Thomas Kickenweiz, MSc Bakk. rer. nat

P. pastoris engineering to metabolize cellobiose and cellulose

DOCTORAL THESIS

to achieve the university degree of

Doktor der Naturwissenschaften

submitted to

Graz University of Technology

Supervisor

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

Institute of Molecular Biotechnology

Graz, January 2018

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.

Date

Signature

Abstract

Cellulose is the most abundant polymer in our biosphere. Cellulose molecules consist of thousands of glucose residues which can be used by cellulolytic microorganisms as carbon source. These microorganisms produce cellulases for cellulose degradation. Cellulases are very important for industries like detergent, energy, paper and pulp or beverage industry. Much effort was done to find novel cellulases to improve the cellulose degradation efficiency. The yeast *Komagataella phaffii* is a broadly used expression host in industrial biotechnology for production of high value compounds and enzymes.

This study was focused to engineer a *K. phaffii* strain which is able to metabolize cellulose. Since cellulose is widely available, it could be used as cheap carbon source for production of high value compounds and enzymes by *K. phaffii*. Furthermore, the study was also focused on creating screenings to find a) *K. phaffii* mutants with improved secretion of expressed cellulases and b) novel cellulases.

Zusammenfassung

Zellulose ist das am häufigsten vorkommene Polymermolekül in unserer Biosphäre. Sie besteht aus tausenden von Glukosemolekülen, die von zellulolytischen Mikroorganismen als Kohlenstoffquelle dient. Diese Mikroorganismen produzieren Cellulasen, die die Zellulose zu Glukose abbauen. Diese Cellulasen haben eine große Bedeutung in verschiedenen Bereichen wie z. B. Detergenz-, Energie-, Papier- und Getränke-Industrie. Große Anstrengungen werden unternommen um neue Cellulasen zu finden, die den Zelluloseverdau verbessern. Die Hefe *Komagataella phaffii* ist ein Mikroorganismus, der in der industriellen Biotechnologie sehr oft zur Produktion von high-value Produkten und Enzymen verwendetet wird.

Diese Arbeit hatte den Schwerpunkt, einen *K. phaffii* Stamm zu erzeugen, der Zellulose verstoffwechseln kann. Die Zellulose-Vorkommen sind sehr reichlich und können als günstige Kohlenstoffquelle dienen, um high-value Produkten und Enzymen mit *K. phaffii* zu erzeugen. Des Weiteren, diese Arbeit hatte auch einen Schwerpunkt Screening-Methoden zu entwickeln um a) *K. phaffii* Mutanten mit verbesserter Sekretion der produzierten Cellulasen zu finden und b) neue Cellulasen zu finden.

Table of Content

Inhalt

Acknowledgments

I want to thank all people who believed in me and supported me during my time as PhD student. Without them, I am sure, it would not have been possible for me to come that far.

I would like to thank my university supervisor, Prof. Anton Glieder for giving me the opportunity to write my PhD thesis in his working group and the chance to do a major part of my PhD thesis research in A*STAR Singapore.

Special thanks are also going to my supervisor in A*STAR Singapore, Dr. Jin Chuan Wu. I want to thank him for giving me the chance to work in his department (Industrial Biotechnology, ICES) on my PhD thesis and for his great supervision during my long stay there.

I would also like to thank A*STAR Singapore and the A*STAR Graduate Academy for giving me an ARAP scholarship. So, I was able to do my research on my PhD thesis in ICES Singapore.

Many thanks are also going to the many colleagues in Singapore and Graz which were a great help to me during my PhD study.
Graz: Andrea, Anna, Astrid, Christian, Clemens, Flo, Julia, Laura, Lukas, Martina, Michi, Thomas, Rui
Singapore: Peiying, Angeline, Chris, Cindy, Crystal, Kimyng, Maggie, Nafang, Songhe, Sufian, Sze Min, Tong Mei, Van, Veeresh, Zhao Hua, Zhibin.
I hope that I have not bothered you too much with my questions.

Especially, I want to thank Lukas with whom I started together the adventure to the unknown when we moved to Singapore at the same time. You have always been a very good friend and there for me when I needed help, in Graz and in Singapore!
I want also to thank my very good friends who are not colleagues but supported me by having always an open door for me when I needed it: Belma, Chamel, Christoph, Johannes, Louis, Markus, Paul and Robert.

Thanks are also going to all my (PhD) student colleagues in ICES. We were there for each other and had also fun during our time in ICES: Agnes, Andrea, Aparna, Azadeh, Chen Ye, Chris, Elmira, Gladys, Ibrahim, Jennifer, Jose, Joshua, Kata, Kumaran, Leonard, Lucy, Maggie, Nana, Parviz, Pei Lin, Poovizhi, Prince, Romen, Ruili, Ulli and Yasmin.

Furthermore, I want to thank my parents and my brother Markus for their fully support during my whole life and especially, when I moved to Singapore. I also want to thank my parents in law for their support during my stay in Singapore and taking care of me.

And finally, I want to thank my wonderful wife Peiying for her help and support during this difficult and challenging time. You were/are always at my side and we went through thick and thin together during my PhD studying time. You cheered me up and made me smile again, every time when I came back home and was down.

Introduction

Cellulose is the most abundant biopolymer in our biosphere. The name "cellulose" was given by the French Academy after the agricultural chemist Anselme Payen had made his observation between 1837 and 1842 that all young plants contain fibrous structures made out of the same chemical substance [Klemm et al. 2005, Suhas et al. 2016]. Cellulose together with lignin and hemicellulose form these fibrous structures, now called lignocellulose. Its cellulose content differs from plant to plant but in general, it is seen as the major cell wall polysaccharide of plant cells. It consists of glucose residues which are interlinked in β 1,4 glycosidic bonds. The cellulose molecules (glucan chains) can be up to 25,000 glucose residues long which are assembled into microfibrils and microfibrils are grouped up to form the fibers. Cellulose fibers contain amorphous and crystalline regions [Juturu and Wu 2014, Suhas et al. 2016].

Its high availability, sustainability and non-toxicity make cellulose to a very interesting material to work with. Naturally, cellulose is part of the agriculture, food and beverage industry but it is also used in many different applications as for example in fibers for textiles, fibers for pulp and paper, or as natural adsorbent [Reddy and Yang 2005; Suhas et al. 2016]. Furthermore, produced cellulose derivatives are used in cosmetics, food, pharmaceutics and surface coatings [Klemm et al. 2005].

