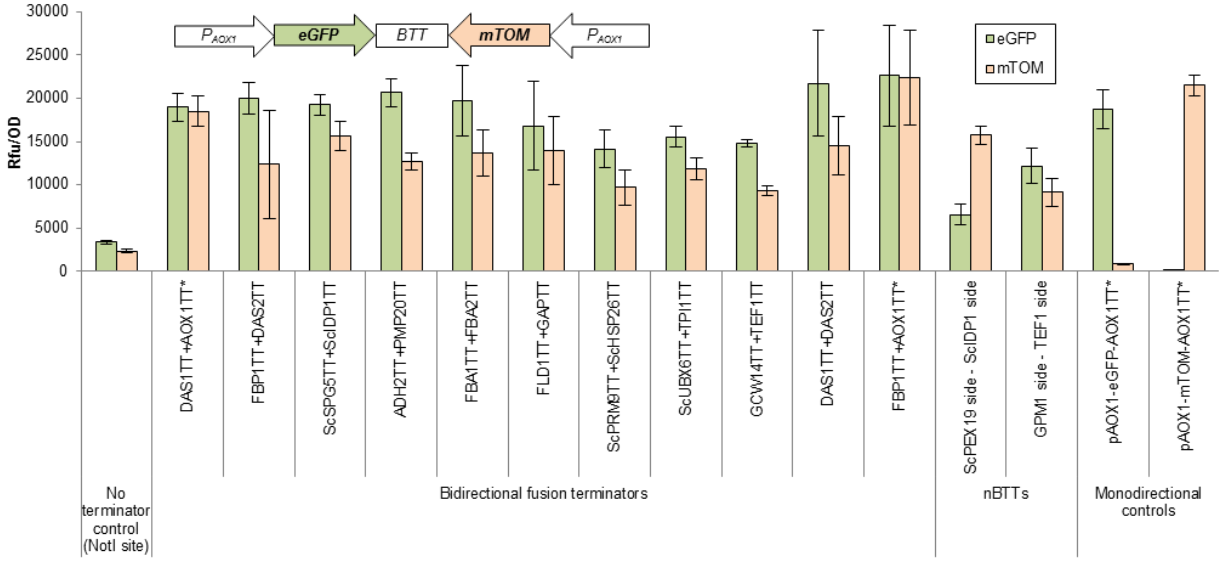
the constitutive $P_{GAP+TEF1}$ fusion had not provided any expression, we did not test histone promoters, since their cell cycle [38] (and thereby growth-associated expression) is similar to constitutive expression.

Concerning methanol inducible expression shown in (Fig. 3a), the construct bearing the natural bidirectional *DAS1-DAS2* promoter gave highest activity, in case the *DAS1* weaker side was expressing CYP2D6 and the stronger *DAS2* side expressing CPR. There is general a trend, that higher activities are achieved from the BDPs when the stronger side is pointed towards expression of the CPR (e.g. in $P_{GAP+AOX1}$, the P_{AOX1} side is stronger on methanol [Fig. 2b], in $P_{CAT1+GAP}$, the P_{CAT1} side is stronger on methanol).

For taxadiene production, we noticed about 100-fold decreased transformation rates if the GGPPS gene was under control of a constitutive promoter. The few colonies growing on the transformation plates did not show any detectable taxadiene production. We assume that strong constitutive GGPPS expression is lethal; the few colonies obtained from transformation of constitutive GGPPS expression integrated likely only the resistance marker cassette or an inactive GGPPS gene (data not shown).

All inducible promoters showed detectable taxadiene levels with most yielding less than 1 mg/L. However, the $P_{GAP-CAT1}$ promoter yielded 6.2 mg/L. This yield was achieved by transformation of a single plasmid and standard cultivation conditions. Even in heavily modified *S. cerevisiae* strains only 8.7 mg/L were achieved [43], highlighting the potency of transcriptional fine tuning using BDPs and *P. pastoris* as a host for pathway expression and terpenoid production. We presume that the exceptionally high yield of this construct is mostly attributable to the use of P_{CAT1} to drive expression of the GGPPS gene. Also in the second best construct ($P_{AOX1-CAT1}$), the GGPPS was under control of P_{CAT1}. P_{CAT1} is a derepressed promoter, i.e. expression is starting once the glucose in the media is depleted. P_{CAT1} is further strongly induced by methanol [36]. So in the top producing strain, the GGPPS gene was at first repressed, partially activated in the depressed phase and fully activated on methanol. Therefore not only the ratio/strength of the promoters, but also the regulatory profile appears to be an important component, that can be easily optimized using the versatile library of BDPs.

S 11



S 11: Bidirectional transcription terminators (BDTs) required for the assembly of bidirectional multi gene coexpression relieve expression loss associated with transcriptional collision. A reporter construct for testing bidirectional transcription termination was assembled (small inlet). Two genes coding for the fluorescent proteins eGfp and dTom were cloned in convergent orientation, separated by a stuffer fragment (not shown). Two *AOX1* promoters were used to drive expression of the reporter genes. The stuffer fragment was cleaved out and replaced with bidirectional terminator sequences. Monodirecitonal terminators (MTTs) were combined into bidirectional fusion terminators and two putative natural BDTs were tested. We included a control lacking termination sequences and bearing solely a NotI restriction site (the stuffer fragment was flanked by two NotI sites, after cutting out the stuffer fragment, the backbone was self-ligated, resulting in eGFP and dTom directly adjacent only separated by the NotI site). In addition control constructs were included, where only a single *AOX1* promoter and a single FP are present. These constructs contain the AOX1* terminator. AOX1TT* denotes the *AOX1* terminator sequence used by Vogl et al., unpublished results. Mean values and standard deviations of fluorescence measurements after pre-growth on glucose followed by methanol induction of biological quadruplicates are shown.

Extended discussion

Background

If multiple genes are to be coexpressed in a single expression cassette using bidirectional promoters, the assembly of more than three genes requires arranging genes in convergent orientation (see the bidirectional assembly in Fig. 3d or S 1d (the transitions between G2+G3, G4+G5 and G6+G7 require a BDT). Lack of a terminator in this context results in transcriptional collision [90] as polymerases transcribing opposite DNA strands in convergent orientation stall upon collision in *S. cerevisiae* [91]. Therefore we aimed to identify efficient BDTs for *P. pastoris*.

Selection of the terminators

We have tested 20 monodirectional *P. pastoris* terminators in previous studies (Vogl *et al*., unpublished results). Also heterologous terminators from *S. cerevisiae* were active in *P. pastoris* and also included in this study. Here we combined selected MDTs into 11 bidirectional fusion terminators by linking them in convergent orientation. This strategy is similar to the design of bidirectional fusion promoters outlined in Fig. 2b,c. Alternatively, natural BDTs (nBDTs) could be used as the *P. pastoris* genome harbors 1461 putative BDTs from genes in tail to tail orientation (Fig. 1b). The *P. pastoris TEF1* terminator (T_{TEF1}) and the *S. cerevisiae* T_{IDP1} appeared to be such nBDTs. In the previously used terminator sequences (approximately 250 bps), the C-terminus of 'Glycine dehydrogenase [decarboxylating], mitochondriala' is annotated at 204 bp in T_{TEF1} and the *ScPEX19* gene ends within the previously used T_{ScIDP1} sequence forming a putative 127 bp BDT. We included these two short nBDTs in the study.

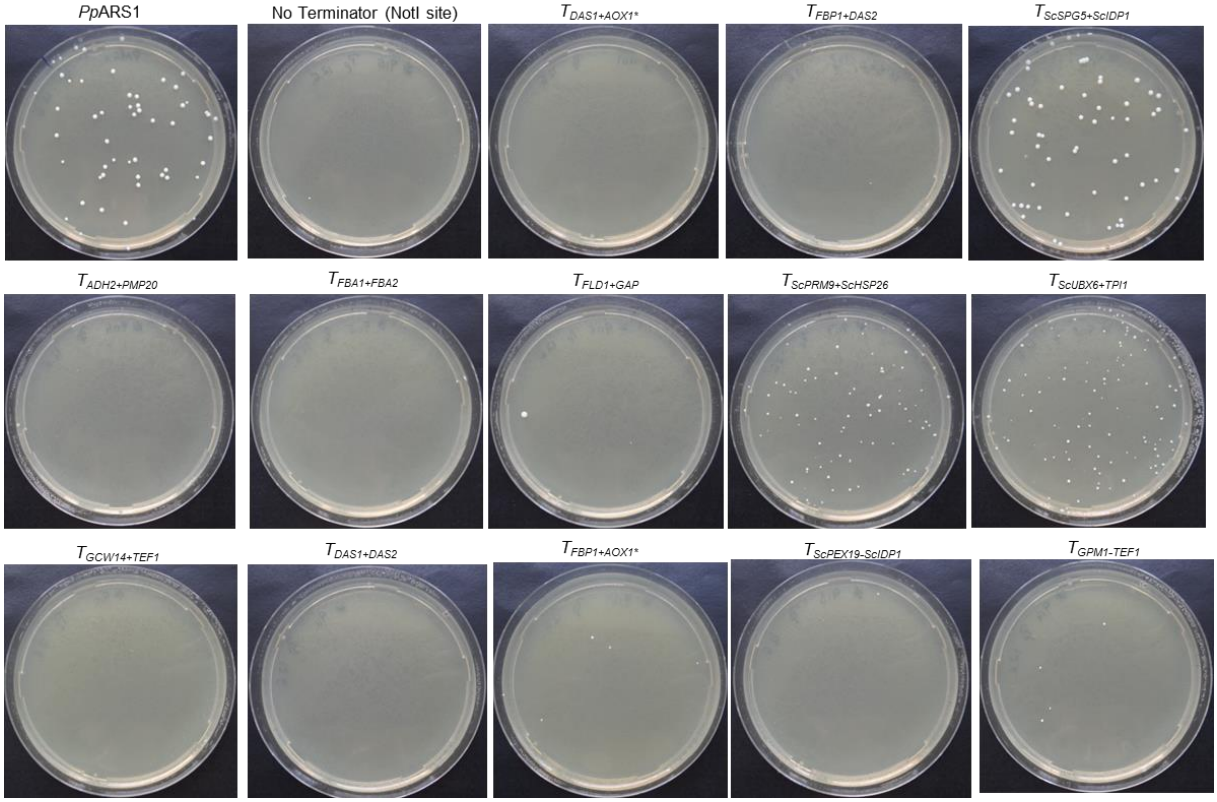
Results

The bidirectional terminators were cloned seamless (*i.e*. maintaining the natural transition between stop codon and terminator without any additional restriction sites) in a reporter vector containing two fluorescent reporter proteins in convergent orientation (see caption). Complete lack of a termination signal in this context (negative control of a NotI site between the reporter genes) resulted in 5.6 to 9.2 fold reduced reporter gene fluorescence (compared to monodirectional single reporter constructs with proper termination signals). These results suggest that transcriptional collision occurs to similar extents in *P. pastoris* as reported in *S. cerevisiae* [90,91], suggesting a relevant issue for bidirectional pathway assembly. Providing either fusion terminators or nBDTs restored reporter protein fluorescence to different extends of the controls. The terminator in the single promoter and single FP controls was T_{AOX1*} the strongest MTT from previous unpublished work. Therefore it is not unexpected that combinations of weaker MTTs reach lower reporter protein fluorescence. All termination signals showed clear improvements towards the no terminator control. $T_{DAS1+AOX1TT*}$, $T_{FBP1+DAS2}$ and $T_{FBP1+AOX1}$ provided on both sides approximately similar reporter fluorescence matching the controls. Also the nBDTs tested showed clear terminator activity. In case of $T_{ScPEX19-IDP1}$, the *IDP1* side is considerably stronger than the *PEX19* side, which gives the weakest reporter fluorescence of all BDTs tested, however still clearly surpassing the no terminator control. $T_{GPM1-TEF1}$ reaches on both sides about similar reporter fluorescence of about half of the single reporter controls. This signal is rather low compared to the best fusion BDTs tested. However the nBDTs tested are with 127 bp and 204 bp considerably shorter than the fusion promoters (approximately 500 bp in length) and may become valuable tools for assembling short bidirectional multi gene expression cassettes. Given the presence of 1461 putative BDTs in the *P. pastoris*

genome and the enrichment of short sequences below 200 bp (Fig. 1b), it should be possible to obtain even more efficient nBDTs.

We were not particularly successful in obtaining flexible nBDPs (Fig. 1c), as they provided only limited combinations of regulatory profiles, cumulative expression levels and ratios. However, we assume that the function of nBDTs is more universal and it should be easier to find suitable nBDTs. To this end, while sBDPs are more flexible and applicable tools for metabolic engineering or synthetic biology than nBDPs, nBDTs appear as sufficient tools for these applications, not necessarily requiring sBDTs.

Autonomously replicating sequence (ARS) function of the terminators generated



In *S. cerevisiae*, transcription termination and autonomously replicating sequence function are associated [92]. In previous unpublished work on *P. pastoris* MTTs we also noticed that some terminators showed ARS function. Terminators with ARS function should be avoided as they may lead to increased background growth and strain instability of episomally replicating sequences (Vogl T. *et al*. unpublished results). Therefore we tested all BDTs reported here for ARS function by transforming 10 ng of the circular plasmids. We included a positive control of a *P. pastoris* ARS sequence [93] cloned into the vector backbone of the *P. pastoris* plasmid used [This control was also included in previous work on MTTs]. The positive control showed pronounced growth of a few dozen colonies. The no terminator/empty vector/negative control (lacking a terminator and containing just a NotI site arising from self-ligating the vector) and most BDTs tested showed no or very few colonies. Few colonies are not a clear evidence for ARS function, as also circular plasmids can integrate into yeast genomes [94]. However circular plasmids show much lower efficiencies than linear DNA providing free ends [94] (which are typically generated for *P. pastoris* transformations by linearizing the plasmids). Interestingly, combinations of *S. cerevisiae* terminators showed clearest ARS function (with $T_{ScSPG5+ScIDP1}$ being somewhat surpassing $T_{ScPRM9+ScHSP26}$ and $T_{ScUBX6+TPI1}$ judging from the colony sizes). These results are in line with characterizations of the MDTs, where T_{ScSPG5} and T_{ScIDP1} had also shown clear ARS function [36]. The monodirectional T_{ScUBX6} had previously also shown termination function. Remarkably, monodirectional versions of T_{ScPRM9} had not shown colonies and $T_{ScHSP26}$ had shown only few colonies that we had not considered ARS function. Combination of these two sequences into $T_{ScPRM9+ScHSP26}$ did however show substantial ARS function, comparable to $T_{ScUBX6+TPI1}$. We recommend therefore testing for ARS function of newly assembled BDTs to avoid issues with ARS background colonies and strain stability.

Chapter 3.5.1
This work was based on the dissertation of Maria Freigassner.

Towards improved membrane protein production in *Pichia pastoris*: General and specific transcriptional response to membrane protein overexpression

Thomas Vogl[1], Gerhard G. Thallinger[2,3,4], Guenther Zellnig[5], David Drew[6], James M. Cregg[7], Anton Glieder[4] and Maria Freigassner[1]

[1] Institute for Molecular Biotechnology, Graz University of Technology, Petersgasse 14/2, 8010 Graz, Austria
[2] Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14/5, 8010 Graz, Austria
[3] Omics Center Graz, Stiftingtalstrasse 24, 8036 Graz, Austria
[4] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14/5, 8010 Graz, Austria
[5] Institute of Plant Sciences, University of Graz, Schubertstrasse 51, 8010 Graz, Austria
[6] Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, United Kingdom
[7] Keck Graduate Institute of Applied Life Sciences, 535 Watson Drive, Claremont, CA 91711, USA

Membrane proteins are the largest group of human drug targets and are also used as biocatalysts. However, due to their complexity, efficient expression remains a bottleneck for high level production. In recent years, the methylotrophic yeast *Pichia pastoris* has emerged as one of the most commonly used expression systems for membrane protein production.

Here, we have analysed the transcriptomes of *P. pastoris* strains producing different classes of membrane proteins (mitochondrial, ER/Golgi and plasma membrane localized) to understand the cellular response and to identify targets to engineer *P. pastoris* towards an improved chassis for membrane protein production.

Microarray experiments revealed varying transcriptional responses depending on the enzymatic activity, subcellular localization and physiological role of the membrane proteins. While an alternative oxidase evoked primarily a response within the mitochondria, the overexpression of transporters entering the secretory pathway had a wide effect on lipid metabolism and induced the upregulation of the UPR (unfolded protein response) transcription factor Hac1p. Coexpression of *P. pastoris* endogenous *HAC1* increased the levels of ER-resident membrane proteins 1.5- to 2.1-fold. Subsequent transcriptome analysis of *HAC1* coexpression revealed an upregulation of the folding machinery correlating with an expansion of the ER membrane capacity, thus boosting membrane protein production. Hence, our study has helped to elucidate the cellular response of *P. pastoris* to the expression of different classes of membrane proteins and led specifically to new insights into the effect of *Pp*Hac1p on membrane proteins entering the secretory pathway.

Introduction

Membrane proteins (MPs) are the most common group of drug targets, with over 50% of prescription drugs targeting human MPs. The in-depth characterization of these MPs has become a major interest of pharmaceutical biotechnology, as solving their structures provides a basis for *in silico* docking studies and structure-based drug design [1]. Additionally, MPs such as cytochrome P450 enzymes offer versatile reactions for bioconversions (e.g. hydroxylation) in industrial biotechnology [2].

However, MP expression is difficult as it requires a suitable balance between translation, folding and trafficking to ensure that functional proteins end up in the membrane. Strong MP

Corresponding author: Freigassner, M. (maria.freigassner@gmail.com)

1871-6784/$ - see front matter http://dx.doi.org/10.1016/j.nbt.2014.02.009

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

overexpression can overwhelm the cellular machinery, thereby resulting in misfolding and degradation [3]. The optimization of MP expression is challenging, as various factors such as the promoter system, sequence-specific properties, the host's cellular milieu, the availability of defined lipids and the cultivation conditions strongly affect MP yields [4]. Especially the host system itself can cause major hurdles for MP overexpression, and lack of understanding of molecular mechanisms and routes in MP biosynthesis often leads to numerous 'trial and error' experiments with little prospect of success. Classic optimization of MP overexpression has mainly relied on case-by-case variations of extrinsic (pH, temperature and aeration) and intrinsic parameters (promoter strength, expression as fusion proteins or with tags, introduction of mutations to increase stability, coexpression of chaperones) rather than attempts to become better acquainted with bottlenecks encountered during expression [5–7]. Recently, -omics methods, most notably transcriptomics based on microarrays, have been used to deepen our understanding of the host response to MP production on a molecular level [1,8,9]. This research has mainly been performed using the 'classic' yeast *Saccharomyces cerevisiae* and led, in several cases, to substantial improvements [10].

The methylotrophic yeast *Pichia pastoris* has been used routinely for the heterologous expression of a large number of industrially relevant biocatalysts and pharmaceutical proteins [11,12]. In the last decade *P. pastoris* has also become one of the most commonly used expression systems for MPs [13] and many structures of MPs have been solved from material produced in *P. pastoris* (recently comprehensively reviewed by [14–20]).

There have also been efforts to study protein production in *P. pastoris* by a systems biology approach using -omics methods, most notably by the labs of D Mattanovich and P Ferrer. A key interest of these systems biology studies was to understand how *P. pastoris* reacts to the overexpression of heterologous secreted proteins. First transcriptome studies were performed with heterologous arrays from *S. cerevisiae*, as *P. pastoris* had not been sequenced at this time [21,22]. After the establishment of these heterologous hybridization methodologies and first results of the effects of heterologous protein expression on *P. pastoris* [21], this system was used to identify novel targets to improve secretion by characterizing the effect of the secretion of human trypsinogen on *P. pastoris* [22]. After sequencing of the *P. pastoris* genome [23–25], the first *P. pastoris* specific microarrays were developed [26] and used to investigate the effect of the overexpression of *HAC1*, the main regulator of the UPR (unfolded protein response) (see [27] for a recent review). The response of *P. pastoris* to misfolded proteins and the role of Hac1p have also been studied using RNA-seq [28], and a proteomics approach was used to investigate the host response to secretory protein production [29,30]. *P. pastoris* specific arrays were also combined with a FACS (fluorescence activated cell sorting) analysis to identify secretion enhancing factors [31]. Further studies on new secretion enhancing factors revealed the importance of membrane homeostasis and the secretory pathway trafficking machinery [32] and also the effects of different oxygen conditions on *P. pastoris* [33,34].

Many MPs also enter the secretory pathway but in contrast to secreted proteins, they accumulate either in the ER, Golgi or the plasma membrane. However, as previous transcriptomics studies have focused on secreted proteins, little information is available specifically on the effects of MP production on *P. pastoris*. To bridge this gap, we aimed to characterize the transcriptional response of *P. pastoris* to MP expression in order to gain an in-depth insight into MP biosynthesis and to track putative bottlenecks and targets for strain improvements. We started from a generic pipeline to rapidly express MPs in *P. pastoris* using C-terminally fused green fluorescent protein (GFP) as the indicator of expression levels. We have used this platform as a basis to express different classes of MPs (mitochondrial, ER/Golgi and plasma-membrane localized) and we studied the transcriptional response to the production of three MPs using DNA microarrays. From the results of this work Hac1p appeared to be a promising target for strain engineering and we have therefore characterized the effect of *PpHAC1* coexpression on MP yields and the cellular transcriptome.

Materials and methods

Strains and chemicals

P. pastoris CBS 7435 (Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands) was used as the host strain for all experiments. Chemicals and detergents were either purchased from Carl Roth GmbH (Karlsruhe, Germany), Anatrace Inc. (Maumee, OH) or Sigma–Aldrich Inc. (St. Louis, MO), unless stated otherwise. For cultivation of *P. pastoris*, full media (YPD: 10 g L^{-1} yeast extract, 20 g L^{-1} peptone, 20 g L^{-1} D-glucose) or buffered minimal media (200 mM potassium phosphate pH 6.0; 13.4 g L^{-1} yeast nitrogen base w/ $(NH_4)_2SO_4$ w/o amino acids; 0.4 mg L^{-1} D-biotin) supplemented with glycerol (0.5%, v/v final concentration; BMG 0.5%) or methanol (1 and 5% (v/v) for BMM2 and BMM10, respectively) were used, with components bought from BD (Franklin Lakes, NJ).

Construction of expression cassettes and plasmids

5′ (*AOX1* promoter) and 3′ flanking regions (*AOX1* terminator and *Sh ble* resistance marker cassette) of the expression cassette were amplified from the plasmid pPICZ B (Invitrogen Corp., Carlsbad, CA), using the primers 5′-gatctaacatccaaagacgaaagg-3′ (P(AOX1)-fw)/5′-catggttgaattcctcgtttcgaataattagttg-3′ (P(AOX1)-rv) and 5′-gtcagatagcgaggtcactcagtccgaacaaaaactcatctcagaagaggatc-3′ (synAOX1TT-fw)/5′-agcttgcaaattaaagccttcgagc-3′ (CycTT-rev) for promoter and terminator-Zeocin regions, respectively. All sequences for the membrane-protein-GFP-His_8 constructs were amplified from *S. cerevisiae* plasmids [35] with primers 5′-gagaagatcaaaaaacaactaattattcgaaacgaggaattcaacc-*gene*-3′ (MP-fw) and 5′-ctgagatgagtttttgttcggactgagtgacctcgctatctgacttaatgatgatgatgatgatgatg-3′ (GFP-rv); '*gene*' denotes a 25-bp gene-specific sequence. The full-length expression cassette was then assembled by overlap extension PCR [36,37] using Phusion High Fidelity DNA polymerase. To create reference strains, GFP-His_8 was expressed as cytoplasmic protein, using the linear expression cassette strategy as described [38].

This study focused on expression of the following MPs: alternative oxidase from *P. pastoris* (*Pp*Aodp) [SwissProt: A4K8T8], human CMP-Sia transporter (*Hs*Cstp) [SwissProt: P78382], copper transporter Ctr3 from *S. cerevisiae* (*Sc*Ctr3p) [SwissProt: Q06686], rice (*Oryza sativa*) CMP-Sia transporter (*Os*Cstp) [SwissProt: Q6ZL17] and human copper transporter Ctr1 (*Hs*Ctr1p) [SwissProt: O15431].

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

For the coexpression of *P. pastoris HAC1*^{i}, the sequence corresponding to the activated *HAC1* gene (Locus ID: chr1-1_0381) was assembled from genomic DNA using primers 5′-gaaagaattcaaccatgcccgtagattcttctcataag-3′ (EcoRI-PpHac1-fw) and 5′-tgcattagcggtaaatggtgctgctggatgatgcaaccgattcgactcg-3′ (PpHac1-rv1) for the 5′ exon. The sequence coding for the 3′ exon was attached by a two-step PCR using primers EcoRI-PpHac1-fw and 5′-gaatacaaagtcatttaaatcaaatgcattagcggtaaatggtgctgc-3′ (PpHac1-rv2) and 5′-ttgagcggccgcttattcctggaagaatacaaagtcatttaaatcaaatgc-3′ (NotI-PpHac1-rv3). *HAC1*^{i} was cloned into plasmid pPpKan, linearized with BamHI and transformed into the respective MP expressing strains, using G418 as selection marker.

Transformation of P. pastoris *and screening of transformants*

The correct full length cassettes or linearized plasmids were used for transformation of *P. pastoris* [39], clones were selected on the respective selective media (Zeocin^{TM} (InvivoGen, San Diego, CA) or G418 (Roth GmbH), final concentrations of 80 μg mL^{-1} or 300 μg mL^{-1}).

After 2 days of incubation at 28°C, transformants were picked and resuspended in 250 μL of BMG 0.5% in 96-well plates. Cells were grown for 48 hours at 28°C, 320 rpm, 80% humidity before methanol induction. Before adding methanol, cells were replica-plated on MD plates (13.4 g L^{-1} yeast nitrogen base w/ $(NH_4)_2SO_4$ w/o amino acids; 2% glucose (w/v); 0.4 mg L^{-1} D-biotin). Methanol was added at a final concentration of 0.5% (v/v) after 48 hours growth on BMG 0.5%. Methanol induction was repeated after 12, 24 and 36 hours. In order to screen for high-level expressors, samples of 50 μL were taken after 0, 12, 24 and 48 hours and diluted in 150 μL ddH_2O. GFP fluorescence was monitored at 512 nm using an excitation wavelength of 488 nm with a Spectramax Gemini XS microplate reader (Molecular Devices, Sunnyvale, CA, USA). In order to compare expression levels, all fluorescence values were normalized by the OD_{595}.

Selected expression clones were analysed by in gel-fluorescence; therefore, cells were collected at 3200 × *g* (10 min; 4°C) and resuspended in 500 μL breaking buffer (50 mM Tris/HCl, pH 7.6; 50 mM EDTA; 10% glycerol (v/v); 1× complete protease inhibitor cocktail, Roche Applied Science, Basel, Switzerland) and transferred to a 2 mL screw-capped tube containing 150 μL of acid-washed glass beads (425–600 μm; Sigma, St. Louis, MO). Cell disruption was performed with a tissue lyser (Qiagen, Chatsworth, CA) for 10 min at 30,000 Hz (4°C). After spinning down residual cell debris and unbroken cells (16,000 × *g*; 4°C, 1 min), the supernatant was collected in a new tube and the disruption was repeated once with the pellet. Crude membranes were isolated from collected supernatants by centrifugation at 16000 × *g*, 4°C for 90 min and resuspended in 10 μL loading dye (50 mM Tris/HCl, pH 7.6; 5% glycerol (v/v); 5 mM EDTA; 50 mM DTT; 0.02% bromophenol blue (w/v)). Resuspended MPs were separated on a 12% Tris/glycine gel (Invitrogen Corp.). Before staining the gel was analysed for GFP-fluorescence using blue light (460 nm) for excitation with a 515-nm filter cut-off on a G:Box bioimager (Syngene, Cambridge, UK).

Laboratory-scale bioreactor cultivation of P. pastoris

Fed-batch cultivation was performed in 1 L computer-controlled DASGIP fedbatch-pro® bioreactors (DASGIP AG, Juelich, Germany), with a final working volume of 500 mL at 28°C. The batch medium consisted of H_3PO_4 85% (21 mL L^{-1}), $CaSO_4{\cdot}2H_2O$ (0.9 g L^{-1}), $MgSO_4{\cdot}7H_2O$ (12.2 g L^{-1}), K_2SO_4 (14.3 g L^{-1}), KOH (3.3 g L^{-1}), mineral trace salt solution (PTM, 4.35 mL L^{-1}) and glycerol at a final concentration of 1% (v/v). The initial batch phase of 12 hours was followed by a 9-hour exponential fed-batch phase with glycerol, at a specific substrate consumption rate of 0.25 g g^{-1} hour^{-1} and a specific growth rate of 0.15 hour^{-1}. Methanol induction was then initiated with a constant feed of pure methanol (3 mL hour^{-1}). Culture pH was maintained at 6.0 by the addition of 25% NH_4OH (v/w) or phosphoric acid 85%, as needed. The inoculum was prepared by inoculating a single colony in 50 mL BMGY in a 250-mL baffled shake flask and subsequent growth for 24 hours at 28°C. From this flask, 10 mL was withdrawn every 6–12 hours to measure biomass and protein yield during the course of the cultivation.

Cell disruption, FSEC, purification of GFP fusion proteins

After harvesting cells at 3000 × *g* (10 min, 4°C), cells were resuspended in disruption buffer (100 mM Tris/HCl pH 7.5; 150 mM NaCl, 10% glycerol, supplemented with protease inhibitors; 10 g CWW/100 mL buffer) and lysed using a Merckenschlager homogenizer. Cell disruption was repeated twice. After removing unbroken cells and cell debris (12,000 × *g*; 15 min; 4°C), membranes were collected at 100,000 × *g* (90 min; 4°C). Pelleted membranes were resuspended in disruption buffer using a Dounce homogenizer, flash frozen in liquid nitrogen and stored at −80°C.

A panel of different detergents (n-decyl-β-D-maltoside, n-dodecyl-β-D-maltoside, n-octyl-β-D-glucoside, LDAO, Fos-choline-12) was tested for their ability to release the GFP fusion protein from the membrane while retaining monodispersity, and fluorescence-based size exclusion chromatography (FSEC) was performed as described previously [35,40].

For purification of GF-MP fusions, Ni-NTA chromatography was employed, as described in [41]. Following isolation, the MP was released from the fusion protein in an overnight digest using equimolar amounts of His_{10}-tagged TEV-protease at 4°C, performed in a dialysis mode to remove imidazole. In a subsequent Ni-NTA purification, TEV-protease, GFP and unprocessed GFP fusion were separated from the recovered MP, which was collected as flowthrough, concentrated to approx. 10 mg mL^{-1} and analysed for its purity and monodispersity by SDS-PAGE and size exclusion chromatography (Superose 6 10/300; GE Healthcare; buffer: 20 mM Tris/HCl, pH 7.5; 200 mM NaCl, 0.03% DDM). Protein concentrations were assessed with the bicinchoninic acid assay (BCA), using bovine serum albumin as standard. Purified proteins were analysed by SDS-PAGE after staining with Coomassie brilliant blue.

Microscopy

Confocal fluorescence microscopy was performed with a Leica AOBS SP2 MP instrument. Before analysis cells were washed and resuspended in 100 mM potassium phosphate buffer, pH 6.0. For staining, MitoTracker® Red CMXRos and the vacuole-sensitive dye FM® 4-64 were used according to the manufacturer's instructions.

For co-immunoprecipitation, cells grown in the presence of methanol were fixated with 3.7% (v/v) formaldehyde (in buffered YPD medium) for 2 hours. Cell lysis was performed with zymolyase 20 T for 60 min at 30°C. Cells were then immobilized on a slide

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

coated with poly-L-lysine and incubated for 30 min with a primary antibody directed against a 75-kDa ER marker protein of *P. pastoris* [42].

For electron microscopy, cells were harvested in the early stationary phase by centrifugation and washed three times with dist. water. Subsequently, cells were fixed for 5 min in a 1% aqueous $KMnO_4$ solution at room temperature, washed with dist. water and fixed in a 1% $KMnO_4$ solution for 20 min again. After four wash steps using dist. water, cells were incubated in 0.5% aqueous uranyl acetate overnight at 4°C and dehydrated in a 4-step process with increasing concentrations of ethanol (50%, 70%, 90%, and 100%). Pure ethanol was then exchanged by propylene oxide, and specimens were gradually infiltrated with increasing concentrations (30%, 50%, 70% and 100%) of Agar 100 epoxy resin (Agar Scientific Ltd., Stansted, England) mixed with propylene oxide for a minimum of 3 hours per step. Samples were left overnight in a 1:1 mixture of Agar 100 epoxy resin and propylene oxide. Infiltration was continued the next day using previously stated mixtures of Agar 100 epoxy resin and propylene oxide. Samples were embedded in pure, fresh Agar 100 epoxy resin and polymerized at 60°C for 48 hours. Ultrathin sections of 80 nm were stained with lead citrate and viewed with a Philips CM 10 transmission electron microscope.

Microarray analysis

For RNA isolation, cells were grown in biological duplicates in 1-l baffled shake flasks (150 ml cultivation volume) to an OD_{595} of 1.0 at 30°C using glycerol 1% (v/v), before protein expression was induced by switching to methanol 0.5% (v/v). Cells were sampled 6 hours after induction and compared to GFP expressors cultivated under identical conditions. Total RNA was isolated using the RiboPure™ Yeast RNA kit (Ambion, Austin, TX). Purity and integrity were assessed with the Agilent Bioanalyzer 2100 using the RNA 6000 Nano Assay kit (Agilent Technologies, CA). Of the total isolated RNA, 10–15 μg was used as starting material for reverse transcription of mRNA and cDNA labelling using the One-Cycle Target Labeling Kit P/N 900493 (Affymetrix, Inc., Santa Clara, CA, USA) to obtain roughly 70–100 μg of labelled cDNA. During RT and labelling, 4 *Bacillus subtilis* genes (*lys*, *phe*, *thr*, *dap*) were spiked as controls at given final ratios of copy numbers (1:100,000; 1:50,000; 1:25,000; 1:6667). Labelled cDNA was quantified photometrically using a ND-1000 Nano-Drop Spectrophotometer (Thermo Scientific, Wilmington, US) and the concentration was adjusted to 0.5 μg μL^{-1}. Before hybridization, cDNA was fragmented by heat shock (94°C, 35 min). After addition of probe array controls (GeneChip Eukaryotic Hybridization Control Kit, P/N 900454, Affymetrix), hybridization of labelled cDNA to the *P pastoris* microarrays (PPA01a520396F) was performed at 45°C for 16 hours while gently rotating. Once hybridization was completed, chips were washed and stained using the EukGE-WS2v5 protocol.

Data evaluation (normalization and scaling) was performed with the package limma [43] within R [44]. Genes were considered to be differentially expressed if expression levels of the averaged biological replicates changed by at least 1.5-fold and Benjamini–Hochberg adjusted *p*-values were below 0.05. Finally, transcripts not annotated were excluded from further analysis. All annotated genes were categorized in GO biological processes according to the SGD GO slim tool (http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl).

Results and discussion

Platform for fast expression of membrane proteins in P. pastoris

We started with the expression of MPs of different classes and organisms to create production strains for the subsequent transcriptome analyses. Namely, we expressed a copper transporter from *S. cerevisiae* (*Sc*Ctr3p), a human CMP-sialic acid transporter (*Hs*Cstp) and the alternative oxidase from *P. pastoris* (*Pp*Aodp). *Sc*Ctr3p, in its natural host, is inserted into the plasma membrane, whereas the *Hs*Cstp resides in the ER/Golgi and *Pp*Aod is a mitochondrial MP (Table 1).

Expression levels of MPs are difficult to assess rapidly, as most enzymatic activity assays for MPs require the isolation and purification of the subcellular membrane fraction containing the MPs. Therefore, MP detection systems have been developed in which GFP is C-terminally linked to the MP, allowing detection of the fusion proteins from intact cells via fluorescence measurements [40,45]. Comparable systems have also been successfully used in *P. pastoris* [46,47]; therefore we set up a similar, universally applicable platform to express our set of target proteins (Fig. 1). For fast cloning, the expression cassettes were assembled by overlap extension PCR, linking the *AOX1* promoter (P_{AOX1}) and the MP-GFP fusion gene to the *AOX1* transcriptional terminator region plus a resistance marker cassette (Zeocin), as described previously [38]. Expression levels and amounts of membrane-integrated protein were estimated from the fluorescence of the GFP fusion proteins from whole-cell cultures, cellular membranes in solution and SDS-polyacrylamide gels [45], while stability and monodispersity of the fusion protein were monitored in crude detergent extracts of membranes before purification by fluorescence size-exclusion

TABLE 1

Properties and expression levels of the membrane proteins used in this study. Information on the origin of the proteins, their localization, number of transmembrane helices (TMH) and volumetric yields of the membrane protein-GFP fusion proteins (c_{MP}) are given. Mean value and standard deviation of three experiments are given

Abbreviation	Protein	Organism of origin	Localization		Number of TMH	c_{MP} (mg L^{-1})
			Native	*P. pastoris*		
***Pp*Aodp**	Alternative oxidase	*P. pastoris*	Mitochondria, *P. pastoris* native protein		2 (parallel)	22.52 ± 0.26
***Hs*Cstp**	CMP-Sia transporter	*Homo sapiens*	Golgi	ER and Golgi	7	14.97 ± 0.28
***Sc*Ctr3p**	Copper transporter	*S. cerevisiae*	Plasma membrane	Plasma membrane, partly vacuolar membranes	3	8.35 ± 1.09

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

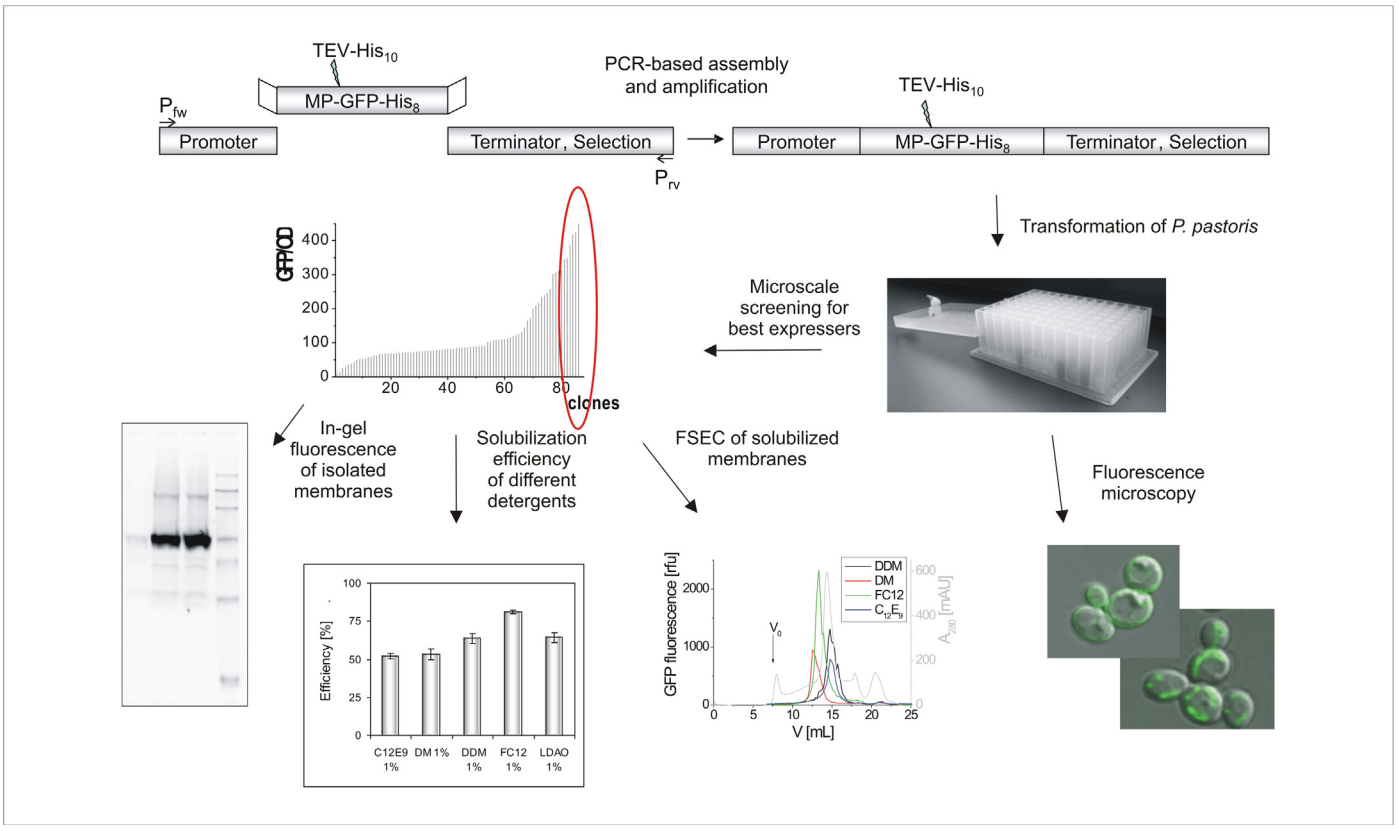


FIGURE 1

olePCR-based micro-scale expression platform for membrane protein expression using *P. pastoris*. DNA sequences coding for the membrane protein-GFP fusion protein (MP-GFP-His_8) were assembled by olePCR (overlap extension PCR) with promoter (P_{AOX1}), terminator regions (*AOX1*TT) and the zeocin selection marker (Zeo), allowing direct transformation of *P. pastoris* cells. Best expressors were selected from 96-well plate cultivations based on GFP fluorescence of growing cells and isolated membranes, using fluorescence microscopy and in-gel fluorescence of SDS-PAGE separated proteins, respectively. Purification of target proteins from cellular membranes was optimized by selecting detergents that efficiently release the GFP fusions while retaining monodispersity, as judged by fluorescence-based size exclusion chromatography.

chromatography (FSEC) [48] (see Fig. 1 for the experimental outline and Supporting Fig. S1 for representative purification data). The expression levels of the three target proteins ranged from 8.4 to 22.5 mg L^{-1}, as outlined in Table 1.