In addition to the use as a versatile and biobased, renewable material industrial biotechnology has a strong focus on the enzymatic degradation of cellulose [Dashtban et al. 2009; Klein-Marcuschamer and Blanch 2015]. Cellulose can be degraded by enzymes called cellulases which are able to cleave the β 1,4 glycosidic bonds. In literature, three types of cellulases are seen as key-enzymes for cellulose degradation. These key-enzymes are β-glucosidases (E.C. 3.2.1.21), endo-glucanases (E.C. 3.2.1.4) and exo-glucanases (E.C. 3.2.1.91 and E.C. 3.2.1.176). Endo-glucanases randomly cleave cellulose molecules within the amorphous region creating more ends which increases the efficiency of exo-glucanases [Kostylev and Wilson 2012; Juturu and Wu 2014; Teeri 1997]. Exo-glucanases cleave off cellobiose from the ends of the cellulose polymer. Cellobiose is a disaccharide containing two glucose molecules. There are two different types of exo-glucanases, one cleaving cellobiose molecules from the reducing end of the cellulose (E.C. 3.2.1.176, with CBHI as a prominent representative) and another one from the non-reducing end (E.C. 3.2.1.91, with CBHII as a

prominent representative) of cellulose molecules. β-glucosidases cleave the liberated cellobiose into two glucose molecules. Cellulolytic microorganisms have developed different strategies for cellulose degradation. Cellulolytic fungi secrete all three key enzymes separately whereas anaerobic cellulolytic bacteria produce cellulosomes. Cellulosomes are multi-enzyme complexes containing endo-glucanases, exo-glucanases and other enzymes. In spite of the high specific efficiency of bacterial cellulosomes for industrial application, cellulases are usually used in separate form [Bayer et al. 2007; Mathew et al. 2008].
In the past decade, the involvement of other important enzymes besides the three key-enzymes in cellulose degradation were found. It has been reported that proteins like swollenin or lytic polysaccharide monooxygenases increase the cellulolytic activity of the key cellulases in fungal systems [Jäger et al. 2011; Morgenstern et al. 2014].

Cellulases are used in many different industries like beverage, pulp and paper, detergent, textile and energy industry. It is said that cellulases make about 20% of the total enzyme market and this market was increasing significantly in the past 5-10 years [Gurung et al. 2013; Singh et al. 2016; Srivastava et al. 2015]. Each industry needs cellulases which show activity and stability at specific conditions. These conditions are in general very harsh and vary from industry to industry. There is high potential in reducing the process costs by improving the cellulose degradation step. Therefore, extensive research is done in screening for novel cellulases which show higher activity and stability under these harsh conditions than commonly used cellulases. Furthermore, known cellulases were modified by mutagenesis experiments to improve the cellulose degradation efficiency [Kuhad et al. 2016]. A main challenge in isolation for novel or improved cellulases is the creation of convenient high throughput screening methods. Conventional screening methods are done by cultivation of the clones in 96 deep-well plates and this strongly limits the number of clones that can be tested [Vervoort et al. 2017; Zhang et al. 2006]. Therefore also high throughput screening methods for cellulases were developed, which however require special equipment like FACS or robotic machineries [Ko et al. 2013; Ostafe et al. 2013].

Another approach for improving efficiency of enzymatic cellulose degradation is to isolate novel cellulolytic microorganisms from nature. The aim is to find microorganisms secreting a mix of cellulases which is more efficient in degrading cellulose than common used ones (e.g. the filamentous fungi *Trichoderma reesei*). This approach is mainly relevant for the energy industry [Srivastava et al. 2015; Zhao et al. 2016]. In energy industry, cellulases are needed

for the degradation of pre-treated lignocellulose to glucose. The degradation of cellulose to glucose is called cellulose saccharification. The liberated glucose is then used by microorganisms as carbon source to produce bioethanol. The yeast *Saccharomyces cerevisiae* is mainly used for fermentation of glucose to bioethanol [Klein-Marcuschamer and Blanch 2015; Mood et al. 2013]. *S. cerevisiae* cannot metabolize cellulose. Therefore, there was the approach to engineer it to co-express a β-glucosidase with an endo-glucanase to improve the conversion of cellulose to bio-ethanol. This engineered strain successfully grew on phosphoric acid swollen cellulose (PASC; amorphous cellulose) and could use PASC as carbon source for bioethanol production [Den Haan et al. 2007]. This approach showed that a non-cellulolytic yeast could be engineered to use cellulose as carbon source.

Also pre-treatment of lignocellulosic biomass, in order to make lignocellulose accessible for enzymatic treatments, makes certain progress due to the efforts done by the energy industry [Klein-Marcuschamer and Blanch 2015; Mood et al. 2013]. Based on the current progress, it might be possible that lignocellulosic biomass can be also used as cheap and non-food derived alternative carbon source for non-cellulolytic microorganisms by other industries. Therefore, it might be worth to engineer other non-cellulolytic microorganisms to use cellulose as carbon source for production of enzymes, chemicals and biobased materials.

One of these candidates is the yeast *Komagataella phaffii* (former *Pichia pastoris*). This yeast is broadly used as expression host for heterologous protein expression in industrial biotechnology and research. As eukaryotic expression host, it is possible to do typical eukaryotic posttranslational modifications on proteins which is very often required for functional expression of eukaryotic proteins. *K. phaffii* is a methylotrophic yeast initially developed for single cell protein production by Phillips Petroleum and adapted for heterologous gene expression by James Cregg and colleagues more than 25 years ago. Among the early tools provided to the research and industrial community by Pichia expression kits by Invitrogen (now Life Technologies/Fischer Scientific) are strong inducible promoters and constitutive promoters. Especially the methanol inducible *AOX1* promoter has been extensively used for reaching high yields of heterologous expressed proteins. Another advantage of *K. phaffii* as expression host is that its cultures have a "plain" supernatant. The secretion of endogenous proteins is usually very low in *K. phaffii*. Therefore, the heterologous expressed protein which is secreted to the medium, makes up the vast part of the total protein in the supernatant [Ahmad et al. 2014; Cereghino and Cregg 2000; Vogl et al. 2013b; Vogl et al. 2016].

K. phaffii, however, is not only used for heterologous protein expression to produce enzymes. It is also used as whole cell biocatalyst for production of bio-pharmaceuticals and other high value compounds [Cereghino and Cregg 2000; Geier et al. 2012, 2013; Vogl et al. 2013a; Wriessnegger et al. 2014] and first pharmaceutical proteins had been FDA (US Food and Drug Federation) approved [RCT *Pichia pastoris* Protein Expression Platform].
Since *K. phaffii* is such a versatile yeast, an engineered strain which is able to use abundant cellulose as carbon source might be very interesting for industry. Different key cellulases have already been separately expressed in *K. phaffii*. The heterologous cellulases made by *K. phaffii* were functionally expressed and secreted to the medium [Chen et al. 2011; Mellitzer et al. 2012; Quay et al. 2011]. This indicated the theoretical possibility of engineering a *K. phaffii* strain which co-expresses all three cellulases, to make use of cellulose as a sole carbon source.

As mentioned before, *K. phaffii* is frequently used as expression host for heterologous protein expression in academia and for industrial manufacturing. As in other eukaryotic expression hosts, major bottlenecks were detected in the secretory pathway of *K. phaffii* during heterologous expression of secreted proteins. Extensive research has been done to understand secretory pathway in yeasts and to improve the efficiency to secrete heterologous expressed proteins in *K. phaffii* and other yeasts [Delic et al. 2014; Idiris et al. 2010; Routenberg Love et al. 2012]. It was possible to increase protein secretion in yeast by overexpression or knock-out of certain genes which are involved in secretory pathway [Idiris et al 2010]. There are also approaches to do mutagenesis experiments on yeasts to increase their protein secretion efficiency [Lin-Cereghino et al. 2013; Zheng et al. 2016]. Since conventional screening methods are lacking high throughput, it is a challenge to establish a simple high throughput screening method for screening mutated clones with better protein secretion [Huang et al. 2015; Vervoort et. al.2017].

Based on the state of the art, this thesis was divided into four parts. Part I of the thesis deals about the basic work to discover tools which are required for stable multiple gene expression of different genes in *K. phaffii*. In this part, the discovery and characterization *of K. phaffii* promoters for protein expression is described. The bidirectional promoters which were used in the other parts of the thesis, had been discovered and characterized in Part I by a major contribution from this PhD thesis. The manuscript co-authored with Thomas Vogl was submitted to Nature Communications and published after additional revision. Part II of the

thesis deals with the possibility of engineering *K. phaffii* to metabolize cellobiose and cellulose. This part was submitted and published recently by the Journal Applied Microbiology and Biotechnology in 2018 [Kickenweiz et al. 2018]. Additional experiments which are not part of the published paper (including supplemental data), are described in the Appendix A1 of this thesis. The methods used for these additional experiments are described in Appendix A2, the construction of vectors and sequences which were used in these additional experiments, are described and listed Appendix A3 and A4, respectively. Part III and Part IV of this thesis deal with novel screening methods using *K. phaffii* as platform strain. These screening methods were based on the results of Part II. The clones were screened according to their growth on cellobiose and cellulose (CMC) as sole carbon source. Furthermore in Part III, a new method to screen for potential *K. phaffii* mutants with increased protein secretion is presented. In Part IV, the described selection method might enable screening of cDNA libraries for novel cellulases in future.

References

Ahmad M, Hirz M, Pichler H, Schwab H (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98:5301–5317. doi: 10.1007/s00253-014-5732-5

Bayer EA, Lamed R, Himmel ME (2007) The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol 18: 237–245. doi: https://doi.org/10.1016/j.copbio.2007.04.004

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24: 45-66. doi: https://doi.org/10.1111/j.1574-6976.2000.tb00532.x

Chen P, Fu X, Ng TB, Ye X-Y (2011) Expression of a secretory β-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. Biotechnol Lett 33: 2475–2479. doi: 10.1007/s10529-011-0724-3

Dashtban M, Schraft H, Qin W (2009) Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci 5: 578-595. doi:10.7150/ijbs.5.578Den Haan

Delic M, Göngrich R, Mattanovich D, Gasser B (2014) Engineering of Protein Folding and Secretion - Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins. Antioxid Redox Signal 21: 414–437. doi: 10.1089/ars.2014.5844

Huang M, Bai Y, Sjostrom SL, Hallström BM, Liu Z, Petranovic D, Uhlén M, Joensson HN, Andersson-Svahn H, Nielsen J (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. PNAS 112: E4689–E4696. doi: https://doi.org/10.1073/pnas.1506460112

Geier M, Braun A, Emmerstorfer A, Pichler H, Glieder A (2012) Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes – a comparative study. Biotechnol J 7: 1346–1358. doi: 10.1002/biot.201200187

Geier M, Braun A, Fladischer P, Stepniak P, Rudroff F, Hametner C, Mihovilovic MD, Glieder A (2013) Double site saturation mutagenesis of the human cytochrome P450 2D6 results in regioselective steroid hydroxylation. FEBS J 280: 3094–3108. doi: 10.1111/febs.12270

Gurung N, Ray S, Bose S, Rai V (2013) A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. Biomed Res Int vol. 2013, Article ID 329121, 18 pages. doi: http://dx.doi.org/10.1155/2013/329121

Idiris A, Tohda H, Kumagai H, Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86:403–417. doi: 10.1007/s00253-010-2447-0

Jäger G, Girfoglio M, Dollo F, Rinaldi R, Bongard H, Commandeur U, Fischer R, Spiess AC, Büchs J (2011) How recombinant swollenin from *Kluyveromyces lactis* affects cellulosic substrates and accelerates their hydrolysis Biotechnol Biofuels 4:33. doi: https://doi.org/10.1186/1754-6834-4-33

Juturu V, Wu JC (2014) Microbial cellulases: Engineering, production and applications. Renew Sust Energ Rev 33: 188–203. doi: https://doi.org/10.1016/j.rser.2014.01.077

Klein-Marcuschamer D, Blanch HW (2015) Renewable fuels from biomass: Technical hurdles and economic assessment of biological routes. AIChE J 61: 2689–2701. doi: https://doi.org/10.1002/aic.14755

Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: faszinierendes Biopolymer und nachhaltiger Rohstoff. Angew Chem 117: 3422 – 3458. doi: 10.1002/ange.200460587

Ko K-C, Han Y, Cheong D-E, Choi JH, Song JJ (2013) Strategy for screening metagenomic resources for exocellulase activity using a robotic, high-throughput screening system J Microbiol Methods 94: 311–316. doi: http://dx.doi.org/10.1016/j.mimet.2013.07.010

Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chang JW, Chavan A, Tsai JW, Lin-Cereghino J (2013) The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene 2013 519: 311–317. doi:10.1016/j.gene.2013.01.062

Kickenweiz T, Glieder A, Wu JC (2018) Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase. Appl Microbiol Biotechnol 102: 1297–1306. doi: https://doi.org/10.1007/s00253-017-8656-z

Kostylev M, Wilson D (2012) Synergistic interactions in cellulose hydrolysis. Biofuels. 3: 61-70. doi: http://dx.doi.org/10.4155/bfs.11.150

Kuhad RC, Deswal D, Sharma S, Bhattacharya A, Jain KK, Kaur A, Pletschke BI, Singh A, Karp M (2016) Revisiting cellulase production and redefining current strategies based on major challenges. Renew Sust Energ Rev 55: 249–272. doi: http://dx.doi.org/10.1016/j.rser.2015.10.132

Mathew GM, Sukumaran RK, Singhania RR, Pandey A (2008) Progress in research on fungal cellulases for lignocellulose degradation. J Sci Ind Res 67: 898-907

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

Mood SH, Golfeshan AH, Tabatabaei M, Jouzani GS, Najafi GH, Gholami M, Ardjmand M (2013) Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. Renewable and Sustainable Energy Reviews 27: 77–93. doi: https://doi.org/10.1016/j.rser.2013.06.033

Morgenstern I, Powlowski J, Tsang A (2014) Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. Brief Funct Genomics 13: 471-481. doi: https://doi.org/10.1093/bfgp/elu032

Ostafe R, Prodanovic R, Commandeur U, Fischer R (2013) Flow cytometry-based ultra-high-throughput screening assay for cellulase activity. Anal Biochem 435: 93–98. Doi: http://dx.doi.org/10.1016/j.ab.2012.10.043

Quay DHX, Bakar FDA, Rabu A, Said M, Illias RM, Mahadi NM, Hassan O, Murad AMA (2011) Overexpression, purification and characterization of the *Aspergillus niger* endoglucanase, *EglA*, in *Pichia pastoris*. Afr J Biotechnol 10:2101-2111. doi: 10.5897/AJB10.1046

RCT *Pichia pastoris* Protein Expression Platform (22nd of April 2018). https://pichia.com/science-center/commercialized-products/

Reddy N, Yang Y (2005) Biofibers from agricultural byproducts for industrial applications. Trends Biotechnol 23: 22–27. doi: 10.1016/j.tibtech.2004.11.002

Routenberg Love K, Politano TJ, Panagiotou V, Jiang B, Stadheim TA, Love JC (2012) Systematic Single-Cell Analysis of *Pichia pastoris* Reveals Secretory Capacity Limits Productivity. PLoS ONE 7: e37915. doi:10.1371/journal.pone.0037915

Singh R, Kumar M, Mittal A, Mehta PK (2016) Microbial enzymes: industrial progress in 21st century. 3 Biotech 6: 174. doi: https://doi.org/10.1007/s13205-016-0485-8