Apart from FSEC and in-gel fluorescence, we also analysed the intracellular localization by fluorescence microscopy, using the intrinsic GFP signal (Fig. 2). The alternative oxidase localized to mitochondrial membranes, as shown by co-staining with the mitochondria-sensitive dye MitoTracker® (Supporting Fig. S2), while the copper transporter *Sc*Ctr3p associated with the plasma membrane but also partly with vacuolar membranes (Supporting Fig. S3). *Hs*Cstp remained in the ER (Supporting Fig. S4) though it localizes to the Golgi in humans. However, neither fold nor stability of *Hs*Cstp was compromised, as judged from its monodispersity and homogeneity observed during FSEC upon solubilization (S1).

The transcriptional response of P. pastoris *to membrane protein expression*

The three heterologous MP expressing strains (*Pp*Aodp, *Sc*Ctr3p, *Hs*Cstp) were used as models for microarray studies to elucidate the transcriptional response of *P. pastoris* to MP expression.

*Pp*Aodp was selected due to its high expression levels and mitochondrial localization, whereas the two transporters were chosen to study the transcriptional response to MPs entering the secretory pathway reaching either the plasma membrane (*Sc*Ctr3p) or residing in the ER (*Hs*Cstp). By choosing model proteins of different structures, functions, origins and intracellular localizations we aimed to gain a better understanding of the different molecular mechanisms of MP biogenesis. A strain producing cytoplasmic soluble GFP was used as control in order to discriminate between effects evoked by production of membrane-embedded and soluble proteins and thus to distinguish between the general response to overexpression and specific protein-dependent transcriptome patterns caused by expression of the MP-GFP fusion partner.

Whole-genome expression profiling was undertaken using custom *P. pastoris* whole genome microarrays (Affymetrix, Santa Clara, CA, United States). The transcript patterns were closely related to both function and cellular localization of the overexpressed MPs. When comparing genes upregulated, downregulated or remaining unchanged in the three MP overexpressing strains (Fig. 3), the two MPs entering the secretory pathway (*Sc*Ctr3p and *Hs*Cstp) showed similar responses (that is, a large number of genes overlapping), but *Pp*Aod showed a divergent response, linked more to its role in the energy metabolism than its MP character. Thus the effects are discussed separately for the mitochondrial protein *Pp*Aod and the two MPs entering the

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

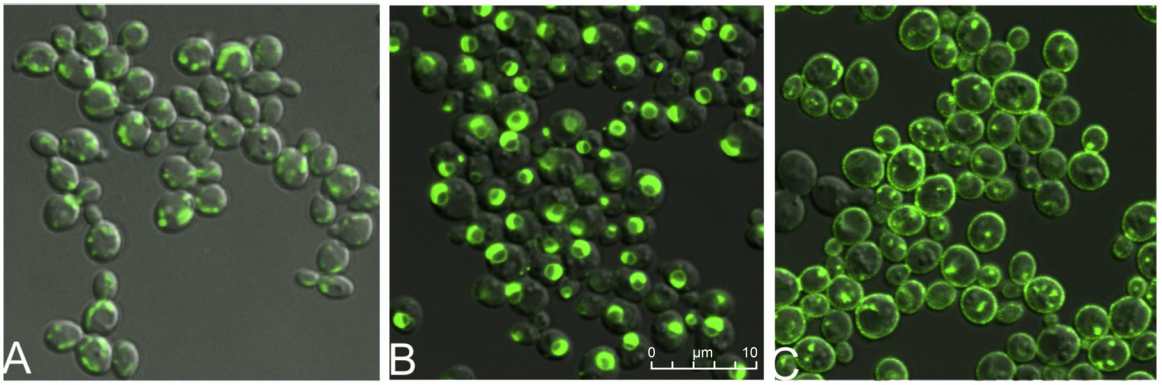


FIGURE 2

Cellular localization of the membrane proteins used in this study assessed by confocal fluorescence microscopy. While *Pp*Aodp localizes to mitochondrial membranes **(a)**, *Hs*Cstp enters the endoplasmic reticulum **(b)**. *Sc*Ctr3p passes through the secretory pathway to reach the plasma membrane **(c)**. See S2, S3, S4 for co-staining images.

secretory pathway (*Sc*Ctr3p and *Hs*Cstp). All microarray hybridization data are provided in the Supporting file S5.

Mitochondrial alternative oxidase Pp*Aodp*

In a comprehensive study outlining expression characteristics of over 550 yeast MPs in *S. cerevisiae*, highest expression levels have been identified for stress response proteins [49]. Therefore we selected the *P. pastoris* mitochondrial alternative oxidase (*Pp*Aodp, involved in the cellular stress response) as a target for MP overexpression. Indeed, the GFP fluorescence of the *Pp*Aod-GFP fusion protein indicated yields of about 22 mg/L in isolated membranes (Table 1). As the main constituent of the alternative respiratory pathway of *P. pastoris*, *Pp*Aodp bypasses the terminal complexes of the electron transport chain (CIII and CIV) by directly transferring electrons from ubiquinol to oxygen. With this system, the cellular efficiency for synthesizing ATP is reduced by two-thirds, as the proton transport across the inner mitochondrial membrane is bypassed [50]. Alternative respiration plays a key role in sustaining growth under metabolic constraints (for example, when the cellular energy demand is low) and in counter-acting cellular stress [50–52].

When studying the transcriptional response to *PpAOD* overexpression under P_{AOX1}, we observed a multitude of cellular changes tightly linked to its biological function. Namely, the expression levels of more than 1100 genes were altered (see Table 2 and S5). While genes involved in amino acid metabolism, vesicular transport, cytoskeleton biogenesis, respiration and cell cycle were compromised, increased transcript levels of genes involved in stress response, protein catabolism and protein modification were observed, thus indicating a transient arrest of normal cellular processes (Table 2). This response partly overlapped with the transcript profile of *S. cerevisiae* cells, which have been exposed to external stress situations [53,54], as demonstrated by an upregulation of genes involved in oxidative stress defense, protein folding and degradation.

Being embedded in the inner mitochondrial membrane, P_{AOX1}-mediated *PpAOD* expression evoked also strong effects within mitochondria, triggering a widespread rearrangement in expression of mitochondrial genes coding for proteins of the folding and import machineries (Table 2). Among import proteins, relative transcript levels of genes coding for transport proteins of the outer and inner mitochondrial membranes (OMM and IMM) and the intermembrane space (IMS) were upregulated upon *Pp*Aodp production. Induced genes include *SSC1*, also known as *mtHSP70*, the ATP-driven core of the PAM complex, which plays a crucial role in protein translocation and mitochondrial quality control [55]. Ssc1p interacts with Mdj1p and Mge1p, two matrix-resident translocation and folding co-chaperones, which were also upregulated during *Pp*Aodp production. The prohibitin complex (*PHB1* and *PHB2*) of the IMS that stabilizes newly imported proteins [56] and the α-subunit of the mitochondrial processing peptidase (*MAS2*) also range among induced genes. Furthermore, mRNA abundance of many proteins assisting in assembly of respiratory proteins, that is, CIII complex, cytochrome c oxidase and ATP synthase, increased as a consequence of *PpAOD* overexpression.

Apart from genes coding for proteins involved in mitochondrial import and folding, genes coding for proteolytic proteins were also upregulated, including the *m*-AAA protease, consisting of subunits Afg3p and Yta12p. This matrix-exposed protease mediates not only assembly, but also mislocation and turnover of misfolded and non-assembled MPs of the IMM, such as the respiratory chain and ATPase complexes. Additionally, genes of the mitochondrial quality control system, which mediates the turnover of oxidized proteins and assists in protein complex assembly under stress conditions, were also upregulated.

Moreover, genes coding for subunits of the proteasome (20S core and 19S regulatory proteasome particles), vacuolar degradation and sorting pathways and the vacuolar protease *PVB1* were strongly upregulated. The induction of these non-mitochondrial genes provides a clear indication that the stress response to *PpAOD* overexpression was cell wide and not limited to the mitochondria. The translocation machinery of mitochondria might have been overloaded, resulting in redirection of superfluous and/or partly folded proteins to the cytosol where they were subsequently subjected to proteasomal degradation as proposed for higher eukaryotes [57]. A downregulation of genes coding for transport proteins residing in the IMM supported the idea that the membrane space of the IMM was restricted and thus mitochondrial

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

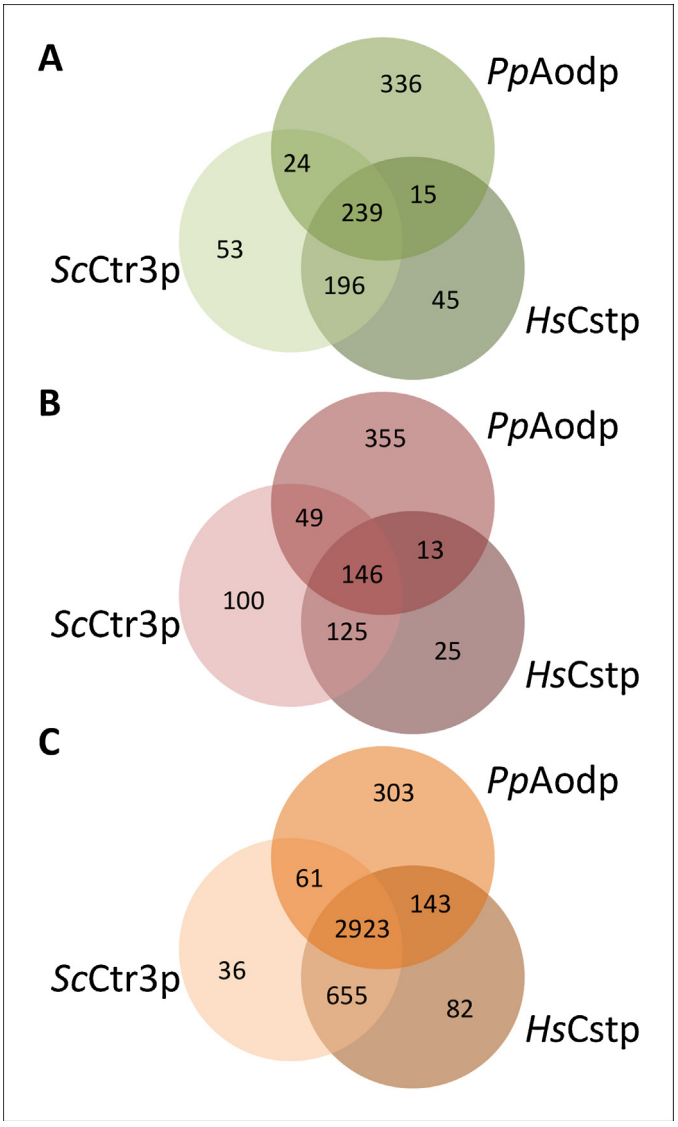


FIGURE 3

Comparative analysis of differential gene expression between *Pp*Aodp-GFP, *Hs*Cstp-GFP and *Sc*CTR3p-GFP fusion protein producing strains. The numbers of unique and overlapping genes showing upregulation **(a)**, downregulation **(b)** or unchanged regulation **(c)** between the three MP producing strains relative to the control strain (expressing GFP alone) are shown. The raw data of S5 were used for the calculations.

membranes were not able to accommodate additional proteins due to the high *Pp*Aodp load.

The transcriptional response also implied that the respiratory capacity was largely rearranged due to *PpAOD* overexpression as genes coding for the respiratory complexes I to IV were down regulated. Genes needed for ATPase biosynthesis (compounds of the membrane-embedded F_0 domain) were also down regulated. Concurrently, a downregulation of genes coding for proteins involved in amino acid metabolism, nucleotide biosynthesis and the TCA cycle was also observed, suggesting that the entire cellular metabolism slowed down during *Pp*Aodp production. This was in line with the previously observed negative effect of *Pp*Aodp overexpression on biomass production [50]. A similar response was observed for genes that were involved in secretion and post-translational protein modification (N-linked glycosylation, GPI-anchor biosynthesis).

Taken together, the transcriptome profile implied that prolonged and strong overexpression of *Pp*Aodp posed a challenge on metabolism and cellular homeostasis. As a consequence of *Pp*Aodp production, ATP was produced in an inefficient way, metabolism slowed down and many biosynthetic pathways were throttled back in order to compensate for the loss of energy caused by alternative respiration. The pattern observed strongly suggests that *Pp*Aodp overproduction induces effects that are more probably linked to its physiological function in the central energy metabolism rather than its MP character.

Membrane proteins Hs*Cstp and* Sc*Ctr3p entering the secretory pathway*

Most eukaryotic MPs (except for mitochondrial and most peroxisomal ones) enter the secretory pathway by translocating into the ER membrane, as exemplified in this study by the model proteins *Sc*Ctr3p and *Hs*Cstp. While *Hs*Cstp localized to the ER membrane, *Sc*Ctr3p was embedded in plasma and partly in vacuolar membranes (Fig. 2). For both proteins production levels in the mg/L range were obtained (Table 1). Upon overexpression of the two proteins, more than 800 genes were differently regulated compared to the GFP expressing control strain. Interestingly, the vast majority of down-regulated and up-regulated genes overlapped between *Hs*Cstp and *Sc*Ctr3p (Fig. 3 and S5), implying that MPs entering the secretory pathway evoke similar responses on the transcriptome level and only minor physiological effects are caused by the nature of the expressed MP itself. Upregulated genes participated in diverse cellular processes including transcription (RNA polymerases I and III), translation (ribosome biogenesis, tRNA metabolism, mitochondrial ribosome biogenesis), stress response, lipid metabolism (sterol biosynthesis, β-oxidation), and nucleotide and amino acid biosynthesis. Downregulated genes were mostly involved in processes such as protein modification (N-glycosylation, GPI anchoring), ubiquitination, fatty acid biosynthesis, respiration and vesicular traffic. Additionally, genes coding for degradation proteins (*UFO1*, *DOA4*, *VPS24*, *VPS25*, *VPS28*, *STP22*, *SNF7*) were repressed.

Altered transcript levels were also identified for genes involved in various aspects of cell surface morphology and biogenesis, including upregulation of flocculation (*MUC1*) and cell wall assembly (*EXG1*, *GAS1*, *WSC3*, *BOT2*), but downregulation of cell wall-associated genes (1,3-β-glucanosyl synthase *FKS3*, *KRE11*, glucan 1,4-α glucosidase *STA1*, cell wall stabilizer *HSP150*), suggesting effects on the cell wall structure. MP production also affected cellular homeostasis considerably as several stress-related genes were upregulated, including genes expressed in response to heat (*TIP1*, *NMA111*, *GAC1*), oxidative stress (*MXR1*) and the general stress factor *MSN2*. These results indicated that heterologous expression of MPs entering the secretory pathway caused a general stress response in *P. pastoris*. However, in contrast to *Pp*Aodp, the overexpression of *Hs*Cstp and *Sc*Ctr3p had no severe impact on energy metabolism.

Concerning lipid metabolism, we found increased transcript levels for genes of β-oxidation (*POX1*) and ubiquinone biosynthesis (*COQ1*), while isoprenoid and phosphatidylethanolamine (*MUQ1*) biosyntheses were repressed according to the transcript response. *Hs*Cstp production triggered specifically an increased

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

TABLE 2

Differently regulated genes upon *Pp*Aodp overexpression in *P. pastoris*. The genes are grouped by cellular function and localization of the respective proteins. The regulation (Reg.) is indicated by an upward-pointing arrow (↑) for upregulated genes, while down-regulated genes are marked by a downward-pointing arrow (↓). Exact fold changes for each gene are provided in S5. Genes reaching a fold changed of ≥1.5, but did not pass the adjusted *p*-value criteria, are written in parentheses.

Function	Genes	Reg.
Mitochondrial genes		
Import machinery	OMM: (*TOM70*), *MIM1*	↑
	IMS: *ERV1*, *MIA40*, *PHB1/2*	
	IMM: *TIM9*, *TIM17*, *TIM22*, *TIM23*, *TIM44*, *TIM50*	
	Matrix: *MDJ1*, *MGE1*, *MAS2*	
Protein folding	*FMC1*, *ATP12*, *COX11*, *PET117*, *SHY1*, *COA1*, *COX15*, *PNT1*, *CBP3*, *CBP4*, *BCS1*, *CYT2*	↑
	COX20	↓
Proteases	*AFG3*, *YTA12*, *PIM1*, *LAP3*, *PCP1*	↑
	MGR3	↓
Transport proteins	*OAC1*, *SFC1*, *MIR1*, *PIC2*, *MDL2*, *ADY2*, *YIA6*, *ACP1*, *RIM2*, *CTP1*, *ORT1*, *PET8*, *CRC1*, *YAT1*, (*LEU5*), (*ORT1*)	↑
Respiration	Complex I: *NUO1*, *NUO10*, *NUO17*, *NUO20*, *NUO21.1*, *NUO24*, *NUO30*, *NUO51*[1]	↓
	NDE1	↓
	Complex II: *SDH1*	↓
	Complex III: *CYT1*, *QCR2*, *QCR6*, *QCR7*, *QCR8*, *QCR9*, *RIP1*	↓
	Complex IV: *COX4*, *COX12*	↓
	Complex IV: *COX5B*	↑
	ATP synthase: *ATP14*, *ATP19*, *ATP20*	↑
Mitochondrial protein biosynthesis	RNA polymerases: *RPO41*	↑
	Translation initiation and elongation factors: *IFM1*; *MEF2*, *TUF1*, *GUF1*	↑
	tRNA synthetases: *DIA1*, *ISM1*, *NAM2*, *MSK1*, *SLM5*, *HER2*	↑
	tRNA synthetases: *MSW1*	↓
	Small mitochondrial ribosomal subunits: *MRPS5*, *MRPS8*, *MRPS9*, *MRPS17*, *MRPS28*, *MRPS35*, *RSM7*, *RSM10*, *RSM18*, *RSM19*, *RSM22*, *RSM23*, *RSM24*, *RSM25*, *RSM26*, *MRP1*, *MRP4*, *MRP51*, *SWS2*, *NAM9*, *PET123*	↑
	Large mitochondrial ribosomal subunits: *MRP7*, *MRP20*, *MRP21*, *MRP49*, *RML2*, *MRPL3*, *MRPL4*, *MRPL6*, *MRPL7*, *MRPL8*, *MRPL9*, *MRPL11*, *MRPL13*, *MRPL15*, *MRPL17*, *MRPL19*, *MRPL20*, *MRPL22*, *MRPL23*, *MRPL24*, *MRPL25*, *MRPL27*, *MRPL28*, *MRPL32*, *MRPL33*, *MRPL36*, *MRPL38*, *MRPL40*, *MRPL44*, *MRPL49*, *MRPL50*, *IMG1*	↑
TCA cycle	*CIT1*, *ACO1*, *IDH1/2*, *FUM1*	↓
Other cellular genes		
Protein catabolism	20S proteasome: *PRE2*, *PRE4*, *PRE7*, *PRE8*, *PUP1*, *PUP2*, *SCL1*, *UMP1*	↑
	19S proteasome: *RPN1*, *RPN2*, *RPN6*, *RPN8*, *RPN10*, *RPT2*, *RPT5*	↑
	Ubiquitination: *UFD1*, *UFD4*, *UBR1*, (*PIB1*), *UBI1-4*, *UBA1*, *RAD6*, *UBP6*, *UBC11*, *UBP12*, *UBP13*	↑
	Ubiquitination: *UBP10*, *UFO1*, *UBC9*	↓
	Retrotranslocation: *CDC48*, *OTU1*, *SHP1*, *DOA1*, *UBX4*, *UFD1*	↑
	Autophagy: *ATG8*, *ATG11*, *ATG15*	↑
	Vacuolar sorting proteins: *VPS13*, *VPS27*, *VPS28*, *VPS64*, *SNF7*, *SNX4*	↑
Protein modification	N-linked glycosylation: *SEC53*, *ALG5*, *ALG14*, *RFT1*, *EOS1*, *MNN2*	↑
	GPI-anchor biosynthesis: *SMP3*, *SPT14*, *GPI19*	↑
Secretion	*SEC23*, *USO1*, *UFE1*, *SVP26*, *ERV25*, *GOT1*, *APM4*, *SPC2*	↑
Amino acid metabolism	Biosynthesis: *ALT1*, *ARG3*, *ARG8*, *ARO3*, *ARO4*, *ARO9*, *GLT1*, *HIS1*, *HIS6*, *LEU1*, *LEU2*, *LEU5*, *LYS9*, *LYS14*, *MET6*, *MET13*, *MET14*, *ORT1*, *PHA2*, *THR1*, *TMT1*, *TRP2*	↓
	Biosynthesis: *ARO10*, *CYS3*, *MET1*, *MET2*, (*MET7*), (*MET16*), *MET28*, (*MET32*), *MMF1*	↑
	Degradation: *CAR1*, *CAR2*, (*GCV2*), *GDH2*, (*PUT1*), *UGA1*	↑
Purine and pyrimidine biosynthesis	*ADE5/7*, *ADE8*, (*URA2*), *URA6*	↑
Lipid metabolism	β-Oxidation: *CRC1*, *FOX2*, *POX1*, *SPS19*, *TES1*	↓
	Fatty acid biosynthesis: *FAA2*, *SUR4*	↓
	Fatty acid biosynthesis: *HTD2*	↑
	Sterol metabolism: *ERG2*, *ERG24*, *CYB5*, *YEH2*	↓
	Sterol metabolism: *ERG1*, *ERG3*, *ERG25*	↑

[a]Nomenclature based on *Schizosaccharomyces pombe* genes.

transcription of fatty acid (*FAA2*), cardiolipin (*PGS1*) and phospholipid (*CDS1*) biosynthesis genes.

Also in *S. cerevisiae* transcript levels of a strain producing a MP entering the secretory pathway have been investigated applying miniarray technology [58]. Namely the transcriptional response of a single strain producing Fps1p, a eukaryotic glycerol facilitator localized to the plasma membrane, was tested under different growth conditions (temperature and pH variations). This experimental setup is different from that of our study, as we compared MP producing strains to a control strain under identical growth conditions. Direct comparison between the study in *S. cerevisiae* [58] and our results is further complicated, since the cultivation conditions and thereby the strain physiology were different. Production of Fps1p in *S. cerevisiae* was found to be optimal before the diauxic shift (that is, under fermentative growth) and these conditions were also applied for the abovementioned transcriptome analysis [58]. In contrast, *P. pastoris* is generally grown under aerobic conditions, also applied in our experiments. *S. cerevisiae*

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

is a Crabtree-positive yeast, so even under aerobic conditions glucose is fermented to ethanol. In contrast *P. pastoris* is Crabtree negative, meaning that glucose is metabolized by respiration under aerobic conditions. Therefore it is not appropriate to compare our *P. pastoris* microarray data to *S. cerevisiae* data from the literature [58] in which different physiological conditions were used for MP production.

Nevertheless, a few similarities were noticed between the response of the *S. cerevisiae* Fps1p producing strain [10,58] and the *P. pastoris* strains of our study. In *S. cerevisiae SPT3* was downregulated under high yield Fps1p production conditions and *spt3* deletion led to a 9- to 69-fold increased Fps1p yield (depending on the growth medium) [10]. As component of the transcription SAGA complex, Spt3p is involved in regulating the transcriptional activity of RNA polymerase II in *S. cerevisiae* [59]. In our study, the *P. pastoris* homolog of *ScSPT3* was also downregulated in the *Hs*Cstp and *Sc*Ctr3p producing strains, suggesting a similar role of this gene for MP production in *P. pastoris*. Additionally, in the *S. cerevisiae spt3* deletion strain, the gene *BMS1* was upregulated and overexpression of *BMS1* resulted in elevated Fps1p yield [10]. *BMS1* plays an important role in biogenesis of the small ribosomal subunit (40S) and its overexpression altered the ratio of ribosomal subunits. Also in the *P. pastoris Sc*Ctr3p and *Hs*Cstp producing strains, *BMS1* was upregulated, suggesting *spt3* deletion and *BMS1* overexpression also as strain engineering targets in *P. pastoris*.

Interestingly, in all *P. pastoris* strains overexpressing MPs (*Pp*Aodp, *Hs*Cstp or *Sc*Ctr3p), *INO1* (encoding inositol-1-phosphate synthase) transcript levels were higher compared to the GFP expressing control strain, suggesting that induction of phosphatidylinositol biosynthesis was a common pattern in all MP producing strains. Also in *S. cerevisiae* increased *INO1* levels were observed during MP production, correlating with membrane proliferation [60]. These changes in transcript patterns of lipid biosynthetic genes suggest a crucial role of the lipid metabolism and a possible target for strain engineering to improve MP expression.

Role of HAC1 *during membrane protein production*

***PpHAC1**^{i} coexpression improves yields of ER resident MPs 1.5- to 2.1-fold*

Interestingly, the gene coding for Hac1p, the central regulator component of the unfolded protein response (UPR), was upregulated in the strain overexpressing *Hs*Cstp. The UPR is the major ER surveillance system and triggers the expression of various proteins, which counteract stress arising from misfolded proteins. Folding stress leads to Hac1p activation by a unique splicing mechanism (see [27] for a recent review on the regulation in *P. pastoris*). Hac1p then activates numerous UPR target proteins involved in protein folding, trafficking and glycosylation, thereby bolstering the folding capacity in the ER [26].

However, Hac1p also affects the lipid composition of ER membranes and the available membrane space. Cox *et al*. [61] showed that increased amounts of phospholipid biosynthesis enzymes, in particular of the phosphatidylinositol pathway, were required to expand the ER by formation of membrane stacks called karmellae. These stacks can potentially lead to an increased ER membrane capacity desirable for the expression of heterologous MPs as already observed for several MPs in *S. cerevisiae* [62–69] and *P. pastoris* [70–74].

Coexpression of the activated, spliced variant of the *HAC1* gene (denoted as *HAC1*^{i}) has been used in several organisms as a target for strain engineering to improve protein secretion. Also in *P. pastoris* several studies have been performed investigating the UPR and the role of Hac1 [27]. Gasser *et al*. [75] reported a 1.3-fold increase in Fab secretion when co-expressing *HAC1* from *S. cerevisiae* in *P. pastoris*. Guerfal *et al*. [74] have tested the effect of *PpHAC1* on secreted, surface displayed proteins and one MP. In general the effects seen varied considerably between various secreted and surface displayed proteins. Namely the expression level of surface displayed proteins decreased in three of four cases, showing an approximately twofold improvement only for one protein. The expression levels of the two tested secreted proteins improved approximately twofold. For the single MP tested (adenosine A2A receptor, a G-protein coupled receptor) positive effects were seen (improved processing of the signal sequence and increased ligand binding).

As the outcome observed by Guerfal *et al*. [74] appeared to be protein dependent for secreted and surface displayed proteins, we wanted to test the effect of *PpHAC1*^{i} coexpression on multiple MPs, to evaluate if *PpHAC1*^{i} coexpression can act as a general strategy to improve MP expression. To this end, we coexpressed *PpHAC1*^{i} in the strains from our initial set of target proteins that were entering the secretory pathway (*Sc*Ctr3p and *Hs*Cstp). To increase the number of model proteins entering the secretory pathway, we tested a human copper transporter (*Hs*Ctr1p) and a CMP-sialic acid transporter from rice (*Os*Cstp), which are also ER associated (Supporting Fig. S6).

As Guerfal *et al*. had shown that constitutive expression of the *PpHAC1*^{i} gene had little to no effect [74], we coexpressed *PpHAC1*^{i} using the methanol inducible P_{AOX1}.

At first we tested the coexpression in shake flask cultures. Strains showing positive effects were scaled up to bioreactor cultivations. For the copper transporter *Sc*Ctr3p, which resides in the plasma membrane, no effect of *HAC1*^{i} coexpression was observed (data not shown). For all proteins localizing to ER membranes, approximately 1.5- to 2.1-fold higher yields were obtained upon *PpHAC1*^{i} coexpression (Table 3), suggesting a generally beneficial effect of

TABLE 3

***PpHAC1**^{i} coexpression improves the yields of ER resident membrane proteins 1.5- to 2.1-fold. Increase in specific yields (μg of recombinant fusion protein per mg of membrane protein) obtained by **HAC1**^{i} coexpression in bioreactor cultivations is given. Mean value and standard deviation of three experiments are given.*

Membrane protein	Membrane protein yield in the progenitor strain	Membrane protein yield upon Hac1p coexpression	Increase (%)
***Hs*Cstp**	4.323 ± 0.010	9.214 ± 0.181	113.2
***Hs*Ctr1p**	0.288 ± 0.005	0.502 ± 0.013	74.0
***Os*Cstp**	2.552 ± 0.052	3.786 ± 0.036	48.4

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

PpHAC1^{i} coexpression to maximize yields for ER resident MPs in *P. pastoris*. This result is somewhat different from previous findings on the effect of the *S. cerevisiae* UPR on MP production, as Griffith *et al.* [76] suggested that there is a negative correlation between UPR upregulation and functional expression of MPs (that is, high UPR induction levels impair MP expression). However, expression of *HAC1*^{i} in *S. cerevisiae* also showed beneficial effects on the secretion of some proteins [77,78]. In addition, recent studies on Hac1p-mediated UPR regulation in *P. pastoris* [28,74] suggest that there is an important difference from *S. cerevisiae*. While *HAC1* expression is regulated by a unique stress-regulated splicing mechanism in *S. cerevisiae*, splicing in *P. pastoris* was shown to occur constitutively, suggesting primarily a UPR regulation by the *HAC1* transcript levels in *P. pastoris*. These differences may explain the dissimilar role of the UPR in MP expression in *S. cerevisiae* and *P. pastoris*, but would require further, more detailed comparative studies.

We aimed to elucidate the mechanisms of how *PpHAC1*^{i} coexpression caused the beneficial effects on MP expression. The folding machinery in the ER may be upregulated, leading to more correctly folded protein and avoiding degradation. Also changes in the ER membrane composition or an extension of the membrane space (by karmellae formation) as previously reported [74] could be the cause.

For further studies to investigate these effects, we focused on the *HsCST* + *PpHAC1*^{i} coexpressing strain, as *PpHAC1*^{i} coexpression had shown the strongest increase in protein yields (2.1-fold) in this context. We performed electron microscopy on the wild type, the *HsCST* expressing strain and the *HsCST* + *PpHAC1*^{i} coexpressing strain to investigate the ER morphology (Fig. 4). Upon *Hs*Cstp production, cells developed additional ER membranes compared to non-producing cells. Coexpression of *PpHAC1*^{i} triggered an even more pronounced proliferation of membrane stacks. So, this correlation suggested that the beneficial effects exerted by *PpHAC1*^{i} coexpression could be caused by an expansion of the ER membrane space.

Transcriptional response to PpHAC1^{i} *coexpression under MP production*

As *PpHAC1*^{i} coexpression had shown a beneficial effect on the production of all ER resident MPs tested (Table 3), we investigated the transcriptional changes using microarrays to get a better understanding of the underlying mechanisms and factors involved.

Recently, a transcriptome study investigated the role of the UPR in *P. pastoris* [26]. Using microarrays, the transcriptome of a strain expressing the *HAC1* homolog from *S. cerevisiae* was compared to the wild-type strain under normal growth and stress conditions (dithiothreitol treatment). A related study had investigated the effect of *ScHAC1*^{i} coexpression in *P. pastoris* on the production of a secreted antibody Fab fragment [79]. The transcriptional regulation of more than 50 genes was analysed by TRAC (transcript analysis with aid of affinity capture) and compared to *S. cerevisiae*. However, neither data set [26,79] provided information on the effect of *P. pastoris* endogenous *HAC1*^{i} on a MP expressing strain.

Therefore we selected the strain with the strongest improvement upon *PpHAC1*^{i} coexpression for transcriptome studies. The *HsCST* + *PpHAC1*^{i} coexpressing strain had shown a 2.1-fold increase in protein yield (Table 3) and exhibited Karmellae formation in electron microscopy (Fig. 4). We compared the transcriptomes of the coexpressing strain with the parental strain expressing only *HsCST*. More than 1300 genes displayed an altered abundance with an approximately equal balance between upregulated (621) and downregulated genes (683) (Table 4 and S5).

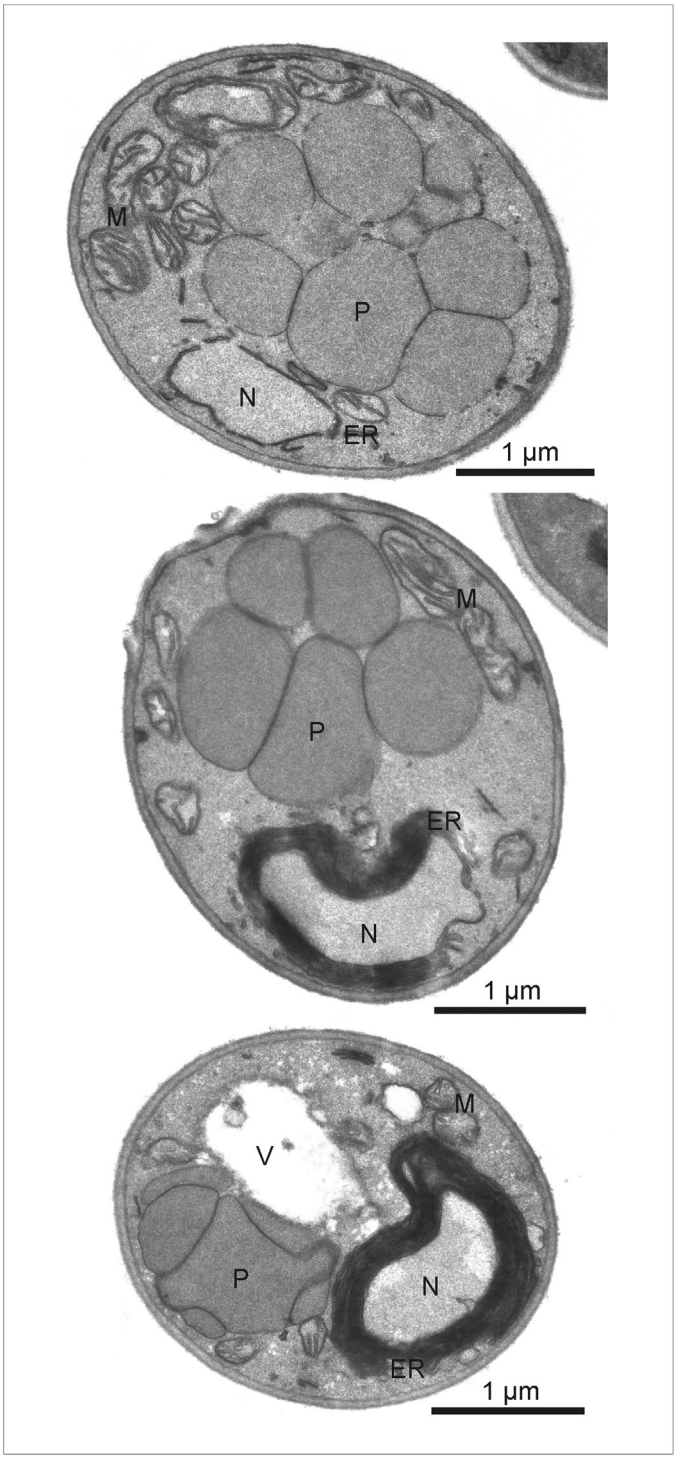


FIGURE 4

Transmission electron micrographs of ultrathin sections of *P. pastoris* wild-type **(a)**, *HsCST* expressing **(b)** and *HsCST* + *PpHAC1*^{i} co-expressing **(c)** cells. Different organelles have been assigned showing peroxisomes (P), endoplasmic reticulum (ER), nucleus (N), mitochondria (M) and vacuoles (V).

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

TABLE 4

Differently regulated genes upon *HAC1*^{i} coexpression in the *Hs*Cstp expressing strain. Same grouping and labelling as Table 2. Superscript numbers indicate genes that have also been observed by other microarray analyses in studies addressing effects of Hac1p production in yeasts and fungi (1: [26], 2: [81], 3: [80], 4: [82]).

Function	Differently regulated genes		Reg.
Protein folding	Folding in the ER	*CNE1*[12], *ERO1*[1234], *PDI1*[1234], *ERV2*, *KAR2/BiP*[123], *LHS1*[1234], *JEM1*[14], *HLJ1*, *SCJ1*[1234], *AHA1*	↑
Protein glycosylation	Core glycosylation	*DIE2*[1], *DPM1*[14]	↑
	Oligosaccharyl transferase complex	*OST1*[12], *OST2*[1234], *OST3*[14], *STT3*[1], *SWP1*[134], *WBP1*[124]	↑
	Glycoprotein processing	*ALG1*, *AGL2*[12], (*ALG7*)[124], *ALG14*, *EOS1*[1], *KTR2*, *MNN2*[1]	↑
		MNN4	↓
	GPI-anchor biosynthesis	*GPI1*, *GPI2*[1], (*GPI14*)[1], *GPI19*, *MCD4*[134], *SMP3*, *GAA1*[4]	↑
	O-glycosylation	*PMT1*[13], *PMT2*[134], *PMT4*[1], *PMT6*[1]	↑
Protein Translocation	Translocon pore	*SEC61*[134]	↑
	Posttranslational translocation	*SEC63* complex: *SEC63*[12], *SEC66*[4], *SEC72*[123]; *LHS1*[1234], *KAR2*[123]	↑
	Cotranslational translocation	*SSH1*	↓
	Signal sequence processing	Signal peptidase complex: *SPT2*, *SEC11*[24]; *STE24*[4]	↑
	ER-to-Golgi transport, COPII-mediated vesicles	*COPII* vesicles: *SEC13*[4], (*SEC16*)[4], *SEC23*[1], *SEC24*[14], *SEC31*, *SFB2*[14], *ERV29*[13], *YIP3*[13], *SAR1*, *EMP24*; TRAPP complex: *BET3*, *TRS33*; SNAREs: *SED5*, *BET1*; other genes: *SEC12*[134], *SEC17*, (*SEC18*), *ERV25*[4], *SLY1*, *GYP1*, *GYP7*, *GYP8*, *COG4*, *GCS1*, *GOT1*[1], *ERP5*	↑
		TRS31, *COG3*	↓
	COPI-mediated retrograde transport	*RET2*[14], *SEC20*, *SEC21*[1], *SEC26*[4], *SEC28*, *GET1*, *GET3*, *USE1*	↑
	Vesicular transport	*VPS55*, *VPS45*, *VPS64*, *VPS62*, *VPS68*, *VPS9*, *VPS15*, *VPS64*, *VPS62*, *PEP8*, *ATG3*, *ATG8*, *ATG13*, *ATG18*, *VTI1*, *YPT52*, *VAC8*, *RCR1*[1]	↑
	Nuclear pore complex	*NMD3*, *NSP1*, *NOP53*, *NUP2*, *NUP84*	↓
	Mitochondrial transport machinery	*SAM50*, *MDM10*, *MMM1*	↑
		TOM20, *TOM40*, *TIM9*, *TIM12*, *TIM13*, *TIM22*, *TIM23*, *TIM44*	↓
Degradation	ERAD	*EPS1*[1], *HLJ1*, *HRD1*[134], *HRD3*[4], *MNL1*[1]	↑
	Proteasome	*RPN8*, *NAS6*, *NAS2*	↑
	Ubiquitination	*DOA4*[4], *VPS25*, *UBP12*, *SNF8*, *RAD6*, *UBC8*, *SLX5*, *HUL4*, *UFO1*	↑
		UBC11, *HUB1*, *BRE5*, *CDC53*	↓
Protein biosynthesis	rRNA processing	*POP4*, *POP3*	↑
		NOP7, *ERB1*, *SPB1*, *FCF2*, *UTP7*, *LCP5*, *UTP13*, *SAS10*, *RCL1*, *MPP10*, *MAK16*, *RRP14*, *UTP5*, *UTP4*, *EBP2*, *RNT1*, *DHR2*, *UTP22*, *PWP2*, *PWP1*, *NOP9*, *ENP1*, *ROK1*, *UTP15*, *UTP6*, *TSR1*, *NSR1*, *NOP10*, *RPF2*, *MRD1*, *RRP9*, *NOP4*, *PXR1*, *EFG1*, *NAN1*, *RRP3*, *SNU13*, *KRR1*, *ESF1*, *IPI1*, *NOP1*, *IPI3*, *UTP30*, *RRP1*, *GAR1*, *BMS1*, *UTP21*, *POP1*, *DIP2*, *RIO1*, *MTR3*, *FHL1*, *TSR2*, *EMG1*, *DBP8*, *RRP5*, *UTP14*, *DBP2*, *FAL1*, *RRP7*	↓
	tRNA processing	*MSY1*, *POA1*, *THG1*	↑
		TRM1, *PUS1*, *TRM2*, *PPM2*, *TRM8*, *TRZ1*, *TRM7*, *LHP1*, *NCL1*, *TRM10*, *TAD2*, *TAD3*, *GUS1*, *TRM11*, *ARC1*, *CDC60*, *TRM112*, *CCA1*, *YDR341C*, *RSC9*, *TRM9*	↓
	Ribosome biogenesis and assembly	*RRS1*, *YTM1*, *SPB1*, *NOP7*, *NUG1*, *UTP23*, *NOB1*, *RCL1*, *NOP2*, *NIP7*, *SPB1*, *RRP14*, *DBP3*, *UTP5*, *UTP4*, *RRP14*, *EBP2*, *BRX1*, *NOC3*, *ALB1*, *NSA1*, *NOG1*, *RPF1*, *NOP6*, *NOP13*, *NOP9*, *JJJ1*, *ENP1*, *NOG2*, *HAS1*, *NSR1*, *RPF2*, *MRD1*, *MAK11*, *RRP12*, *NOP4*, *SSF2*, *KRI1*, *ARX1*, *NMD3*, *NOP53*, *KRR1*, *NOP14*, *RLP24*, *SIK1*, *UTP30*, *DRS1*, *RRP1*, *DBP9*, *DBP10*, *BMS1*, *UTP21*, *ARB1*, *LTV1*, *TMA23*, *NOP16*, *RRB1*, *RPL6B*, *GAR2*, *EMG1*, *DBP8*, *UTP14*, *RLP7*, *MRT4*, *NSA2*, *FAP7*, *MAK21*, *RRP17*, *RRP7*, *TIF6*	↓
	Splicing, snRNPs	*CLF1*, *PRP4*, *MSL1*, *LUC7*, *CEF1*, *ISY1*, *SAP49*, *SNP1*	↑
		PRP42, *PRP24*, *PRP43*, *SNU114*, *SMB1*, *PRP21*	↓
	Translation initiation factor	*TIF4631*, *NIP1*, *PRT1*, *RLI1*, *RPG1*, *SUI2*, *SUI3*, *TIF11*, *TIF3*, *TIF34*, *TIF6*, *EIF3H*, *EIF3M*, *FUN12*, *GCD11*, *GCN3*, *INT6*	↓
	Transcription initiation factor	*TAF9*, *TAF13*, *TFA1*, *TAF6*, *TFC3*, *RRN11*, *SPT15*	↑
		RRN3, *RRN7*	↓
	Mitochondrial transcription/translation	*PUS9*, *MSW1*	↑
		MSS116, *DSS1*, *NAM2*, *PET127*, *RPO41*, *MRPL9*, *YML6*, *MST1*, *MRPS28*, *SUV3*, *PET309*, *NUC1*	↓
	RNA polymerases	*CDC36*, *RPB11*, *TFC3*, *CTK3*, *RPB7*, *LSM1*, *MED11*, *SSN8*, *PSH1*, *RRN11*, *SOH1*, *SSN3*	↑
		RPA43, *RPC53*, *RPA49*, *RPA2*, *RRN3*, *RPA34*, *RPA12*, *RPC37*, *RPC10*, *RPB5*, *RPO31*, *RPC31*, *RPB8*, *RRN7*	↓
	mRNA processing	*PFS2*, *POP2*, *SPT2*	↑
		RRP45, *STO1*, *PAP2*, *NAM7*, *CDC39*, *RAT1*	↓
	DNA repair	*RAD9*	↑
		PRI1	↓
	DNA polymerase	*POL2*, *POL5*	↓

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

TABLE 4 (*Continued*)

Function	Differently regulated genes		Reg.
Lipid biosynthesis	Fatty acid biosynthesis	*FAS2*, *CEM1*, (*FAS3*), *MUQ1*	↑
		SUR4, (*FAT1*), *FAS1*, *CPT1*	↓
	GPI biosynthesis	*GPI1*, *GPI2*[1], (*GPI14*)[1], *GPI19*, *MCD4*[34], *SMP3*, *GAA1*[4]	↑
	Ergosterol biosynthesis	(*ERG1*[2]), *ERG2*, *ERG3*, *ERG5*, (*ERG6*), *ERG13*, *NCP1*	↓
	β-Oxidation	*POX1*	↓
	Phospholipid biosynthesis	*SLC14*, *INO4*, *IPK1*, *DPP1*, *VPS34*, *VAC14*, *FAB1*, *SEC14*	↑
		INO1	↓

Research Paper

We noticed that many secretory genes including the classical UPR targets (as described in [26,80–82]) showed altered regulation upon *PpHAC1*^{i} coexpression. As expected, we observed a general upregulation of genes coding for ER-resident chaperones assisting in protein folding, including calnexin *CNE1*, thiol-oxidases *ERO1*, *PDI1* and *ERV2* (all being involved in formation of disulfide bridges), peptidyl-prolyl-*cis-trans* isomerase (*CPR5*), *KAR2*/*BiP*, *SCJ1* and *LHS1*, which play a role in protein translocation and folding, the Hsp40 chaperone *JEM1* and the stress-related co-chaperone *AHA1* [83–88]. Apart from the folding machinery, genes participating in ER-associated protein degradation (ERAD) were induced, including *EPS1*, the Hsp40 co-chaperone *HLJ1* that mediates degradation of MPs, *HRD1*, *HRD3* and *MNL1* [89–94]. Despite the upregulation of these genes leading to proteolysis, *Hs*Cstp yields have not been negatively affected but even 2.1-fold increased (Table 3). The upregulation of the chaperones mentioned above could compensate for the ERAD, or the ERAD was induced after the membrane space in the ER had been filled with *Hs*Cstp, leading only to a degradation of excessive protein that could not be accommodated in the ER. Remarkably, a few genes involved in protein folding were repressed during *HsCST* + *PpHAC1*^{i} expression. Repressed genes included *EGD2*, coding for the α-subunit of the heteromeric nascent polypeptide-associated complex (NAC), which is involved in protein sorting and translocation in *S. cerevisiae*; the Hsp40 chaperone *ERJ5* and *FMO1* encoding a flavin monooxygenase that maintains the ER redox buffer ratio for correct folding of disulfide-bonded proteins; the cytoplasmic co-chaperone *CNS1* and the cytoplasmic ribosome-associated Hsp40 homologue *ZUO1* and its HSP70-like interaction partner *SSZ1* which stabilize nascent polypeptide chains.

Besides these chaperones, many induced genes were involved in various aspects of glycosylation, including both N-linked and O-linked glycosylation, GPI-anchor biosynthesis and folding of glycoproteins (Table 4). Transcription of many other endogenous MPs of the ER was also induced. Furthermore, many genes assisting in various aspects of secretion were induced during *PpHAC1*^{i} coexpression. The protein translocation machinery in the ER was upregulated including components of the SEC translocon pore (*SEC61*), post-translational (*SEC63*, *SEC66*, *SEC72*) and co-translational (*SSH1*) translocation across the ER membrane and ER-resident import proteins (*LHS1*, *KAR2*). Among genes that were involved in protein processing in the ER, the HDEL receptor *ERD2* that mediates retention of ER-targeted proteins, and genes whose products process signal sequences (*SPT2*, *SEC11*, *STE24*) were more abundant during *PpHAC1*^{i} coexpression. Along the secretory pathway, genes involved in the ER-to-Golgi transport via COPII vesicles, clathrin complex compounds, compounds of the post-Golgi secretory vesicles and the exocyst complex were induced (Table 4). The upregulation of these genes involved in secretion would be a possible mechanistic explanation for the approximately twofold improved yields of secretory proteins upon *PpHAC1*^{i} co-expression observed previously by Guerfal *et al.* [74].

Apart from increased transcription of genes of the secretory pathway, targeting to endosomal and vacuolar compartments was upregulated as well, judging from transcript abundance of the ESCRTII and ESCRTIII complexes which mediate sorting of aberrant transmembrane proteins to the endosome and multivesicular body path for degradation. Many genes of vacuolar protein sorting and autophagy families and of other vacuolar proteins were induced during *PpHAC1*^{i} coexpression (Table 4). The upregulation of genes involved in vacuolar processes upon *PpHAC1*^{i} coexpression correlated also with their marked formation seen in electron micrographs (Fig. 4).

Though many genes assisting in the secretion process were induced as a consequence of *PpHAC1*^{i} coexpression, genes coding for secreted or cell wall associated proteins were repressed, similar to the response in *S. cerevisiae* [95]. The increased protein flux into the ER, caused by heterologously expressed *Hs*Cstp and *Pp*Hac1p-induced target proteins, might have led to a downregulatation of endogenous secretory genes to reduce the protein load in the ER, thus alleviating ER stress.

No clear trend regarding protein catabolism could be observed, as compounds of the proteasome biogenesis and ubiquitination pathways were partly upregulated and partly downregulated.

With regard to general cellular 'housekeeping' processes, carbohydrate metabolism, protein folding, transport, transcription, vesicular transport and protein catabolism were upregulated, whereas anabolic processes (for example, amino acid biosynthesis) and respiration were repressed. In this regard, the overall effect of *PpHAC1*^{i} coexpression largely overlapped with the transcript profile observed in *P. pastoris* during constitutive expression of *ScHAC1*^{i}, though different genes and promoter systems have been employed [26].

However, unlike Graf *et al.* [26], we observed a downregulation of genes with functions in organelle biogenesis, ribosome biogenesis including those involved in rRNA processing (RRP family genes) and export and assembly of ribosome subunits. Mitochondrial genes (transport machineries, ribosomes, initiation and elongation factors) were downregulated as well. The repression of ribosome biogenesis and translation with the concomitant decrease in organelle biogenesis indicated a global slowdown in cellular metabolism.

Concerning lipid metabolism, we noticed a downregulation of sterol and ergosterol biosynthesis genes (*ERG2*, *ERG3*, *ERG5*, *ERG6*, *ERG13*, *NCP1*), which has also been reported by Graf *et al.*, and an

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

upregulation of PI and GPI biosynthesis (Table 4). Surprisingly, *INO1*, which had been upregulated in *Pp*Aodp, *Hs*Cstp and *Sc*Ctr3p producing strains, was downregulated in the *HsCST* + *PpHAC1*^{i} coexpressing strain.

In summary, neither translocon compounds nor ER-resident chaperones were upregulated during initial *Hs*Cstp production, whereas *PpHAC1*^{i} coexpression led to a significant upregulation of genes involved in the respective processes. Therefore, we suggest that the improved yields caused by *PpHAC1*^{i} coexpression can be attributed to a synergistic upregulation of the folding machinery and an expansion of the ER membrane capacity, thereby boosting MP expression.

Conclusions

Heterologous expression of eukaryotic MPs in reasonable amounts is still a bottleneck in both structural and functional studies. We have therefore analysed the transcriptional responses evoked by the overexpression of MPs with different functions and localizations (*Pp*AODp, *Hs*Cstp, *Sc*Ctr3p) in *P. pastoris* to gain better understanding of the underlying cellular processes. Microarrays of strains expressing the mitochondrial protein *Pp*Aod revealed strong transcriptional changes, implying that the entire cell metabolism was reorchestrated to the needs of the *Pp*Aodp-associated shift in energy metabolism. From a prominent upregulation of genes involved in proteolysis and a concomitant downregulation of respiration and TCA cycle genes, the picture emerged that high level *Pp*Aodp production clearly perturbed the cellular homeostasis.

The majority of eukaryotic MPs, including the *Hs*Cstp and *Sc*Ctr3p transporters, enter the secretory pathway, where the ER plays a crucial role in folding, membrane insertion and posttranslational modifications. Apart from protein-related functions, the ER serves as the major site of lipid and sterol biosynthesis.

Overexpression of the ER-resident *Hs*Cstp triggered upregulation of the gene coding for the transcription factor Hac1p, which plays a key role in lipid metabolism apart from its well-known function in the ER resident quality control system. In yeast, simultaneous Hac1p production is known to increase production of secreted heterologous proteins [74,75,77]. Regarding MPs, so far *HAC1* coexpression in *P. pastoris* had only been proven useful in terms of improving homogeneity (for example, better processing of signal sequences), but no effect on yields have been observed [74]. Here we have investigated to what extent and how coexpression of *PpHAC1*^{i} affects productivity during MP production in *P. pastoris*. Upon *PpHAC1*^{i} coexpression, a 1.5-fold to 2.1-fold increase relative to the progenitor strains was observed for the ER-resident MPs tested (*Hs*Cstp, *Hs*Ctr1p and *Os*Cstp). By investigating the transcriptional effect of *PpHAC1*^{i} coexpression on MP production using microarrays, we noticed a coordinated transcriptional upregulation of genes responsible for both folding and membrane capacity within the ER.

In conclusion, this transcriptome study led to new insights into how *P. pastoris* responds to cellular constraints caused by the overexpression of integral MPs. In addition to elucidation of the beneficial role of *HAC1* coexpression for ER resident MP expression, the cellular lipid metabolism and organelle homeostasis emerged as promising targets, bearing considerable potential for rational strain engineering to further improve MP production in *P. pastoris*.

Acknowledgements

Financial support for this project was offered by the Fonds zur Wissenschaftlichen Forschung (FWF) through the PhD program 'DK Molecular Enzymology' (W901-B05) and in part by the Austrian Ministry of Science and Research GEN-AU project BIN (FFG Grant 820962) and the Austrian COMET program managed by FFG. Antibodies were a generous gift of Prof. Guenther Daum (Institute of Biochemistry, Graz University of Technology). We gratefully acknowledge the cooperation of A. Jay Sunga (Keck Graduate Institute, Claremont Colleges) and Joel Kreps (Verenium Corp., San Diego) in microarray studies. We also thank Rui M. C. Portela (Universidade Nova de Lisboa, Faculty of Science and Technology) for the generation of the Venn diagrams and Peter Krempl (Institute for Genomics and Bioinformatics, Graz University of Technology) for assistance in submitting the microarray data. Furthermore we highly appreciate the fruitful collaboration with Dr. Heimo Wolinski and Prof. Sepp-Dieter Kohlwein (Institute of Molecular Biosciences, University of Graz) for their invaluable support in fluorescence microscopy studies. Finally, the authors thank Prof. Nico Callewaert for providing information on the *PpHAC1* sequence.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.nbt.2014.02.009.

References

[1] Ashe MP, Bill RM. Mapping the yeast host cell response to recombinant membrane protein production: relieving the biological bottlenecks. Biotechnol J 2011;6:707–14.
[2] Bernhardt R. Cytochromes P450 as versatile biocatalysts. J Biotechnol 2006;124:128–45.
[3] Loll PJ. Membrane protein structural biology: the high throughput challenge. J Struct Biol 2003;142:144–53.
[4] Freigassner M, Pichler H, Glieder A. Tuning microbial hosts for membrane protein production. Microb Cell Fact 2009;8:69–90.
[5] White MA, Clark KM, Grayhack EJ, Dumont ME. Characteristics affecting expression and solubilization of yeast membrane proteins. J Mol Biol 2007; 365:621–36.
[6] André N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, et al. Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. Protein Sci 2006;15:1115–26.
[7] Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, et al. Consequences of membrane protein overexpression in *Escherichia coli*. Mol Cell Proteomics 2007;6:1527–50.
[8] Bawa Z, Bland CE, Bonander N, Bora N, Cartwright SP, Clare M, et al. Understanding the yeast host cell response to recombinant membrane protein production. Biochem Soc Trans 2011;39:719–23.
[9] Bonander N, Bill RM. Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production. Expert Rev Proteomics 2009;6:501–5.
[10] Bonander N, Darby RA, Grgic L, Bora N, Wen J, Brogna S, et al. Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. Microb Cell Fact 2009;8:10.
[11] Daly R, Hearn MTW. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J Mol Recognit 2005;18:119–38.
[12] Vogl T, Hartner FS, Glieder A. New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr Opin Biotechnol 2013;24:1094–101.
[13] Öberg F, Sjöhamn J, Conner MT, Bill RM, Hedfalk K. Improving recombinant eukaryotic membrane protein yields in *Pichia pastoris*: the importance of codon optimization and clone selection. Mol Membr Biol 2011;28:398–411.

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

[14] Ramón A, Marín M. Advances in the production of membrane proteins in *Pichia pastoris*. Biotechnol J 2011;6:700–6.
[15] Bornert O, Alkhalfioui F, Logez C, Wagner R. Overexpression of membrane proteins using *Pichia pastoris*. Curr Protoc Protein Sci 2012. Chapter 29:Unit 29.2.
[16] Singh S, Gras A, Fiez-Vandal C, Martinez M, Wagner R, Byrne B. Large-scale production of membrane proteins in *Pichia pastoris*: the production of G protein-coupled receptors as a case study. Methods Mol Biol 2012;866:197–207.
[17] Krettler C, Reinhart C, Bevans CG. Expression of GPCRs in *Pichia pastoris* for structural studies. Methods Enzymol 2013;520:1–29.
[18] Oberg F, Hedfalk K. Recombinant production of the human aquaporins in the yeast *Pichia pastoris*. Mol Membr Biol 2013;30:15–31 [Invited Review].
[19] Hedfalk K. Further advances in the production of membrane proteins in *Pichia pastoris*. Bioengineered 2013;4:1–5.
[20] Gonçalves AM, Pedro AQ, Maia C, Sousa F, Queiroz JA, Passarinha LA. *Pichia pastoris*: a recombinant microfactory for antibodies and human membrane proteins. J Microbiol Biotechnol 2013;23:587–601.
[21] Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, et al. Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. Microb Cell Fact 2004;3:17–30.
[22] Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D. Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl Environ Microbiol 2007;73:6499–507.
[23] Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. Microb Cell Fact 2009;8:29.
[24] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 2009;27:561–6.
[25] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of *Pichia pastoris* CBS7435. J Biotechnol 2011; 154:312–20.
[26] Graf A, Gasser B, Dragosits M, Sauer M, Leparc GG, Tüchler T, et al. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. BMC Genomics 2008;9:390.
[27] Vogl T, Glieder A. Regulation of *Pichia pastoris* promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.
[28] Hesketh AR, Castrillo JI, Sawyer T, Archer DB, Oliver SG. Investigating the physiological response of *Pichia* (Komagataella) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. Appl Microbiol Biotechnol 2013;97:9747–62.
[29] Vanz AL, Lünsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N, et al. Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. Microb Cell Fact 2012;11:103.
[30] Vanz AL, Nimtz M, Rinas U. Decrease of UPR- and ERAD-related proteins in *Pichia pastoris* during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. Microb Cell Fact 2014;13:23.
[31] Stadlmayr G, Benakovitsch K, Gasser B, Mattanovich D, Sauer M. Genome-scale analysis of library sorting (GALibSo): isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*. Biotechnol Bioeng 2010;105:543–55.
[32] Baumann K, Adelantado N, Lang C, Mattanovich D, Ferrer P. Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. Microb Cell Fact 2011;10:93.
[33] Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, et al. A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. BMC Syst Biol 2010;4:141.
[34] Baumann K, Dato L, Graf AB, Frascotti G, Dragosits M, Porro D, et al. The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* – a comparative analysis. BMC Genomics 2011;12:218.
[35] Newstead S, Kim H, von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 2007; 104:13936–41.
[36] Pscheidt B, Liu Z, Gaisberger R, Avi M, Skranc W, Gruber K, et al. Efficient biocatalytic synthesis of (R)-pantolactone. Adv Synth Catal 2008;350:1943–8.
[37] Pscheidt B, Glieder A. Yeast cell factories for fine chemical and API production. Microb Cell Fact 2008;7:25.
[38] Liu Z, Pscheidt B, Avi M, Gaisberger R, Hartner FS, Schuster C, et al. Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. Chembiochem 2008;9:58–61.
[39] Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, et al. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 2005;38(44):46–8.
[40] Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. Nat Protoc 2008;3:784–98.
[41] Drew D, Slotboom D-J, Friso G, Reda T, Genevaux P, Rapp M, et al. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci 2005;14:2011–7.
[42] Ivashov VA, Grillitsch K, Koefeler H, Leitner E, Baeumlisberger D, Karas M, et al. Lipidome and proteome of lipid droplets from the methylotrophic yeast *Pichia pastoris*. Biochim Biophys Acta 2013;1831:282–90.
[43] Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005. p. 397–420.
[44] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
[45] Drew D, Lerch M, Kunji E, Slotboom D-J, de Gier J-W. Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods 2006;3:303–13.
[46] Brooks CL, Morrison M, Joanne Lemieux M. Rapid expression screening of eukaryotic membrane proteins in *Pichia pastoris*. Protein Sci 2013;22:425–33.
[47] Mizutani K, Yoshioka S, Mizutani Y, Iwata S, Mikami B. High-throughput construction of expression system using yeast *Pichia pastoris*, and its application to membrane proteins. Protein Expr Purif 2011;77:1–8.
[48] Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 2006; 14:673–81.
[49] Osterberg M, Kim H, Warringer J, Melén K, Blomberg A, von Heijne G. Phenotypic effects of membrane protein overexpression in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 2006;103:11148–53.
[50] Kern A, Hartner FS, Freigassner M, Spielhofer J, Rumpf C, Leitner L, et al. *Pichia pastoris* just in time alternative respiration. Microbiology 2007;153:1250–60.
[51] Veiga A, Arrabaça JD, Loureiro-Dias MC. Stress situations induce cyanide-resistant respiration in spoilage yeasts. J Appl Microbiol 2003;95:364–71.
[52] Yukioka H, Inagaki S, Tanaka R, Katoh K, Miki N, Mizutani A, et al. Transcriptional activation of the alternative oxidase gene of the fungus *Magnaporthe grisea* by a respiratory-inhibiting fungicide and hydrogen peroxide. Biochim Biophys Acta 1998;1442:161–9.
[53] Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 2000;11:4241–57.
[54] Godon C, Lagniel G, Lee J, Buhler JM, Kieffer S, Perrot M, et al. The H_2O_2 stimulon in *Saccharomyces cerevisiae*. J Biol Chem 1998;273:22480–89.
[55] Wagner I, Arlt H, van Dyck L, Langer T, Neupert W. Molecular chaperones cooperate with PIM1 protease in the degradation of misfolded proteins in mitochondria. EMBO J 1994;13:5135–45.
[56] Nijtmans LG, de Jong L, Artal Sanz M, Coates PJ, Berden JA, Back JW, et al. Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins. EMBO J 2000;19:2444–51.
[57] Radke S, Chander H, Schäfer P, Meiss G, Krüger R, Schulz JB, et al. Mitochondrial protein quality control by the proteasome involves ubiquitination and the protease Omi. J Biol Chem 2008;283:12681–85.
[58] Bonander N, Hedfalk K, Larsson C, Mostad P, Chang C, Gustafsson L, et al. Design of improved membrane protein production experiments: quantitation of the host response. Protein Sci 2005;14:1729–40.
[59] Wu P-YJ, Ruhlmann C, Winston F, Schultz P. Molecular architecture of the *S. cerevisiae* SAGA complex. Mol Cell 2004;15:199–208.
[60] Becker F, Block-Alper L, Nakamura G, Harada J, Wittrup KD, Meyer DI. Expression of the 180-kD ribosome receptor induces membrane proliferation and increased secretory activity in yeast. J Cell Biol 1999;146:273–84.
[61] Cox JS, Chapman RE, Walter P. The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Mol Biol Cell 1997;8:1805–14.
[62] Wright R, Basson M, D'Ari L, Rine J. Increased amounts of HMG-CoA reductase induce karmellae: a proliferation of stacked membrane pairs surrounding the yeast nucleus. J Cell Biol 1988;107:101–14.
[63] Menzel R, Vogel F, Kärgel E, Schunck WH. Inducible membranes in yeast: relation to the unfolded-protein-response pathway. Yeast 1997;13:1211–29.
[64] Schunck WH, Vogel F, Gross B, Kärgel E, Mauersberger S, Köpke K, et al. Comparison of two cytochromes P-450 from *Candida maltosa*: primary structures, substrate specificities and effects of their expression in *Saccharomyces cerevisiae* on the proliferation of the endoplasmic reticulum. Eur J Cell Biol 1991;55:336–45.
[65] Takewaka T, Zimmer T, Hirata A, Ohta A, Takagi M. Null mutation in IRE1 gene inhibits overproduction of microsomal cytochrome P450Alk1 (CYP 52A3) and proliferation of the endoplasmic reticulum in *Saccharomyces cerevisiae*. J Biochem 1999;125:507–14.
[66] Wiedmann B, Silver P, Schunck WH, Wiedmann M. Overexpression of the ER-membrane protein P-450 CYP52A3 mimics sec mutant characteristics in *Saccharomyces cerevisiae*. Biochim Biophys Acta 1993;1153:267–76.
[67] Wanker EE, Sun Y, Savitz AJ, Meyer DI. Functional characterization of the 180-kD ribosome receptor in vivo. J Cell Biol 1995;130:29–39.
[68] Vergères G, Yen TS, Aggeler J, Lausier J, Waskell L. A model system for studying membrane biogenesis. Overexpression of cytochrome b5 in yeast results in marked proliferation of the intracellular membrane. J Cell Sci 1993;106(Pt 1):249–59.
[69] Nishikawa S, Hirata A, Nakano A. Inhibition of endoplasmic reticulum (ER)-to-Golgi transport induces relocalization of binding protein (BiP) within the ER to form the BiP bodies. Mol Biol Cell 1994;5:1129–43.
[70] Schiller H, Haase W, Molsberger E, Janssen P, Michel H, Reiländer H. The human ET(B) endothelin receptor heterologously produced in the methylotrophic yeast *Pichia pastoris* shows high-affinity binding and induction of stacked membranes. Recept Channels 2000;7:93–107.
[71] Grünewald S, Haase W, Molsberger E, Michel H, Reiländer H. Production of the human D2S receptor in the methylotrophic yeast *P. pastoris*. Recept Channels 2004;10:37–50.
[72] Licata L, Haase W, Eckhardt-Strelau L, Parcej DN. Over-expression of a mammalian small conductance calcium-activated K+ channel in *Pichia pastoris*: effects of trafficking signals and subunit fusions. Protein Expr Purif 2006;47:171–8.

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

[73] Shukla AK, Haase W, Reinhart C, Michel H. Heterologous expression and characterization of the recombinant bradykinin B2 receptor using the methylotrophic yeast *Pichia pastoris*. Protein Expr Purif 2007;55:1–8.
[74] Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, et al. The HAC1 gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Fact 2010;9:49–60.
[75] Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D. Engineering of *Pichia pastoris* for improved production of antibody fragments. Biotechnol Bioeng 2006;94:353–61.
[76] Griffith DA, Delipala C, Leadsham J, Jarvis SM, Oesterhelt D. A novel yeast expression system for the overproduction of quality-controlled membrane proteins. FEBS Lett 2003;553:45–50.
[77] Valkonen M, Penttilä M, Saloheimo M. Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. Appl Environ Microbiol 2003;69:2065–72.
[78] Breinig F, Diehl B, Rau S, Zimmer C, Schwab H, Schmitt MJ. Cell surface expression of bacterial esterase A by *Saccharomyces cerevisiae* and its enhancement by constitutive activation of the cellular unfolded protein response. Appl Environ Microbiol 2006;72:7140–7.
[79] Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, et al. Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. BMC Genomics 2007;8:179.
[80] Kimata Y, Ishiwata-Kimata Y, Yamada S, Kohno K. Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. Genes Cells 2006;11:59–69.
[81] Guillemette T, van Peij NNME, Goosen T, Lanthaler K, Robson GD, van den Hondel CAMJJ, et al. Genomic analysis of the secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. BMC Genomics 2007;8:158.
[82] Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell 2000;101:249–58.
[83] Parlati F, Dominguez M, Bergeron JJ, Thomas DY. *Saccharomyces cerevisiae* CNE1 encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus. J Biol Chem 1995;270:244–53.
[84] Frand AR, Kaiser CA. The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. Mol Cell 1998;1:161–70.
[85] Silberstein S, Schlenstedt G, Silver PA, Gilmore R. A role for the DnaJ homologue Scj1p in protein folding in the yeast endoplasmic reticulum. J Cell Biol 1998;143:921–33.
[86] Baxter BK, James P, Evans T, Craig EA. SSI1 encodes a novel Hsp70 of the *Saccharomyces cerevisiae* endoplasmic reticulum. Mol Cell Biol 1996;16:6444–56.
[87] Saris N, Holkeri H, Craven RA, Stirling CJ, Makarow M. The Hsp70 homologue Lhs1p is involved in a novel function of the yeast endoplasmic reticulum, refolding and stabilization of heat-denatured protein aggregates. J Cell Biol 1997;137:813–24.
[88] Nishikawa S, Endo T. The yeast JEM1p is a DnaJ-like protein of the endoplasmic reticulum membrane required for nuclear fusion. J Biol Chem 1997;272:12889–92.
[89] Lotz GP, Lin H, Harst A, Obermann WMJ. Aha1 binds to the middle domain of Hsp90, contributes to client protein activation, and stimulates the ATPase activity of the molecular chaperone. J Biol Chem 2003;278:17228–35.
[90] Wang Q, Chang A. Substrate recognition in ER-associated degradation mediated by Eps1, a member of the protein disulfide isomerase family. EMBO J 2003;22:3792–802.
[91] Wang Q, Chang A. Eps1, a novel PDI-related protein involved in ER quality control in yeast. EMBO J 1999;18:5972–82.
[92] Huyer G, Piluek WF, Fansler Z, Kreft SG, Hochstrasser M, Brodsky JL, et al. Distinct machinery is required in *Saccharomyces cerevisiae* for the endoplasmic reticulum-associated degradation of a multispanning membrane protein and a soluble luminal protein. J Biol Chem 2004;279:38369–78.
[93] Hampton RY, Gardner RG, Rine J. Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. Mol Biol Cell 1996;7:2029–44.
[94] Nakatsukasa K, Nishikawa S, Hosokawa N, Nagata K, Endo T. Mnl1p, an alpha-mannosidase-like protein in yeast *Saccharomyces cerevisiae*, is required for endoplasmic reticulum-associated degradation of glycoproteins. J Biol Chem 2001;276:8635–8.
[95] Kimata Y, Kimata YI, Shimizu Y, Abe H, Farcasanu IC, Takeuchi M, et al. Genetic evidence for a role of BiP/Kar2 that regulates Ire1 in response to accumulation of unfolded proteins. Mol Biol Cell 2003;14:2559–69.

Please cite this article in press as: Vogl, T. et al., Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression, New Biotechnol. (2014), http://dx.doi.org/10.1016/j.nbt.2014.02.009

Chapter 3.5.2
Overexpression of transcription factors from a derepressed promoter enables bypassing of methanol dependent carbon source regulation in the methylotrophic yeast *Pichia pastoris*

This work was based on the Diploma thesis of Lukas Sturmberger.

Title page

Title

Overexpression of transcription factors from a derepressed promoter enables bypassing of methanol dependent carbon source regulation in the methylotrophic yeast *Pichia pastoris*

Author list

Thomas Vogl#, Lukas Sturmberger#, Patrick Hyden, Pia C. Fauland, Gerhard Thallinger, Anton Glieder*

# These authors contributed equally.
* Corresponding author

Abstract

Carbon source regulated promoters are well-studied standard tools to control gene expression. In the commonly used protein production host *Pichia pastoris*, methanol-inducible promoters are used because of their tight regulation and exceptional strength. Yet, induction with toxic and flammable methanol is a considerable safety risk in some bigger production plants. Acquiring control over the natural regulation of promoters is important for metabolic engineering and synthetic biology applications.
Here we studied new regulatory circuits based on the *alcohol oxidase 1* promoter (P_{AOX1}), which is tightly repressed in presence of repressing carbon sources and strongly induced by methanol. We overexpressed putative carbon source dependent regulators identified by a homology search in related yeasts. While constitutive overexpression showed only marginal or detrimental effects, derepressed expression (activated when the repressing carbon source is depleted) showed that three factors are singlehandedly suitable to strongly activate P_{AOX1}. Transcriptome analyses demonstrated that these three factors regulate partly overlapping and unique sets of genes. Overexpression of a single factor is sufficient to convert existing P_{AOX1} based expression strains into glucose/gylcerol regulated, methanol-free systems. Given the wide applicability of carbon source regulated promoters and the ease of adopting carbon source feed rates in large scale bioreactors, similar approaches as in *P. pastoris* may also be useful in other organisms.

Introduction

Studies on carbon source regulated promoters in microorganisms have shed light on the basic mechanisms of transcriptional regulation. Early studies on the *Escherichia coli lac* operon (1) and galactose and ethanol regulated promoters in *Saccharomyces cerevisiae* (2) have established the basic paradigms of prokaryotic and eukaryotic gene regulation. With the advent of recombinant protein production, carbon source regulated promoters have also been adapted and widely used to drive the expression of heterologous genes. Typical on/off dynamics achievable by repressing and inducing carbon sources (or non-metabolized surrogates such as Isopropyl-β-D-thiogalactopyranosid, IPTG) are ideal tools for creating biphasic production processes: Cells are at first cultivated on a repressing carbon source without expression of the heterologous gene affecting cell growth. Once a high cell density is reached, switching to an inducing carbon source starts expression of the gene of interest (3). In metabolic engineering and synthetic biology, the expression of multiple genes commonly needs to be regulated either simultaneously, consecutively or even in an inverse mode. Acquiring control over the natural regulation of promoters, the design of synthetic promoters and

artificial transcription factors have proven imperative for optimizing protein production, metabolic engineering and synthetic biology applications (4–9).

The methylotrophic yeast *Pichia pastoris* is a commonly used protein production host and according to a recent literature search even more frequently applied for heterologous protein production than *S. cerevisiae* (10). *P. pastoris* provides favorable properties such as growth to high cell densities and efficient secretion of heterologous proteins (facilitating purification and downstream processing) (11). Glycoengineered *P. pastoris* strains are the only yeast expression platform enabling the production of humanized glycoproteins (12, 13). *P. pastoris* provides also strong and tightly regulated promoters to regulate gene expression (14). Most commonly the methanol inducible promoter of the *alcohol oxidase 1* gene (P_{AOX1}) is used for heterologous protein expression. P_{AOX1} shows three layers of regulation typical for carbon source regulated promoters (14, 15): 1. Repression: The natural Aox1 protein and the *AOX1* transcript are undetectable when cells are grown on a repressing carbon source (glucose, glycerol, ethanol) (14). 2. Derepression: Once the repressing carbon source in the media is depleted, P_{AOX1} shows weak expression of about 2-4% of methanol induced levels. 3. Methanol induction fully activates P_{AOX1} reaching naturally up to 30 % of total intracellular protein. With heterologous proteins, yields up to 22 g/l of intracellular (16) and >18 g/l secreted protein (17) have been achieved. However, especially in large scale bioreactors, substantial amounts of toxic and flammable methanol are required, which is a considerable safety risk. Other promoters of the *P. pastoris* methanol utilization pathway show similarly tight repression as P_{AOX1}, but stronger derepression up to approximately one third of methanol inducible levels (18)/(unpublished results). Synthetic variants of P_{AOX1} were also derepressed (6) and in bioreactor cultivations the derepressed state could be maintained by applying limited feed rates of repressing carbon sources (19). The carbon source is immediately used up by the cells providing energy for protein production but not repressing the promoter. Therefore derepressed expression is a desirable alternative to methanol induction, as it allows simple design of biphasic production processes by varying only the feed rate: Cell growth under repressing conditions can be achieved by providing the repressing carbon source in excess, expression of the derepressed promoter is triggered by limiting the feed rate generating derepressed conditions (6, 19).

However, compared to the model organisms *E. coli* and *S. cerevisiae* relatively little basic information is available on *P. pastoris* (20) and the molecular regulation of methanol inducible promoters is still only partially understood (reviewed in (14)).
In this study, we aimed to alter the regulation of P_{AOX1} by coexpressing transcriptional regulators involved in methanol dependent carbon source regulation. Overexpression of three factors individually activated P_{AOX1} between 7 to 183 % of methanol induced levels (depending on the gene of interest). Based on this finding we generated conversion and enhancer plasmids that can be used to transform existing P_{AOX1} based, methanol dependent strains. Thereby derepressed strains producing the protein of interest (POI) with higher space time yields (depending on the POI) were generated.

Results

Selection of putative regulators and promoters for overexpression

While regulation of bacterial carbon source regulated promoters is typically relying on self-sufficient activators and repressors (e.g. catabolite activator protein [*CAP*] and the lac repressor [*lacI*] for the *E. coli lac* operon), the regulation of eukaryotic promoters such as yeast *GAL* or *ADH* promoters involves an interconnected set of activators, repressors, chromatin remodeling machinery and master regulators such as multi sub unit kinases and phosphatases (2). Knowledge on the exact

factors involved in methanol dependent regulation is limited in *P. pastoris*, a few regulating factors have been reported, yet their exact mechanism are incompletely understood (14). Known activators include Mxr1p (methanol expression regulator 1) (21, 22), Trm1p (transcriptional regulator of methanol induction 1) (23) and Prm1p (positive regulator of methanol 1) (24). Mxr1p is regulated by a 14-3-3 protein (25) and Rop1p (repressor of PEPCK) (26, 27) is antagonistically binding to the same DNA sequence as Mxr1p (27). In addition an involvement of glucose sensors (28) and a hexose transporter (29) in repression of P_{AOX1} was shown.
Methanol dependent gene regulation has also been studied in the related methylotrophic yeasts (15, 30) *Hansenula polymorpha* (31–33) and *Candida boidinii* (34–37), where in part orthologous activators were reported and different gene names assigned ((33, 37, 38), compared in S 2). Also additional factors involved in methanol dependent gene regulation, so far not reported in *P. pastoris*, were found: Chromatin remodeling factors Swi1p and Snf2p are required for methanol dependent gene expression in *H. polymorpha* (32) and repressors orthologous to *S. cerevisiae* Mig1p/Mig2p (Multicopy Inhibitor of *GAL* gene expression) were reported (31, 36).

The conventional yeast *S. cerevisiae* is more distantly related to *P. pastoris* than the methylotrophic yeasts *H. polymorpha* and *C. boidinii* (39, 40). Yet, carbon source regulated gene expression has been extensively studied in this prominent eukaryotic model organism, focusing on the response towards the carbon sources glucose, galactose, ethanol and fatty acids (2). Glucose is the preferred carbon source in *S. cerevisiae* and genes required for the metabolism of other carbon sources (galactose, fatty acids, glycerol) are catabolite-repressed on glucose. The *P. pastoris* methanol master regulator Mxr1p and *S. cerevisiae* Adr1p (alcohol dehydrogenase regulator 1) required for activation of ethanol, glycerol and fatty acid metabolism genes are homologs (21). This relation suggests that the carbon source regulating factors in *P. pastoris* (methanol regulated genes) and *S. cerevisiae* (various glucose repressed genes) have developed from related mechanisms in their last common ancestor. To this end, we extended the search for factors involved in carbon source regulation also to *S. cerevisiae*. Aside the aforementioned Mig repressors, additional carbon source related activators such as Cat8p (for derepression of various genes, in part coregulation with Adr1p) and Gal4p (activation of genes required for galactose utilization) are known. Furthermore the activators/repressors are partly regulated antagonistically by the master kinase Snf1p and the phosphatase Glc7p (requiring its regulatory subunit Reg1p). Also the chromatin remodeling histone deacetylases Rpd3p and Hda1p affected binding of the activators Adr1 and Cat8 (41).
P. pastoris contains clear homologs of all the above mentioned factors from *H. polymorpha*, *C. boidinii* and *S. cerevisiae* (S 2), yet their function and involvement in methanol dependent gene regulation is unknown. We issued a hypothetical model (Fig. 1A, S 3) combing confirmed *P. pastoris* regulators and the well-studied regulation of these homologs in *S. cerevisiae* (S 3A). Promising factors were overexpressed and checked for effects on the regulation of P_{AOX1}.In case of Mxr1p and its *S. cerevisiae* homolog Adr1p, phosphorylation of serine residues has been reported to be involved in regulating activity (25). Therefore we tested also a Mxr1 variant, in which the two serines were mutated to alanine (termed Mxr1c).

We used a reporter construct in which P_{AOX1} is driving expression of a reporter gene (a red florescent protein variant termed dTomato, dTOM (42)). The putative regulators were expressed from the same construct from two different promoters and the plasmids were screened under methanol free conditions for P_{AOX1} activation (Fig. 1).We used the well-established promoter of the glyceraldehyde-3-phosphate dehydrogenase gene (P_{GAP}) (14) and the novel promoter of the *catalase 1* gene (P_{CAT1}), recently reported (18) to drive expression of the putative regulators (Fig. 1B). P_{GAP} is on glucose constitutively active and moderately downregulated on methanol, whereas P_{CAT1} is repressed on glucose and expression starts when the glucose in the medium is depleted (derepression) (Fig. 1C). We included negative controls in which the putative regulator was replaced by a second fluorescent protein (an enhanced green fluorescent protein variant, eGFP). This allowed to rule out a

derepressing effect just by the fusion of P_{AOX1} to a second, differently regulated promoter (Fig. 1B). In addition, the expression profile of the second promoter could be monitored (Fig. 1D). This design of using a weaker/differently regulated promoter to control the strong promoter driving expression of the gene of interest is reminiscent of transcriptional amplification strategy (TAS) used in higher eukaryotes to amplify cell-type-specific regulation (43).

Expression of *TRM1*, *MXR1* or *PRM1* from P_{CAT1} activates P_{AOX1}

The reporter plasmids bearing the putative regulators under control of P_{GAP} and P_{CAT1} were used for transformation of the *P. pastoris* wildtype strain and reporter protein fluorescence was measured under repressed, derepressed and methanol induced conditions (Fig. 1D, S 4). The control constructs containing a second fluorescent protein (eGFP) instead of an activator maintained tight repression under derepressed conditions and could be induced with methanol to identical levels demonstrating that the fusion of P_{AOX1} to the constitutive P_{GAP} or derepressed P_{CAT1} did not affect regulation.