Srivastava N, Srivastava M, Mishra PK, Singh P, Ramteke PW (2015) Application of Cellulases in Biofuels Industries: An Overview. J Biofuels Bioenergy 1: 55-63. doi: 10.5958/2454-8618.2015.00007.3

Suhas, Gupta VK, Carrott PJM, Singh R, Chaudhary M, Kushwaha S. (2016) Cellulose: A review as natural, modified and activated carbon adsorbent. Bioresour Technol 216: 1066–1076. doi: https://doi.org/10.1016/j.biortech.2016.05.106

Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol 15: 160-167. doi: https://doi.org/10.1016/S0167-7799(97)01032-9

Vervoort Y, Linares AG, Roncoroni M, Liu C, Steensels J, Verstrepen KJ (2017) High-throughput system-wide engineering and screening for microbial biotechnology. Curr Opin Biotechnol 46: 120-125. doi: https://doi.org/10.1016/j.copbio.2017.02.011

Vogl T, Hartner FS, Glieder A (2013a) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr Opin Biotechnol 24: 1094–1101. doi: https://doi.org/10.1016/j.copbio.2013.02.024

Vogl T, Glieder A (2013b) Regulation of *Pichia pastoris* promoters and its consequences for protein production. New Biotechnol 30: 385-404. doi: https://doi.org/10.1016/j.nbt.2012.11.010

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. **doi:** 10.1021/acssynbio.5b00199

Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, Schürmann M, Mink D, Zellnig G, Schwab H, Pichler H (2014) Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris***.** Metab Eng 24: 18–29. doi: https://doi.org/10.1016/j.ymben.2014.04.001

Zhang YHP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. Biotechnol Adv 24: 452-481. doi: https://doi.org/10.1016/j.biotechadv.2006.03.003

Zhao S, Yan Y-S, He Q-P, Yang L, Yin X, Li C-X, Mao L-C, Liao L-S, Huang J-Q, Xie S-B, Nong Q-D, Zhang Z, Jing L, Xiong Y-R, Duan C-J, Liu J-L, Feng J-X (2016) Comparative genomic, transcriptomic and secretomic profiling of *Penicillium oxalicum* HP7-1 and its cellulase and xylanase hyper-producing mutant EU2106, and identification of two novel regulatory genes of cellulase and xylanase gene expression. Biotechnol Biofuels **9**:203. doi: https://doi.org/10.1186/s13068-016-0616-9

Zheng X, Fang B, Han D, Yang W, Qi F, Chen H, Li S (2016) Improving the Secretory Expression of an α-Galactosidase from *Aspergillus niger* in *Pichia pastoris*. doi: https://doi.org/10.1371/journal.pone.0161529

Part I

Engineered bidirectional promoters enable rapid multi-gene co-expression optimization

Authors

Thomas Vogl[1*~], Thomas Kickenweiz[1*], Julia Pitzer[2], Lukas Sturmberger[2], Astrid Weninger[1], Bradley W. Biggs[3,t], Eva-Maria Köhler[1], Armin Baumschlager[1§], Jasmin Elgin Fischer[1], Patrick Hyden[1], Marlies Wagner[1], Martina Baumann[4,5], Nicole Borth[4,5], Martina Geier[2], Parayil Kumaran Ajikumar[3], Anton Glieder[1#]

*Contributed equally

[#]Corresponding author (a.glieder@tugraz.at)

[1] Institute of Molecular Biotechnology, NAWI Graz, Graz University of Technology, Petersgasse 14, Graz 8010, Austria

[2] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, Graz 8010, Austria

[3] Manus Biosynthesis, 1030 Massachusetts Avenue, Suite 300, Cambridge, MA 02138

[4] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, Vienna 1190, Austria

[5] Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, Vienna 1190, Austria

[~] Current address: Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

[§] Current address: Department of Biosystems Science and Engineering, ETH Zürich, Mattenstrasse 26, 4058 Basel, Switzerland

[t] Current address: Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208

Manuscript was published in Nature Communications **9**, Article number: 3589 (2018)

Comment:

The author of this PhD thesis (Thomas Kickenweiz) and the first author of this manuscript (Thomas Vogl) contributed equally to this manuscript. This manuscript was published in Nature Communications.

The work done by Thomas Kickenweiz in this manuscript was part of a planned Master's Thesis which was stopped for being able to switch the Master's thesis to a PhD Thesis and to use the results of the promoter research as basis for this PhD thesis here. The results gained during the planned Master's Thesis were finally not used for a written Master's Thesis submitted by Thomas Kickenweiz.

The research on promoters to discover suitable tools for *Komagataella phaffii* (formerly: *Pichia pastoris*) cell engineering was essential for being able to create stable strains expressing different genes. The bidirectional histone promoters which are described in this manuscript, were chosen as promoters among the other discovered promoters for expression of the cellulases by *K. phaffii* in this PhD thesis.

Following experiments were done by Thomas Kickenweiz as part of this manuscript:

1) Establishing the screening system for identification/characterization of bidirectional promoters in *Pichia pastoris* by co-expression of eGFP and dTomato.

2) Construction of the first synthetic bidirectional promoter (AOX1+GAP) for/in *Pichia pastoris* and its characterization.

3) Selection of the natural bidirectional promoters in *Pichia pastoris* genome. Furthermore, the isolation and characterization of the discovered natural bidirectional promoters.

4) Bioinformatic approach for deletion studies on DAS1 and DAS2 promoters and making of deletion variants of these promoters. (Comment: These deletion variants were the base for further studies since they were used for fusion experiments to create synthetic promoters in the manuscript)

However, the results of the promoter research were also part of the PhD thesis of Thomas Vogl (title: "Synthetic biology to improve protein expression in *Pichia pastoris*") in form of an older version of the manuscript (title: "A library of bidirectional promoters facilitates fine-tuning of gene coexpression").

The corresponding author (Prof. Anton Glieder) and the first author (Dr. Thomas Vogl) of the manuscript entitled "Engineered bidirectional promoters enable rapid multi-gene co-expression optimization" accepted and confirmed that Thomas Kickenweiz may use this

manuscript for his PhD-thesis by referring to the significant contributions and parts/work he had contributed to this paper.

Abstract

Numerous synthetic biology endeavors require well-tuned co-expression of functional components for success. Classically, monodirectional promoters (MDPs) have been used for such applications, but MDPs are limited in terms of multi-gene co-expression capabilities. Consequently, there is a pressing need for new tools with improved flexibility in terms of genetic circuit design, metabolic pathway assembly, and optimization. Motivated by nature's use of bidirectional promoters (BDPs) as a solution for efficient gene co-expression, we have generated a library of 168 synthetic BDPs in the yeast *Komagataella phaffii* (syn. *Pichia pastoris*), leveraging naturally occurring BDPs as a parts repository. This library of synthetic BDPs allows for rapid screening of diverse expression profiles and ratios to optimize gene co-expression, including for metabolic pathways (taxadiene, β-carotene). The modular design strategies applied for creating the BDP library could be relevant in other eukaryotic hosts, enabling a myriad of metabolic engineering and synthetic biology applications.

Introduction

Efficient and well-tuned co-expression of multiple genes is a common challenge in metabolic engineering and synthetic biology, wherein protein components must be optimized in terms of cumulative expression, expression ratios, and regulation [1–4]. When co-expressing multiple proteins, not only their ratios to each other, but also their total (cumulative) amounts summed together matter. Too excessive loads of heterologous proteins may overburden the cellular machinery of recombinant expression hosts. Hence, in addition to balancing the proteins relative to each other, their total (cumulative) expression strength needs to be adjusted. Else, burdensome overexpression of proteins or accumulation of toxic intermediate metabolites may prove detrimental to the cellular host and undermine engineering goals. One remedy has been to restrict protein overexpression to only certain times through dynamic or regulated transcription ("inducibility") [1]. A second is to balance pathway expression to prevent toxic metabolite accumulation [3,5], mimicking natural pathways' balanced protein stoichiometries [6].

Though effective, these methods' ability to improve pathway performance by controlling gene expression is constrained to the tools available. To date, and especially in the context of eukaryotic microbes, this has primarily been restricted to monodirectional promoters (MDPs), which possess limits in terms of cloning and final pathway construction. Interestingly, nature has encountered similar gene expression challenges, developing its own set of solutions. This includes the use of bidirectional promoters (BDPs) to expand expression flexibility, exemplified by multi-subunit proteins such as histone forming nucleosomes [7].

Natural BDPs (nBDPs) and divergent transcription have been characterized in all model organisms [8,9,18,10–17], with RNAseq studies even indicating that eukaryotic promoters are intrinsically bidirectional [9,10,12,19]. Moreover, nBDPs with non-cryptic expression in both orientations frequently co-regulate functionally related genes [20,21]. Inspired by these circuits, biological engineers have recently utilized BDPs to improve designs for gene co-expression in *Escherichia coli* [22], *Saccharomyces cerevisiae* [23], plants [24], and mammals [25,26]. These studies offer promise, but larger sets of readily available BDPs remain limited, and the reported strategies have lacked generalizability. To our knowledge, *S. cerevisiae's* less than dozen BDPs represent the largest collection [23] and do not provide the desired spectrum of different expression ratios or consecutive induction.

BDPs offer the ability to dramatically improve pathway design, with applicability in numerous and even emerging hosts. In contrast to monodirectional expression cassettes in tandem, bidirectional cloning offers a simple and quick solution to identify optimal promoter contributions for co-expression in a single cloning-expression-screening experiment. But, for BDPs to be fully utilized a much larger set must be engineered, with the ideal library representing different expression levels and regulatory profiles varied per expression direction. Such a library could halve cloning junctions compared to conventional MDPs, facilitating rapid assembly of combinatorial libraries that efficiently explore broad expression landscapes. In addition, development of tools such as these could help to unlock the use of emerging hosts, such as *Pichia pastoris* (syn. *Komagataella phaffii*), which have the potential not only for industrial and pharmaceutical enzyme production, but food and diary protein production and as chemical factories [27].

Here, we have generated a collection of 168 BDPs in the methylotrophic yeast *P. pastoris*, using its natural histone promoters as an engineering template. Our library covers a 79-fold range of cumulative expression, has variable expression ratios ranging from parity to a 61-fold difference between sides, and combines different regulatory profiles per side including the possibility for consecutive induction. The utility of these BDPs was demonstrated through the optimization of

multi-gene co-expression, and the conserved nature of the framework histone promoters suggests the generalizability of this approach for other eukaryotes.

Results and discussion

Expression capabilities and limitations of natural BDPs

Our study began by searching for nBDPs that might satisfy various engineering needs (Fig. 1a), targeting our search to the yeast *P. pastoris*. Long favored as a host for heterologous protein production [28], *P. pastoris* has recently emerged as a promising chassis for metabolic engineering applications owing to its growth to high cell densities and its excellent protein expression capabilities [29]. In addition, its methanol utilization (MUT) pathway represents one of the largest sets of tightly co-regulated genes in nature, offering transcriptional repression via glucose and inducibility via methanol [30], making it an ideal target for BDP mining. Bioinformatics approaches (S 1) identified 1462 putative BDPs in *P. pastoris'* genome (Fig. 1b), with a subset of 40 BDPs selected for detailed characterization due to their expected high expression as housekeeping genes or previous application as MDPs (Fig. 1c, S 2 for a list of the promoters tested).

All putative MUT pathway [30] and housekeeping gene nBDPs were tested to identify potential regulated and constitutive promoters, respectively. Our promoter screening involved green and red fluorescent protein (FP) reporters (Fig. 1c), normalized with respect to their different relative fluorescence units (rfu), which vary, due to their dependence on the specific quantum yields of the FPs and spectrometer settings, to allow direct comparison of the two promoter sides in our experimental setting (S 4). This normalization factor was applied to all promoter measurements reported in this work. Among MUT promoters, only the *DAS1-DAS2* promoter ($P_{DAS1-DAS2}$) showed strong expression on both sides, matching the most frequently used monodirectional *AOX1* (*alcohol oxidase 1*) promoter, concurring with a previous study ([30] and S 5a,b). Other MUT promoters showed only strong monodirectional expression (Fig. 1c). Several putative nBDPs of housekeeping genes showed detectable expression on both sides, but weaker than the classical and most frequently applied monodirectional *GAP (glyceraldehyde-3-phosphate-dehydrogenase)* promoter (P_{GAP}), one of the strongest constitutive promoter in *P. pastoris* [31], which was used as a benchmark (Fig. 1c). Though the majority of nBDPs mined provided limited engineering applicability, the histone promoters (P_{HTX1}, P_{HHX1} and P_{HHX2}) showed promise due to their equally strong expression on both sides, matching (Fig. 1c) the P_{GAP} benchmark during growth on glucose as a carbon source.

Bidirectional histone promoters as useful parts repository

Based on the results from the nBDPs screening (Fig. 1c), we focused subsequent engineering efforts on the three bidirectional histone promoters P_{HTX1}, P_{HHX1} and P_{HHX2}, where *HTX* refers to the bidirectional promoter at the *HTA+HTB* locus and *HHX* represents *HHT-HHF*. These promoters regulate the expression ratios of highly conserved multimeric histone proteins, which are required for packaging DNA into chromatin[7]. They are required to be produced in equimolar amounts in the cell and evolutionary conserved BDPs control these ratios. Note that, *P. pastoris* contains in contrast to *S. cerevisiae* [7] only a single *HTA+HTB* locus (*HTX1*) and two *HHT+HHF* loci (*HHX1*, *HHX2)*.

The function, structure, involvement in gene regulation and modifications of histones have been extensively investigated in several model organisms, with an emphasis on the cell-cycle regulated expression of histone promoters [32,33]. Histone promoter have even been utilized to drive heterologous gene expression in fungi [34,35] and plants [36], but these studies focused solely on monodirectional expression from histone promoters without evaluating their bidirectional potential.

For our studies, because *P. pastoris* reaches higher specific growth rates and biomass on glycerol compared to glucose [37,38], we tested the histone BDPs on both carbons sources. The monodirectional P_{GAP} benchmark performed better on glucose than glycerol [31,39]. However, the histone BPDs performed better on glycerol and even outperformed the P_{GAP} benchmark by up to 1.6-fold (Fig. 2a).

Notably, the bidirectional *P. pastoris* histone promoters condense the regulatory elements needed for strong bidirectional expression compared to monodirectional benchmark promoters (Fig. 2b). This is exemplified in the length of these promoters (365 to 550 bp) compared to the monodirectional P_{GAP} (486 bp) and P_{AOX1} (940 bp). Nonetheless, both sides of the bidirectional promoters reached expression levels comparable to MDPs, reflected by a higher relative expression efficiency (a term defined here as expression strength per promoter length, discussed in greater detailed below).