Under derepressed conditions, overexpression of *TRM1*, *MXR1* and *PRM1* from P_{CAT1} led to activation of P_{AOX1} to different extents. P_{CAT1}-*TRM1* and P_{CAT1}-*MXR1* reached surprisingly high levels of the methanol induced control construct, 44 % and 25 % respectively. P_{CAT1}-*TRM1* reached 7 %, still clearly exceeding the background signal. Constitutive expression of some factors from P_{GAP} drastically reduced transformation rates suggesting detrimental effects (discussion and S 4). Only P_{GAP}-*TRM1* and P_{GAP}-*PRM1* seem to show under derepressed conditions weak effects on P_{AOX1} (below 5 % of methanol induced levels). For the other putative regulators, no clear effects on P_{AOX1} under derepressed conditions were noticed.
Under methanol induced conditions, most strikingly expression of *TRM1* either from P_{GAP} or P_{CAT1} showed increased reporter protein fluorescence (1.82- and 1.66-fold of the methanol induced control constructs expressing eGFP). Expression of *MXR1* from P_{CAT1} led to a decrease to 37 % of the methanol induced control construct. We also noticed a growth arrest of this strain (S 5), suggesting detrimental effects of the strong overexpression. This effect is in line with previous reports (21), where overexpression of *MXR1* from P_{AOX1} also abolished growth. The other putative regulators showed on methanol similar reporter protein fluorescence as the control constructs on methanol. In related methylotrophic yeasts the extent of derepression depends on the carbon source present (15). Testing the *P. pastoris* activator strains form Fig. 1D on alternative carbon sources such as glycerol, ethanol and oleic acid did not show any striking differences for derepressed P_{AOX1} activation (S 6). We also did not observe striking differences between *MXR1* and *MXR1c* (two putatively phosphorylated serines mutated), except that P_{CAT1} driven coexpression of *MXR1c* from P_{CAT1} appeared less detrimental on methanol than *MXR1* (83 vs. 37% reporter protein fluorescence of control).
Based on this data, in the further studies we focused our interest on the three *P. pastoris* endogenous activators *TRM1*, *MXR1* and *PRM1*.

We were interested in the natural regulation of the three factors and therefore cloned the promoters of *TRM1*, *MXR1* and *PRM1* upstream of the dTomato reporter gene. Fluorescence was measured under glucose repressed, derepressed and methanol induced conditions (Fig. 2). Under glucose repressed conditions, only P_{TRM1} gave a weak signal clearly above the detection limit, suggesting constitutive expression. P_{MXR1} and P_{PRM1} gave rather low inconclusive values. However, once the glucose in the medium was depleted, fluorescence signals from all promoters increased (about four-fold for P_{TRM1}; for P_{MXR1} and P_{PRM1} clearly above background signal). Upon methanol induction P_{TRM1} and P_{PRM1} reporter fluorescence increased about four- and two-fold, whereas the P_{MXR1} signal remained approximately the same.

This data suggests also that these TFs have rather weak promoters, as P_{AOX1} with the same reporter protein reached on methanol considerably higher levels (Fig. 1D). Methanol induced P_{TRM1} reached less than 10 %, the other factors less than 2 % of methanol induced P_{AOX1} (Fig. 2).

Genome wide transcriptional response towards *TRM1*, *MXR1* and *PRM1* overexpression

We aimed to elucidate the sets of genes regulated by *TRM1*, *MXR1* and *PRM1,* all of which had activated P_{AOX1} to different extents under derepressed conditions (if under control of P_{CAT1}) using microarrays. We used the strain expressing eGFP instead of an activator as control. The strains were grown in bioreactors to achieve more uniform growth than in deep well plates or shake flasks. Fluorescence of the dTomato reporter protein under control of P_{AOX1} was measured (Fig. 3A). Similar to experiments in DWPs (Fig. 1D), the control construct bearing an eGFP gene instead of an activator remained tightly repressed under glucose depleted conditions. Strains coexpressing the activators *TRM1*, *MXR1* and *PRM1* showed reporter protein fluorescence to different extents.
Samples for RNA isolation were taken after glucose depletion and analyzed using custom Affymetrix microarray chips previously reported (18, 44). The results are summarized in Fig. 3 and complete lists of differentially regulated genes are provided in S 7. P_{CAT1}-*MXR1* showed the highest number of deregulated genes (2920) compared to the P_{CAT1}-*eGFP* control strain with more genes down- than upregulated (Fig. 3B). P_{CAT1}-*TRM1* and P_{CAT1}-*PRM1* showed lower numbers of deregulated genes (2061 and 1075) with approximately equal numbers of genes up- and downregulated. Given the total number of probe sets (5869) these numbers account for 50 % (P_{CAT1}-*MXR1)* to 18 % (P_{CAT1}-*PRM1)* of *P. pastoris* genes.
There were overlapping sets of up and down-regulated genes between all three activator expressing strains (Fig. 3C). Yet, *MXR1* and *TRM1* and showed more overlaps with one another than with the *PRM1* expressing strain. We also analyzed the data in regard to biological processes (classified by COG terms (45), Fig. 3D). As suggested from the summarized genome wide response, *MXR1* overexpression showed stronger effects than *TRM1* and *PRM1*.

Conversions plasmids allow to transform existing P_{AOX1} based expression strains into methanol free systems

Aside the basic characterization of the effects of P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*PRM1* overexpression (Fig. 2, Fig. 3), the strong derepression effects observed (Fig. 1D), appeared also valuable to generate methanol free production strains. Expression of *TRM1* and *MXR1* allowed derepressed expression to up to 44 % of methanol induced levels in small scale cultivations. In these constructs the reporter gene (dTomato) was on the same vector as the activators. We envisioned conversion plasmids that could be transformed into existing strains already producing GOIs under the control of P_{AOX1} and thereby making them methanol free (Fig. 4A).
To this end we switched the Zeocin marker of the aforementioned activator plasmids (P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1)* to Geneticin markers, allowing retransformation of existing Zeocin selection based expression strain. We also switched the marker of the P_{CAT1}-*eGFP* control plasmid and included this plasmid as 'empty' vector control (to rule out general beneficial effects of a second transformation event). Note that these plasmids still contain the *AOX1* promoter controlling the dTomato gene. We reasoned that this system could be used to screen for outstanding production clones based on the dTomato signal (dTomato expression is also visible to the naked eye as the cells turn pink, *i.e.* if transformants turn bright pink under methanol free conditions, they should also strongly express the pre-integrated GOI under control of P_{AOX1}).

We transformed the linearized conversion plasmids (P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*eGFP*) into three existing P_{AOX1} based strains available from previous studies (18). These strains were expressing intracellular eGFP (Fig. 4C), secretory *Candida antarctica* lipase B (Fig. 4D) and secretory horseradish peroxidase (Fig. 4E).
The intention of the transformation of an existing eGFP producing strain with the vectors, was to reproduce the screening results obtained where P_{AOX1}-dTomato was present on the same vector (Fig. 1D). CalB and HRP are commercially relevant enzymes for biocatalysis and signal amplification in immunodetection respectively.
We tested multiple transformants of the constructs for fluorescence and enzyme activity (Fig. 4B, S 8). In the majority of cases, more than 50 % of the transformants were active and showed dTomato fluorescence. Systematically comparing these transformants to the untransformed 'parental' strains and the control vectors, showed derepression effects to different extents (10 to 183 % of methanol induced levels). Transformation of the *eGFP* control plasmid did not affect derepression in any of the strains tested, but methanol induced activities compared to the parental strain were in part mildly effected (Fig. 4D,E). The strains were also compared to control constructs expressing the same genes from the constitutive *P. pastoris* state of the art promoter of the *glyceraldehyde-3-phosphate dehydrogenase* gene (P_{GAP}) (14). For the P_{AOX1}-*eGFP* strain 68 % of methanol induced levels were reached (Fig. 4C) even outperforming the constitutive P_{GAP} reference. Note that in this case the eGFP control plasmid is leading to ambiguous effects, as it also contains an eGFP gene (under control of the derepressed P_{CAT1}). To this end we have also included a control, where the same plasmid is transformed into a strain bearing a dTomato gene under control of P_{AOX1}, showing consistent results (Fig. 4C). In addition the control plasmid and the two activator plasmids were also transformed in the wildtype strain (Fig. 4C). In case of the CalB strain, the conversion plasmids led to an activation of up to 183 % of methanol induced levels, however still remaining lower than the constitutive P_{GAP} control (Fig. 4D). For HRP, only 10 % of methanol induced levels were reached (Fig. 4E). This is also the only case where *MXR1* expression reached higher effects than *TRM1* expression.

Discussion

We had initially anticipated, that overexpression of single factors would not be sufficient to activate the *P. pastoris AOX1* promoter. Typically not only the amount of a TF but also its subcellular localization (cytosol vs. nucleus) and posttranslational modifications such as phosphorylation affect its activity (21, 25). We used regulatory models on the *S. cerevisiae ADH2* and *GAL* glucose-repressed promoters (S 3) as a basis for this assumption. Glucose-repressed genes in *S. cerevisiae* are repressed by Mig1p and Mig2p. Binding of the activator Adr1p (homolog of *P. pastoris* Mxr1p (21)) is additionally regulated by intricate mechanisms of phosphorylation and inhibitory binding of a 14-3-3 protein (46). Effects of a 14-3-3 protein have also been studied for *P. pastoris* Mxr1p (25). We had planned to combine multiple factors (in part as phosphomimetic variants), possibly in a knockout strain of repressors to strongly activate P_{AOX1} under methanol free conditions. However expression of *TRM1*, *MXR1* and *PRM1* from the derepressed *CAT1* promoter was sufficient to activate P_{AOX1} between 7 to 44 % of methanol induced levels (using eGFP as reporter gene, Fig. 1D).
These activating effects are not limited to the *AOX1* promoter: Microarray data (Fig. 3) showed that 50 % (P_{CAT1}-*MXR1)* to 18 % (P_{CAT1}-*PRM1)* of *P. pastoris* genes are differently regulated (compared to the P_{CAT1}-*eGFP* control strain). *MXR1* overexpression activates most strikingly about 40 % of genes involved in defense mechanisms (Fig. 3D). We assume that this response is naturally geared towards reducing toxicity associated with methanol metabolism requiring detoxification of radical oxygen species generated by oxidation of methanol to formaldehyde (47, 48). About 40 % of genes involved in nuclear structure are downregulated upon *MXR1* overexpression. Interestingly, also more than 30 % of nuclear structure genes are downregualted upon *TRM1* overexpression, and barely any genes upregulated.

When transforming the plasmids used in this study into *P. pastoris*, we noticed on two occasions decreased transformation rates: On the one hand for constructs bearing certain activators under P_{GAP} and on the other hand for conversion plasmids requiring Zeocin and Geneticin co-selection.
While P_{CAT1} based activator expression plasmids with Zeocin selection were giving uniformly high transformation rates, some P_{GAP} plasmids (Mxr1, Mxr1c, Reg1, Gal4) showed repeatedly approximately 10- to 100-fold reduced transformation efficiencies and different colony sizes (data no shown). We presume that these issues are caused by detrimental effects of strong constitutive overexpression of these regulators from the *GAP* promoter. For *MXR1*, it has previously been reported that overexpression on methanol from P_{AOX1} abolishes growth completely (21). *MXR1* could only be successfully overexpressed from a weak methanol inducible promoter (P_{PEX8}) (21). However in our work, even transformation plates of the putatively lethal P_{GAP}-*MXR1* construct contained a few colonies of different sizes. *P. pastoris* is known to show clonal variation even for the same plasmid (49), possibly owing to different copy numbers or integration loci. We transformed only low amounts of DNA favoring single copy integration and we screened 42 colonies per construct (S 4) to rule out effects of clonal variation. In most cases, these expression landscapes showed uniform expression. However, for some constructs (e.g. P_{GAP}-*MXR1* and P_{GAP}-*REG1*) the landscapes were rather distorted (S 4). More strikingly about only half of these transformants showed dTomato fluorescence upon methanol induction suggesting that the cassettes are not always functionally expressed. It has previously been noticed that under metabolic burden or stress conditions (50, 51) *P. pastoris* may lose parts or entire expression cassettes. Constructs based on repressible expression of the same factors by P_{CAT1} did not show these issues, as P_{CAT1} is due to the high glucose levels repressed upon growth on the transformation plates. Yet, upon methanol induction some P_{CAT1} based expression strains showed detrimental effects (e.g. P_{CAT1}-*MXR1* abolished growth and decreased reporter protein fluorescence), in line with previously reported effects of strong overexpression by P_{AOX1} (21).

These findings suggest that the dosage of the TFs plays a crucial role in the balancing the desired beneficial effects of activation during derepression and undesired effects abolishing cell growth. The natural transcriptional regulation of *MXR1* and *TRM1* (Fig. 2) is further supporting this notion: The natural promoter of *MXR1* gives weaker reporter protein fluorescence than P_{TRM1}. P_{TRM1} is also 4.5-fold induced on methanol. These results suggest that Trm1p is typically present at higher amounts and may explain why overexpression of *TRM1* is tolerated better than overexpression of *MXR1*.

A key question is how an excess of TFs may be sufficient for P_{AOX1} activation. One major factor appears to be the expression of the activators under derepressed conditions using P_{CAT1}. Upon glucose depletion the cellular machinery has alleviated glucose repression and is geared towards expression of catabolite repressed genes. Initially we had also considered knockouts of the genes coding for repressors of glucose repressed genes (Mig1p and Mig2p, S 2). Since Mig1/2p homologs in *S. cerevisiae* and related methylotrophic yeasts (31, 36) release glucose repression upon glucose depletion, such knockouts appear to become obsolete under derepressed conditions. Once derepressed conditions are reached, overexpression of the factors appears to mimic methanol induction. Negative regulators may be titrated by the strong overexpression. Transporters for nuclear export of the factors may be overloaded or negative posttranslational modifications such as phosphorylation/dephosphorylation may not be performed on all transcription factor molecules, because the kinase/phosphatases are only expressed at their wild type levels and are overloaded dealing with highly increased levels of the TFs.
However, activating P_{AOX1} under presence of glucose (e.g. making it constitutively active) appears to be considerably more effort and would likely require knockouts of the repressors and possibly modifications of additional regulators.

Understanding the exact underlying molecular mechanisms of the effects observed in this study will require extensive further studies. *S. cerevisiae* Adr1p (homolog of *P. pastoris* Mxr1p (21)) has been widely studied since more than three decades (52), yet the exact molecular regulation and all factors involved remain incompletely understood (53).
We see our work rather as a proof of concept to bypass carbon source regulation easily even without complete mechanistic understanding. P_{CAT1}-Activator conversion plasmids were found to be valuable tools to bypass P_{AOX1} regulation transforming existing strains into methanol free systems (Fig. 4).
Also when transforming these conversion plasmids into existing strains, the transformation rates were reduced and in part high clonal variation evident (Fig. 4B, S 8). These effects are presumably not only attributable to Zeocin and Geneticin co-selection as transforming the wildtype strain (requiring only Geneticin selection) gave also only a few dozen colonies (Fig. 4B), while on Zeocin we obtained with similar sized plasmids a few hundred colonies (data not shown). In some cases (CalB, HRP) transformations had to be repeated to obtain a sufficient number of transformants for screening. The low transformation rates are not limited to plasmids bearing the activators *TRM1* and *MXR1* but also occurred with the eGFP control plasmid. Therefore we assume that reduced transformation rates are not caused by detrimental effects of the activators.

Aside of the different transformation rates, the conversion plasmids showed also different rates of active transformants: When transforming the conversion plasmids into the wildtype strain and an eGFP expressing strain about 90 % of the transformants showed derepressed activity (Fig. 4B). However, for CalB and HRP expressing strains only approximately about 50% of the transformants were active. In experiments providing the activator and GOI on the same plasmid (Fig. 1D), nearly all transformants of P_{CAT1}-*TRM1,* P_{CAT1}-*MXR1* and P_{CAT1}-*eGFP* showed uniformly derepressed activity (S 4). So these effects seem to be caused by the second transformation event (possibly inactivating the existing integrated expression cassette) and/or Zeocin and Geneticin coselection. The difference between the wildtype and eGFP strains compared to CalB and HRP strains suggests also an influence of the GOI present (possibly as the latter proteins exert with secretion or metabolite production more stress on the cells than fluorescent proteins).

We checked the literature for similar issues on transformation rates and inactive transformants in *P. pastoris*, but to the best of our knowledge we could not find explicit descriptions or an explanation. Either the effects are only limited to the expression of the activators (and the eGFP gene) used here, or previous studies did not delve into comparing hundreds of transformants. Yet understanding these issues and including empty vector controls could be important for future studies in *P. pastoris*. So far, somewhat inconclusive data has been reported from attempts to activate P_{AOX1} by TF overexpression (24), reviewed in (14). In a patent application (24) only constitutive expression of *MXR1* and *PRM1* was tested and no absolute expression data (or comparison to methanol induction) was shown. In our hands, constitutive expression of *MXR1* and *PRM1* either failed or gave only marginal results. However, we can confirm an activating effect of constitutive *TRM1* expression under methanol induced conditions (54).

If undesired integration events with identical sequences on the conversion plasmids should cause these issues, their removal may alleviate the issues observed. We had included the *AOX1* promoter and terminator driving expression of the dTomato reporter on the conversion plasmids to screen by naked eye for derepressed colonies (as dTomato expression turns cells pink). Since approximately half of the transformants were derepressed active, this feature could be omitted (given a medium throughput screening assay for the protein to be expressed is available). Also the relatively large activator genes (*MXR1*: 3468 bp, *TRM1*: 2667 bp) may be prone to recombination with the genome (still containing the natural copy) and could be sequence diversified by using different codons.

Despite the issues with transformation rates and inactive transformants, the conversion plasmid strategy proved useful to transform existing P_{AOX1} based strains into methanol free systems. However, the feasibility has to be tested on a case by case basis as some derepressed strains even outperformed the methanol induced parental strain and constitutive promoters, while others performed poorly (Fig. 4C-E).
Since we used small scale cultivations with unmodified standard protocols (55), cultivation conditions can be optimized for maintaining the derepressed conditions for example by using a continuous enzymatic release of low levels of glucose from a glucose-based polymer, as reported for P_{AOX1} applications (56). Thereby constantly carbon source for protein production and cell growth could be provided, while not repressing P_{CAT1}-activator and thereby P_{AOX1}. In bioreactor cultivations, derepression can be maintained by a limited carbon source feed (as demonstrated for synthetic derepressed P_{AOX1} variants (6, 19)), which would allow more efficient expression, than achievable in small scale cultivations applied here.

Our work suggests that with the availability of suitable parts as derepressed promoters and potent activators, it can be surprisingly simple to bypass carbon source regulation making induction obsolete. In *P. pastoris*, overexpression of a single factor is sufficient to convert existing strains into glucose/gylcerol regulated, methanol-free systems. These strains may become valuable tools to increase space time yields and to design safer, more environmentally friendly production processes.
Given the wide applicability of carbon source regulated promoters in basic research and biotechnology (1–3, 9), our approach in *P. pastoris* may also be useful in other organisms.

Acknowledgments

The authors gratefully acknowledge support from NAWI Graz. This project received funding by the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 289646 (Kyrobio). T.V. was supported by the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) while performing this research.

Materials and methods

Strains and cultivation conditions

Materials such as enzymes and cloning kits were used as recently reported (18, 57). The *P. pastoris* CBS7435 wildtype strain was used for most experiments. We also transformed strains available from previous studies with conversion plasmids. These strains include P_{AOX1}-*eGFP*, P_{AOX1}-*CalB* and P_{AOX1}-*HRP* strains from Vogl *et al*. (18). For CalB and HRP, mutS strains (58) were used, as higher productivity on methanol has been reported (59).

We used the standard small scale 96 deep well plate (DWP) cultivation protocol reported by Weis *et al*. (55). In short, 250 µL BMD1 (buffered minimal dextrose medium, as reported) were inoculated with a single colony from transformation plates and grown for 60 h on glucose. For induction a final methanol concentration of 0.5 % (v/v) was used. First induction was performed after 60 h of growth on glucose with 250 µL of buffered media with 1% methanol (BMM2). Subsequently after 12 h, 24 h and 48 h 50 µL of BMM10 (with 5% methanol) were added for further induction. Glucose levels were determined with Combur test strips (Roche Diagnostics, Rotkreuz, Switzerland) or a Glucose UV kit by DIPROmed (Vienna, Austria). The following antibiotic concentrations were used: *E. coli*: LB-medium containing 25 µg/ml Zeocin, 50 µg/ml Kanamycin; *P. pastoris*: 100 µg/ml Zeocin, 300 µg/ml Geneticin.

The screening and rescreening procedures to compare the *P. pastoris* strains have previously been reported (57). In short, in the initial screening round 42 colonies/transformants per construct were picked and cultivated in 96 well deep-well plates. Under derepressed conditions, activities were measured and glycerol stocks of the entire DWPs made. After methanol induction and data analysis, four representative clones were chosen for the rescreening and therefore streaked for single colonies from the glycerol stocks to avoid mixed populations. Transformants for the rescreening were cultivated in seven-fold replicates. One representative clone was used for the comparisons shown in Fig. 1 and Fig. 4. Due the unexpectedly high clonal variation encountered during this study, the screening and rescreening data for the activator and conversion plasmids are provided as supporting information S 4 and S 8.

Cloning

The Zeocin based activator plasmids bearing a single activator and cloning of the promoters of *MXR1*, *TRM1* and *PRM1* has been reported in the master thesis of Lukas Sturmberger alongside the experimental procedures of the microarray experiments.
To allow transformation of the activators into existing production strains already bearing a plasmid with Zeocin selection, the Zeocin marker of selected activator plasmids was replaced with a Kanamycin/Geneticin marker (respective antibiotics for selection in *E. coli*/*P. pastoris*). The Zeocin resistance marker cassettes of the P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*eGFP* vectors was cut out by digestion with *Not*I and *Swa*I. There were no unique restriction sites available to remove solely the Zeocin marker, therefore also the adjacent *AOX1* transcriptional terminator (*AOX1TT*) and the *E. coli* origin of replication (pUC ori) were removed. The Kan/Gen selection cassette flanked by the *AOX1TT* and the pUC ori was amplified form the pPpKan_S vector (58) using primers URA3-pUCori-Gib and sTomato-AOX1TT-Gib (S 1). These primer sequences contained overhangs for Gibson assembly (60) and where thereby linked with the *SwaI*/*NotI* digested vector backbones. The entire inserted sequences were confirmed by Sanger sequencing. Subsequently the *CAT1* promoter sequences of these vectors were shortened to 500 bp length. To this end, vector fragments containing the shortened P_{CAT1} and activators (or eGFP) were amplified using primers pAOX1-5prime-pCAT1-500-Gib and intURA3-642..673-fwd-Gib. These PCR fragments were used in a Gibson assembly with the universally applicable pAOX1fwd and intURA3-642..673-rev-Gib PCR product obtained from the Kan-P_{CAT1} vectors previously generated. The entire vectors were confirmed by sequencing.

Fluorescence and enzyme activity measurements

eGFP and dTom reporter protein fluorescence was measured on a Biotek Synergy MX plate reader. Samples were diluted in 96-well fluorescence microtiter-plates to a total volume of 200 µL. eGFP fluorescence was measured at 488/507 nm (excitation/emission), dTomato at 554/581 nm. The relative fluorescence unit (rfu) values were normalized to the OD_{600} value.
HRP and CalB activity assays were performed as described in the literature using 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) and para-nitrophenyl butyrate (pNPA) as substrates (59).

Figures

Fig. 1

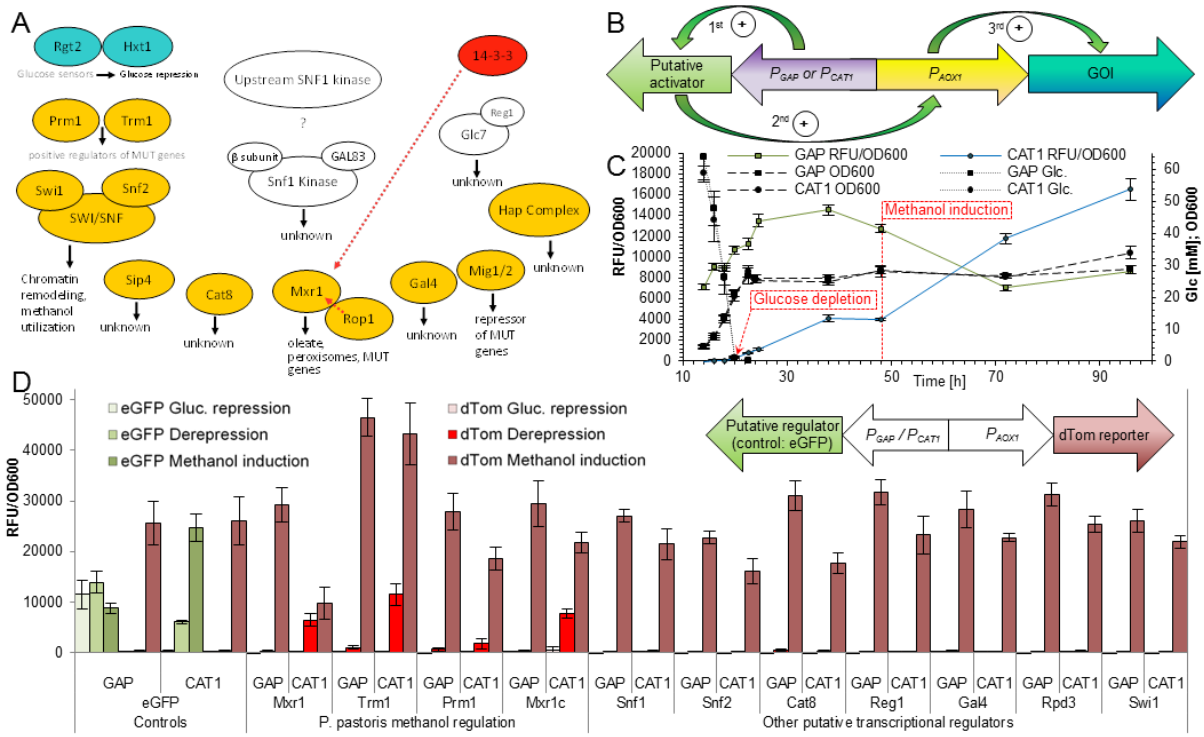


Fig. 1: Expression of *TRM1*, *MXR1* or *PRM1* from the derepressed *CAT1* promoter activates the *AOX1* promoter under derepressed conditions.

A) Experimentally confirmed and hypothetical regulators involved in MUT gene regulation in *P. pastoris*. Homologs to factors involved in catabolite repression and expression of glucose-repressed genes in *S. cerevisiae* (S 3A) and factors from studies in *P. pastoris* and other methylotrophic yeasts are shown (S 2, S 3B). See S 3 for extended discussion.

B) Transcriptional regulatory circuit design to activate P_{AOX1}. P_{GAP} or P_{CAT1} drive expression of a putative activator that is activating P_{AOX1} thereby leading to methanol free expression of the gene of interest (GOI).

C) Regulation of the *GAP* and *CAT1* promoters used to drive expression of the putative regulators. The promoters were cloned upstream of an eGFP reporter protein and transformed into *P. pastoris*. The strains were cultivated in shake flasks and inoculated to a low starting OD_{600}. Reporter protein fluorescence, OD_{600} and glucose levels were measured. Fluorescence/OD_{600} values at t=0 are not shown, as the starting OD_{600} of 0.05 was outside the linear range of the spectrometer used. The initial glucose concentration of the media was 55.5 mM (10 g/l). Mean values (MVs) and standard deviatons (SDs) of biological triplicates are shown.

D) Testing for effects of 11 putative regulators on P_{AOX1}. A reporter plasmid bearing a red fluorescent protein variant (dTomato, dTom) under control of P_{AOX1} was generated (schematic on right side). Putative regulators or an enhanced green fluorescent protein (eGFP) as control were expressed from P_{GAP} or P_{CAT1}. Reporter protein fluorescence (normalized per OD_{600}) was measured under repressed, derepressed and methanol induced conditions from cells grown in deep well plates. Note that the signals of the two FPs used are not directly comparable due to different quantum yields and fluorescence spectrometer settings used. MVs and SDs of biological quadruplicates are shown. See S 4 for screening and rescreening data.

Fig. 2

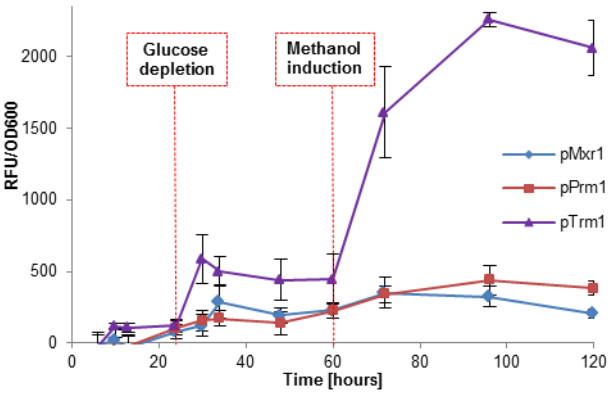


Fig. 2: Regulation of the promoters of *TRM1*, *MXR1* and *PRM1* under glucose repressed, derepressed and methanol induced conditions. The promoter sequences were cloned upstream of a red fluorescent protein variant (dTomato) and transformed into the *P. pastoris* wildtype strain. The recombinant strains were grown in deep well plates on glucose and induced with methanol after 60h. Glucose depletion was assayed by glucose detection strips. Reporter protein fluorescence was measured at the time points indicated and normalized per biomass (measured by OD_{600}). Mean value and standard deviations of biological triplicates are shown.

Fig. 3

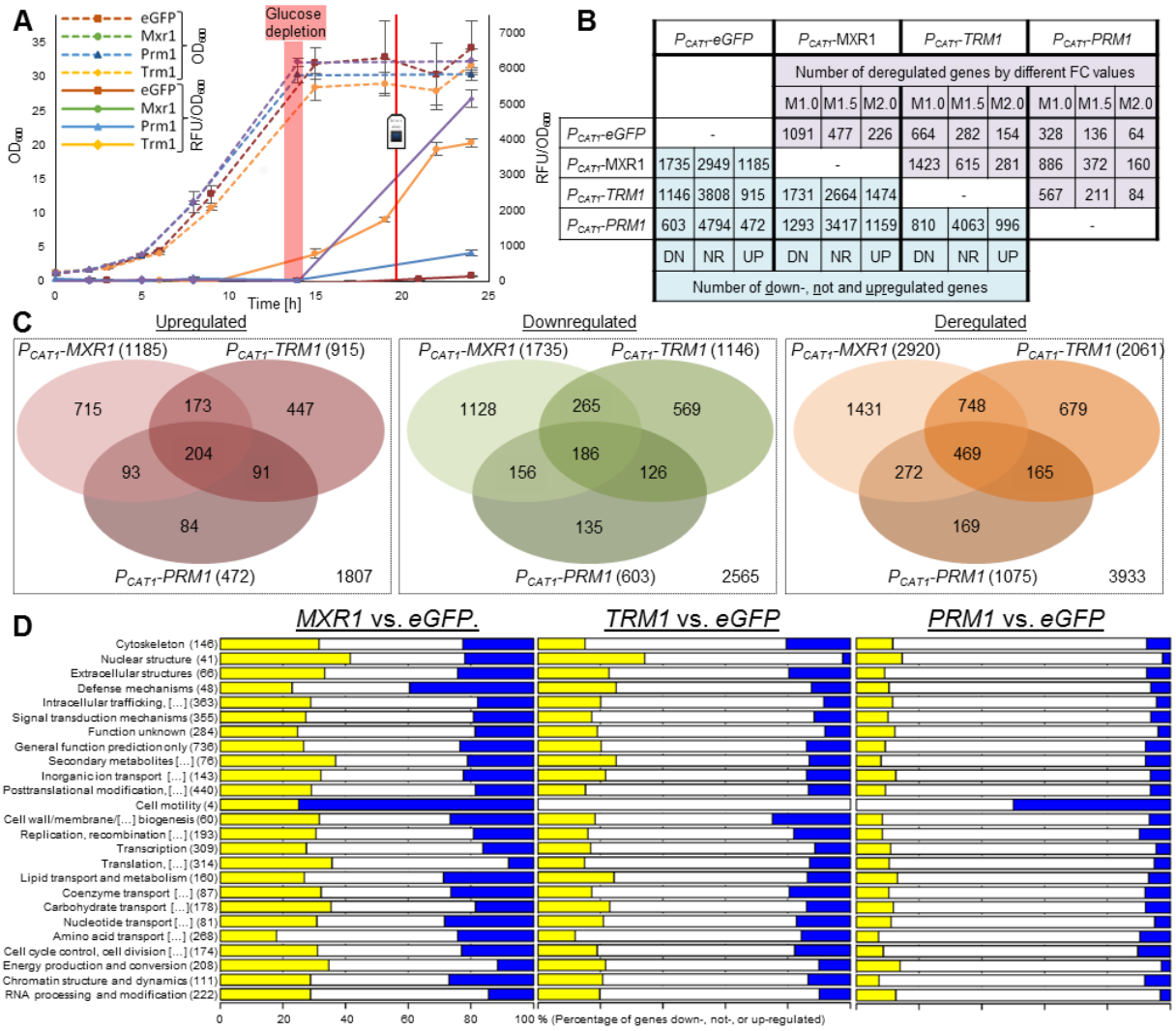


Fig. 3: Genome wide transcriptional response towards *TRM1*, *MXR1* and *PRM1* overexpression under control of P_{CAT1} (compared to an eGFP expressing control strain). This analysis was performed similar to (18). Lists of differently regulated genes between pairwise comparisons are provided in S 7.

A) The strains P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1*, P_{CAT1}-*PRM1* and P_{CAT1}-*eGFP* were cultivated in bioreactors and grown past glucose depletion for microarray analyses in biological duplicates (sampling point indicated by red line). Growth curves (OD_{600}) and reporter gene fluorescence (dTomato under the control of P_{AOX1}, normalized per OD_{600}) are shown. MVs and SDs of six measurements (technical triplicates of biological duplicates) are shown. Glucose depletion was deduced by the oxygen peaks and confirmed by glucose measurements.

B) Comparison of the transcriptional response of the four strains. In the lower left corner the number of down- (DN), not regulated (NR) and upregulated (UP) genes is given ($p < 0.01$) [total number of probe sets: 5869]. In the upper right corner deregulated (DN or UP) genes are listed by different fold change (FC) criteria [M1.0, 1.5 and 2.0 denoting the respective log2 values].

C) Comparative analysis of differential gene expression between P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*PRM1*. Each strain was at first normalized to the P_{CAT1}-*eGFP* control. The numbers of unique and overlapping genes showing up-, down- and deregulation are given (FC >log2(1.0), $p < 0.05$, the same criteria and normalization were applied for the analyses shown in panel D).

D) Functional grouping of differentially regulated genes to biologic processes by COG terms (45). The relative numbers of down-, not- and upregulated genes assigned to the same COG terms are shown (each strain was compared to the P_{CAT1}-*eGFP* control as in panel C). The total number of genes assigned to each COG term is given in parentheses. Full COG terms if abbreviated : "Intracellular trafficking, secretion, and vesicular transport", "Secondary metabolites biosynthesis, transport and catabolism", "Inorganic ion transport and metabolism", "Posttranslational modification, protein turnover, chaperones", "Translation, ribosomal structure and biogenesis", "Replication, recombination and repair", "Translation, ribosomal structure and biogenesis", "Coenzyme transport and metabolism", "Carbohydrate transport and metabolism", "Nucleotide transport and metabolism", "Amino acid transport and metabolism" and "Cell cycle control, cell division, chromosome partitioning".

Fig. 4

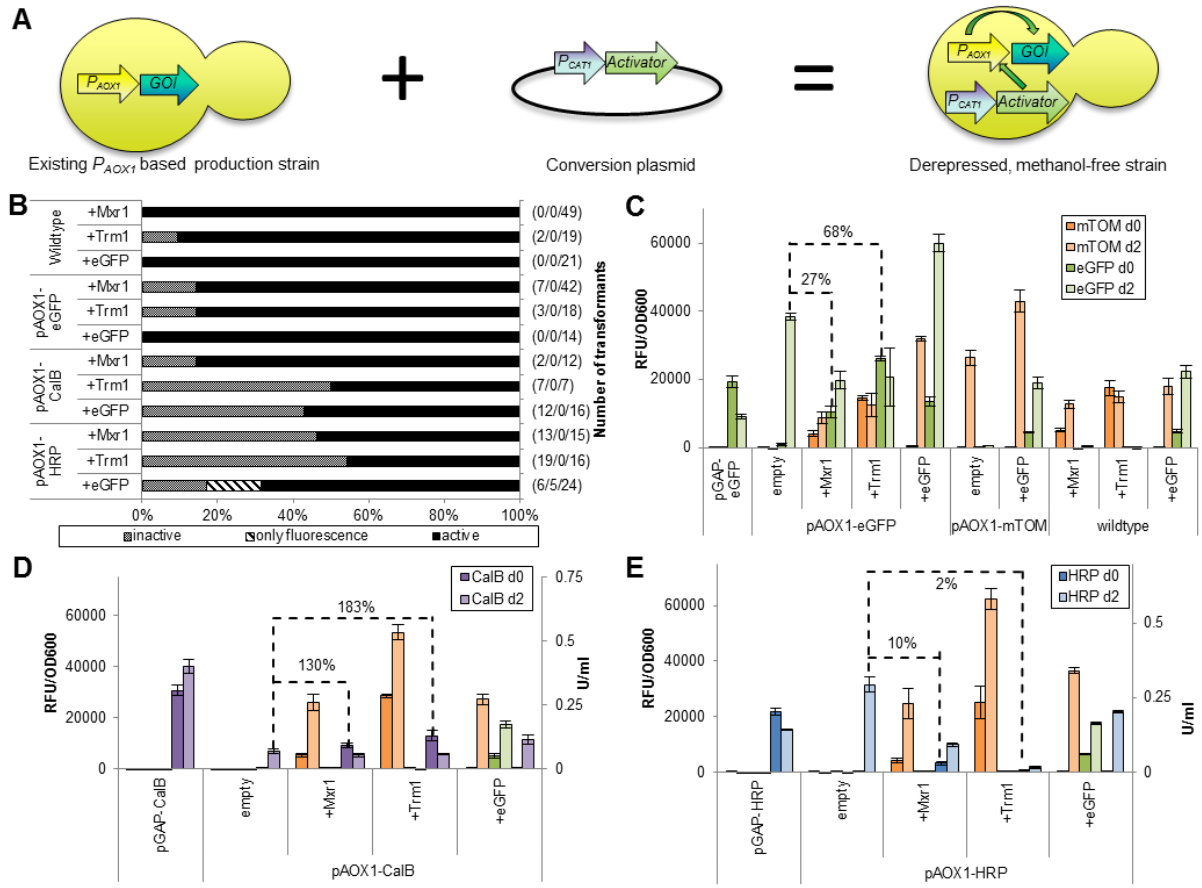


Fig. 4: Conversions plasmids allow to transform existing P_{AOX1} based expression strains into methanol free systems.

A) A *P. pastoris* strain bearing a gene of interest (GOI) under control of P_{AOX1} is transformed with conversion plasmids containing the activators *TRM1* or *MXR1* (in the control construct *eGFP*) under control of P_{CAT1} resulting after coselection in derepressed methanol free strains.

B) The conversion plasmids P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*eGFP* were transformed into the wildtype strains and strains bearing eGFP (panel C), CalB (panel D) and HRP (panel E) under control of P_{AOX1} (pAOX1). The number of inactive transformants, transformants showing only dTomato reporter fluorescence and activity and fluorescence are given. The screening landscapes are provided in S 8.

C)-E) Enzyme activity (if applicable) and fluorescence measurements of the aforementioned strains under derepressed (60 h growth on glucose, 'd0') and 48 methanol induced ('d2') conditions were measured. Derepressed values of activator strains are compared to the methanol induced levels of the parental (empty) strain. MVs and SDs of biological triplicates are shown.

Supporting information

S 1

S 1: Primer sequences used in this study.

Primer name	Sequence
URA3-pUCori-Gib	gagcccaatcgacaatttttcggattttgcatttaaatacatgtgagcaaaaggccagcaaaaggccag
sTomato-AOX1TT-Gib	cttgtacggtatggacgagctttataagtaagcggccgctcaagaggatgtcagaatgccatttgc
pAOX1fwd	AGATCTAACATCCAAAGACGAAAGGTTGAATGAAAC
intURA3-642..673-rev-Gib	ctatgaggcttcgtttcaggcgatacttccac
pAOX1-5prime-pCAT1-500-Gib	gtttcattcaacctttcgtctttggatgttagatctTAATCGAACTCCGAATGCGGTTCTCCTG
intURA3-642..673-fwd-Gib	gtggaagtatcgcctgaaacgaagcctcatag

S 2

S 2: (A) Identities of proteins involved in the methanol dependent gene regulation in the methylotrophic yeasts *P. pastoris*, *H. polymorpha* and *C. boidinii* and galactose dependent gene regulation in *S. cerevisiae* by a BLAST search. (B) Detailed information on Tables taken from Lukas Sturmberger's master thesis (Characterization of synthetic regulatory elements for protein expression in Pichia pastoris, Master thesis, Graz University of Technology 2013).

A

BLAST results within four yeast species. Protein sequences from *S. cerevisiae* (taxid: 4932) were blasted (NCBI blastp) against the genome of *P. pastoris* (taxid: 4922)*, C. boidinii* (taxid: 5477) and *H. polymorpha* (taxid: 870730). Due to the lack of complete genome sequences in the case of *H. polymorpha* and *C. boidinii* only single deposited files could be used to perform a BLAST search. For reasons of avoidance of false positive hits, sequences from the latter two species without any corresponding experimental reference were omitted from the table. Three-letter protein abbreviations in *H. polymorpha*, *C. boidinii* and *S. cerevisiae* were taken from referenced publications and SGD (Saccharomyces Genome Database), respectively. All BLAST queries resulted in E-values <1e-25. n.s.a. no sequence available. n.h. no homologue. * PpZta1 is a medium-chain dehydrogenase/reductase, *S. cerevisiae* harbors 20 proteins.

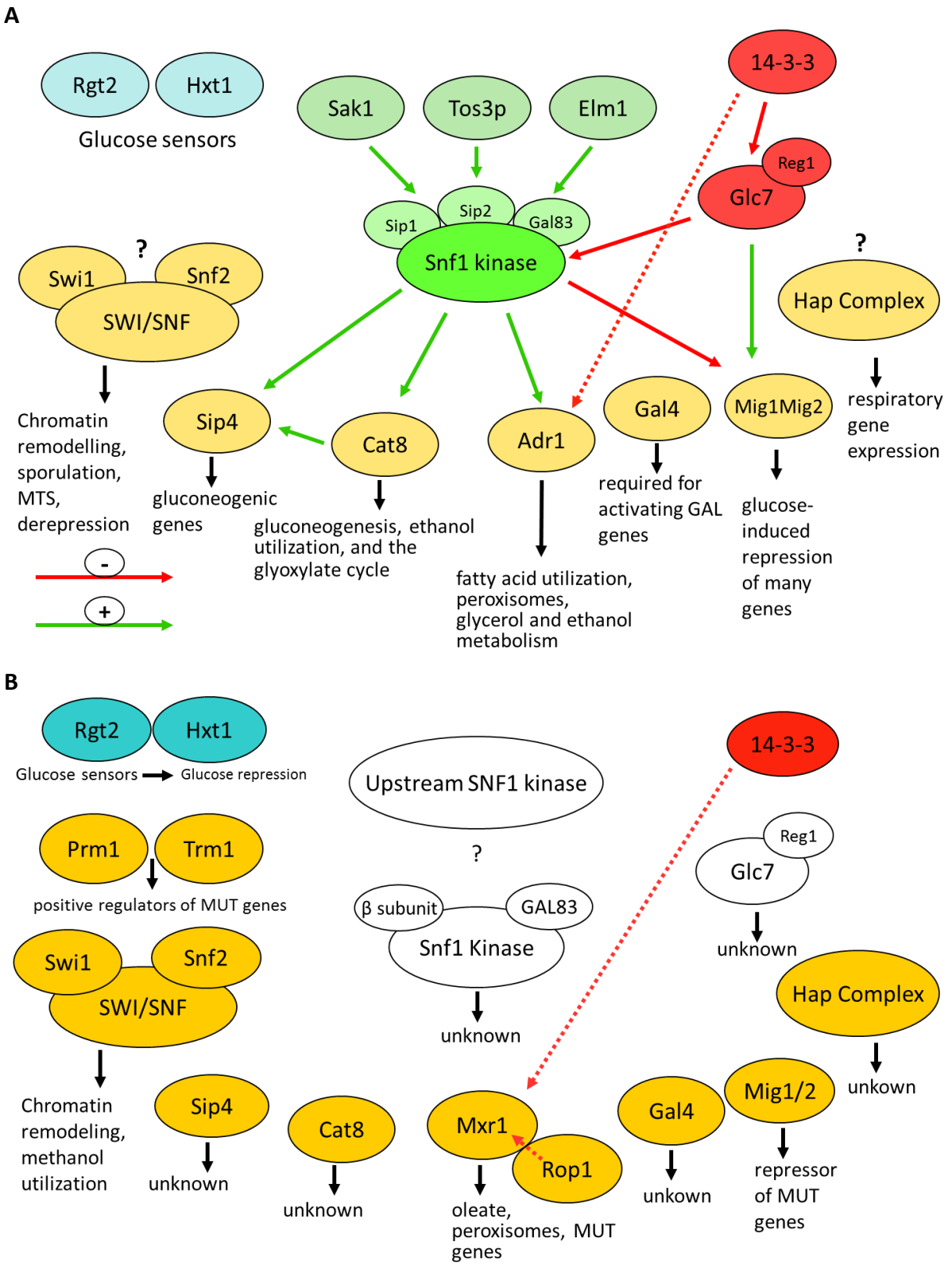
	Protein name (if annotated) - GenBank accession number			
Gene name	***P. pastoris***	***H. polymorpha***	***C. boidinii***	***S. cerevisiae***
Mxr1	Mxr1 - CCA40655.1	n.s.a.	Trm2 - BAJ07608.1	Adr1 - AAA73863.1
cov./seq.ident.			88%/37%	34%/55%
Prm1	Prm1 - CCA40959.1	Mut3 - AAK84946.1	Trm1 - BAF99700.1	Asg1 - NP_012136.1
cov./seq.ident.		96%/53%	75%/74%	56%/53%
Swi1	Swi1 - CCA37890.1	Swi1 - AAQ75382.1	n.s.a.	Swi1 - NP_015309.1
cov./seq.ident.		99%/29%		75%/29%
Snf1	CCA38457.1	AAN84785.1	n.s.a.	Snf1 - NP_010765.3
cov./seq.ident.		51%/56%		98%/65%
Cat8	Cat8 - CCA38204.1	n.s.a.	n.s.a.	Cat8 - NP_014007.1
cov./seq.ident.				74%/32%
Reg1	Hex2 - CCA36537.1	n.s.a.	n.s.a.	Reg1 - NP_010311.1
cov./seq.ident.				44%/61%
Gal4	Lac9 - CCA37633.1	n.s.a.	n.s.a.	Gal4 - NP_015076.1
cov./seq.ident.				55%/29%
Trm1	CCA39317.1	Mpp1-AAO72735.1	n.s.a.	n.h
cov./seq.ident.		68%/33%		
Snf2	CCA40198.1	n.s.a.	n.s.a.	Snf2 - NP_014933.3
cov./seq.ident.				76%/54%
Rpd3	Rpd3 - CCA37028.1	n.s.a.	n.s.a.	Rpd3 - NP_014069.1
cov./seq.ident.				90%/87%
Hda1	Hda1 - CCA38680.1	n.s.a.	n.s.a.	Hda1 - NP_014377.1
cov./seq.ident.				92%/62%
Zta1	CCA38647.1	ADH - ADM49192.1	n.s.a.	Zta1 - NP_009602.1
cov./seq.ident.		99%/28%*		99%/59%
Rop1	Zms1 - CCA39607.1	n.s.a.	n.s.a.	Tda9 - NP_013630.1
cov./seq.ident.				78%/32%
Mig1	CCA40819.1	Mig2 - ABU63593.1	Mig1 - BAM38481.1	Mig1 - CAA39084.1
cov./seq.ident.	16%/74%	82%/41%	23%/87%	
Mig2	Mig1 - CCA37444.1	Mig2 - ABU63593.1	Mig1 - BAM38481.1	Mig2 - NP_011306.1
cov./seq.ident.	16%/63%	29%/46%	30%/66%	

B

Table 4.1.: Synopsis of putative factors involved in the regulation of carbon metabolism listed with their respective function in methylotrophic yeasts. Listed in this table is condensed information of published data in methylotrophic yeasts *P. pastoris*, *H. polymorpha* and *C. boidinii*. In case we were unable to find experimental data from methylotrophic yeast species, we would reference to the homologous protein function in *S. cerevisiae.*

Protein name	Function	Size [bp]	References
Mxr1	Master regulator of methanol and peroxisomal genes in *P. pastoris* Mxr1p and *C. boidinii* (Trm2p), the *S. cerevisiae* homolog Adr1p is also involved in activation of catabolite repressed genes	3468	(21, 25, 38)
Prm1	Prm1p in *P. pastoris* and its homologue Trm1p in *C. boidinii* and Mut3p in *H. polymorpha* are positive regulators of MUT genes	2970	(24, 61, 62)
Trm1	Regulator of methanol inducible genes, homolog Mpp1 (methylotrophic peroxisomal protein1) in *H. polymorpha* important for growth on methanol	2667	(23, 63)
Swi1	ΔSwi1ΔSnf2 double knockout showed defective methanol utilization in *H. polymorpha*	2454	(32)
Snf2		5009	
Snf1	*S. cerevisiae* Snf1 kinase is a master regulator of carbon catabolite-derepression	1664	(64)
Cat8	Zinc cluster transcriptional activator involved in *S. cerevisiae* in activation of catabolite repressed genes.	3111	(65)
Reg1	Annotated as regulatory subunit of Glc7p master phosphatase in *P. pastoris*, Glc7p acts antagonistically to master kinase Snf1 by dephosphorylation	1995	(66)
Gal4	*S. cerevisiae* regulator of galactose metabolism, in *P. pastoris* annotated as Lac9, unknown function as *P. pastoris* is unable to grow on galactose	2988	(67)
Rpd3	Histone deacetylases from *S. cerevisiae*, regulate transcription, silencing, autophagy and other processes by influencing chromatin remodeling, *Δrpd3* and *Δhda1* genes allowed constitutive promoter binding of Adr1 and Cat8	1516	(41, 68)
Hda1		2216	

S 3



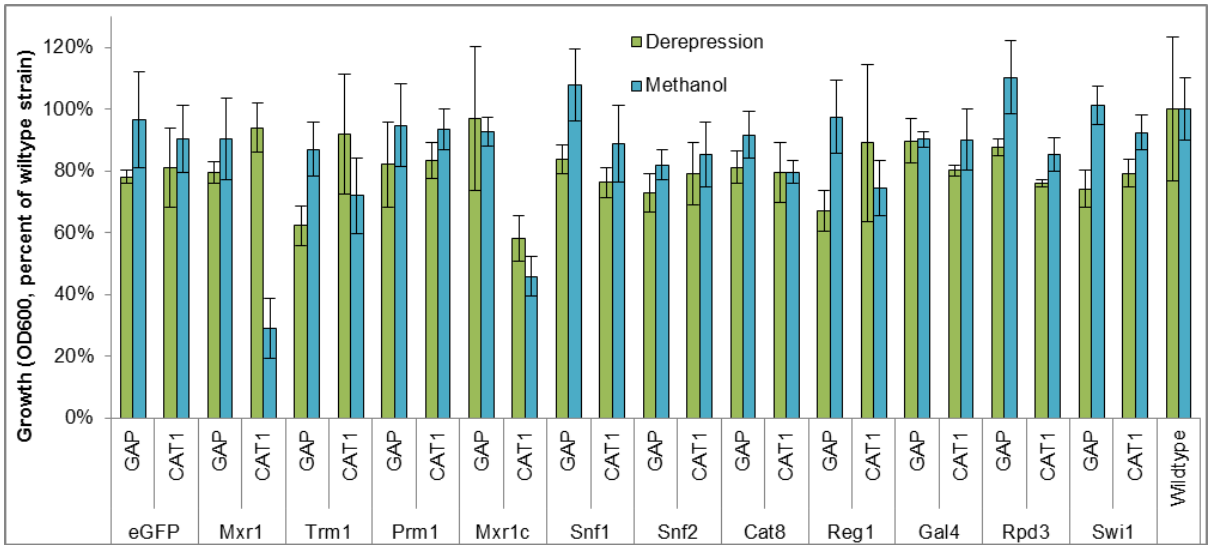
S 3: (A) Factors involved in catabolite repression and expression of glucose-repressed genes in *S. cerevisiae*. (B) This information and studies in *P. pastoris* and other methylotrophic yeasts was used to generate a model of hypothetical factors involved in *P. pastoris* MUT gene regulation (also shown in the main manuscript as Fig. 1A).

In *S. cerevisiae*, not only presence of activators such as Adr1p and Cat8p are required to trigger expression from glucose-repressed genes, also their phosphorylation status controls their activity. The master kinase Snf1 and associated subunits are activated upon glucose depletion, phosphorylating and thereby activating a number of proteins required for glucose-repressed genes. The phosphatase Glc7 (requiring the regulatory subunit Reg1) is the antagonist of Snf1 required for maintaining glucose repression. The red and green arrows indicate activation and inactivation, respectively. Snf1 mediated phosphorylation is not directly activating all factors, but favoring expression of derepressed genes. The repressors Mig1p/Mig2p are inactivated by phosphorylation, thereby alleviating glucose repression. Possible interactions that still remain to be elucidated are marked with a question mark as in the case of the SWI/SNF and the Hap complex.
Although homologs to the *S. cerevisiae* master regulators exist in methylotrophic yeasts, their exact regulation is unknown. Only experimentally verified interactions are shown as arrows, unclear and hypothetical interactions are marked with a question mark. See S 1 for further details

S 4

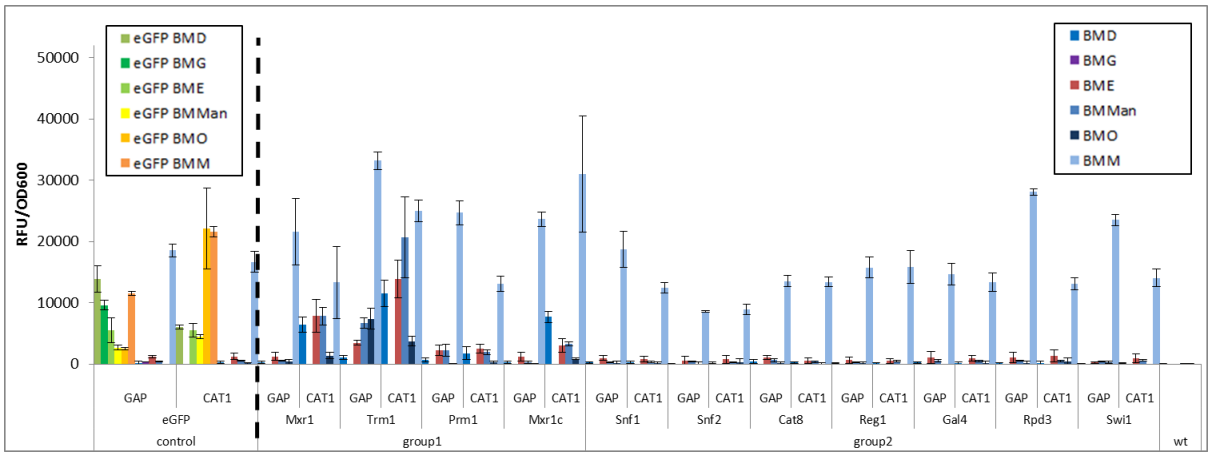
S 4: Activator constructs form Fig. 1D: Screening landscapes comparing approximately 42 transformants per construct under derepressed and methanol induced conditions and rescreening data. Transformants selected for rescreening are marked with red boxes in the screening data. These transformants were streaked for single colonies to rule out mixed populations and tested in a rescreening in 7-fold replicates for reproducible expression. One representative transformant (again marked by a red box) was used for the results shown in Fig. 1D.

S 5



S 5: P_{CAT1}-*MXR1* and P_{CAT1}-*MXR1c* strains show growth defects compared to the wildtype strain. OD_{600} as a measure for cell growth were sampled under derepressed conditions and methanol induced conditions. The measurements of the same strains from the same experiment shown in Fig. 1D are shown. The OD_{600} values are shown as percent of the wildtype strain.

S 6



S 6: Growth of the strains shown in Fig. 1D on different carbon sources. The RFU/OD_{600} values are given for the red fluorescent protein dTom, except for the controls which expressed eGFP additionally. Strain were grown for 60 h in DWPs on the following carbon sources (concentrations as reported previously (18)): glucose (BMD), glycerol (BMG), ethanol (BME), mannitol (BMMan), oleate (BMO) and methanol (BMM).

S 7

S 7: Lists of genes differentially regulated in the microarray data.

Each sheet contains the data of the pairwise comparisons of two conditions (see also Fig. 3). Significantly differentially regulated genes ($p < 0.01$) are shown. In addition to the comparisons under derepressed conditions (summarized in Fig. 3), the TF overexpressing strains P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1*, P_{CAT1}-*PRM1* are also compared to the control strain P_{CAT1}-*eGFP* under methanol induced conditions. The data on the convtrol strain P_{CAT1}-*eGFP* under glucose repressed, derepressed, methanol induced and glucose induced conditions has previously been published (18). In the same bioreactor cultivation runs, RNA isolations, reverse transcription and micro array hybridization experiments we had also included P_{CAT1}-*TRM1*, P_{CAT1}-*MXR1* and P_{CAT1}-*PRM1*. Therefore there is no additional bias of separate experiments arising.

Abbreviations:
COG: Clusters of Orthologous Groups of proteins (45); COGCat: COG category; COGCatDesc: COG category description; EC: Enzyme Commission number; GO: Gene Ontology term; log2FC: logarithm to the basis 2, calculated for fold changes; Abslog2FC: absolute log2FC; P.Value and adj.P.Val: probability values calculated from two biological replicates.

S 8

S 8: Screening and rescreening data of transformations of the conversion plasmids P_{CAT1}-*MXR1*, P_{CAT1}-*TRM1* and the control plasmid P_{CAT1}-*eGFP* into the wildtype strain (wt) and strains bearing CalB, HRP and eGFP under control of P_{AOX1} (pAOX1). Fluorescence measurements (in all strains indicative of the dTomato gene on the conversion plasmid) and enzymes activities are shown (enzyme activities are only shown as a relative comparison as mili absorption units per minute). Identical to S 3, transformants selected for rescreening are marked with red boxes in the screening data. These transformants were streaked for single colonies to rule out mixed populations and tested in a rescreening in 7-fold replicates for reproducible expression. One representative transformant (again marked by a red box) was used for the results shown in. Fig. 4C-E. For the screening data, derepressed conditions (60 h growth on glucose) are shown. For the rescreening derepressed conditions are in part referred to as ‘d0’. Also 24 h methanol induced data are in part shown and referred to as ‘d1’.

References

1. Jacob,F. and Monod,J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–56.
2. Hahn,S. and Young,E.T. (2011) Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, **189**, 705–36.
3. Blazeck,J. and Alper,H.S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.*, **8**, 46–58.
4. Khalil,A.S., Lu,T.K., Bashor,C.J., Ramirez,C.L., Pyenson,N.C., Joung,J.K. and Collins,J.J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, **150**, 647–58.
5. Blazeck,J., Garg,R., Reed,B. and Alper,H.S. (2012) Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnol. Bioeng.*, **109**, 2884–95.
6. Hartner,F.S., Ruth,C., Langenegger,D., Johnson,S.N., Hyka,P., Lin-Cereghino,G.P., Lin-Cereghino,J., Kovar,K., Cregg,J.M. and Glieder,A. (2008) Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Res.*, **36**, e76.
7. Gardner,T.S., Cantor,C.R. and Collins,J.J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**, 339–42.
8. Da Silva,N.A. and Srikrishnan,S. (2012) Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. *FEMS Yeast Res.*, **12**, 197–214.
9. Weinhandl,K., Winkler,M., Glieder,A. and Camattari,A. (2014) Carbon source dependent promoters in yeasts. *Microb. Cell Fact.*, **13**, 5.
10. Bill,R.M. (2014) Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.*, **5**, 85.
11. Vogl,T., Hartner,F.S. and Glieder,A. (2013) New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. *Curr. Opin. Biotechnol.*, **24**, 1094–1101.
12. Hamilton,S.R., Davidson,R.C., Sethuraman,N., Nett,J.H., Jiang,Y., Rios,S., Bobrowicz,P., Stadheim,T.A., Li,H., Choi,B.-K., *et al.* (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science (80-.).*, **313**, 1441–3.
13. Li,H., Sethuraman,N., Stadheim,T.A., Zha,D., Prinz,B., Ballew,N., Bobrowicz,P., Choi,B.-K., Cook,W.J., Cukan,M., *et al.* (2006) Optimization of humanized IgGs in glycoengineered Pichia pastoris. *Nat Biotechnol*, **24**, 210–5.
14. Vogl,T. and Glieder,A. (2013) Regulation of Pichia pastoris promoters and its consequences for protein production. *New Biotechnol.*, **30**, 385–404.
15. Hartner,F.S. and Glieder,A. (2006) Regulation of methanol utilisation pathway genes in yeasts. *Microb. Cell Fact.*, **5**, 39–59.
16. Hasslacher,M., Schall,M., Hayn,M., Bona,R., Rumbold,K., Lückl,J., Griengl,H., Kohlwein,S.D. and Schwab,H. (1997) High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. *Protein Expr. Purif.*, **11**, 61–71.
17. Mellitzer,A., Ruth,C., Gustafsson,C., Welch,M., Birner-Grünberger,R., Weis,R., Purkarthofer,T. and Glieder,A. (2014) Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks. *J. Biotechnol.*, 10.1016/j.jbiotec.2014.08.035.
18. Chapter 3.1 of this thesis (Vogl et al. manuscript in preparation).
19. Ruth,C., Zuellig,T., Mellitzer,A., Weis,R., Looser,V., Kovar,K. and Glieder,A. (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. *Syst. Biol. Synth. Biol.*, **4**, 181–91.
20. Dikicioglu,D., Wood,V., Rutherford,K.M., McDowall,M.D. and Oliver,S.G. (2014) Improving functional annotation for industrial microbes: a case study with Pichia pastoris. *Trends Biotechnol.*, 10.1016/j.tibtech.2014.05.003.
21. Lin-Cereghino,G.P., Godfrey,L., de la Cruz,B.J., Johnson,S., Khuongsathiene,S., Tolstorukov,I., Yan,M., Lin-Cereghino,J., Veenhuis,M., Subramani,S., *et al.* (2006) Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. *Mol. Cell. Biol.*, **26**, 883–97.
22. Kranthi,B.V., Kumar,R., Kumar,N.V., Rao,D.N. and Rangarajan,P.N. (2009) Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. *Biochim. Biophys. Acta*, **1789**, 460–8.
23. Sahu,U., Krishna Rao,K. and Rangarajan,P.N. (2014) Trm1p, a Zn(II)2Cys6-type transcription factor, is essential for the transcriptional activation of genes of methanol utilization pathway, in Pichia pastoris. *Biochem. Biophys. Res. Commun.*, 10.1016/j.bbrc.2014.07.094.

24. Takagi S, Tsutsumi N, Terui Y, Kong XY, inventors; Novozymes A/S, Bagsvaerd (DK), assignee. Method for methanol independent induction from methanol inducible promoters in Pichia. United States patent US 8,143,023. 2012 Jun 28.

25. Parua,P.K., Ryan,P.M., Trang,K. and Young,E.T. (2012) Pichia pastoris 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction. *Mol. Microbiol.*, **85**, 282–98.

26. Kumar,N.V. and Rangarajan,P.N. (2011) Catabolite repression of phosphoenolpyruvate carboxykinase by a zinc finger protein under biotin- and pyruvate carboxylase-deficient conditions in Pichia pastoris. *Microbiol. (Reading, U. K.)*, **157**, 3361–9.

27. Kumar,N.V. and Rangarajan,P.N. (2012) The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in Pichia pastoris. *J. Biol. Chem.*, **287**, 34465–73.

28. Polupanov,A.S., Nazarko,V.Y. and Sibirny,A.A. (2012) Gss1 protein of the methylotrophic yeast Pichia pastoris is involved in glucose sensing, pexophagy and catabolite repression. *Int. J. Biochem. Cell Biol.*, **44**, 1906–18.

29. Zhang,P., Zhang,W., Zhou,X., Bai,P., Cregg,J.M. and Zhang,Y. (2010) Catabolite repression of Aox in Pichia pastoris is dependent on hexose transporter PpHxt1 and pexophagy. *Appl. Environ. Microbiol.*, **76**, 6108–18.

30. Yurimoto,H., Oku,M. and Sakai,Y. (2011) Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int. J. Microbiol.*, **2011**, 101298.

31. Stasyk,O.G., van Zutphen,T., Ah Kang,H., Stasyk,O. V, Veenhuis,M. and Sibirny,A. a (2007) The role of Hansenula polymorpha MIG1 homologues in catabolite repression and pexophagy. *FEMS Yeast Res.*, **7**, 1103–13.

32. Ozimek,P., Lahtchev,K., Kiel,J.A.K.W., Veenhuis,M. and van der Klei,I.J. (2004) Hansenula polymorpha Swi1p and Snf2p are essential for methanol utilisation. *FEMS Yeast Res.*, **4**, 673–82.

33. Leao-Helder,A.N., Krikken,A.M., van der Klei,I.J., Kiel,J. a K.W. and Veenhuis,M. (2003) Transcriptional down-regulation of peroxisome numbers affects selective peroxisome degradation in Hansenula polymorpha. *J. Biol. Chem.*, **278**, 40749–56.

34. Yurimoto,H. (2009) Molecular Basis of Methanol-Inducible Gene Expression and Its Application in the Methylotrophic Yeast Candida boidinii. *Biosci. Biotechnol. Biochem.*, **73**, 793–800.

35. Yurimoto,H. and Sakai,Y. (2009) Methanol-inducible gene expression and heterologous protein production in the methylotrophic yeast Candida boidinii. *Biotechnol. Appl. Biochem.*, **53**, 85–92.

36. Zhai,Z., Yurimoto,H. and Sakai,Y. (2012) Molecular characterization of Candida boidinii MIG1 and its role in the regulation of methanol-inducible gene expression. 10.1002/yea.

37. Sasano,Y., Yurimoto,H., Yanaka,M. and Sakai,Y. (2008) Trm1p, a Zn(II)2Cys6-type transcription factor, is a master regulator of methanol-specific gene activation in the methylotrophic yeast Candida boidinii. *Eukaryot. Cell*, **7**, 527–36.

38. Sasano,Y., Yurimoto,H., Kuriyama,M. and Sakai,Y. (2010) Trm2p-dependent derepression is essential for methanol-specific gene activation in the methylotrophic yeast Candida boidinii. *FEMS Yeast Res.*, **10**, 535–44.

39. Hanson,S.J., Byrne,K.P. and Wolfe,K.H. (2014) Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus Saccharomyces cerevisiae system. *Proc. Natl. Acad. Sci. U. S. A.*, 10.1073/pnas.1416014111.

40. Ravin,N. V, Eldarov,M.A., Kadnikov,V. V, Beletsky,A. V, Schneider,J., Mardanova,E.S., Smekalova,E.M., Zvereva,M.I., Dontsova,O.A., Mardanov,A. V, *et al.* (2013) Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1.

41. Tachibana,C., Biddick,R., Law,G.L. and Young,E.T. (2007) A poised initiation complex is activated by SNF1. *J. Biol. Chem.*, **282**, 37308–15.

42. Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.G., Palmer,A.E. and Tsien,R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.*, **22**, 1567–72.

43. Liu,B., Paton,J.F. and Kasparov,S. (2008) Viral vectors based on bidirectional cell-specific mammalian promoters and transcriptional amplification strategy for use in vitro and in vivo. *BMC Biotechnol.*, **8**, 49.

44. Vogl,T., Thallinger,G.G., Zellnig,G., Drew,D., Cregg,J.M., Glieder,A. and Freigassner,M. (2014) Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression. *N. Biotechnol.*, **00**.

45. Tatusov,R.L., Galperin,M.Y., Natale,D. a and Koonin,E. V (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, **28**, 33–6.

46. Parua,P.K., Ratnakumar,S., Braun,K.A., Dombek,K.M., Arms,E., Ryan,P.M. and Young,E.T. (2010) 14-3-3 (Bmh) proteins inhibit transcription activation by Adr1 through direct binding to its regulatory domain. *Mol. Cell. Biol.*, **30**, 5273–83.

47. Yano,T., Yurimoto,H. and Sakai,Y. (2009) Activation of the oxidative stress regulator PpYap1 through conserved cysteine residues during methanol metabolism in the yeast Pichia pastoris. *Biosci., Biotechnol., Biochem.*, **73**, 1404–11.

48. Yano,T., Takigami,E., Yurimoto,H. and Sakai,Y. (2009) Yap1-regulated glutathione redox system curtails accumulation of formaldehyde and reactive oxygen species in methanol metabolism of Pichia pastoris. *Eukaryot. Cell*, **8**, 540–9.

49. Cregg,J.M., Tolstorukov,I., Kusari,A., Sunga,J., Madden,K. and Chappell,T. (2009) Expression in the yeast Pichia pastoris. *Methods Enzymol.*, **463**, 169–89.

50. Zhu,T., Guo,M., Sun,C., Qian,J., Zhuang,Y., Chu,J. and Zhang,S. (2009) A systematical investigation on the genetic stability of multi-copy Pichia pastoris strains. *Biotechnol. Lett.*, **31**, 679–84.

51. Geier,M., Fauland,P., Vogl,T. and Glieder,A. (2014) Compact multi-enzyme pathways in P. pastoris. *Chem. Commun. (Camb).*, **62**.

52. Denis,C.L. and Young,E.T. (1983) Isolation and characterization of the positive regulatory gene ADR1 from Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **3**, 360–70.

53. Ratnakumar,S., Kacherovsky,N., Arms,E. and Young,E.T. (2009) Snf1 controls the activity of adr1 through dephosphorylation of Ser230. *Genetics*, **182**, 735–45.

54. Nishi T, Ishiguro K, inventors; Kaneka corporation (JP), assignee. Yeast for transformation and process for producing protein. WO2012102171 A1 patent application.

55. Weis,R., Luiten,R., Skranc,W., Schwab,H., Wubbolts,M. and Glieder,A. (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. *FEMS Yeast Res.*, **5**, 179–89.

56. Panula-Perälä,J., Vasala,A., Karhunen,J., Ojamo,H., Neubauer,P. and Mursula,A. (2013) Small-scale slow glucose feed cultivation of Pichia pastoris without repression of AOX1 promoter: towards high throughput cultivations. *Bioprocess Biosyst. Eng.*, 10.1007/s00449-013-1098-9.

57. Vogl,T., Ruth,C., Pitzer,J., Kickenweiz,T. and Glieder,A. (2014) Synthetic Core Promoters for Pichia pastoris. *ACS Synth. Biol.*, **3**, 188–91.

58. Näätsaari,L., Mistlberger,B., Ruth,C., Hajek,T., Hartner,F.S. and Glieder,A. (2012) Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One*, **7**, e39720.

59. Krainer,F.W., Dietzsch,C., Hajek,T., Herwig,C., Spadiut,O. and Glieder,A. (2012) Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. *Microb Cell Fact*, **11**, 22.

60. Gibson,D.G., Young,L., Chuang,R., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–5.

61. Candida,Y., Sasano,Y., Yurimoto,H., Yanaka,M. and Sakai,Y. (2008) Trm1p , a Zn (II) 2 Cys 6 -Type Transcription Factor , Is a Master Regulator of Methanol-Specific Gene Activation in the Methylotrophic. **7**, 527–536.

62. Vallini,V., Berardi,E. and Strabbioli,R. (2000) Mutations affecting the expression of the MOX gene encoding peroxisomal methanol oxidase in Hansenula polymorpha. *Curr. Genet.*, **38**, 163–70.

63. Van Zutphen,T., Baerends,R.J.S., Susanna,K.A., de Jong,A., Kuipers,O.P., Veenhuis,M. and van der Klei,I.J. (2010) Adaptation of Hansenula polymorpha to methanol: a transcriptome analysis. *BMC Genomics*, **11**, 1.

64. Wilson,W.A., Hawley,S.A. and Hardie,D.G. (1996) Glucose repression/derepression in budding yeast: SNF1 protein kinase is activated by phosphorylation under derepressing conditions, and this correlates with a high AMP:ATP ratio. *Curr. Biol.*, **6**, 1426–34.

65. Heo,J., Hong,W., Cho,E., Kim,M., Kim,J., Kim,C., Rhee,S. and Kang,H. (2003) Properties of the Hansenula polymorpha-derived constitutive GAP promoter, assessed using an HSA reporter gene. *FEMS Yeast Res.*, **4**, 175–184.

66. Zaman,S., Lippman,S.I., Schneper,L., Slonim,N. and Broach,J.R. (2009) Glucose regulates transcription in yeast through a network of signaling pathways. *Mol. Syst. Biol.*, **5**, 245.

67. Traven,A., Jelicic,B. and Sopta,M. (2006) Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep.*, **7**, 496–9.

68. Rundlett,S.E., Carmen,A.A., Kobayashi,R., Bavykin,S., Turner,B.M. and Grunstein,M. (1996) HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 14503–8.

Chapter 3.6

Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*

Short title: Restriction site free cloning (RSFC) plasmids

This work was performed in equal contribution with Mudassar Ahamd.

Thomas Vogl[1*¶], Mudassar Ahmad[1¶], Florian W. Krainer[1], Helmut Schwab[1], Anton Glieder[1]

[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz A-8010, Austria

* Corresponding author
E-mail: thomas.vogl@aon.at

¶ These authors contributed equally.

Table of contents

Abstract

Tagging proteins is a standard method facilitating protein detection, purification or targeting. When tagging a certain protein of interest, it is challenging to predict which tag will give optimal results and will not interfere with protein folding, activity or yields. Here we present a vector family of 40 plasmids allowing simple, seamless fusions of a single PCR product with various N- and C-terminal tags, signal sequences and promoters. In conventional cloning, tags are either added on PCR primers, requiring a distinct primer per tag, or provided on the vector, leaving a restriction site scar. The restriction site free cloning (RSFC) strategy presented in this paper relies on seamless cloning using type IIS restriction endonucleases. After cutting out a stuffer (placeholder) fragment from the vectors, a single PCR product can be directly inserted in frame into all 40 plasmids using blunt end or TA ligations, requiring only verification of the orientation.

We have established a RSFC vector family for the commonly used protein expression host *Pichia pastoris* and demonstrated the system with the secretory expression of horseradish peroxidase (HRP). HRP fusions to four tags (Myc, FLAG, His, Strep) and two fusion proteins (GFP and MBP) showed a 31 fold difference in volumetric activities. C-terminal tagging caused in some cases almost a complete loss of function, whereas N-terminal tags showed moderate differences. The results obtained with HRP underline the importance of comparing different tags to maximize activities of fusion proteins. In a similar fashion the RSFC strategy could be applied to screen for optimal tags, promoters or signal sequences in other expression systems or to facilitate the evaluation of (iso-) enzyme families.

Introduction

Protein tags are commonly applied tools facilitating purification (affinity tags), enabling immuno-detection (epitope tags) or increasing solubility. Fusions to fluorescent proteins help elucidating the cellular localization and fusions to signal sequences provide specific intracellular targeting or secretion [1,2]. However, as an extrinsic addition to a protein of interest (POI), such fusions may also show detrimental effects by affecting protein conformation, yields, activity or stability [1,3]. The specific interactions of the POI with a certain tag are generally hard to foresee and may also depend on the position of the tag (N- or C-terminal). Unknown proteolytic processing or intracellular targeting of the POI may also influence the suitability of a specific fusion site. In addition, the same tagged protein may behave differently depending on the host system used (e.g. bacteria, yeast, higher eukaryotes) [4]. As there are large numbers of affinity, epitope tags and fusion proteins available it is challenging to predict the optimal choice for a certain POI. Therefore, commonly multiple tags are tested in N- or C- terminal positions and screened for optimal results [4–7].

However, preparing expression constructs containing multiple tags may require tedious cloning work. Tags are commonly provided on the plasmid adjacent to the multiple cloning site (MCS). This requires unique vectors for each tag and N-/C-terminal position. The gene of interest (GOI) needs to be cloned into the MCS via unique restriction endonuclease (RE) recognition sites. These restriction site scars remain in the protein coding sequence (CDS) and are later translated into additional amino acids, which may interfere with the POI's properties. Also cloning strategies based on *in vivo* recombination such as Gateway (e.g. [8]) leave the recombination sequence as a scar in the CDS.

Ideally, tags should be fused seamlessly to the GOI i.e. without any restriction site scars or additional sequences from the MCS. Seamless cloning can be achieved by various strategies [9]. Frequently,

tags are directly added by PCR as a 5' overhang of a primer and thereby seamlessly attached to the CDS. This approach requires however a unique primer for each tag, N-/C-terminal position and each GOI.

We aimed to design a simple, seamless system to facilitate testing of multiple tags in N-/C-terminal position at minimal cost and effort (e.g. without the need to order numerous primers).

Several novel cloning methods are completely independent of REs and allow simple assembly of multiple fragments solely by short overlaps (around 25 bp) relying on *in vitro* 'recombination' (e.g. annealing of single stranded overhangs and enzymatic linkage). These methods include SLIC (sequence and ligation–independent cloning) [10], SLiCE (Seamless Ligation Cloning Extract) [11], Gibson assembly [12], CPEC (Circular Polymerase Extension Cloning) [13] and are concisely compared on the website of the Joint BioEnergy Institute (JBEI), Emeryville, CA ,USA [14,15]. All these methods may be used to seamlessly add a tag to a protein by adding the tag sequence to a PCR primer. However, there is an additional overhang required for *in vitro* recombination with the vector, requiring relatively long primers. Most inconveniently a new primer is needed for each tag, each position and each POI to be tested.

Therefore we have based our strategy on type IIS REs. In contrast to type II REs, which recognize and cut within a palindromic sequence, type IIS REs cut outside of a non-palindromic recognition sequence [16,17]. Thereby RE site scars can be circumvented making type IIS REs prominent tools for seamless cloning [9]. There are various type IIS enzymes available that create different types of overhangs including up to 4 bp overhangs suitable for sticky end cloning (e.g. *Eam*1104I [18], *Bsa*I [19,20]), single base pair overhangs that can be applied for TA cloning (e.g. *Xcm*I [21,22], *Eam*1105I [23,24], *BciVI* [25]) or blunt end cloning (*Mly*I/*Sch*I [25,26]), see Fig. 1 A.

In this study we have evaluated type IIS REs for blunt end and TA cloning and designed a restriction site free cloning (RSFC) strategy that enables simple, seamless cloning of a PCR product in frame with any desired upstream or downstream sequence in a vector. Based on this strategy, we have designed a RSFC vector family of 40 plasmids for the methylotrophic yeast *Pichia pastoris*, a commonly used protein production host for industrially relevant biocatalysts and biopharmaceuticals [27–29]. The vectors feature different epitope and affinity tags (Myc, FLAG, His, Strep) and fusion proteins (eGFP and MBP) in N- and C-terminal position that are provided for intracellular and secretory expression.

Results and Discussion

Restriction site free cloning (RSFC)

Blunt end vs. TA cloning concept

We aimed to design a vector system in which a single PCR product of a GOI can be directly fused, sequence independently to various N- or C-terminal tags provided on different plasmids. Thereby only two primers are required to test seamless fusions of multiple tags with the GOI. This design is achieved by inserting a stuffer (placeholder) fragment flanked by two type IIS RE sites in opposite orientations in all vectors (Fig. 1 B-D). The CDSs of different N- and/or C-terminal tags or fusion proteins are provided upstream/downstream of the stuffer fragment. By digestion using the respective type IIS RE, the stuffer fragment including the RE sites is cut out, resulting in RE site free vector backbones that can be directly ligated with the same PCR product (Fig. 1 B).

Commonly used type IIS RE based cloning strategies such as Golden Gate cloning [19,20] cannot be used for this purpose as they rely on type IIS enzymes creating short overhangs such as *Eam*1104I or

*Bsa*I (Fig. 1 A). The use of these enzymes requires also RE digestion of the PCR product and the overhangs created on the vectors would differ between tags and impede seamless fusions.
Direct, sequence independent cloning of PCR products is in this context only possible by using TA cloning or blunt end ligations. These methods are in general not directional (with a few exceptions e.g. [25,26]) and require verification of the orientation (e.g. by colony PCR, cPCR). TA cloning is based on the property of *Taq*-Polymerase to add a single deoxyadenine (dA) nucleotide at the 3' ends of amplified DNA [21]. These PCR products can be directly cloned using a vector with a single 3' deoxythymidine (dT) overhang. TA cloning works more efficiently than blunt end cloning [21], however the required dA nucleotide complicates seamless fusions to tags as it must be universally incorporated in the transitions between tag and vector. In this respect, blunt end ligations, that are completely sequence independent, are more favorable.
We designed test vectors based on type IIS REs for blunt end and TA cloning to compare their suitability. There is only one blunt end type IIS RE available that cuts outside of its recognition sequence (*Mly*I). *Mly*I has also been established for directional blunt end ligations of PCR products using a *lac*O, *lac*Z based blue-white screening in *Escherichia coli* [25,26]. There are several type IIS REs available, that create a single base 3' overhang (e.g. *Bmr*I, *Bci*VI, *Hph*I, see Fig. 1 A). We tested commercially available preparations of these three enzymes all of which showed sufficient cleavage efficiencies (data not shown). *Hph*I and *Bci*VI have been previously used for TA cloning [21,25], yet these restriction sites were present more frequently in the vector backbones we wanted to use. Therefore we used *Bmr*I.
The basic sequence design of the transitions between the vector, the type IIS restriction sites and the stuffer fragment are shown in Fig. 1 C,D. For blunt end cloning using *Mly*I, the design is completely sequence independent (Fig. 1 C). For TA cloning, 3' dT residues must be provided on the vector backbone and incorporated in the transition between vector and GOI. We solved this by using the dT nucleotide of the start codon (ATG) and the dA nucleotide of a partial stop codon (TAX), creating a 3' dT base on the reverse strand (Fig 1. D). Depending on the desired sequence context, 'X' may be provided on the vector side as A/G for a stop codon (translation termination) or T/C (coding for tyrosine, for linkage of C-terminal tags).

Cloning efficiencies

We compared the basic blunt end and TA cloning based system at first with expression vectors for *Schizosaccharomyces pombe* as these plasmids required fewer modifications in the vector backbones than the *P. pastoris* plasmids we intended to use. See Figure S 2 for plasmid maps and the Materials and Methods section for details on the design. After cutting out the stuffer fragment using *Mly*I or *Bmr*I, the vector backbones were dephosphorylated to counter act self-ligation. Primers for insert amplification were phosphorylated prior to ligation (see Material and Methods section for experimental details and a simple, cost effective protocol). Both cloning strategies resulted in similar transformation efficiencies (via electroporation), approximately 10^2 to 10^3 colony forming units (cfu)/µg DNA (in the ligation reaction) with self-made competent cells (competence with circular, supercoiled plasmids: 10^6 to 10^7 cfu/µg DNA) and in both cases all 10 out of 10 clones tested contained an insert. We verified the orientation by cPCR; as statistically expected approximately half the clones contained an insert in the correct orientation (blunt end/*Mly*I: 5 of 10, TA cloning/*Bmr*I: 7 of 10). Supplementary figure S 3 outlines a simple cPCR strategy to test the correct orientation (using sequencing primers of the vector and the primers used for amplifying the insert). The vector/insert transitions were confirmed by sequencing and did not show any mutations. However, when cloning

an insert into a larger set of vectors using blunt end ligations (see *P. pastoris* vectors below) we noticed occasionally single bp deletions of the insert adjacent to the vector transition (<5% of constructs). Sequencing of additional transformants resulted in all cases in correct sequences. Notably, the deletions were always in the 5' ends of the insert and occurred more often after repeated freeze/thaw cycles of the PCR product. We therefore recommend aliquoting the PCR product and vector backbones and using them only once.