Noticeably all *P. pastoris* histone promoters contain clear TATA box motifs (Fig. 2b), meaning they are grouped with a class of yeast promoters that rely on TATA-binding protein to initiate transcription instead of alternative factors [40]. TATA box containing promoters are typically tightly regulated and involved with cellular stress response genes [40], including with *P. pastoris* MUT genes [30], whereas TATA-less promoters are typically constitutively active [40]. Hence, the TATA boxes in the histone promoters concur with their tight cell cycle associated expression [7].

Using the TATA boxes as a hallmark for determining the core promoter length, we observed exceptionally short core promoters in all histone BDPs (55-81 bp, compared to 160 bp in case of the well-studied P_{AOX1} [41]). Core promoters are the basic region needed for transcription initiation and bound by general transcription factors (TFs) and RNA polymerase II (RNAPII). It is worth nothing that histone core promoter sequences contain the 5' untranslated region (5'UTRs) of the natural histone mRNAs, as these cannot easily be functionally separated from the core promoter [42,43]. Regardless of this complication, the short core promoters/5' UTRs identified here are desirable tools for promoter engineering as they can be simply provided on PCR primers [41–43]. Concurringly, these short histone core promoters turned out to be an excellent repository of parts for promoter bidirectionalization and the creation of synthetic hybrid promoters.

Creation of BDPs with varied expression strength

Their strong bidirectional expression and short length provided opportunity to use the histone BDPs as a template for mutagenesis strategies [44] to create a library of variants with greater expression flexibility. To expand the expression capabilities of the natural histone BDPs beyond only a fixed ratio and cumulative expression strength, we utilized truncation and deletion strategies of P_{HHX2} (Fig. 2c,d) to construct a synthetic BDP (sBDP) library with diversified expression strengths and ratios (Fig. 3c-d). Interestingly, removing the core promoter from one side of a bidirectional promoter (Fig. 2c,d, S 7) increased monodirectional expression on the other side up to 1.5-fold, hinting a regulatory model in which two core promoters are competing for transcription initiation by general TFs or RNAPII (extended discussion in S 7). The 31 variants generated from *HHX2* histone promoter deletions (Fig. 2c,d) spanned a more than 15-fold range in cumulative expression levels and up to 39-fold expression ratio between sides.

Creation of inducible sBDPs by MDP bidirectionalization

We next sought to introduce inducibility to this library of promoters with varied expression strength and ratios by incorporating design elements from the inducible MUT pathway. As mentioned, MUT promoters such as $P_{DAS1\text{-}DAS2}$ (S 5) showed promise because of their expression capacity (Fig. 1c), but are cumbersome to work with due to size (2488 bp). To solve this, we aimed to generate shorter and more flexible inducible BDPs by bidirectionalizing MDPs, fusing a second core promoter in reverse orientation to an MDP (Fig. 3a). As core promoters in eukaryotes typically

provide little expression on their own, strong expression generally upstream activating sequences (UAS), which are also referred to as enhancers, or *cis*-regulatory modules (CRMs) [45], with the CRM terminology including repressor binding sites (Fig. 3a illustration). Here the previously identified short core promoter/5' UTRs of the histone promoters held utility (Fig. 2b). We hypothesized that adding a short, nonregulated core promoter in reverse orientation upstream of an MDP could duplicate the expression and regulation of the native orientation [24,25].

Accordingly, we fused six histone core promoters to twelve monodirectional *P. pastoris* promoters, partly varying the lengths of the core promoters and the MDPs (Fig. 3a). Two thirds of the 30 constructs were successfully bidirectionalized, showing detectable expression from the second core promoter. In the case of three promoters (P_{AOX1}, P_{FLD1} and P_{DAS2}), bidirectionalized expression greater than 50% of the native monodirectional side was reached. The construct $Pcore_{HTA1\text{-}81}$+$P_{DAS2\text{-}699}$ even outperformed strong MDPs. Different core promoter lengths only moderately affected expression, while MDP length 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). This was perhaps surprising in light of milestone bidirectionalization studies in higher eukaryotes [24,25] where testing only a 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.

Creation of fusion sBDPs with varied regulation

All BDPs to this point possessed the same regulation on both sides. Having varied regulation can allow for expression cascades, which can be beneficial when it is necessary to express one gene before another, such as a chaperone before its protein folding target. We generated fusions of constitutive, derepressed, and inducible MDPs [30], creating 30 fusion sBDPs with distinct regulation on each side (Fig. 3b,c; S 6). These fusions generally maintained each side's original regulation and individual expression levels, allowing for the creation of variably regulated BDPs with a range of expression ratios between sides (0.16 to 0.96). A subset of the fusion promoters (Fig. 3c) consisted of combinations of *DAS1* and *DAS2* deletion variants (S 5) demonstrating that separately engineered MDPs maintain their individual expression levels and can be rationally combined to generate BDPs with desired expression ratios. Some fusion variants showed synergistic effects, such as the 1.8-fold increase in expression for a *GAP-DAS2* fusion promoter. Others showed antagonistic effects, such as the 40% repression of a *HTA1-TAL2* fusion promoter, suggesting a transcriptional 'spillover' between promoters (S 6). These findings contrast previous MDP fusion studies in *S. cerevisiae* [23,46–50], potentially due to the greater number of promoters and combinations tested here. It is known that binding of insulator proteins can decouple regulation of BDPs per side in *S. cerevisiae* [17], and thus the properties of fusion promoters are difficult to predict. These synergistic effects, though, can be harnessed to design shorter, more efficient promoters and so we expanded this principle to the design of hybrid promoters (Fig. 4 and Fig. 5c), ultimately finding it successful.

sBDP library provides unprecedented relative expression efficiency

Through the creation of this sBDP library, it became clear that we had little ability to predict function based on promoter length and core promoter properties alone. To help improve our understanding, we assembled short defined CRMs (30-175 bp, S 5, S 7) with histone core promoters (Fig. 2b) into compact bidirectional hybrid [51] promoters (Fig. 4). The CRMs were selected from methanol regulated promoters based on literature data available on P_{AOX1} ([31], S 5) and deletion studies on P_{DAS1} and P_{DAS2} (S 5). Each CRM was characterized with a single core promoter (S 7b), two core promoters, and combinations of CRMs in different positions and orientations (Fig. 4). To create combinations of regulatory profiles we fused a truncated histone promoter variant ($P_{HHT2\text{-}T3}$, Fig. 2c,d) to a single CRM and one core promoter.

Inducible synthetic hybrid BDPs matched expression from the monodirectional *AOX1* reference promoter (bottom of Fig. 4). However, the generated sBDPs were considerably shorter (179 to 457 bp) than P_{AOX1} (940 bp). To illustrate this length advantage, we characterized their 'relative expression efficiency', which we define as normalized fluorescence per bp in this study. As the expression output depends on the reporter protein, these relative expression efficiencies are dependent upon the fluorescence reporter proteins and even spectrometers used. Hybrid BDPs showed up to 3.3-fold higher relative expression efficiencies than typically used nMDPs and were 2.1-fold more efficient than the most efficient nBDP (Fig. 5c). In addition, sMDP controls were up to 2.4-fold more efficient than nMDPs (S 7). The length of the core promoters and the orientation of the CRMs only marginally affected expression of the hybrid BDPs. Orientation independency in yeast CRMs has long been known [40], and our results demonstrate that this property can also be harnessed to generate strong BDPs.

In summary, the modular design strategies outlined (Fig. 2b-d, Fig. 3, Fig. 4) produced a versatile library of 168 BDPs offering 1.) different regulatory profiles, 2.) providing a 79-fold range of cumulative expression, and 3.) up to 61-fold expression ratio between sides, meeting the intended design requirements for our library (Fig. 5a,b).

The library of BDPs facilitates dual gene co-expression optimization

After developing a cloning strategy to insert the library of BDPs into a cloning junction between genes of interest (S 8), we next aimed to demonstrate the utility of our BDP library for optimizing multi-gene co-expression. First, we optimized dual gene co-expression for production of taxadiene (Fig. 6a), the first committed precursor of the potent anticancer drug Taxol (paclitaxel), which requires expression of geranylgeranyl diphosphate synthase (GGPPS) and taxadiene synthase (TXS) [3]. Second, we evaluated co-expression of a human cytochrome P450s (CYP2D6) and its electron donating NADPH-dependent reductase partner (CPR) using a subset of strong, differently regulated BDPs from the library (Fig. 6b). Third, we evaluated the effect of the chaperone protein-disulfide-isomerase (PDI) on secretion of the disulfide-bond-rich biocatalyst *Candida antarctica* lipase B (CalB, Fig. 6c).

Our results showed that constitutive expression worked only for CalB. Constitutive expression of ER localized CYP2D6/CPR may exert too much stress on the cells, leading possibly stress responses and degradation driving its activity below the limit of detection. For taxadiene production, we noticed an approximately 100-fold decrease in transformation rates when the GGPPS gene was under control of a constitutive promoter, with the few candidate colonies showing no detectable taxadiene production. For the three gene pairs tested (Fig. 6a-c), there was a 5.2 to 50-fold difference in activity/yields of the best and worst performing promoter choice. Most strikingly, for taxadiene production, the worst strain produced only 0.1 mg/L, whereas the best strain (bearing a $P_{GAP+CAT1}$ fusion promoter) reached 6.2 mg/L, in range with engineered *S. cerevisiae* strains (8.7±0.85 mg/L) [52].

We presume that the high yield of this strain is mostly attributable to the use of P_{CAT1} to drive expression of the GGPPS gene, as also the second-best design ($P_{AOX1-CAT1}$) had GGPPS under the control of the same promoter. P_{CAT1} is a derepressed promoter, meaning expression starts once the glucose in the media is depleted, and is further strongly induced by methanol [30]. So, in the best taxadiene producing strains, the GGPPS gene was at first repressed, partially activated in the derepressed phase, and then fully activated on methanol. This demonstrates, in addition to the importance of the ratio and strength of the promoters, that the regulatory profile is critical and can be easily optimized using this versatile library of sBDPs. Tailoring cultivation conditions towards each side of a BDP may further help to optimize yields **[53]**. Worth noting, each application had a different best promoter (GGPPS+TDS: $P_{GAP+CAT1}$, CYP2D6+CPR: $P_{DAS1-DAS2}$, CalB+PDI: $P_{CAT1-AOX1}$) and the obtained

titers/activities did not necessarily correlate with reporter protein fluorescence measured previously for these BDPs (S 9), highlighting gene pair specific effects and the importance of screening a diverse library (Fig. 6a-c). Once optimized expression profiles were known, they could be quickly recreated with MDPs (Fig. 6b,c), demonstrating that even if MDPs should be used for the final design, BDPs can be used to identify optimal expression profiles with faster and simplified cloning techniques, as previously discussed (S 8).

BDPs alongside BDTs simplify multi-gene pathway assembly and fine-tuning

Finally, we wanted to assemble a pathway with greater than two components. In doing so, we quickly found that with increasing numbers of genes, inclusion of bidirectional terminators (BDTs) was necessary. Lack of BDTs in this context results in transcriptional collision as polymerases transcribing opposite DNA strands in convergent orientation stall upon collision [54–56]. We combined selected MDTs, including heterologous *S. cerevisiae* terminators shown to be active in *P. pastoris* [30], into 11 bidirectional fusion terminators by linking them in convergent orientation (Fig. 7). Additionally, natural BDTs (nBDTs) can be used as the *P. pastoris* genome harbors 1461 putative BDTs from genes in tail to tail orientation (Fig. 1b). We included two such short nBDTs from both *P. pastoris* and *S. cerevisiae*.

The bidirectional terminators were cloned, maintaining the natural transition between stop codon and terminator without any additional restriction sites, into a reporter vector containing two fluorescent proteins in convergent orientation (Fig. 7). Complete lack of a termination signal in this context, created by leaving only an 8 bp *Not*I restriction between the reporter genes resulted in an ~8-fold reduced reporter gene fluorescence suggesting that transcriptional collision occurs to similar extents in *P. pastoris* as reported in *S. cerevisiae* [54–56]. Providing either fusion terminators or nBDTs showed clear improvements compared to the no terminator control, restoring 50-90% of reporter protein fluorescence. As in previous work on *P. pastoris* MDTs [30], we also noticed that some BDTs functioned as autonomous replicating sequences (ARS) (S 10), which may lead to increased background growth and strain instability for episomally replicating sequences. We therefore recommend screening new BDTs for ARS function, as fusion terminators behaved in part differently from the originating MDTs (S 10).

With these novel BDTs available, we tested combinations of BDPs (constitutive, inducible, expression ratios) to optimize expression of the four-gene carotenoid pathway for β-carotene synthesis (Fig. 6d). Monodirectional cassettes using P_{AOX1} (inducible) and P_{GAP} (constitutive) were included as reference. The bidirectional constructs showed a 12.1-fold range in β-carotene yields, with the highest β-carotene yield coming from the methanol inducible bidirectional designs (C2/C7, Fig. 6d). This construct surpassed the monodirectional P_{AOX1} design 2-fold and matched the best MDP-based inducible construct previously reported in *P. pastoris* (5.2±0.26 mg/g CDW [cell dry weight]) [30]. Regarding constitutive/growth-associated expression of the pathway, the best bidirectional design based on histone promoters (C11) yielded 14.9-fold higher β-carotene titers than the monodirectional standard P_{GAP} design. This improvement may be explained by the regulation of the promoters used. P_{GAP} is constitutively expressed and constitutive expression of the β-carotene pathway from this promoter may present too great a metabolic burden. Core histone genes, in contrast are cell cycle regulated and typically only activated in the late G1 phase to provide sufficient histones for the newly replicated DNA in the S phase [7]. It appears plausible, that cell cycle associated expression from histone promoters exerted less metabolic burden than entirely constitutive expression from P_{GAP}, leading to their improved function.

Conclusion

Constructing efficiently expressed and well-balanced pathways is paramount for harnessing biology to its full industrial potential. Here, using the natural histone bidirectional promoters of *P. pastoris* as template, we combined multiple engineering strategies, including truncation and MDP bidirectionalization, to develop a library of sBDPs with a broad range of expression levels and ratios and with different regulation profiles. We found that this library not only covers diverse expression profiles, but also is highly efficient in terms of the expression output. Even more, we demonstrated its utility for multi-gene pathway optimization, highlighted by simple optimization experiments for taxadiene and β-carotene production. By screening of our large 168 member library, we identified a subset of highly useful BDPs and compiled a minimal set of 12 BDPs (6 BDPs to be tested in both orientations, Tab. 1 and S 3 for annotated sequence files). These promoters have regulatory diversity, different strengths and ratios. In addition, this subset offers extended diversity if cultivated with different carbon sources (glucose/glycerol, methanol). Screening with this initial set provides a foundation for subsequent fine-tuning.