In general these RSFC ligations resulted in lower efficiencies (cfu/µg DNA) than comparable sticky end ligations, but still yielded sufficient numbers of transformants for our standard cloning applications. *Mly*I based blunt end ligations worked similarly efficient as *Bmr*I based TA cloning. Previously, TA cloning has been reported to be more efficient than blunt end cloning [21], however the difference may arise from the different enzymes used for vector preparation in our study.
We mutated *Mly*I sites present in the vector backbones to enable the stuffer removal (see Material and Methods section for details). All mutations but one resulted in no differences in DNA yields compared to the parental plasmids. Mutating a *Mly*I site in the *E. coli* pUC origin of replication to a sequence previously reported [25,26] decreased plasmid yields to approximately 30 % of the unmutated parental plasmid (wildtype pUC: ~400 ng/µl, *Mly*I mutated pUC: ~120 ng/µl). The *MlyI* site appears also in other high copy number origins of replication and switching to a lower copy number ORI would also result in lower plasmid yields. We intended to use the RSFC plasmids only for sub cloning and aimed to perform expression in *P. pastoris*. To this end mini prep yields (typically >5µg) were sufficient. However if similar plasmids should be used for expression in *E. coli*, we would recommend to screen other possible mutations of the *Mly*I site using degenerate primers to restore wild type plasmid yields.
However, the blunt end/*Mly*I based strategy allowed completely seamless cloning whereas seamless TA cloning was hindered by the requirement for dT/dA bases in the insert/vector transition. This problem is similar to the use of typeIIS enzymes creating longer sticky end overhangs that need to be complementary between all plasmids of a vector family (for example in plasmids by BioGrammatics, Inc., Carlsbad, CA, USA and 'Electra' plasmids by DNA2.0, Inc., Menlo Park, CA, USA). As outlined in Fig. 1 D, the TA strategy can be still used for fusion of the same PCR product to different tags, however N-terminal tags are always linked via an ATG (coding for methionine/start), whereas C-terminal tags must be linked via tyrosine codons. Tyrosine is naturally a relatively rarely occurring and bulky amino acid, making it structurally rather unfavorable as a linker to an adjacent tag. In 'Electra' plasmids by DNA2.0 this issue is solved by adding an additional C-terminal 'linker' amino acid to all vectors, whereas in the RSFC strategy only vectors with C-terminal tags require a linker amino acid. Still we have solely focused on the blunt end/*Mly*I based strategy in the following plasmid design for *P. pastoris*. The blunt end/*Mly*I based ligations required no A-tailing step of PCR products but reached similar ligation efficiencies as TA cloning and allowed completely sequence independent fusions.
In summary, our cloning approach, relying on blunt end or TA ligations between a phosphorylated PCR product and a dephosphorylated vector backbone created by type IIS RE digestion, allowed seamless, sequence independent cloning at reasonable efficiencies. PCR products can be directly used for ligations and do not need RE digestion, therefore any insert sequence can be used (TA cloning with proof reading polymerases requires a separate dA-tailing step). There have previously been type IIS based cloning efforts using blunt end and TA ligations for the cloning of PCR fragments [21–26]. However, these strategies did not allow seamless fusions and are in part with *lac*O, *lac*Z

based blue white screening [25,26], despite the convenience of directional cloning, even incompatible with seamless fusions. To distinguish our approach from these efforts and other type IIS based strategies such as Golden Gate cloning [19,20], we have termed our approach restriction site free cloning (RSFC).

RSFC plasmids for *P. pastoris* as toolbox for optimizing protein production

Tags and fusion proteins

We applied the RSFC cloning strategy to design a plasmid family for *P. pastoris* allowing seamless fusions of a GOI with various tags and fusion proteins in N- and C-terminal position. There are different expression plasmids available for *P. pastoris* based on various cloning strategies such as Gateway [8], TA cloning [22,25], sticky end type IIS ligations (plasmids by BioGrammatics, 'Electra' plasmids by DNA2.0) and 'classical' typeII RE/ligation based systems ([30–32] and *P. pastoris* plasmids by Life Technologies, Carlsbad, CA, USA). The pCri vector family [32] is a multi-host platform, allowing to clone a single PCR product via restriction digestion and a MCS into different vectors. For *P. pastoris* only three pCri plasmids with a His tag are available. Therefore none of the vector systems currently available for *P. pastoris* offer different tags and only the BioGrammatics and Electra plasmids by DNA2.0 vectors allow seamless, yet sequence dependent cloning still requiring restriction digestion of the insert.

We designed a set of 40 RSFC plasmids for *P. pastoris* (termed pPpRSFC) offering different tags (Myc, FLAG, His, Strep) and fusion proteins (enhanced green fluorescent protein, eGFP and maltose binding protein, MBP) in N- and C-terminal position, see Tab. 1 for exact properties and Fig. 2 for a schematic vector map. We have assigned numbers (#1 to #40) to the plasmids and are using them hereafter when referring to a specific construct.

After stuffer removal by *Mly*I digestion, a single PCR product can be cloned in a seamless and sequence independent fashion into all vectors, fused to tags and fusion proteins ranging from 18 to 1101 bp in length. Epitope and affinity tags are included and constructs with affinity tags contain a TEV protease cleavage site to allow tag removal. The hexameric His tag is provided with and without TEV protease cleavage site. MBP is provided as a fusion protein with the potential to improve solubility and act as a purification tag, although in *P. pastoris* problems with proteolytic degradation have been reported [33]. eGFP is an enhanced version of GFP allowing simple fluorescence detection of tagged proteins.

When cloning a GOI into the pPpRSFC vectors, the forward primer must be designed starting at the DNA sequence coding for the 2nd amino acid of the POI (omitting the ATG start codon). The reverse primer must be designed starting (on the reverse strand) at the DNA sequence coding for the last amino acid/penultimate codon (omitting the stop codon). Especially a stop codon on the PCR product would interfere with tag fusions, therefore the start and stop codon are always provided on the vectors and must be omitted from PCR inserts.

P. pastoris is not only suitable for intracellular expression but can also produce secreted heterologous proteins at high titers while secreting only little endogenous protein [27–29]. Therefore we designed all plasmids also for secretory expression using the *S. cerevisiae* mating factor alpha pre-pro signal sequence (MF alpha), the most commonly applied signal sequence in *P. pastoris*. The MF alpha sequence is processed by two proteases (Ste13 and Kex2) that cleave the

amino acid sequence KREAEA at the end of MF alpha [34]. Kex2 cleaves efficiently after KR whereas the Ste13 cleavage after the EA repeat may be incomplete, depending on the following amino acids of the POI. In several cases removal of the EAEA repeats has led to a more homogenous product [35,36]. Therefore we designed the basic MF alpha pPpRSFC plasmids (#2, 3, 5, 6, 36, 37, 39, 40) with and without the EAEA sequence. Plasmids bearing tags always contain the EAEA repeat (Tab. 1).

Promoters, integration events and resistance markers

The pPpRSFC plasmids are based on the pPpT4 vector family reported by Näätsaari *et al*. [30] and also used as a platform for the *P. pastoris* Electra vectors by DNA2.0. The majority of pPpRSFC plasmids contain the promoter of the *alcohol oxidase 1* gene (p*AOX1*). This strong, tightly regulated methanol inducible promoter is most commonly used in *P. pastoris* [37]. We have also designed basic plasmids bearing the *glyceraldehyde-3-phosphate dehydrogenase* promoter (p*GAP*) to enable methanol free, constitutive expression (see Tab. 1).

In contrast to the yeast *S. cerevisiae*, where stable, autonomously replicating plasmids are available, circular plasmids bearing a yeast ARS (autonomously replicating sequence) are not stable in *P. pastoris* and genomic integration of plasmid cassettes is the method of choice for heterologous gene expression [27,38]. Most commonly *P. pastoris* integration cassettes are created by linearizing plasmids or generation of linear cassettes by PCR [39,40] and targeted to the *AOX1* locus via homologous sequences. Depending on the linearization site in the plasmid, different homologous recombination events can be targeted [38]. The pPpRSFC plasmids allow linearization to target gene replacement at the *AOX1* locus. Thereby the endogenous *AOX1* gene is deleted and the minor *AOX2* gene must take over the function of oxidizing methanol to formaldehyde. Due to the lower expression levels of *AOX2*, these *aox1* knockout strains show a Mut^S (methanol utilization slow) phenotype, which may result in higher yields than a Mut^+ phenotype [41,42]. This can be achieved by linearization using *Bgl*II. If the *Bgl*II site is present in the insert, the vectors can still be linearized using the rare 8 bp *Swa*I sites as a failsafe backup. If a Mut^+ phenotype is desired, the vectors can be linearized using unique REs cleaving in the 5' or 3' homologous sequence (e.g. *Pme*I, *Ase*I or *Eco*NI, *Bsr*BI). However, due to low homologous recombination frequencies in *P. pastoris* wildtype strains [30], even when targeting a gene replacement at the *AOX1* locus (expected Mut^S phenotype), still the majority of transformants are Mut^+. Therefore it is necessary to validate the Mut phenotype by growth on methanol plates.

The RSFC plasmids are based on a modular design, the promoter, N- or C-terminal tags, terminator, the resistance marker and the 3' homologous sequence can be easily exchanged using unique restriction enzyme sites (Fig. 2). Most plasmids are based on Zeocin selection, however basic expression plasmids (#35-40) were also constructed with a histidine marker to be used with auxotrophic strains. The tagged expression cassettes from the Zeocin plasmids can be easily shuffled to the histidine plasmid backbones using unique *Pci*I and *Bam*HI sites.

In the pPpRSFC plamids not only the transition between the insert and the vector is seamless, also the transition of the promoter to the start codon and the stop codon to the terminator are seamless. In standard RE based cloning, the MCS may interfere with translation initiation [43] and this appears relevant for the *AOX1* promoter as extensions of the 5' UTR (also caused by a MCS) were shown to negatively affect expression [44].

Applications of RSFC vectors to optimize HRP expression in *P. pastoris*

Effects of tags and fusions proteins in N- and C-terminal position

With the set of pPpRSFC plasmids available, we aimed to validate the system with a typical application. We tested expression of horseradish peroxidase (HRP), a commonly used reporter enzyme for signal amplification in diagnostic kits and immunohistochemistry. Secretory expression of HRP has been previously demonstrated in *P. pastoris* [42,45–47]. Cytoplasmic expression promised little chance of success as HRP is a secretory plant peroxidase that requires formation of disulfide bridges and is typically glycosylated in the secretory pathway [48,49]. Still, we tested the basic pPpRSFC plasmid (#1, untagged, p*AOX1*) for cytoplasmic HRP expression. This construct showed neither activity in the supernatant (Fig. 3) nor in the cytoplasm (data not shown). Therefore different tags were only evaluated for secretory expression. A single PCR fragment of HRP was cloned into the vectors as outlined above. All p*AOX1* plasmids were linearized via *Bgl*II sites to target a gene replacement event at the *AOX1* locus, and screened for a Mut^S phenotype, which has been reported to be more favorable for HRP expression than Mut^+ [42].

The different tags and positions had diverse effects on volumetric HRP activities (Fig. 3) and led to valuable insights. For all tags, the N-terminal version was giving higher activities than the C-terminal version. For the larger fusion proteins (eGFP and MBP), C-terminal tagging even led to almost complete loss of activity (#10 and #30). Comparing the tagged construct with the highest activity (#21) with the construct of the lowest activity (#10) gives a 31 fold difference. Surprisingly, N-terminal tagging with the relatively large eGFP (and MBP) did not strongly affect activity, whereas shorter tags (Myc, FLAG, His, Strep) showed varying effects. The N-terminally His tagged construct with TEV protease cleavage site (#25) showed the lowest activity of all N-terminally tagged proteins. However, the N-terminal His tagged construct without TEV protease cleavage site (#21) showed activity similar to other tags, hinting a negative effect of the TEV protease cleavage site in this context. Changes of the MFalpha sequence by removal of the EAEA sequence decreased activity 1.6 fold with the methanol inducible *AOX1* promoter (#2 vs. #3). With the constitutive *GAP* promoter (#5 vs. #6), removal of the EAEA sequence even led to a 17 fold decrease in activity. A possible mechanistic explanation would be that the EAEA repeats improved secretion due to increased Kex2 cleavage efficiencies ([50,51]). p*GAP* driven HRP expression was therefore, depending on the presence of EAEA repeats, competitive to the methanol inducible p*AOX1*. Due to shorter process times (not requiring methanol induction) p*GAP* driven expression may even be more favorable for HRP production regarding space time yields and process setup.

The effects seen on volumetric activities by using different tags may be caused by various reasons. On the one hand the tags may have interfered to a different extent with protein folding or access of the substrate to the active site thereby negatively affecting activity. On the other hand they also may have affected the protein yields by altering the protein stability, interfering with the secretion process or even on the mRNA level with transcript stability. Also the tags or linker sequence may have targeted proteolytic degradation, as previously reported for MBP in *P. pastoris* [33]. However, as we aimed only to evaluate the suitability of the RSFC strategy for screening different tags, we did not further investigate the underlying causes. The pPpRSFC plasmid family proved to be a simple tool to optimize volumetric activities of tagged HRP, showing that especially the tag positions and presence of EAEA repeats are crucial factors.

Fluorescence microscopy of strains expressing eGFP tagged HRP

GFP has routinely been used in *P. pastoris* as an intracellular reporter for comparing promoter activities [52–54] and to facilitate screening of protein production by testing fusions proteins [55], especially for membrane proteins [56–58]. Concerning GFP fusions of secretory proteins, conflicting results were obtained. In some cases GFP was successfully used as secretion reporter and for protein fusions [59–62]. In other cases problems with secretion (e.g. intracellular retention) were noticed [63–66]. As we had also designed N- and C-terminal fusions with eGFP (including the MFalpha signal sequence for secretion, #9 and #10), we performed fluorescence microscopy to investigate possible cellular retention and bottlenecks in the HRP secretion process.

The N-terminal eGFP-HRP fusion exhibited largely unchanged HRP activity, whereas the C-terminally tagged version had almost completely lost activity (Fig. 3). We also included controls of intracellular eGFP expression (#1) and secretory eGFP alone (without an HRP fusion, created by self-ligating #9). Fluorescence microscopy images of methanol grown cells are shown in (Fig. 4). While cytoplasmic expression showed bright fluorescence of the whole cell (D), all secretory constructs (A,B,C) showed punctate structures. These structures appeared somewhat similar to ER or Golgi mistargeting observed previously when expressing a GFP tagged membrane protein (human μ-opioid receptor, a G-protein coupled receptor) [65]. Most notably also the control of eGFP alone (C), without an HRP fusion showed this retention. eGFP was apparently poorly secreted in *P. pastoris* and effects evoked by HRP may be masked and outweighed by the poor eGFP secretion.

We also measured eGFP fluorescence in the supernatant and the cell fraction (Figure S 4). Fluorescence in the supernatant could be detected for secretory constructs (A,B,C), while the cytoplasmic eGFP expression control (D) showed only marginal fluorescence in the supernatant. However, also for the secretory constructs (A,B,C) intracellular fluorescence surpassed fluorescence in the supernatant approximately 5 to 12 fold. These results suggested together with the microscopy images, that large amounts of eGFP were withheld in the secretion process. In this respect, eGFP fusion proteins may be used with caution when investigating secretory processes in *P. pastoris*. However, these effects may also be protein dependent, as there were cases reported where GFP was successfully used to evaluate signal sequences [59,60] and some GFP fusion proteins were sufficiently secreted [61,62].

Conclusions and outlook

The RSFC cloning strategy outlined here and the pPpRSFC plasmid family are simple tools to optimize expression of tagged proteins with little cloning efforts. RSFC requires at first the design and assembly of the vector family to be used. However, subsequent screening is drastically facilitated as large amounts of vector backbones can be prepared at once by *Mly*I digestion. Subsequently, the backbones ready for cloning can be frozen as aliquots and thawed when needed.

There have been systems reported previously that allow testing of the expression of a POI in different expression hosts by using only two PCR products [67]. This approach is based on ligation-independent cloning (LIC) similar to [10–13,68]. While these methods allow highly efficient, seamless cloning, they rely on the annealing of single stranded overhangs, thereby requiring identical sequences between vector and insert. Therefore these methods are not suitable for seamless, sequence independent fusions possible with RSFC. However, as a downside of RSFC the blunt end ligations work less efficiently than annealing based *in vitro* recombination methods [10–13] and confirmation of the correct orientation is required. Otherwise only about 50% of the transformants

show the desired orientation which is a disadvantage for library approaches. Nevertheless, after stuffer removal, inserts can also be cloned directionally into RSFC plasmids by *in vitro* recombination methods. However this task requires the design of a separate primer for every tag and position to be tested as the overhang required for annealing changes between the vectors. We recommend using *in vitro* recombination methods (such as Gibson assembly [12]) with pPpRSFC plasmids when testing only a few constructs. When testing a larger number of constructs, the increased costs for primers and materials associated with *in vitro* recombination methods outweigh the costs for cPCRs to test the orientation of blunt end ligations.

A limitation of the RSFC system reported here is the use of *Mly*I, the only type IIS enzyme performing a blunt end cleavage. The recognition sequence of *Mly*I is five bp long (Fig. 1 A), thereby posing a problem as it appears statistically once per 512 bp ($4^5/2$) [69]. This may require frequent removal of *MlyI* sites in the vector backbones to be used. *Mly*I sites in CDSs of tags, fusion proteins and resistance markers can be easily removed by mutating the *Mly*I sequence to synonymous codons. However, mutating sites present in promoters, terminators or origins of replication have to be validated for unchanged functionality (or must be exchanged for parts free of *Mly*I sites). These issues could be solved by using artificial type IIS REs with longer recognition sequences. The catalytic domain of the archetypical type IIS enzyme *Fok*I has been fused to I-*Sce*I, a homing endonuclease with an 18 bp recognition sequence. This chimeric meganuclease showed sufficient cleavage resulting in 4 bp 'sticky' overhangs that could be ligated at 90% fidelity [69]. Following this strategy, the catalytic domain of *Mly*I (which is similar to *Fok*I [70]) could also be fused to I-*Sce*I. Statistically an 18 bp recognition sequence would appear approximately once in 10^{11} (4^{18}) bp, however I-*Sce*I recognizes also slightly degenerate sequences leading to an estimated appearance once in 10^8 bp [69,71,72], which would still surpass the specificity of wildtype *Mly*I by several orders of magnitude.

Most vectors for *P. pastoris* have been conceptualized solely as straightforward expression vectors ([8,22,25,30,31] and *P. pastoris* plasmids by Life Technologies, BioGrammatics and DNA2.0) and few plasmid families allow to fine-tune expression [30,31]. The 40 plasmids reported here extend the scope of applications and facilitate characterization and optimization of the production of heterologous proteins in *P. pastoris*. The RSFC strategy outlined here is not limited to tags and fusions proteins, but could also be applied to compare different promoters or signal sequences in other expression systems. Similarly, isoenzymes or families of homologous enzymes can be fused to tags to screen for better expression, solubility or other properties to identify enzymes combining desired biological, chemical and technological features.

Materials and Methods

Chemicals and media

Phusion DNA Polymerase, restriction endonucleases and other DNA modifying enzymes were acquired from Thermo Fisher Scientific (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). Miscellaneous chemicals were purchased from Becton, Dickinson and Company (Franklin Lakes, NJ, USA), Carl Roth (Karlsruhe, Germany) and Fresenius Kabi Austria (Graz, Austria).

Plasmids were isolated using a GeneJET Plasmid Miniprep Kit by Thermo Fisher Scientific. The standard protocol was optimized for *Mly*I based constructs to compensate the decreased plasmid yields. A single colony of a strain bearing the respective plasmid was streaked on an agar plate containing the respective antibiotic. After incubation overnight, a cell pellet (approximately 0.1 g wet cells) was scratched of the plate and used for the isolations (final elution volume: 100 µl of ddH_2O).

Agarose embedded DNA, restriction digests and PCRs were purified using a Wizard SV Gel and PCR Clean-Up System by Promega.

P. pastoris strains were grown on full medium (yeast extract, peptone, 2 % glucose, YPD), buffered minimal dextrose (BMD) and buffered minimal methanol medium with 0,5% methanol (BMM) as described by Weis *et al*. [16]. As only exception we used 2 % glucose in the BMD medium and for HRP expression, media were supplemented with 1 mmol/l ferrous sulfate heptahydrate ($FeSO_4.7H_2O$). *Escherichia coli* strains were selected on LB-medium containing 25 µg/ml Zeocin™ (Life Technologies, Carlsbad, CA, USA). *P. pastoris* transformants were selected on YPD agar plates containing 100 µg/ml Zeocin. Primers were ordered from Integrated DNA Technologies (Leuven, Belgium), see Supplementary Table S 5 for the sequences.

Plasmid construction

S. pombe RSFC test vectors pGAZ2-TA-BmrI-stuffer and pGAZ2-Blunt-MlyI-stuffer

For all cloning work an *E. coli* Top10 F' strain was used. The vectors for initially comparing blunt end and TA cloning were based on a replicative *S. pombe* vector pGAZ2 (S 2, unpublished results). For the TA-cloning vector 'pGAZ2-TA-BmrI-stuffer', a stuffer fragment was amplified using primers TA_fwd_HindIII+BmrI+stuffer and TA_rev_BamHI+BmrI+stuffer (see Supplementary Table S 5) and cloned into pGAZ2 via *Hind*III and *Bam*HI sites. The stuffer fragment was selected as a sequence that has no significant homology to *E. coli* and *S. pombe* genomes and lacks *Mly*I, *Bmr*I, *Hind*III and *Bam*HI RE sites; we used a part of a *P. pastoris* alpha, alpha trehalase gene. The 'pGAZ2-Blunt-MlyI-stuffer' vector required mutating two *Mly*I sites in the vector backbone. This was done by PCR amplifying the vector using primers pUC_mut_MlyI_fwd + pUC_mut_MlyI_rev and ZeoCDS_mut_MlyI_fwd + ZeoCDS_mut_MlyI_rev using Pfu Ultra polymerase (Agilent Technologies, Santa Clara, CA) followed by *Dpn*I digestion to remove template vector. The *Mly*I site in the pUC was mutated to the sequence reported by Rao *et al*. [25], the *Mly*I site in the zeocin resistance gene was mutated to a synonymous codon. After transformation, introduction of the correct mutations were confirmed by Sanger sequencing. Both plasmids do not provide seamless fusions, as the stuffer fragments were for convenience inserted via *Hind*III and *Bam*HI sites leaving RE site scars. For test purposes the gene coding for *Thermomyces lanuginosus* endo-beta-1,4-D-xylanase was amplified using primers Xyla_fwd and Xyla_rev and cloned into the two vectors (detailed protocol see below).

P. pastoris pPpRSFC plasmid family

The expression plasmids used in this study are based on the pPpT4 *P. pastoris*/*E. coli* shuttle vector family (e.g. GenBank accession number JQ519690.1) reported by Näätsaari *et al*. [30]. Two *Mly*I sites in the backbone (pUC and zeocin resticane gene) were mutated in the same way as in the *S. pombe* plasmids (same primers) and confirmed by sequencing. The *AOX1* promoter, *Mly*I stuffer fragment and *AOX1* terminator were amplified in separate PCR reactions using primers PAOX1_PciIF/OePAox1StufferR, OestufferF/OeStufferR and OeAox1TTstufferF/Aox1TT_BamHIR respectively. In the subsequent overlap extension PCR reactions the fragments were joined together using primer pair PAOX1_PciIF/ AOX1TT_BamHIR followed by restriction with *Pci*I/*Bam*HI and were cloned in a vector backbone with mutated *Mly*I sites to create an intermediatory plasmid backbone termed 'pPp'. The stuffer fragment sequence was selected from as a sequence that has no homology to *E. coli* and *P. pastoris* and lacks unique RE used in the pPpRSFC plasmid family. We selected a part of a gene involved in the *S. cerevisiae* biotin metabolism. An *Eco*RI site in the stuffer was mutated using primers pairs OeEcoRIF and OeEcoRIR. There appeared a few additional mutations in the stuffer that had no functional consequences and where therefore left unchanged (see plasmid sequences in Supplementary file S 1).

For constitutive plasmids, the *GAP* promoter was amplified via primers GAP_PciIF/OeGapStuffR and was cloned into the pPp backbone using PciI/EcoRI to create #4 (pPpRSFC-pGAP). The 3' AOX1 homologous sequence was amplified via primers 3'AOX1_PstIASCIF/3'AOX1_KpnISwaIR and was cloned into pPp using *Kpn*I/*Pst*I restriction sites to create #1 (pPpRSFC). For secretory expression plasmids, the MFAlpha sequence was amplified using primer pair AlphaFSSF/AlphaEcoRIR (or aEAEAEcoRIR for insertion of Glu-Ala repeats). The *AOX1*/*GAP* promoters were amplified via primers PAOX1_PciIF + OeAlphaPAox1R / GAP_PciIF+ OeGapAlphaR. The MFAlpha sequence was fused with p*AOX1*/p*GAP* using primers PAOX1_PciIF+ AlphaEcoRIR (expression cassette for #3) or PAOX1_PciIF+ aEAEAEcoRIR (expression cassette for #2), / GAP_PciIF+ AlphaEcoRIR (expression cassette for #6) or GAP_PciIF+ aEAEAEcoRIR (expression cassette for #5). The pAOX1-MFAlpha PCR products were cloned into pPpRFSC via PciI/EcoRI sites to create #3 and #2. The pGap-MFAlpha fusion construct was cloned into pPpRSFC-pGAP via PciI/EcoRI restriction site to construct pPpRSFC-#6 and #5.

p*GAP* plasmids do not contain the 3' *AOX1* sequence for homologous integration in the *AOX1* locus. Plasmids #1 to #6 were made initially and completely sequenced. In the creation of the following plasmids, only newly inserted parts (and the RE sites used for cloning) were confirmed by sequencing. A full description of how the 28 plasmids (#7-#40) with the N- and C- terminal tags were created would be beyond the scope of this section and is provided in the Supplementary file S 5 (spreadsheets on plasmid construction). For further details consult the annotated plasmid sequences provided in Supplementary file S 1.

The HRP gene (isoenzyme A2A [46,47]) was amplified using primers HRP-A2-RSFC-fwd and HRP-A2-RSFC-rev and cloned in the respective vectors (detailed protocol see below).

RSFC cloning of inserts and colony PCRs

For blunt end cloning, the vector backbone was dephosphorylated using either Thermo Scientific shrimp alkaline phosphatase or FastAP according to the manufacturer's recommendations. The backbone was gel purified and used for ligations with phosphorylated PCR products. Prior, PCR primers were phosphorylated using Thermo Scientific T4 Polynucleotide Kinase according to the manufacturer's recommendations. Subsequently the reaction mixtures containing the

phosphorylated primers were desalted on nitrocellulose filters (Merck Millipore, Darmstadt, Germany) and added to the PCR (Phusion polymerase). Ligations were performed using the blunt end protocol provided for Thermo Scientific T4 DNA Ligase.

For TA cloning, phosphorylated Phusion PCR products were purified (Promega Wizard SV Gel and PCR Clean-Up System) and dA-tailed using Taq-Polymerase (GoTaq Flexi, Promega [Fitchburg, WI, USA], standard buffer, 0.2 mmol/l dATP, 30 min incubation at 72° C) and directly used for ligation (blunt end protocol provided for Thermo Scientific T4 DNA Ligase).

To verify the correct orientation by colony PCR, primers were selected as outlined in Supplementary file S 3. A tiny amount of an *E. coli* colony from a transformation of the respective ligation was added to a GoTaq Flexi reaction. The manufacturer's protocol was followed except reducing the reaction volume to 20 µl and increasing the initial denaturing step to five min to break the cells.

P. pastoris transformations and screening

For testing the pPpRSFC plasmids, the *P. pastoris* CBS7435 wildtype strain was used. Plasmids bearing the *AOX1* promoter were linearized with *Bgl*II, plasmids with p*GAP* were linearized with *Swa*I. All linearized plasmids were transformed into competent *P. pastoris* cells prepared by the condensed protocol reported by Lin-Cereghino *et al*. [73]. Only low amounts of DNA (0.5 µg) were used for the transformations to avoid multicopy integration. A landscape of 80 clones was screened and checked for the desired Mut^S phenotype on minimal methanol plates. Ten Mut^S clones were rescreened for uniform expression; a single representative clone was used for the subsequent characterizations. Screenings, rescreening and characterizations were performed in deep well plates as described previously [74]. BMD 2% was used instead of BMD 1% (giving higher yields, data not shown) and the methanol induction was performed in 12 h intervals for 72 h.

HRP activity assay, eGFP fluorescence microscopy and measurements

HRP activity assays with 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium salt (ABTS) were performed as described previously [42]. For intracellular HRP activity measurements, cells were broken using Yeast Protein Extraction Reagent (Y-PER from Thermo Scientific).

The cell suspensions of eGFP expressing strains were centrifuged and washed in an equal amount of water before fluorescence microscopy (Leica Microsystems, Germany, DM LB2, DFC350FX) at 1000 fold magnification, fluorescence images were taken using filter set 'I3' [excitation filter BP 450- 490]. eGFP fluorescence (ex/em 488/507 nm) and OD_{600} were measured in a Synergy MX plate reader (Biotek, Winooski, VT, USA) using micro titer plates (Nunc MicroWell 96-Well Optical-Bottom Plates with Polymer Base, Black; Thermo Fisher Scientific). Cell suspensions were diluted to be within the linear range. The background measurements of diluted medium were subtracted. Subsequently the relative fluorescence units were normalized per OD_{600}.

Acknowledgments

We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) and the Higher Education Commission of Pakistan. We would like to thank Harald Pichler and Christine M. Winkler for fruitful discussions regarding eGFP tagging experiments and Sarah Trunk for excellent technical assistance. We would also like to thank Rafael E. Luna for critical reading of the manuscript and helpful comments in frame of a scientific story telling workshop held by the DK 'Molecular Enzymology' Graz.

References

1. Malhotra A (2009) Tagging for protein expression. Methods Enzymol 463: 239–258. doi:10.1016/S0076-6879(09)63016-0.
2. Su WW (2005) Fluorescent proteins as tools to aid protein production. Microb Cell Fact 4: 12. doi:10.1186/1475-2859-4-12.
3. Arnau J, Lauritzen C, Petersen GE, Pedersen J (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 48: 1–13. doi:10.1016/j.pep.2005.12.002.
4. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41: 98–105. doi:10.1016/j.pep.2005.01.019.
5. Niiranen L, Espelid S, Karlsen CR, Mustonen M, Paulsen SM, et al. (2007) Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli. Protein Expr Purif 52: 210–218. doi:10.1016/j.pep.2006.09.005.
6. Gordon E, Horsefield R, Swarts HGP, de Pont JJHHM, Neutze R, et al. (2008) Effective high-throughput overproduction of membrane proteins in Escherichia coli. Protein Expr Purif 62: 1–8. doi:10.1016/j.pep.2008.07.005.
7. Mohanty AK, Wiener MC (2004) Membrane protein expression and production: effects of polyhistidine tag length and position. Protein Expr Purif 33: 311–325. doi:10.1016/j.pep.2003.10.010.
8. Sasagawa T, Matsui M, Kobayashi Y, Otagiri M, Moriya S, et al. (2011) High-throughput recombinant gene expression systems in Pichia pastoris using newly developed plasmid vectors. Plasmid 65: 65–69. doi:10.1016/j.plasmid.2010.08.004.
9. Lu Q (2005) Seamless cloning and gene fusion. Trends Biotechnol 23: 199–207. doi:10.1016/j.tibtech.2005.02.008.
10. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4: 251–256. doi:10.1038/nmeth1010.
11. Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 40: e55. doi:10.1093/nar/gkr1288.
12. Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6: 343–345. doi:10.1038/nmeth.1318.
13. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4: e6441. doi:10.1371/journal.pone.0006441.
14. Online manual to 'j5 automated DNA assembly'. Available: http://j5.jbei.org/j5manual/pages/22.html Accessed: 27 November 2014
15. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1: 14–21. doi:10.1021/sb2000116.
16. Bath AJ, Milsom SE, Gormley NA, Halford SE (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. J Biol Chem 277: 4024–4033. doi:10.1074/jbc.M108441200.
17. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38: D234–D236. doi:10.1093/nar/gkp874.
18. Padgett KA, Sorge JA (1996) Creating seamless junctions independent of restriction sites in PCR cloning. Gene 168: 31–35.
19. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3: e3647. doi:10.1371/journal.pone.0003647.
20. Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4: e5553. doi:10.1371/journal.pone.0005553.
21. Mead DA, Pey NK, Herrnstadt C, Marcil RA, Smith LM (1991) A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology (N Y) 9: 657–663.
22. Lan D, Wang W, Sun L, Wang Y, Yang B (2011) A novel vector for expression cloning of large numbers of polymerase chain reaction products in Pichia pastoris. 5: 211–216. doi:10.5897/AJMR10.407.
23. Ido E, Hayami M (1997) Construction of T-tailed vectors derived from a pUC plasmid: a rapid system for direct cloning of unmodified PCR products. Biosci Biotechnol Biochem 61: 1766–1767.
24. Reisinger C, Kern A, Fesko K, Schwab H (2007) An efficient plasmid vector for expression cloning of large numbers of PCR fragments in Escherichia coli. Appl Microbiol Biotechnol 77: 241–244. Available: http://www.ncbi.nlm.nih.gov/pubmed/17786428. Accessed 22 August 2011.
25. Rao B, Zhong X, Wang Y, Wu Q, Jiang Z, et al. (2010) Efficient vectors for expression cloning of large numbers of PCR fragments in P . pastoris. Yeast: 285–292. doi:10.1002/yea.

26. Wu Q, Zhong X, Zhai C, Yang J, Chen X, et al. (2010) A series of novel directional cloning and expression vectors for blunt-end ligation of PCR products. Biotechnol Lett 32: 439–443. doi:10.1007/s10529-009-0166-3.
27. Cregg JM, Tolstorukov I, Kusari A, Sunga J, Madden K, et al. (2009) Expression in the yeast Pichia pastoris. Methods Enzymol 463: 169–189. doi:10.1016/S0076-6879(09)63013-5.
28. Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, et al. (2013) Pichia pastoris: protein production host and model organism for biomedical research. Futur Microbiol 8: 191–208. doi:10.2217/fmb.12.133.
29. Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 24: 1094–1101. doi:10.1016/j.copbio.2013.02.024.
30. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, et al. (2012) Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7: e39720. doi:10.1371/journal.pone.0039720.
31. Sears IB, O'Connor J, Rossanese OW, Glick BS (1998) A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. Yeast 14: 783–790. doi:10.1002/(SICI)1097-0061(19980615)14:8<783::AID-YEA272>3.0.CO;2-Y.
32. Goulas T, Cuppari A, Garcia-Castellanos R, Snipas S, Glockshuber R, et al. (2014) The pCri System: A Vector Collection for Recombinant Protein Expression and Purification. PLoS One 9: e112643. doi:10.1371/journal.pone.0112643.
33. Li Z, Leung W, Yon A, Nguyen J, Perez VC, et al. (2010) Secretion and proteolysis of heterologous proteins fused to the Escherichia coli maltose binding protein in Pichia pastoris. Protein Expr Purif 72: 113–124. doi:10.1016/j.pep.2010.03.004.
34. Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984) Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. Cell 37: 1075–1089.
35. Briand L, Perez V, Huet JC, Danty E, Masson C, et al. (1999) Optimization of the production of a honeybee odorant-binding protein by Pichia pastoris. Protein Expr Purif 15: 362–369.
36. Treerattrakool S, Eurwilaichitr L, Udomkit A, Panyim S (2002) Secretion of Pem-CMG, a peptide in the CHH/MIH/GIH family of Penaeus monodon, in Pichia pastoris is directed by secretion signal of the alpha-mating factor from Saccharomyces cerevisiae. J Biochem Mol Biol 35: 476–481.
37. Vogl T, Glieder A (2013) Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 30: 385–404. doi:10.1016/j.nbt.2012.11.010.
38. Daly R, Hearn MTW (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. J Mol Recognit 18: 119–138. doi:10.1002/jmr.687.
39. Liu Z, Pscheidt B, Avi M, Gaisberger R, Hartner FS, et al. (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. Chembiochem 9: 58–61. doi:10.1002/cbic.200700514.
40. Pscheidt B, Liu Z, Gaisberger R, Avi M, Skranc W, et al. (2008) Efficient Biocatalytic Synthesis of (R)-Pantolactone. Adv Synth Catal 350: 1943–1948. doi:10.1002/adsc.200800354.
41. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the Pichia pastoris expression system. Yeast 22: 249–270. doi:10.1002/yea.1208.
42. Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, et al. (2012) Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 11: 22. doi:10.1186/1475-2859-11-22.
43. Crook NC, Freeman ES, Alper HS (2011) Re-engineering multicloning sites for function and convenience. Nucleic Acids Res 39: e92. doi:10.1093/nar/gkr346.
44. Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, et al. (2012) Analysis of the 5' untranslated region (5'UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in Pichia pastoris. Gene 496: 118–127. doi:10.1016/j.gene.2012.01.006.
45. Morawski B, Lin Z, Cirino P, Joo H, Bandara G, et al. (2000) Functional expression of horseradish peroxidase in Saccharomyces cerevisiae and Pichia pastoris. Protein Eng 13: 377–384.
46. Krainer FW, Pletzenauer R, Rossetti L, Herwig C, Glieder A, et al. (2013) Purification and basic biochemical characterization of 19 recombinant plant peroxidase isoenzymes produced in Pichia pastoris. Protein Expr Purif. doi:10.1016/j.pep.2013.12.003.
47. Näätsaari L, Krainer FW, Schubert M, Glieder A, Thallinger GG (2014) Peroxidase gene discovery from the horseradish transcriptome. BMC Genomics 15: 227. doi:10.1186/1471-2164-15-227.
48. Veitch NC (2004) Horseradish peroxidase : a modern view of a classic enzyme. Phytochemistry 65: 249–259. doi:10.1016/j.phytochem.2003.10.022.
49. Lin-Cereghino GP, Stark CM, Kim D, Chang JWJ, Shaheen N, et al. (2013) The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. Gene 519: 311–317. doi:10.1016/j.gene.2013.01.062.
50. Kjeldsen T (2000) Yeast secretory expression of insulin precursors. Appl Microbiol Biotechnol 54: 277–286.

51. Thim L, Hansen MT, Norris K, Hoegh I, Boel E, et al. (1986) Secretion and processing of insulin precursors in yeast. Proc Natl Acad Sci U S A 83: 6766–6770.
52. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, et al. (2008) Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 36: e76. doi:10.1093/nar/gkn369.
53. Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, et al. (2009) An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. FEMS Yeast Res 9: 1271–1282. doi:10.1111/j.1567-1364.2009.00571.x.
54. Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A (2014) Synthetic Core Promoters for Pichia pastoris. ACS Synth Biol 3: 188–191. doi:10.1021/sb400091p.
55. Al-Samarrai TH, Kirk CA, Jones WT, Harvey D, Templeton MD (2010) A simple and rapid method for selecting high producers of recombinant proteins in individual clones of P. pastoris. Comb Chem High Throughput Screen 13: 377–382.
56. Brooks CL, Morrison M, Lemieux JM (2013) Rapid expression screening of eukaryotic membrane proteins in Pichia pastoris. Protein Sci 22: 425–433. doi:10.1002/pro.2223.
57. Mizutani K, Yoshioka S, Mizutani Y, Iwata S, Mikami B (2011) High-throughput construction of expression system using yeast Pichia pastoris, and its application to membrane proteins. Protein Expr Purif 77: 1–8. doi:10.1016/j.pep.2010.12.009.
58. Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, et al. (2014) Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression. N Biotechnol 00. doi:10.1016/j.nbt.2014.02.009.
59. Eiden-Plach A, Zagorc T, Heintel T, Carius Y, Breinig F, et al. (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by Candida glabrata, Pichia pastoris, Saccharomyces cerevisiae, and Schizosaccharomyces pombe. Appl Environ Microbiol 70: 961–966. doi:10.1128/AEM.70.2.961.
60. Kottmeier K, Ostermann K, Bley T, Rödel G (2011) Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in Pichia pastoris. Transformation. doi:10.1007/s00253-011-3246-y.
61. Tang J-B, Zhu P, Yang H, Sun L-M, Song S, et al. (2008) Expression and secretion of recombinant ZZ-EGFP fusion protein by the methylotrophic yeast Pichia pastoris. Biotechnol Lett 30: 1409–1414. doi:10.1007/s10529-008-9714-5.
62. Panjideh H, Coelho V, Dernedde J, Fuchs H, Keilholz U, et al. (2008) Production of bifunctional single-chain antibody-based fusion proteins in Pichia pastoris supernatants. Bioprocess Biosyst Eng 31: 559–568. doi:10.1007/s00449-008-0203-y.
63. Cha HJ, Dalal NN, Bentley WE (2005) Secretion of human interleukin-2 fused with green fluorescent protein in recombinant Pichia pastoris. Appl Biochem Biotechnol 126: 1–11.
64. Passolunghi S, Brocca S, Cannizzaro L, Porro D, Lotti M (2003) Monitoring the transport of recombinant Candida rugosa lipase by a green fluorescent protein-lipase fusion. Biotechnol Lett 25: 1945–1948.
65. Sarramegna V, Talmont F, Seree de Roch M, Milon A, Demange P (2002) Green fluorescent protein as a reporter of human mu-opioid receptor overexpression and localization in the methylotrophic yeast Pichia pastoris. J Biotechnol 99: 23–39.
66. Surribas A, Resina D, Ferrer P, Valero F (2007) Rivoflavin may interfere with on-line monitoring of secreted green fluorescence protein fusion proteins in Pichia pastoris. Microb Cell Fact 6: 15. doi:10.1186/1475-2859-6-15.
67. Dortay H, Akula UM, Westphal C, Sittig M, Mueller-Roeber B (2011) High-throughput protein expression using a combination of ligation-independent cloning (LIC) and infrared fluorescent protein (IFP) detection. PLoS One 6: e18900. doi:10.1371/journal.pone.0018900.
68. De Jong RN, Daniëls MA, Kaptein R, Folkers GE (2006) Enzyme free cloning for high throughput gene cloning and expression. J Struct Funct Genomics 7: 109–118. doi:10.1007/s10969-006-9014-z.
69. Lippow SM, Aha PM, Parker MH, Blake WJ, Baynes BM, et al. (2009) Creation of a type IIS restriction endonuclease with a long recognition sequence. Nucleic Acids Res 37: 3061–3073. doi:10.1093/nar/gkp182.
70. Besnier CE, Kong H (2001) Converting MlyI endonuclease into a nicking enzyme by changing its oligomerization state. EMBO Rep 2: 782–786. doi:10.1093/embo-reports/kve175.
71. Scalley-Kim M, McConnell-Smith A, Stoddard BL (2007) Coevolution of a homing endonuclease and its host target sequence. J Mol Biol 372: 1305–1319. doi:10.1016/j.jmb.2007.07.052.
72. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128: 2477–2484. doi:10.1021/ja057519l.
73. Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, et al. (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38: 44, 46, 48.
74. Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, et al. (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5: 179–189. doi:10.1016/j.femsyr.2004.06.016.

Figures

Fig. 1

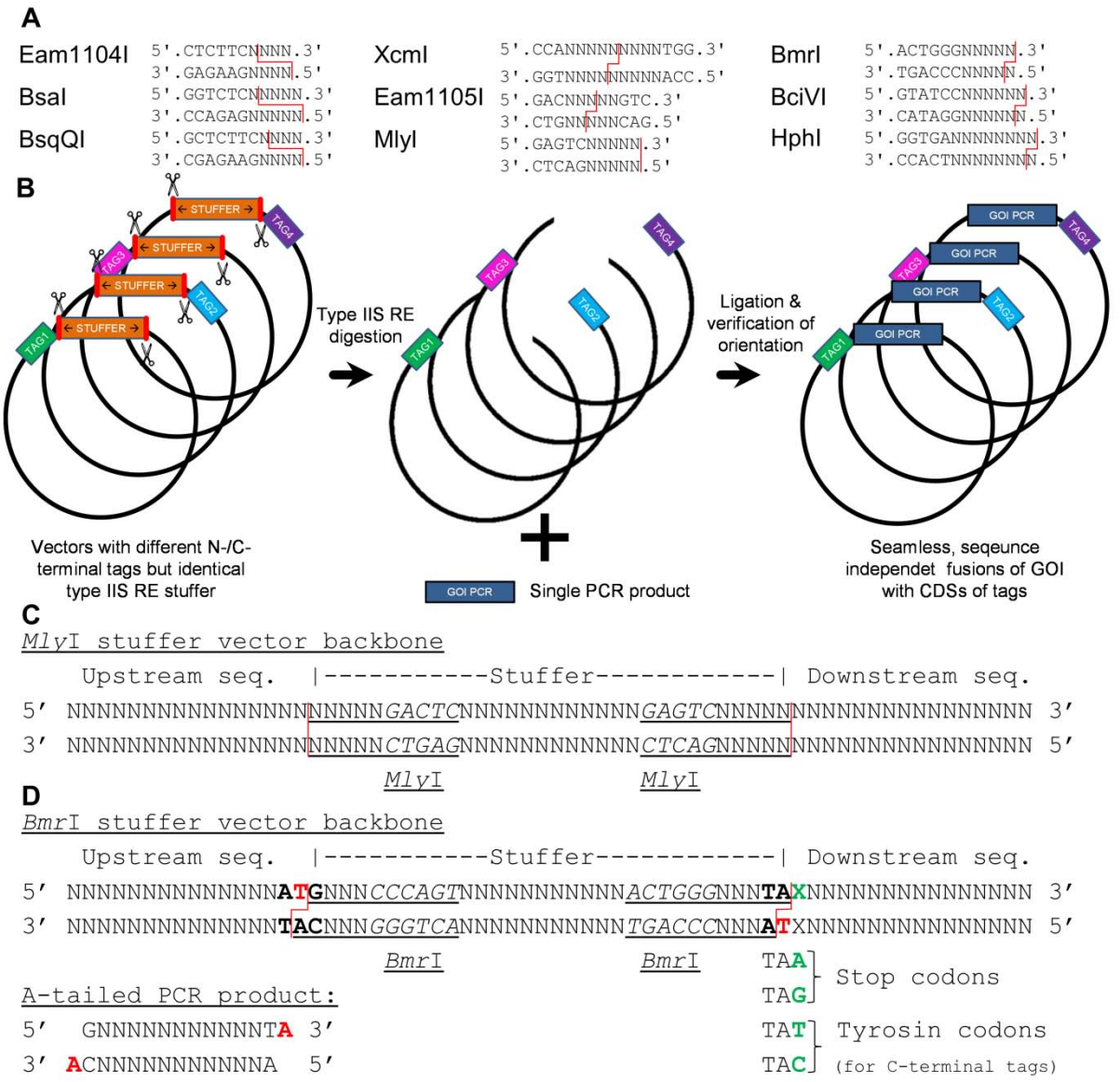


Fig. 1: Detailed outline of the restriction site free cloning (RSFC) strategy.

(A) Recognition sites of various type IIS REs. The cleavage patterns are indicated as red lines.

(B) Schematic workflow of restriction site free cloning. After removal of a stuffer fragment using type IIS REs, a single PCR product can be ligated into all vectors in a seamless, sequence independent fashion. The strategy is shown for four vectors but can be extended to as many as desired.

(C) Design of the *Mly*I stuffer fragment for blunt end ligations. The *Mly*I recognition sequence is written in italics, the entire cleavage pattern is underlined. Variable bp are denoted as 'N'. Upstream sequences may include promoters, N-terminal tags and signal sequences, downstream sequences may include C-terminal tags and stop codons.

(D) Design of the *Bmr*I stuffer fragment for TA cloning. Same explanation as (C), in addition the incorporation of the dA and dT residues for TA cloning via Start- and Stop/Tyr-codons are shown (red). By varying the last nucleotide 'X' of the Stop/Tyr codon, either translation can be terminated or a C-terminal tag linked in frame. A dA-tailed PCR product suitable for ligation is also shown.

Fig. 2

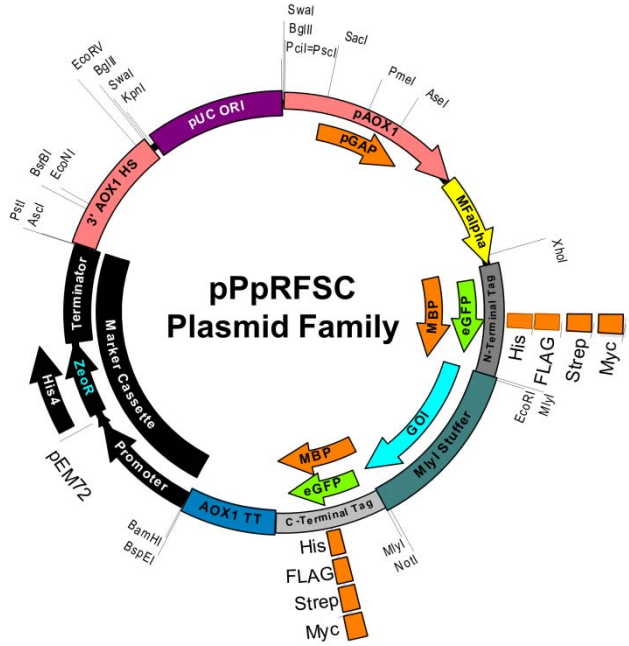


Fig. 2: Representative map of *P. pastoris* RSFC plasmids.

The features of all RSFC plasmids designed are summarized in this schematic map. Different promoters, N/C-terminal tags, resistance markers are shown. p*GAP* plasmids do not contain the 3' *AOX1* homologous sequence for recombination (3' AOX1 HS). HIS4 vectors contain in addition an ampicillin resistance cassette. The mating factor alpha signal sequence (MF alpha) is optional and only present in secretory plasmids. See Tab. 1 for the part combinations created in this study. Features are not drawn to scale. Exact plasmid maps are provided in the Supplementary file S 1 in GenBank format.

Fig. 3

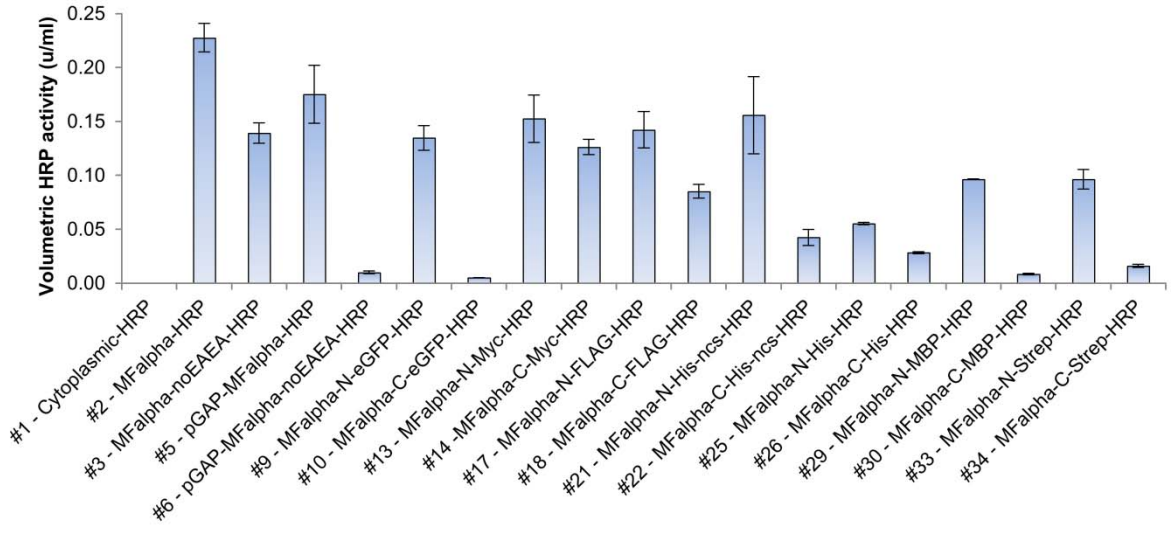


Fig. 3: Type of tag and position (N/C-terminal) strongly affect volumetric HRP activities.

Volumetric HRP activities in the supernatant with ABTS as substrate after cultivation on methanol for 72 h are shown. The pPpRSFC plasmids used are indicated via the numbers given in Tab. 1, the use of a signal sequence (MFalpha), different promoter than p*AOX1* and if applicable tag and position (N/C) are given. Mean values and standard deviations of biological triplicates are shown.

Fig. 4

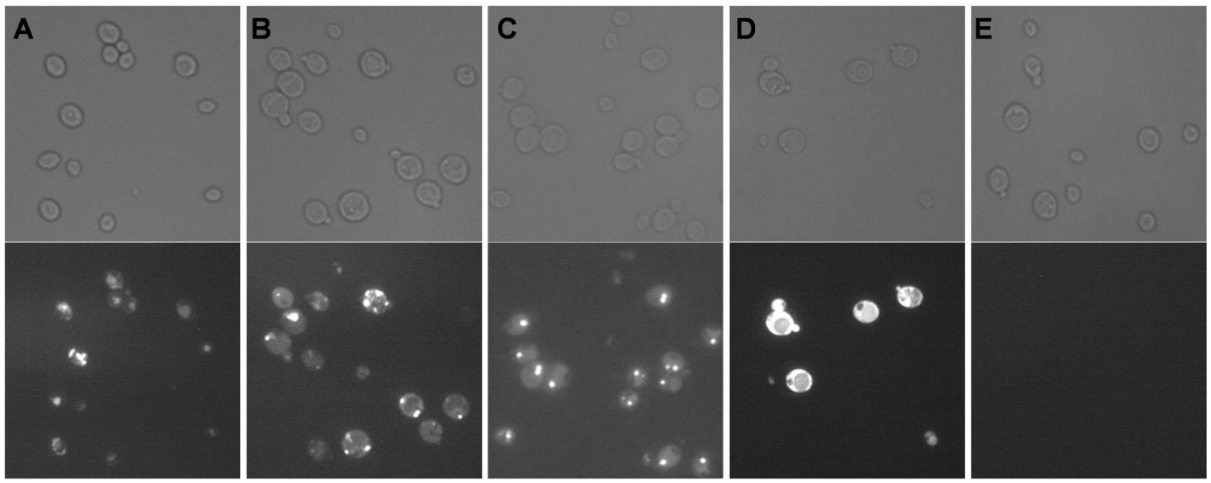


Fig. 4: Fluorescence microscopy of fusions of HRP to GFP.
Bright field images are shown on top, fluorescence images below.
(A) HRP N-terminally tagged with eGFP (#9-MFalpha-N-eGFP-HRP),
(B) HRP C-terminally tagged with eGFP (#10-MFalpha-C-eGFP-HRP),
(C) control of eGFP with MFalpha (self-ligated #9),
(D) control of cytoplasmic eGFP expression (#1-eGFP),
(E) negative control of empty Mut^S strain.
For the bright field image of panel (C) brightness was decreased -11 %, contrast was increased +44 % for better comparability with the other panels.

Table 1

Tab. 1: RSFC vector family designed for *P. pastoris*.

#	Name	Tag/Fusion protein, position[a] & length[b]			TEV protease cleavage site	Mode of expression	EAEA repeat	Selection marker[c]	Plasmid size (bp)
1	pPpRSFC	-	-	-	NA[d]	intracellular	NA	Zeocin	4840
2	pPpRSFC-MFalpha	-	-	-	NA	secretory	yes	Zeocin	5104
3	pPpRSFC-MFalpha-noEAEA	-	-	-	NA	secretory	no	Zeocin	5092
4	pPpRSFC-pGAP	-	-	-	NA	intracellular	NA	Zeocin	3771
5	pPpRSFC-pGAP-MFalpha	-	-	-	NA	secretory	yes	Zeocin	4035
6	pPpRSFC-pGAP-MFalpha-noEAEA	-	-	-	NA	secretory	no	Zeocin	4023
7	pPpRSFC-N-eGFP	eGFP	N	240	no	intracellular	NA	Zeocin	5584
8	pPpRSFC-C-eGFP	eGFP	C	239	no	intracellular	NA	Zeocin	5584
9	pPpRSFC-MFalpha-N-eGFP	eGFP	N	239	no	secretory	yes	Zeocin	5848
10	pPpRSFC-MFalpha-C-eGFP	eGFP	C	239	no	secretory	yes	Zeocin	5848
11	pPpRSFC-N-Myc	MYC	N	11	no	intracellular	NA	Zeocin	4870
12	pPpRSFC-C-Myc	MYC	C	10	no	intracellular	NA	Zeocin	4870
13	pPpRSFC-MFalpha-N-Myc	MYC	N	10	no	secretory	yes	Zeocin	5134
14	pPpRSFC-MFalpha-C-Myc	MYC	C	10	no	secretory	yes	Zeocin	5134
15	pPpRSFC-N-FLAG	FLAG	N	9	no	intracellular	NA	Zeocin	4864
16	pPpRSFC-C-FLAG	FLAG	C	8	no	intracellular	NA	Zeocin	4864
17	pPpRSFC-MFalpha-N-FLAG	FLAG	N	8	no	secretory	yes	Zeocin	5128
18	pPpRSFC-MFalpha-C-FLAG	FLAG	C	8	no	secretory	yes	Zeocin	5128
19	pPpRSFC-N-His-ncs[e]	His	N	7	no	intracellular	NA	Zeocin	4858
20	pPpRSFC-C-His-ncs	His	C	6	no	intracellular	NA	Zeocin	4858
21	pPpRSFC-MFalpha-N-His-ncs	His	N	6	no	secretory	yes	Zeocin	5122
22	pPpRSFC-MFalpha-C-His-ncs	His	C	6	no	secretory	yes	Zeocin	5122
23	pPpRSFC-N-His	His	N	7	yes	intracellular	NA	Zeocin	4879
24	pPpRSFC-C-His	His	C	6	yes	intracellular	NA	Zeocin	4879
25	pPpRSFC-MFalpha-N-His	His	N	6	yes	secretory	yes	Zeocin	5143
26	pPpRSFC-MFalpha-C-His	His	C	6	yes	secretory	yes	Zeocin	5143
27	pPpRSFC-N-MBP	MBP	N	367	yes	intracellular	NA	Zeocin	5959
28	pPpRSFC-C-MBP	MBP	C	366	yes	intracellular	NA	Zeocin	5959
29	pPpRSFC-MFalpha-N-MBP	MBP	N	366	yes	secretory	yes	Zeocin	6223
30	pPpRSFC-MFalpha-C-MBP	MBP	C	366	yes	secretory	yes	Zeocin	6223
31	pPpRSFC-N-Strep	Strep	N	9	yes	intracellular	NA	Zeocin	4885
32	pPpRSFC-C-Strep	Strep	C	8	yes	intracellular	NA	Zeocin	4885
33	pPpRSFC-MFalpha-N-Strep	Strep	N	8	yes	secretory	yes	Zeocin	5149
34	pPpRSFC-MFalpha-C-Strep	Strep	C	8	yes	secretory	yes	Zeocin	5149
35	pPpRSFC-HIS	-	-	-	NA	intracellular	NA	HIS4	7683
36	pPpRSFC-HIS-MFalpha	-	-	-	NA	secretory	yes	HIS4	7947
37	pPpRSFC-HIS-MFalpha-noEAEA	-	-	-	NA	secretory	no	HIS4	7935
38	pPpRSFC-HIS-pGAP	-	-	-	NA	intracellular	NA	HIS4	6614
39	pPpRSFC-HIS-pGAP-MFalpha	-	-	-	NA	secretory	yes	HIS4	6878
40	pPpRSFC-HIS-pGAP-MFalpha-noEAEA	-	-	-	NA	secretory	no	HIS4	6866

[a] = N- or C- terminal fusion to the POI

[b] = Length in amino acids (intracellular N- terminal tags are because of the start codon one aa longer, the TEV protease cleavage site (seven aa) is not included in this number)

[c] = Zeocin selection is applicable for *E. coli* and *P. pastoris*, HIS4 plasmids are based on ampicillin selection in *E. coli* and used in combination with a histidine auxotrophic (*his4*) *P. pastoris* strain

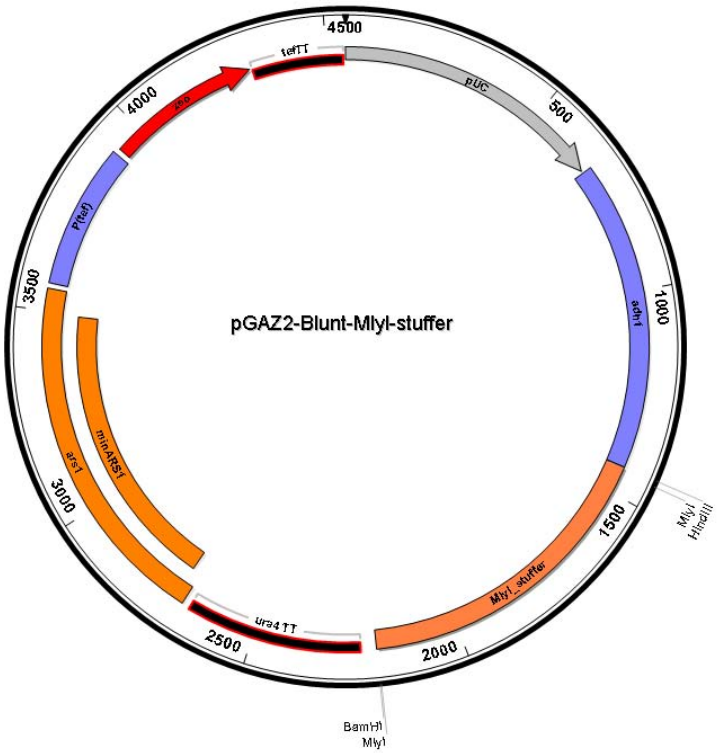
[d] = NA: not applicable

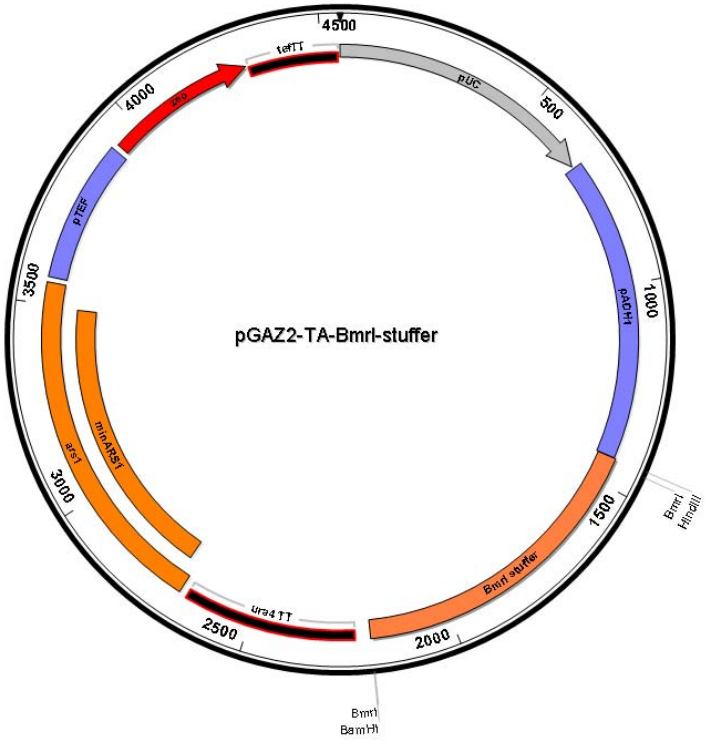
[e] = ncs: no TEV protease cleavage site

Supporting information

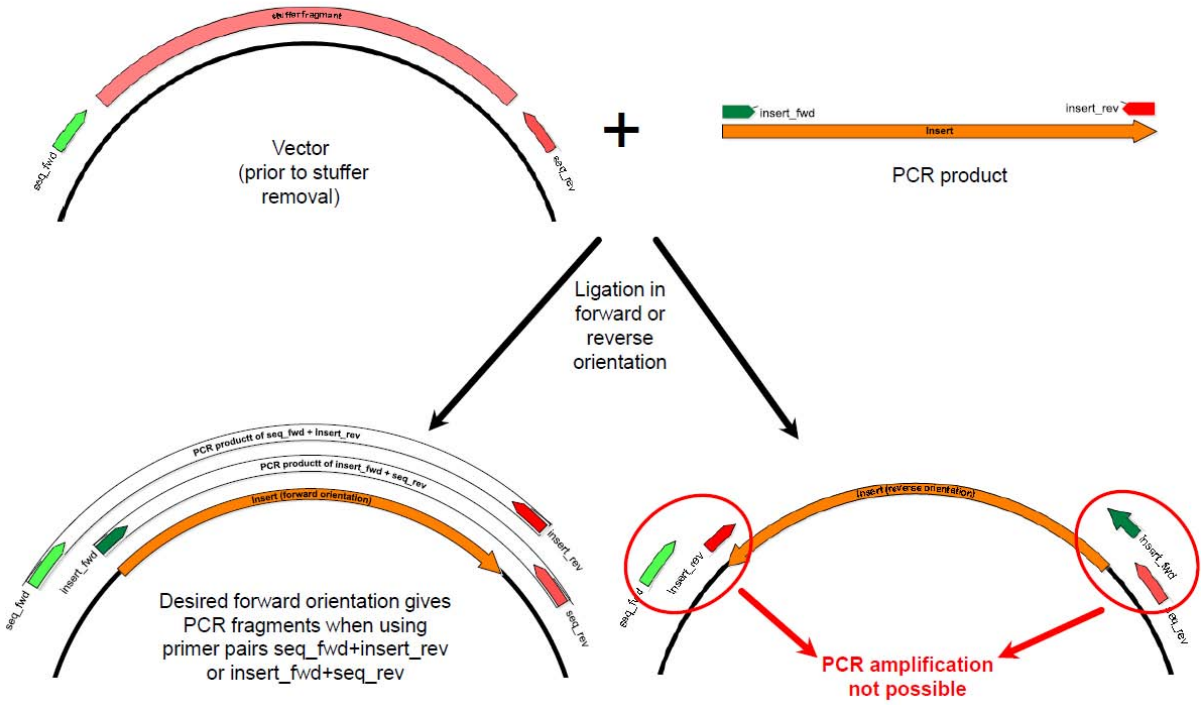
<u>S 1:</u> Plasmid sequences of the constructs used in this study in Genbank format.

<u>S 2:</u> Vector maps of the *S. pombe* vectors used in this study.

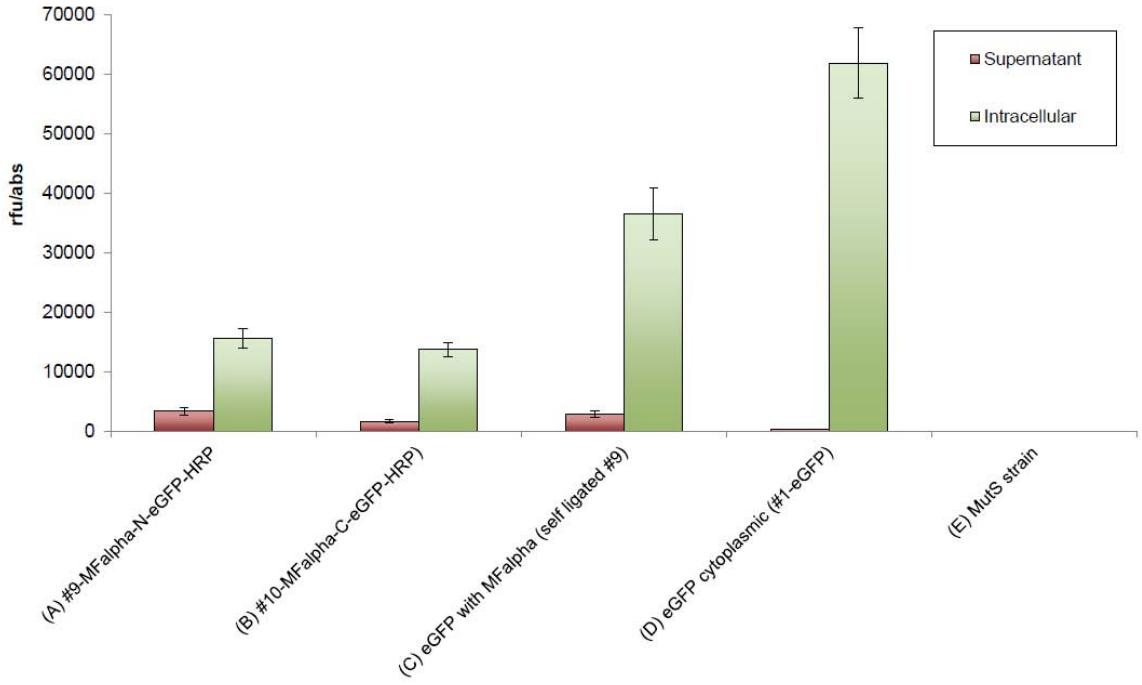




S 3: Simple strategy for confirming the orientation of the insert. The forward or reverse primer used for amplifying the insert can be used together with the forward or reverse sequencing primer of the vector to confirm the correct orientation. Upon correct primer choice only the forward orientation gives a PCR fragment. The sequencing primers designed for Sanger sequencing allow sequencing of the insert from both sides. Depending on the vector, different primers should be used (e.g. when the MFalpha signal sequence or a fusion protein is present, see the primer list for all sequencing primers available).



S 4: Fluorescence measurements of fusions of HRP to eGFP.
Samples are labeled in the same way as in Fig. 4. eGFP fluorescence of supernatants and cell pellets of methanol induced cells were normalized per cell density (OD_{600}).



S 5: Primers used in this study and detailed plasmid construction. Primers used for creating the *S. pombe* and *P. pastoris* plasmids are separated. Also primers for construction of the plasmids are separated from primers for sequencing and insertion of GOIs. In addition separate spreadsheets are providing information on the exact construction of the plasmids by listing the PCR products and restriction enzymes used for assembly (.xlsx file).

4. Summary and conclusion

In this thesis, synthetic biology approaches were applied to engineer the methylotrophic yeast *Pichia pastoris*. Starting from the characterization of natural promoters (Chapters 3.1 to 3.3) unprecedented synthetic bidirectional promoters (3.4.2) and basic transcriptional activating circuits (3.5.2) were generated. In addition also molecular tools for gene expression such as synthetic core promoters (3.4.1) and expression vectors simplifying optimization of gene expression (3.6) were characterized.

Several dozen genes have been annotated to be involved in methanol utilization in *P. pastoris* [1–3] and studies on the transcriptomes of *P. pastoris* [4,5] and other methylotrophic yeasts [6,7] have indicated that several of these genes follow a similar regulation as P_{AOX1}. Those are promising targets to identify additional strong, methanol regulated promoters (Chapter 3.1). Previously, only five methanol regulated promoters had been characterized in *P. pastoris* (P_{AOX1}, P_{DAS2}, P_{FLD1}, P_{AOX2}, P_{PEX8} [8]). For this thesis 'canonical' reactions (i.e. direct metabolization of methanol) and non-canonical reactions (of associated processes such as the defense against reactive oxygen species, ROS and the pentose phosphate pathway, PPP) were considered (Fig. 1). The differential expression of genes coding for the respective enzymes in presence of different carbon sources was analyzed by microarray experiments and 45 promoters were cloned upstream of the gene coding for an enhanced green fluorescent reporter protein (eGFP). Thereby 15 methanol regulated promoters were characterized. About half of these promoters showed very high expression levels in the range of the state of the art *AOX1* benchmark promoter. This is to the best of my knowledge the largest set of tightly co-regulated promoters available in any commonly used microbial expression host. In *S. cerevisiae* only a few copper (P_{CTR1}, P_{CTR3}, P_{CUP1}) and galactose (P_{GAL1}, P_{GAL3}, P_{GAL7}, P_{GAL10}) regulated promoters are available [9,10]. A set of promoters with different sequences but tight regulation, activated by the same stimulus, is a valuable tool for the coexpression of multiple genes, suggesting *P. pastoris* as a potent platform for metabolic engineering endeavors involving the expression of large heterologous pathways.

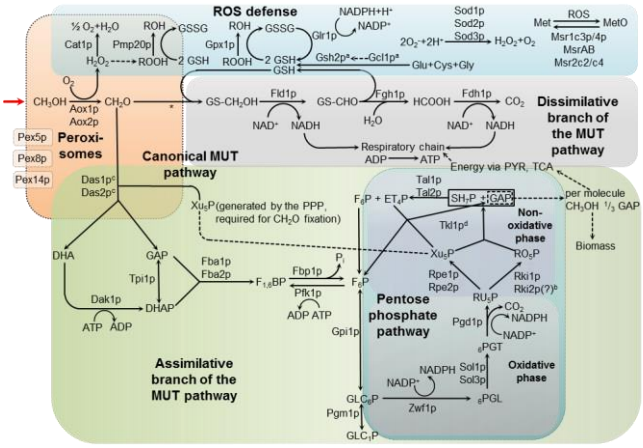


Fig. 1: Canonical and non-canonical parts of the *P. pastoris* methanol utilization pathway. See Chapter 3.1 – Figure 1 for a larger version and a detailed caption.

One of the key findings of the systematic characterization of the *P. pastoris* MUT pathway (Chapter 3.1), was the derepressed regulation of the promoter of the *catalase 1* gene (P_{CAT1}). Derepression means that the promoter is repressed on repressing carbon sources, and activated when the carbon source in the media is depleted. For *P. pastoris* P_{AOX1} this derepression effect is very weak (2-4% of methanol induced levels) requiring methanol for full induction. In related methylotrophic yeasts, derepressed promoters reach however up to 70 % of methanol induced levels [11]. Also derepressed synthetic variants of P_{AOX1} were reported [12,13]. P_{CAT1} is the first naturally methanol regulated and derepressed *P. pastoris* promoter reaching under glucose limited growth conditions about one third of methanol induced levels (Chapter 3.1). On methanol, P_{AOX1} and P_{CAT1} reach about similar expression levels. In addition P_{CAT1} can be induced with oleic acid to similar levels as on methanol,

while P_{AOX1} remains tightly repressed on oleic acid. Therefore P_{CAT1} renders two induction strategies possible: The derepressed regulation can be harnessed by applying limited carbon source feeds in bioreactor cultivations. Thereby the cells are maintained in a derepressed state while still providing new energy for protein production. Alternatively, oleic acid can be used to replace methanol resulting in biphasic production processes similar to already established methanol dependent processes applied for P_{AOX1}. The derepressed regulation of P_{CAT1} was also essential for setting up activating circuits for P_{AOX1} (discussed in Chapter 3.5.2).

However, in the first work with P_{CAT1} we suffered massively from background colonies and strain instabilities. Especially transformations of P_{CAT1} bearing plasmids with Geneticin resistance markers were initially showing stable, expected phenotypes in only approximately 1 % of the transformants. It took almost two years to elucidate that these issues are caused by the presence of an autonomous replicating sequence (ARS) in the 5'end of P_{CAT1} (Chapter 3.2). The initial promoter length of 692 bp contained this ARS sequence ($P_{CAT1\text{-}692}$). Shortening the promoter to 500 bp ($P_{CAT1\text{-}500}$) (Fig. 2) removed the ARS sequence, abolishing background growth completely, allowing to use P_{CAT1} also with Geneticin resistance markers (required for co-transformation of existing Zeocin selection based expression strains). Surprisingly a second effect of the ARS was noticed: Under selective pressure the combination of P_{CAT1} and its endogenous ARS in an episomal expression strategy enabled up to seven fold higher expression levels (alongside ~100 fold higher transformation rates) than genomic integration. This effect was not only noticed for the reporter gene eGFP, but also for two industrially relevant hydroxynitrile lyases from *Manihot esculenta* and *Linum usitatissimum*, showing also higher landscape uniformity of the circular ARS plasmids than genomic integration.

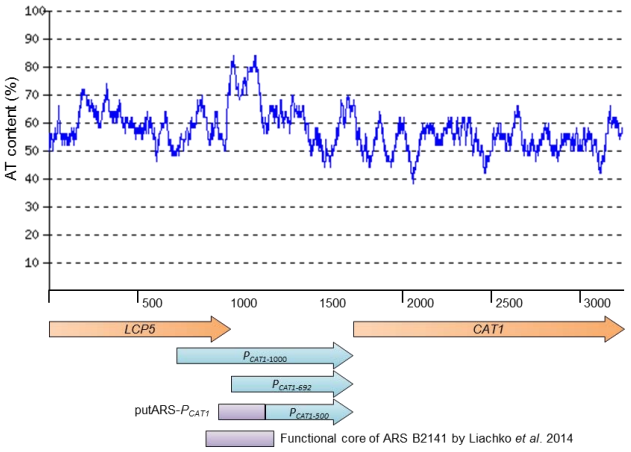


Fig. 2: The upstream region of the CAT1 gene contains an AT-rich ARS. See Chapter 3.2– Figure 1 for details.

Derepressed expression employing P_{CAT1} was a useful feature, allowing shortened, methanol-free screenings. In related methylotrophic yeasts, orthologous MUT promoters were even derepressed to up to 70 % of methanol induced levels [11]. To this end, six orthologous promoters from *H. polymorpha*, *C. boidinii* and *P. methanolica* were tested in *P. pastoris* (Chapter 3.3). Most strikingly the *HpFMD* promoter reached in *P. pastoris* derepressed expression similar to the constitutive *GAP* promoter using an eGFP reporter gene. On methanol P_{HpFMD} even outperformed P_{AOX1} about twofold, reaching the highest reporter gene fluorescence of any single promoter tested in this thesis. These effects could be reproduced for additional proteins. This result is surprising, as orthologous promoter sequences had not maintained their endogenous regulation in new hosts in previous studies (reviewed in [11]).

One of the strongest methanol regulated promoters acts in opposite direction driving the expression of the two *dihydroxyacetone synthase* genes *DAS1* and *DAS2* [14]. A special emphasis of this thesis rested on such bidirectional promoters, BDPs (Chapter 3.4.2). BDPs allow by divergent expression the simple coexpression of two genes of interest (Fig. 3). Dual gene coexpression is for example required when producing dimeric proteins such as antibodies (consisting of a heavy and a light chain), or when co-expressing any gene of interest (GOI) with a folding helper/chaperone. Multi gene expression is required when producing multi subunit proteins or large metabolic pathways. In such expression endeavors the cumulative expression levels, the ratios of the proteins in regard to each other and the optimal expression profile need to be optimized. We envisioned an expression strategy, in which a library of BDPs can be used for optimizing dual and multi gene coexpression. These BDPs should provide different expression levels and regulatory profiles on both sides facilitating optimization of the aforementioned cumulative expression levels, ratios and regulatory profiles. Three and a half years and 168 BDPs later, this goal was achieved. Because of the limitations of natural BDPs (nBDPs), synthetic BDPs (sBDPs) were generated following bidirectionalization, fusion and hybrid BDP design strategies. Surprisingly natural bidirectional histone promoters were found to be exceptionally strong promoters and helpful tools for sBDP design in *P. pastoris*. Histones range amongst the most highly conserved eukaryotic proteins and also their genomic organization flanking a BDP is conserved. Due to their high conservation and universal role, bidirectional histone promoters may provide a valuable general engineering framework for bidirectional gene expression and the generation of sBDPs in any eukaryotic organism. This notion has been confirmed in *S. cerevisiae* and *Schizosaccharomyces pombe* (personal communication Pitzer J.) and Chinese hamster ovary cells (personal communication Weinhandl K.).

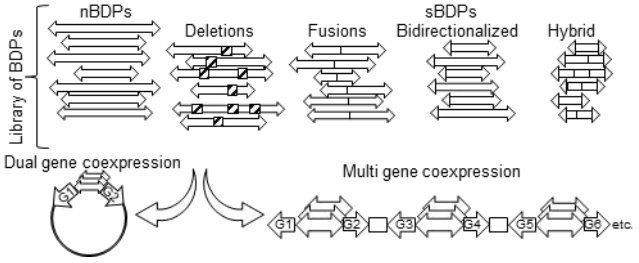


Fig. 3: A library of diversely regulated natural and synthetic bidirectional promoters (nBDPs and sBDPs) covering a wide range of regulatory profiles facilitates optimization of dual gene coexpression and the assembly of multi gene coexpression cassettes. See Chapter 3.4.2 for details.

In addition to these solely *cis*-acting sequence based approaches, also the involvement of *trans*-acting factors in methanol dependent gene expression was tested (Chapter 3.5.2, Master thesis by Lukas Sturmberger [15]) in order to establish new regulatory circuits. Based on previous studies in *P. pastoris* (summarized in Chapters 2.2/[8] and 2.3) and knowledge on homologous factors in *S. cerevisiae*, a set of 11 factors was overexpressed and tested for their effect on P_{AOX1} regulation. Constitutive expression of the factors, as previously reported in patents (discussed in detail in chapter 2.2/[8]), showed little or even negative effects. However, overexpression of three genes (*MXR1*, *TRM1*, *PRM1*) from the derepressed *CAT1* promoter activated P_{AOX1} up to 50 % of methanol induced levels using a fluorescent reporter protein. Apparently, the time point of expression of these activators is critical. Presumably they can only exert a beneficial effect, if the cellular machinery is already geared towards MUT gene expression and repression is partially alleviated. Based on these findings, conversion plasmids were generated that allow transforming existing P_{AOX1} based strains into methanol free, derepressed systems.

The expression of MUT pathway genes (Chapter 3.1) and the effect of transcription factor overexpression (3.5.2) was analyzed by DNA microarrays. In addition, also the work on the

transcriptomes of membrane protein overexpressing strains by Maria Freigassner and the effect of *HAC1* coexpression on MP yields [16] was finally evaluated and published (Chapter 3.5.1, [17]), providing insight and potential solutions to the complex challenge of membrane protein expression. Many human drug targets and also a range of industrially relevant biocatalysts are membrane proteins. For example in the milestone publication of the Keasling group on artemisinic acid production in *S. cerevisiae* [9], a key factor was the optimization of the expression of the membrane associated cytochrome P450 enzymes and their associated reductases.

Looking to the future needs and opportunities also synthetic core promoters (Chapter 3.4.1, [18]) were characterized and found to be promising tools for transcriptional fine-tuning (Fig. 4). This work had been initiated by Claudia Ruth and Julia Pitzer [19] and is now further continued in a model based study and design by Rui Portela. The core promoter work in this thesis relied on a small set of semi-rationally designed promoters, showing the opportunity to regulate gene expression by simple variations in core promoters.

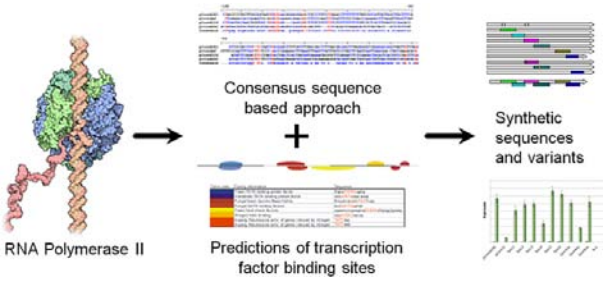


Fig. 4: Synthetic core promoters for *P. pastoris*. Taken from [18] (see Chapter 3.4.1).

All promoters and terminators characterized in this thesis were cloned seamlessly, *i.e.* the transition of the promoter to the start codon (or of the stop codon to the terminator) maintained the natural sequence context without introducing additional restriction endonuclease (RE) recognition sites or linker sequences. This was done either by PCR or overlap directed in vitro recombination based technologies [20,21] and also by special type IIS restriction enzyme based technologies [22] We had initially adapted type IIS REs for the cloning of monodirectional and bidirectional promoters (Chapters 3.1 and 3.4.2). In time, we realized that this cloning strategy allows also seamless fusions of any gene of interest to various N- or C-terminal tags (Chapter 3.6). Protein tagging is a valuable tool for purification, detection and targeting of proteins. However, as an exogenous addition to the protein of interest, they may also negatively affect folding, activity or stability. With the restriction site free cloning (RSFC) strategy developed here in collaboration with Mudassar Ahmad, various tags can be tested in N- or C-terminal position to find the optimal fusion partner for a certain protein of interest.

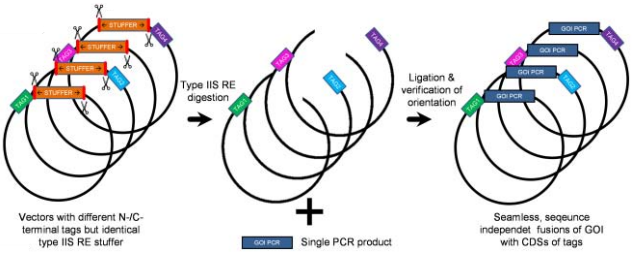


Fig. 5: Schematic workflow of RSFC (restriction site free cloning). See Chapter3.6 for details.

The work of this thesis has provided novel, efficient strategies for single protein production in *P. pastoris*. More strikingly, also the first toolbox suitable for heterologous pathway expression and metabolic engineering endeavors was established. Natural monodirectional promoters of the MUT pathway alongside monodirectional terminators provide a unique set of tightly regulated promoters supporting the controlled expression of more than 10 genes which was studied in more detail by Geier M. and Fauland P. [23]. The library of synthetic bidirectional promoters alongside bidirectional terminators extends the potential even further, easing transcriptional fine-tuning and the assembly

of even larger multi-gene pathways. To the best of our knowledge, this library of BDPs is the largest and most versatile collection of promoters available in any microbial expression host, including the classic model organism and metabolic engineering workhorse *S. cerevisiae*.

On the one hand, the parts reported here make *P. pastoris* an outstanding platform for metabolic engineering applications, paving the way for the tightly regulated expression of large heterologous pathways. On the other hand, the underlying strategies applied including histone promoters as engineering framework for eukaryotic BDP design and the benefits of derepressed promoters, may fuel similar approaches in other expression systems.

References

[1] Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb Cell Fact 2009;8:29.

[2] De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6.

[3] Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 2011;154:312–20.

[4] Liang S, Wang B, Pan L, Ye Y, He M, Han S, et al. Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. BMC Genomics 2012;13:738.

[5] Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, et al. Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation. Microb Cell Fact 2004;3:17–30.

[6] Van Zutphen T, Baerends RJS, Susanna KA, de Jong A, Kuipers OP, Veenhuis M, et al. Adaptation of Hansenula polymorpha to methanol: a transcriptome analysis. BMC Genomics 2010;11:1.

[7] Ravin N V, Eldarov MA, Kadnikov V V, Beletsky A V, Schneider J, Mardanova ES, et al. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1 2013.

[8] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol 2013;30:385–404.

[9] Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 2013;496:528–32.

[10] Da Silva NA, Srikrishnan S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res 2012;12:197–214.

[11] Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact 2006;5:39–59.

[12] Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, et al. Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks. J Biotechnol 2014.

[13] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 2008;36:e76.

[14] Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 2012;11:22.

[15] Lukas Sturmberger, Characterization of synthetic regulatory elements for protein expression in Pichia pastoris, Master thesis, Graz University of Technology 2013

[16] Maria Freigassner, Alternative oxidases: Elucidation of structural and functional differences between plants and fungi, Dissertation, Graz University of Technology 2009

[17] Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, Glieder A, et al. Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression. N Biotechnol 2014;00.

[18] Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic Core Promoters for Pichia pastoris. ACS Synth Biol 2014;3:188–91.

[19] Claudia Ruth, Expression of recombinant proteins in Pichia pastoris employing synthetic promoters, Dissertation, Graz University of Technology 2010

[20] Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.

[21] Mead DA, Pey NK, Herrnstadt C, Marcil RA, Smith LM. A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology (N Y) 1991;9:657–63.

[22] Lu Q. Seamless cloning and gene fusion. Trends Biotechnol 2005;23:199–207.

[23] Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in P. pastoris. Chem Commun (Camb) 2014;62.

5 Appendix

Plasmids generated (in *E. coli* TOP 10 F')

Name	#	Alternate name/Comment	Marker	Chapter
Chapter 3.1 (related natural bidirectional BDPs of chapter 3.5 also shown here)				
pPpT4mutZeoMlyI-intArg4-eGFP-BmrIstuffer	TV0082		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-AOX1BglII	TV0020		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-AOX1bidi	TV0021	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-AOX1bidi reverse	TV0022	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-AOX2	TV0023		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FLD1	TV0025		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FLD1 reverse	TV0026	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-FLD1 Shen et al.	TV0027		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-DAS 1/2	TV0028	DAS1, bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-DAS 1/2 reverse	TV0029	DAS2, bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-DAS 2 NEFBidi	TV0030	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-DAS1	TV0031		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FBA1	TV0032		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FBA1 reverse	TV0033	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-FBA2	TV0034		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FBA2 reverse	TV0035	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-ADH2	TV0036		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-ADH2 reverse	TV0037	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-pCAT1-692	TV0038		Zeo	3.1, 3.2
pPpT4mutZeoMlyI-intArg4-eGFP-FGH1	TV0039		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FGH1 reverse	TV0040	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-FDH1	TV0041		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FDH1 reverse	TV0042	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-TPI1 Stadelmayr et al.	TV0044		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-DAK1	TV0046		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-DAK1 reverse	TV0047	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-FBP1	TV0048		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-FBP1 reverse	TV0049	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-TKL1	TV0050		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-TKL1 reverse	TV0051	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-PEX14	TV0052		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PEX8	TV0053		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PEX5	TV0054		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PEX5 reverse	TV0055	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-GAP(short)	TV0056		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-GAPbidi	TV0057		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-GAPbidi reverse	TV0058	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-PFK	TV0059		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PFK reverse	TV0060	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-GPM1	TV0061		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-GPM1reverse	TV0062	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-PGIshort	TV0063		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-GLR1 down	TV0105	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-GLR1 up (methyltransferse)	TV0106	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-GPX	TV0107		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-GSH1	TV0108		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-GSH2	TV0109		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-MSR1c3 down	TV0110	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-MSR1c3 up	TV0111	bidirectional	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-MSR1c4	TV0112		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-MSR2c2 (long)	TV0113		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-MSR2c4	TV0114		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-MSRAB	TV0115		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PXR down	TV0116	PMP20 bidirectional	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-PXR up	TV0117		Zeo	
pPpT4mutZeoMlyI-intArg4-eGFP-SOD1down (=SOD)	TV0118	bidirectional	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-SOD1up (=hyph prot)	TV0119	bidirectional	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-SODcytoSODmito down	TV0120	bidirectional	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-SODcytoSODmito up	TV0121	bidirectional	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pDAS1-1000	TV0122		Zeo	3.1, 3.4
pPpT4mutZeoMlyI-intArg4-eGFP-pDAS2-1000	TV0123		Zeo	3.1, 3.4
pPpT4mutZeoMlyI-intArg4-eGFP-pPGD	TV0124		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pRKI1	TV0125		Zeo	3.1

pPpT4mutZeoMlyI-intArg4-eGFP-pRKI2	TV0126		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pRPE1 up	TV0127		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pRPE2 up	TV0128		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pSOL1	TV0129		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pSOL3	TV0130		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pTAL1 up	TV0131		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pTAL2 up	TV0132		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pTKL1	TV0133	pTKL1-1000	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pZWF1	TV0134		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pGUT1 down	TV0135		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-eGFP-pGUT1 up	TV0136		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pAOX1-CalB	TV1020	Just single gene expressed, but based on bidirectional vector, hence the name	Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pCAT1-CalB	TV0935		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pGAP-CalB	TV0936		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pAOX1-HRPA2	TV0939		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pCAT1-HRPA2	TV1021		Zeo	3.1
pPpT4mutZeoMlyI-intArg4-bidi-pGAP-HRPA2	TV0940		Zeo	3.1
Chapter 3.2				
pPpT4mutZeoMlyI-intArg4-EGFP- pCAT1-500	TV0973		Zeo	3.2
pPpT4mutZeoMlyI-intArg4-EGFP- pCAT1-692	TV0038		Zeo	3.2, 3.1
pPpT4mutZeoMlyI-intArg4-EGFP- pCAT1-1000	TV0974		Zeo	3.2
pPpT4mutZeoMlyI-intArg4-EGFP-putARS-pCAT1-500..764	TV1022		Zeo	3.2
pPpT4-KAN-intARG4-eGFP-pCAT1-500	TV1035		Kan/Gen	3.2
pPpT4-KAN-intARG4-eGFP-pCAT1-692	TV1036		Kan/Gen	3.2
pPpT4-KAN-intARG4-eGFP-pCAT1-1000	TV1037		Kan/Gen	3.2
pPpT4-KAN-intARG4-eGFP-putARS-pCAT1-500..764	TV1038		Kan/Gen	3.2
pPpT4-GUT1-intARG4-eGFP-pCAT1-500	TV1070		Amp/Δgut1	3.2
pPpT4-GUT1-intARG4-eGFP-pCAT1-692	TV1039		Amp/Δgut1	3.2
pPpT4-GUT1-intARG4-eGFP-pCAT1-1000	TV1040		Amp/Δgut1	3.2
pPpT4-GUT1-intARG4-eGFP-putARS-pCAT1	TV1071		Amp/Δgut1	3.2
pPpT4mutZeoMlyI-intArg4-MeHNLoptTV-pCAT1-692	TV1187	CAM1	Zeo	3.2
pPpT4mutZeoMlyI-intArg4-MeHNLoptTV-pCAT1-500	TV1188	CAM2	Zeo	3.2
pPpT4-GUT1-intArg4-MeHNLoptTV-pCAT1-692	TV1189	CAM3	Amp/Δgut1	3.2
pPpT4-GUT1-intArg4-MeHNLoptTV-pCAT1-500	TV1190	CAM4	Amp/Δgut1	3.2
pPpT4mutZeoMlyI-intArg4-LuHNL-pCAT1-692	TV1191	CAL1	Zeo	3.2
pPpT4mutZeoMlyI-intArg4-LuHNL-pCAT1-500	TV1192	CAL2	Zeo	3.2
pPpT4-GUT1-intArg4-LuHNL-pCAT1-692	TV1193	CAL3	Amp/Δgut1	3.2
pPpT4-GUT1-intArg4-LuHNL-pCAT1-500	TV1194	CAL4	Amp/Δgut1	3.2
Chapter 3.3				
pPpT4mutZeoMlyI-intArg4-eGFP-HpFMD	TV0064		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-eGFP-HpMOX	TV0065		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-eGFP-CbFLD1	TV0066		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-eGFP-PmMOD1	TV0067		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-eGFP-PmMOD2	TV0068		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-eGFP-CbAOD1	TV0069		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-CalB-pHpFMD	TV1165		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-HRPA2A-pHpFMD	TV1166		Zeo	3.3
pPpT4mutZeoMlyI-intArg4-MeHNLoptTV-pHpFMD	TV1167		Zeo	3.3
Chapter 3.4				
Screening vectors				
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-BmrIstuffer	TV0159	Short name: pPpT4-bidi-sTomato-EGFP- BmrIstuffer Should be called dTomato (This vector was used for 99% of BDPs)	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomatoNheI_EGFP-BmrIstuffer	TV0158	Should be called dTomato	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_Citrine_EGFP-Bmrstuffer	TV0156		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_dTomato_EGFP-Bmrstuffer	TV0157	Actually tandem dTomato	Zeo	3.4
Natural bidirectional promoters				
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi1	TV0181	two genes involved in transcription initiation	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi2	TV0182	HTX1	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi3	TV0183	HHX2	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi4	TV0184	PFK and TATA box associated factor	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi5	TV0185	two ribosomal proteins	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi6	TV0186	RNA pol + transcription factor	Zeo	3.4

pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi7	TV0187	ribosomal protein and SER3	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi8	TV0188	FTR1 + FET3 (high affinity iron uptake)	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi9	TV0189	CYC3+ PDA1	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-singleTomatoNheI_EGFP-natbidi10	TV0190	FDH1 + autophagy thing	Zeo	3.4
Natural histone promoters and control constructs				
pPpT4-bidi-sTomato-EGFP-pHTX1-WT-fwd	TV0253		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHTX1-WT-rev	TV0254		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX1-WT-fwd	TV0255		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX1-WT-rev	TV0256		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-WT-fwd	TV0257		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-WT-rev	TV0258		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHO1-WT-bidi	TV0259		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHO1-WT-1000	TV0260		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHTZ-WT-243	TV0261		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHTZ-WT-1000	TV0262		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCSE4-WT-bidi	TV0527		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCSE4-WT-1000	TV0528		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pGAP-fwd	TV0263		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pGAP-rev	TV0264		Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-fwd	TV0453	C7	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-rev	TV0454	C8	Zeo	3.4
pPpT4-bidi-pHHX2-fwd-eGFP	TV0265		Zeo	3.4
pPpT4-bidi-pHHX2-rev-eGFP	TV0266		Zeo	3.4
pPpT4-bidi-pHHX2-fwd-sTomato	TV0267		Zeo	3.4
pPpT4-bidi-pHHX2-rev-sTomato	TV0268		Zeo	3.4
HHX2 deletions				
pPpT4-bidi-sTomato-EGFP-pHHX2-D1	TV0269	D1	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D2	TV0270	D2	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D3	TV0271	D3	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D4	TV0272	D4	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D5	TV0273	D5	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D6	TV0274	D6	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D7	TV0275	D7	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D8	TV0276	D8	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D9	TV0277	D9	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D10	TV0278	D10	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D11	TV0279	D11	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D12	TV0280	D12	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D13	TV0281	D13	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D14	TV0282	D14	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D15	TV0283	D15	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D9-D12	TV1099	BF10 (sic!)	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D7-D12	TV1100	BF11 (sic!)	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-D8-D11	TV1101	BF12 (sic!)	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHT2-T292	TV0284	T1	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHT2-T219	TV0285	T2	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHT2-T146	TV0286	T3	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHT2-T73	TV0287	T4	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHF2-T292	TV0288	T5	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHF2-T219	TV0289	T6	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHF2-T146	TV0290	T7	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHF2-T73	TV0291	T8	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SDA1	TV0431	S1	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SDA2	TV0432	S2	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SDA3	TV0433	S3	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SDA4	TV0434	S4	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SDA5	TV0435	S5	Zeo	3.4
Bidirectionalization				
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-73+pAOX1BglII (BZ1)	TV0830		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-73+pAOX1-711 (BZ2)	TV0831		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-73+pAOX1-643 (BZ3)	TV0832		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF2-76+pDAS1-552 (BZ4)	TV0833		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF2-76+pDAS1-1000 (BZ5)	TV0834		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTA1-	TV0835		Zeo	3.4

81+pDAS2-699 (BZ6)				
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTA1-81+pDAS2-1000 (BZ7)	TV0836		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-86+pPXR1-478CBS (BZ8)	TV0837		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-86+pPXR1-392CBS (BZ9)	TV0838		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-86+pPXR1-480GS (BZ10)	TV0839		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT1-91+pFLD1-366 (BZ11)	TV0840		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-80+pFDH1-564 (BZ12)	TV0841		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT1-91+pFBA2-500 (BZ13)	TV0842		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT1-91+pFBA2-704 (BZ14)	TV0843		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-80+pTAL2-1000 (BZ15)	TV0844		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-80+pTAL2-500 (BZ16)	TV0845		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-73+pCAT1-692 (BZ17)	TV0846		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-73+pCAT1-500 (BZ18)	TV0847		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF2-76-pGAP-486 (BZ19)	TV0848		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTA1-81-pTEF1-424 (BZ20)	TV0849		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-86-pADH2-500 (BZ21)	TV0850		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-MCM5-TAL2bidi (made together with BZ series))	TV0851		Zeo	3.4
BZ23 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-89+pAOX1-711	TV0916		Zeo	3.4
BZ24 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT2-105+pAOX1-711	TV0917		Zeo	3.4
BZ25 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-106 +pPXR1-392CBS	TV0918		Zeo	3.4
BZ26 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHTB1-126+pPXR1-392CBS	TV0919		Zeo	3.4
BZ27 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT1-111+pFLD1-366	TV0920		Zeo	3.4
BZ28 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHT1-131+pFLD1-366	TV0921		Zeo	3.4
BZ29 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-80+pAOX1-711	TV0922		Zeo	3.4
BZ30 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-100+pAOX1-711	TV0923		Zeo	3.4
BZ31 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pCoreHHF1-121+pAOX1-711	TV0924		Zeo	3.4
Fusion promoters				
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS1,2nat-fwd K2 [shortcut number 11]	TV0204		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pAOX1BglII-pCAT1-FUSION-fwd K1 [shortcut number 12]	TV0205		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pGAPshort-pCAT1-FUSION-fwd K3 [shortcut number 13]	TV0206		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pGAPshort-pTEF1-FUSION-fwd K3 [shortcut number 14]	TV0207		Zeo	3.4
BZF1 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFBA2-500+pTAL2-500	TV0925		Zeo	3.4
BZF2 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFDH1-564+pDAS1-552	TV0926		Zeo	3.4
BZF3 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFDH1-564+pCAT1-500	TV0927		Zeo	3.4
BZF4 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS2-699+pDAS1-552	TV0928		Zeo	3.4
BZF5 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFDH1-564+pPXR1-392	TV0929		Zeo	3.4
BZF6 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFLD1-366+pAOX1-643	TV0930		Zeo	3.4
BZF7 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pAOX2-500+pCAT1-500	TV0931		Zeo	3.4
BZF8 - pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pFLD1-	TV0932		Zeo	3.4

366+ pPXR1-392				
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS1-552+pGAP-486	TV1090	BF1	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS1-552+pHTA1-464	TV1091	BF2	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS1-552+pHTB1-469	TV1092	BF3	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS2-699+pGAP-486	TV1093	BF4	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS2-699+pHTA1-464	TV1094	BF5	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pDAS2-699+pHTB1-469	TV1095	BF6	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pTAL2-501+pGAP-486	TV1096	BF7	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pTAL2-501+pHTA1-464	TV1097	BF8	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-pTAL2-501+pHTB1-469	TV1098	BF9	Zeo	3.4
Monodirectional DAS1,2 deletion variants				
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del1	TV0137		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del2	TV0138		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del2+del5	TV0139		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del3	TV0140		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del4	TV0141		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del5	TV0142		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del6	TV0143		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del7	TV0144		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-del8	TV0145		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS1-WT	TV0146		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del1	TV0147		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del2	TV0148		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del3	TV0149		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del4	TV0150		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del5	TV0151		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del6	TV0152		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del7	TV0153		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-del8	TV0154		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-eGFP-SbfI-pDAS2-WT	TV0155		Zeo	3.4
Bidirectional fusions of monodirectional DAS1,2 deletions				
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-1000+pDAS1-1000	TV0778	DDC1	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del8+pDAS1-del2del5	TV0779	DDC2	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del2+pDAS1-del2del5	TV0780	DDC3	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del6+pDAS1-del6	TV0781	DDC4	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del8+pDAS1-del6	TV0782	DDC5	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del6+pDAS1-del2del5	TV0783	DDC6	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del5+pDAS1-del6	TV0784	DDC7	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-del6+pDAS1-del7	TV0785	DDC8	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-trunc386+pDAS1-del6	TV0786	DDC9	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi_singleTomato_EGFP-pDAS2-trunc261+pDAS1-del6	TV0787	DDC10	Zeo	3.4
Synthetic hybrid BDPs (shBDPs) and controls				
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi1	TV0436	B1; shBDP21	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi2	TV0437	B2; shBDP22	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi3	TV0438	B3; shBDP23	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi4	TV0439	B4; shBDP24	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi5	TV0440	B5; shBDP19	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi6	TV0441	B6; shBDP17	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi7	TV0442	B7; shBDP20	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi8	TV0443	B8; shBDP18	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi9	TV0444	B9; shBDP15	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi10	TV0445	B10; shBDP16	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi11	TV0446	B11; shBDP37	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi12	TV0447	B12; shBDP36	Zeo	3.4

pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi13	TV0690	B13; shBDP32	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi14	TV0691	B14; shBDP33	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi15	TV0692	B15; shBDP34	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-SynBidi16	TV0693	B16; shBDP28	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCoreHHF2-76	TV0448	C1; CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCoreHHF2-86	TV0449	C2; CF2-86	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-2-pCoreHHF2TATA(61)	TV0450	C3; CF2-61	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCoreHHT2-73	TV0287	C4=T4; CT2-73	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCoreHHT2-89	TV0451	C5; CT2-89	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pCoreHHT2-TATA(55)	TV0452	C6; CT2-55	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-1..302	TV0572	V1; pAOX1-(-302..-1)	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pDAS1-TATA-105+pAOX1-1..302	TV0573	V2; pDAS1-D5-D7L+pAOX1-(-320..-1)	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-160..940+pCoreHHF2-TATA(61)	TV0574	V3; pAOX1-(-940..-160)+pCoreHHF2-61	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-TATA-160->pCoreHHF2-TATA(61)	TV0575	V4; TL-CF2-61	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-TATA-188->pCoreHHF2-76	TV0576	V5; TS-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-XuanD->pCoreHHF2-76	TV0577	V6; HX-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-Rap1ext-->pCoreHHF2-76	TV0578	V7; R1-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pAOX1-D6ext-->pCoreHHF2-76	TV0579	V8; H6-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pDAS1-TATA-105->pCoreHHF2-TATA(61)	TV0580	V9; 5L-CF2-61	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pDAS1-TATA-126->pCoreHHF2-76	TV0581	V10; 5S-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pDAS2-D6toD8->pCoreHHF2-76	TV0582	V11; D6-CF2-76	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pHHT2-T146+pAOX1-TATA-188->pCoreHHF2-76-->	TV0583	V12; shBDP11	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pHHT2-T146+pDAS1-TATA-126->pCoreHHF2-76-->	TV0584	V13; shBDP13	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pHHT2-T146+pAOX1-XuanD->pCoreHHF2-76-->	TV0585	V14; shBDP9	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pHHT2-T146+pAOX1-Rap1ext->pCoreHHF2-76-->	TV0586	V15; shBDP10	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pHHT2-T146+<-pAOX1-TATA-188|pDAS1-TATA-126->pCoreHHF2-76->	TV0587	V16	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pDAS1-TATA-126|pAOX1-TATA-188->pCoreHHF2-76->	TV0588	V17; 2 CRMs +1 CP, supporting info	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pCoreHHT2-73|pDAS1-TATA-126->|pAOX1-TATA-188->|pCoreHHF2-76->	TV0589	V18; shBDP25	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pCoreHHT2-73|pDAS1-TATA-126<-|pAOX1-TATA-188<-|pCoreHHF2-76->	TV0590	V19; shBDP26	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pCoreHHT2-73|pDAS1-TATA-126->|<-pAOX1-TATA-188|pCoreHHF2-76->	TV0591	V20; shBDP27	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-pHHX2-V21	TV0592	V21	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pAOX1-TATA-188|pDAS1-TATA-126->pCoreHHF2-76	TV0593	V22; 2 CRMs +1 CP, supporting info	Zeo	3.4
pPpT4-bidi-sTomato-EGFP<-pAOX1-XuanD(+Hartner-D2)|pDAS2-D6toD8->pCoreHHF2-76	TV0594	V23; 2 CRMs +1 CP, supporting info	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<-pAOX1-TATA-160|pDAS1-TATA-105->pCoreHHF2-TATA(61)->	TV0595	V24; 2 CRMs +1 CP, supporting info	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-TATA(55)<-pAOX1-TATA-160->pCoreHHF2-TATA(61)	TV0661	V25; shBDP3	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pAOX1-TATA-188->pCoreHHF2-76	TV0662	V26; shBDP4	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pAOX1-XuanD->pCoreHHF2-76	TV0663	V27; shBDP1	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pAOX1-Rap1ext-->pCoreHHF2-76	TV0664	V28; shBDP2	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pAOX1-D6ext->pCoreHHF2-76	TV0665	V29; shBDP5	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-TATA(55)<-pDAS1-TATA-105->pCoreHHF2-TATA(61)	TV0666	V30; shBDP6	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pDAS1-TATA-126->pCoreHHF2-76	TV0667	V31; shBDP7	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-<--pCoreHHT2-73<-pDAS2-D6toD8->pCoreHHF2-76	TV0668	V32; shBDP8	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V33	TV0694	V33; shBDP12	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V34	TV0695	V34; shBDP14	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V35	TV0775	V35; shBDP29	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V36	TV0776	V36; shBDP30	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V37	TV0777	V37; shBDP31	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V38	TV1195	V38; V30 'rev'	Zeo	3.4

pPpT4-bidi-sTomato-EGFP-V39	TV1196	V39; V31 'rev'	Zeo	3.4
pPpT4-bidi-sTomato-EGFP-V40	TV1197	V40; V32 'rev'	Zeo	3.4
Bidirectional terminators				
pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-NotI-stuffer-NotI-sTomato-pAOX1 (=bidi terminator cloning vector)	TV0760	TBC2	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-NotI-sTomato-pAOX1 (no stuffer control)	TV0955	TBC1	Zeo	3.4
TBF1 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-DAS1TT+AOX1TTstar wo NotI K-2A	TV0941		Zeo	3.4
TBF2 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-FBP1TT+DAS2TT K-11B	TV0942		Zeo	3.4
TBF3 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-ScSPG5TT+ScIDP1TT K-3D	TV0943		Zeo	3.4
TBF4 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-ADH2TT+PXR1TT K-7D	TV0944		Zeo	3.4
TBF5 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-FBA1TT+FBA2TT K-4G	TV0945		Zeo	3.4
TBF6 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-FLD1TT+GAPTT K-9D	TV0946		Zeo	3.4
TBF7 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-ScPRM9TT+ScHSP26TT K-1A	TV0947		Zeo	3.4
TBF8 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-ScUBX6TT+TPI1TT K-11D	TV0948		Zeo	3.4
TBF9 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-FDH1TT+TAL2TT K-6E	TV0949		Zeo	3.4
TBF10 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-GCW14TT+TEF1TT K-11C	TV0950		Zeo	3.4
TBF11 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1- DAS1TT+DAS2TT K-3A	TV0951		Zeo	3.4
TBF12 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-FBP1TT+AOX1TTstar wo NotI K-10B	TV0952		Zeo	3.4
TBN1 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-ScIDP1TT (+ScPEX19TT) K-6C	TV0953		Zeo	3.4
TBN2 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-sTomato-pAOX1-TEF1TT (+gcvP-TT) K-8D	TV0954		Zeo	3.4
TBC1 - pPpT4mutZeoMlyI-intArg4-pAOX1-eGFP-NotI-sTomato-pAOX1 (no stuffer control) K-2E	TV0955		Zeo	3.4
Applications (dual gene coexpression)				
CalB+sPDI				
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-BmrIstuffer(fwd)	TV0208		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pDAS1,2nat-fwd K1 [short 1]	TV0209		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pDAS1,2nat-rev K1 [short 2]	TV0210		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pAOX1+pGAP-fwd K1 [short 3]	TV0211		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pAOX1+pGAP-rev K1 [short 4]	TV0212		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pAOX1+pCAT1-fwd K1 [short 5]	TV0213		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pAOX1+pCAT1-rev K2 [short 6]	TV0214		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pGAP+pCAT1-fwd K1 [short 7]	TV0215		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-GAP+pCAT1-rev K1 [short 8]	TV0216		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pHTA1,HTB2-fwd K1 [short 9]	TV0217		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pHTA1,HTB2-rev K1 [short 10]	TV0218		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pHistH3,H4-fwd K1 [short 11]	TV0219		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CalB_VTU-synPDImutBmrI-pHistH3,H4-rev K1 [short 12]	TV0220		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-pCAT1-CalB_VTU-pAOX1-synPDImutBmrI	TV0933	CalB C1 (control)	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-pCAT1-CalB_VTU-pGAP-synPDImutBmrI	TV0934	CalB C2 (control)	Zeo	3.4
HRP+sPDI				
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-BmrIstuffer(fwd) K7	TV0222	Entry vector	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pDAS1,2nat-fwd (abbr. 1) K1	TV0223		Zeo	3.4

pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pDAS1,2nat-rev (abbr. 2) K1	TV0224		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pAOX1+pGAP-fwd (abbr. 3) K2	TV0225		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pAOX1+pGAP-rev (abbr. 4) K1	TV0226		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pAOX1+pCAT1-fwd K1(abbr. 5) K2	TV0227		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pAOX1+pCAT1-rev (abbr. 6) K1	TV0228		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pGAP+pCAT1-fwd (abbr. 7) K2	TV0229		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pGAP+pCAT1-rev (abbr. 8) K1	TV0230		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pHTA1,HTB2-fwd (abbr. 9) K1	TV0231		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pHTA1,HTB2-rev (abbr. 10) K2	TV0232		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pHistH3,H4-fwd (abbr. 11) K1	TV0233		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HRPA2-synPDImutBmrI-N314H-pHistH3,H4-rev (abbr. 12) K1	TV0234		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-pAOX1-HRPA2-pCAT1-synPDImutBmrI-N314H	TV0937	HRP C1	Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-pAOX1-HRPA2-pGAP-synPDImutBmrI-N314H	TV0938	HRP C2	Zeo	3.4
HsCYP2D6+CPR				
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-BmrIstuffer K3	TV0236		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pDAS1,2nat-fwd [short 1] K3	TV0317		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pDAS1,2nat-rev [short 2] K3	TV0318		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pAOX1+pGAP-fwd [short 3] K1	TV0319		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pAOX1+pGAP-rev [short 4] K3	TV0320		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pAOX1+pCAT1-fwd [short 5] k5	TV0321		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pAOX1+pCAT1-rev [short 6] k1	TV0322		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pGAP+pCAT1-fwd [short 7] K1	TV0323		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pGAP+pCAT1-rev [short 8] K2	TV0324		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pGAP+pTEF1-fwd [short 9] K2	TV0325		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-HsCYP2D6-HsCPRmutBmrI-pGAP+pTEF1-rev [short 10] K1	TV0326		Zeo	3.4
CtCYP52A13+CPR				
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-BmrIstuffer(fwd) (=pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-BmrIstuffer(fwd) K2)	TV0193		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pDAS1,2nat-fwd K2 [shortcut number 1]	TV0194		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pDAS1,2nat-fwd K3 [shortcut number 2]	TV0195		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pAOX1BglII-pGAPshort-FUSION-fwd K2 [shortcut number 3]	TV0196		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pAOX1BglII-pGAPshort-FUSION-rev K3 [shortcut number 4]	TV0197		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pAOX1BglII-pCAT1-FUSION-fwd K3 [shortcut number 5]	TV0198		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pAOX1BglII-pCAT1-FUSION-rev K3 [shortcut number 6]	TV0199		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pGAPshort-pCAT1-FUSION-fwd K3 [shortcut number 7]	TV0200		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pGAPshort-pCAT1-FUSION-rev K1	TV0201		Zeo	3.4

[shortcut number 8]				
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pGAPshort-pTEF1-FUSION-fwd K1 [shortcut number 9]	TV0202		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-bidi-CtCYP52A13mutBmrI-Ct_NCP_AmutBmrI-pGAPshort-pTEF1-FUSION-fwd K1 [shortcut number 10]	TV0203		Zeo	3.4
Taxadiene (dual gene coexpression application)				
pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-BmrIstuffer	TV1221 TV1222	Taxadiene entry vector	Zeo	3.4
TX1-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pDAS1,2-fwd KC	TV1223		Zeo	3.4
TX2-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pDAS1,2-rev KB	TV1224		Zeo	3.4
TX3-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pDAS2-d8|pDAS1-d2d5 KA	TV1225		Zeo	3.4
TX4-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-B16 fwd KB	TV1226		Zeo	3.4
TX5-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-B16 rev KC	TV1227		Zeo	3.4
TX6-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pAOX1+pCAT1-500 fwd KC	TV1228		Zeo	3.4
TX7-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pAOX1+pCAT1-500 rev KB	TV1229		Zeo	3.4
TX8-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pAOX1+pGAP fwd KB	TV1230		Zeo	3.4
TX9-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pAOX1+pGAP rev KA	TV1231		Zeo	3.4
TX10-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pGAP+pCAT1-500 fwd KB	TV1232		Zeo	3.4
TX11-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pGAP+pCAT1-500 rev KB	TV1233		Zeo	3.4
TX12-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pHTX1 KA	TV1234		Zeo	3.4
TX13-pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-pHHX2 KC	TV1235		Zeo	3.4
Entry vectors for pathways				
pPpT4_S-DAS1TT-NotI-AOX1TT Clone A	TV0975		Zeo	3.4
pPpT4mutZeoMlyI-intArg4-DAS1TT-NotI-AOX1TT Clone F	TV0976		Zeo	3.4
Chapter 3.5				
For Zeocin selection based screening vectors bearing the factors under control of P_{GAP}/P_{CAT1} and the promoters of certain TFs cloned upstream of eGFP, see also the master thesis of Lukas Sturmberger. Still all strains are listed here.				
pPpT4mutZeoMlyI-intUra3-pGAP-Mxr1-pAOX1-sTomato	TV0677		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAt1-Mxr1-pAOX1-sTomato	TV0678		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Prm1-pAOX1-sTomato	TV0679		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Prm1-pAOX1-sTomato	TV0680		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Mpp1-pAOX1-sTomato	TV0681	Trm1	Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Mpp1-pAOX1-sTomato	TV0682	Trm1	Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Mxr1c-pAOX1-sTomato	TV0687		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT1-Mxr1c-pAOX1-sTomato	TV0688		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Snf1-pAOX1-sTomato	TV0761		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Snf2-pAOX1-sTomato	TV0762		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Cat8-pAOX1-sTomato	TV0763		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Reg1-pAOX1-sTomato	TV0764		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Gal4-pAOX1-sTomato	TV0765		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Snf2-pAOX1-sTomato	TV0766		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Rpd3-pAOX1-sTomato	TV0767		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Hda1-pAOX1-sTomato	TV0768		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Swi1-pAOX1-sTomato	TV0769		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Snf1-pAOX1-sTomato	TV0770		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Reg1-pAOX1-sTomato	TV0771		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Gal4-pAOX1-sTomato	TV0772		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT-Swi1-pAOX1-sTomato	TV0773		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-Rpd3-pAOX1-sTomato	TV0774		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP1-Cat8-pAOX1-sTomato	TV0720		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP1-Hda1-pAOX1-sTomato	TV0721		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pGAP-eGFP-pAOX1-sTomato	TV0305	In some lists mislabeled intArg4 instead of intUra3	Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pCAT1-eGFP-pAOX1-sTomato	TV0306		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pMxr1-sTomato	TV0683		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pPrm1-sTomato	TV0684		Zeo	3.5
pPpT4mutZeoMlyI-intUra3-pMpp1-sTomato	TV0685	Trm1	Zeo	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-500-Mxr1	TV0977		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-500-Mpp1	TV0978	Trm1	Kan/Gen	3.5

pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-500-Prm1	TV0979		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-500-eGFP	TV0980		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-Mxr1 (pCAT1-692)	TV0889	These are pCAT1-692 based conversion plasmids initially made, encountered massive instability problems upon Gen selection. I can't think of a possible scenario where those should be needed. Listed as backup	Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-Mpp1 (pCAT1-692) (Mpp1 = Trm1)	TV0890		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-Prm1 (pCAT1-692)	TV0891		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pGAP-Mxr1c (pCAT1-692)	TV0892		Kan/Gen	3.5
pPpKan-intURA3-sTomato-AOX1BglII-pCAT1-eGFP (pCAT1-692)	TV0893		Kan/Gen	3.5
Chapter 3.6.1				
pPpT4_SB-sync1-eGFP	TV0070		Zeo	3.6.1
pPpT4_SB-sync2-eGFP	TV0071		Zeo	3.6.1
pPpT4_SB-sync3-eGFP	TV0072		Zeo	3.6.1
pPpT4_SB-sync4-eGFP	TV0073		Zeo	3.6.1
pPpT4_SB-sync6-eGFP	TV0074	sync5 in paper	Zeo	3.6.1
pPpT4_SB-sync7-eGFP	TV0075	sync6 in paper	Zeo	3.6.1
pPpT4_SB-sync3_ins-eGFP	TV0076		Zeo	3.6.1
pPpT4_SB-sync4_ins-eGFP	TV0077		Zeo	3.6.1
pPpT4_SB-sync7_ins-eGFP	TV0078	sync6_ins in paper	Zeo	3.6.1
pPpT4_SB-pAOX1-eGFP	TV0079		Zeo	3.6.1
pPpT4_SB-pCore11-eGFP	TV0080		Zeo	3.6.1
pPpT4_SB-truncatedpAOX1-eGFP	TV0081	simplifying cloning	Zeo	3.6.1
pPpT4_SB-pCore1-eGFP	TV0689		Zeo	3.6.1
Chapter 3.6.3 (Stain collection numbers of the IMBT Graz provided)				
pPpRSFC [1P]	6480	#1	Zeo	3.6.3
pPpRSFC_alpha [2P]	6481	#2	Zeo	3.6.3
pPpRSFC_alpha.noEAEA [3P]	6482	#3	Zeo	3.6.3
pPpRSFC_GAP [4P]	6483	#4	Zeo	3.6.3
pPpRSFC_GAP_alpha [5P]	6484	#5	Zeo	3.6.3
pPpRSFC_GAP_alpha.noEAEA [6P]	6485	#6	Zeo	3.6.3
pPpRSFC_N.EGFP [7P]	6712	#7	Zeo	3.6.3
pPpRSFC_C.EGFP [8P]	6713	#8	Zeo	3.6.3
pPpRSFC_alpha_N.EGFP [9P]	6714	#9	Zeo	3.6.3
pPpRSFC_alpha_C.EGFP [10P]	6715	#10	Zeo	3.6.3
pPpRSFC_N.MYC [11P]	6716	#11	Zeo	3.6.3
pPpRSFC_C.MYC [12P]	6717	#12	Zeo	3.6.3
pPpRSFC_alpha_N.MYC [13P]	6718	#13	Zeo	3.6.3
pPpRSFC_alpha_C.MYC [14P]	6719	#14	Zeo	3.6.3
pPpRSFC_N.FLAG [15P]	6720	#15	Zeo	3.6.3
pPpRSFC_C.FLAG [16P]	6721	#16	Zeo	3.6.3
pPpRSFC_alpha_N.FLAG [17P]	6722	#17	Zeo	3.6.3
pPpRSFC_alpha_C.FLAG [18P]	6723	#18	Zeo	3.6.3
pPpRSFC_N.HIS.ncs [19P]	6724	#19	Zeo	3.6.3
pPpRSFC_C.HIS.ncs [20P]	6725	#20	Zeo	3.6.3
pPpRSFC_alpha_N.HIS.ncs [21P]	6726	#21	Zeo	3.6.3
pPpRSFC_alpha_C.HIS.ncs [22P]	6727	#22	Zeo	3.6.3
pPpRSFC_N.HIS [23P]	6728	#23	Zeo	3.6.3
pPpRSFC_C.HIS [24P]	6729	#24	Zeo	3.6.3
pPpRSFC_alpha_N.HIS [25P]	6730	#25	Zeo	3.6.3
pPpRSFC_alpha_C.HIS [26P]	6731	#26	Zeo	3.6.3
pPpRSFC_N.MBP [27P]	6732	#27	Zeo	3.6.3
pPpRSFC_C.MBP [28P]	6733	#28	Zeo	3.6.3
pPpRSFC_alpha_N.MBP [29P]	6734	#29	Zeo	3.6.3
pPpRSFC_alpha_C.MBP [30P]	6735	#30	Zeo	3.6.3
pPpRSFC_N.STREP [31P]	6736	#31	Zeo	3.6.3
pPpRSFC_C.STREP [32P]	6737	#32	Zeo	3.6.3
pPpRSFC_alpha_N.STREP [33P]	6738	#33	Zeo	3.6.3
pPpRSFC_alpha_C.STREP [34P]	6739	#34	Zeo	3.6.3
pPpRSFC_HIS [1PHIS]	6762	#35	Amp/Δhis4	3.6.3
pPpRSFC_HIS_alpha [2PHIS]	6763	#36	Amp/Δhis4	3.6.3
pPpRSFC_HIS_alpha.noEAEA [3PHIS]	6764	#37	Amp/Δhis4	3.6.3
pPpRSFC_HIS_GAP [4PHIS]	6765	#38	Amp/Δhis4	3.6.3
pPpRSFC_HIS_GAP_alpha [5PHIS]	6766	#39	Amp/Δhis4	3.6.3
pPpRSFC_HIS_GAP_alpha.noEAEA [6PHIS]	6767	#40	Amp/Δhis4	3.6.3