Generating similar BDP libraries in other organisms will require species specific engineering, especially for obtaining inducible promoters. Methanol inducible promoters are rather unique to *P. pastors* and other methylotrophic yeasts [57], whereas other systems will require species specific promoters such as galactose regulated promoters in *S. cerevisiae* [58]. In higher eukaryotes, where carbon source regulated promoters are scarce, inducible BDPs based on synthetic TFSs [26] could be generated relying on strategies developed for MDPs [59,60].

However, as this library strategy relies on parts from the highly-conserved histone BDP architecture, with homologs in *S. cerevisiae*, *Schizosaccharomyces pombe*, and even Chinese Hamster Ovary cells (manuscript in preparation), we have reason to believe that the promoter engineering and cloning strategies outlined in this work will be generalizable to other eukaryotes. Hence, the use of similar BDP libraries is likely to expand to many hosts, and allow for efficient and rapid pathway optimization, expanding the possibilities of synthetic biology and metabolic engineering.

Author Contributions

T.V. and T.K. contributed equally to this work. T.V. and T.K. selected the nBDPs. T.V. discovered the histone promoters and designed all sBDPs. T.K., L.S., B.A., E-M.K., P.H., M.B. and T.V. performed the promoter experiments. A.G. recognized the need for an innovative co-expression strategy. T.V. selected the nBDTs and designed the sBDTs. E-M.K. performed the terminator experiments. The applications of the BDP library for dual gene co-expression were designed by T.V. and performed by J.E.F. and B.W.B. (taxadiene), A.W. (CYP2D6) and T.V. (PDI co-expression). A.G. selected the expression targets. M.G. and T.V. designed the pathway experiment. J.P., M.W. and L.S. performed the pathway experiment. A.G. and T.V. conceived of the study. P.K.A. conceived of the taxadiene experiment. T.V., B.W.B, P.K.A. and A.G. wrote the manuscript. A.G., M.G., P.K.A. and N.B. supervised the research. All authors read and approved the final version of the manuscript.

Acknowledgements

Individual parts of this study received funding by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 289646 (Kyrobio) and grant agreement no. 266025 (Bionexgen) and from the Innovative Medicines Initiative Joint Undertaking project CHEM21 under grant agreement n°115360, resources of which are composed of financial contribution from the European Union´s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies´ in kind contribution. In addition, it has been supported by the Austrian BMWFW, BMVIT, SFG, Standortagentur Tirol, Government of Lower Austria and ZIT through the Austrian FFG-COMET-Funding Program. T.V. was supported by Austrian Science Fund (FWF) project no. W901 (DK “Molecular Enzymology” Graz) while performing this research. The authors gratefully acknowledge support from NAWI Graz. We would like to thank Clemens Farnleitner and Alexander Korsunsky for excellent technical assistance.

Conflict of interest

T.V., T.K., L.S. and A.G. are inventors on a patent application entitled "Bidirectional promoter" (EP2862933). T.V., A.G. and P.K.A. have filed a patent application entitled “Production of terpenes and terpenoids”.

References

1. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–32 (2013).
2. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349,** 1095–100 (2015).
3. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330,** 70–4 (2010).
4. Tan, S. Z. & Prather, K. L. Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Chem. Biol.* **41,** 28–35 (2017).
5. Xu, P. Production of chemicals using dynamic control of metabolic fluxes. *Curr. Opin. Biotechnol.* **53,** 12–19 (2018).
6. Lalanne, J., Taggart, J. C., Guo, M. S., Schieler, A. & Li, G. Evolutionary Convergence of Pathway-specific Enzyme Expression Stoichiometry. *Cell* **In Press,** 1–13 (2018).
7. Eriksson, P. R., Ganguli, D., Nagarajavel, V. & Clark, D. J. Regulation of histone gene expression in budding yeast. *Genetics* **191,** 7–20 (2012).
8. Wei, W., Pelechano, V., Järvelin, A. I. & Steinmetz, L. M. Functional consequences of bidirectional promoters. *Trends Genet.* **27,** 267–76 (2011).
9. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–7 (2009).
10. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457,** 1038–42 (2009).
11. Park, D., Morris, A. R., Battenhouse, A. & Iyer, V. R. 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.* **42,** 3736–49 (2014).
12. Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14,** 880–93 (2013).
13. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499,** 360–3 (2013).
14. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–7 (2010).
15. Chen, Y. *et al.* Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.* **48,** 984–94 (2016).
16. Masulis, I. S., Babaeva, Z. S., Chernyshov, S. V. & Ozoline, O. N. Visualizing the activity of Escherichia coli divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. *Sci. Rep.* **5,** 11449 (2015).
17. Yan, C., Zhang, D., Raygoza Garay, J. A., Mwangi, M. M. & Bai, L. Decoupling of divergent gene regulation by sequence-specific DNA binding factors. *Nucleic Acids Res.* 1–12 (2015). doi:10.1093/nar/gkv618
18. Mostovoy, Y., Thiemicke, A., Hsu, T. Y. & Brem, R. B. The Role of Transcription Factors at Antisense-Expressing Gene Pairs in Yeast. *Genome Biol. Evol.* **8,** 1748–61 (2016).
19. Jin, Y., Eser, U., Struhl, K. & Churchman, L. S. The Ground State and Evolution of Promoter Region Directionality. *Cell* 1–10 (2017). doi:10.1016/j.cell.2017.07.006

20. Adachi, N. and Lieber, M. R. Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome. *Cell* **109,** 807–809 (2002).

21. Trinklein, N. D. *et al.* An abundance of bidirectional promoters in the human genome. *Genome Res.* **14,** 62–6 (2004).

22. Yang, S., Sleight, S. C. & Sauro, H. M. Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits. *Nucleic Acids Res.* 1–7 (2012). doi:10.1093/nar/gks972

23. Öztürk, S., Ergün, B. G. & Çalık, P. Double promoter expression systems for recombinant protein production by industrial microorganisms. *Appl. Microbiol. Biotechnol.* **101,** 7459–7475 (2017).

24. Xie, M., He, Y. & Gan, S. Bidirectionalization of polar promoters in plants. *Nat. Biotechnol.* **19,** 677–9 (2001).

25. Amendola, M., Venneri, M. A., Biffi, A., Vigna, E. & Naldini, L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nat. Biotechnol.* **23,** 108–16 (2005).

26. Fux, C. & Fussenegger, M. Bidirectional expression units enable streptogramin-adjustable gene expression in mammalian cells. *Biotechnol. Bioeng.* **83,** 618–25 (2003).

27. Wagner, J. M. & Alper, H. S. Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. *Fungal Genet. Biol.* **89,** 126–136 (2016).

28. Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **98,** 5301–17 (2014).

29. Schwarzhans, J., Luttermann, T., Geier, M., Kalinowski, J. & Friehs, K. Towards systems metabolic engineering in Pichia pastoris. *Biotechnol. Adv.* **35,** 681–710 (2017).

30. Vogl, T. *et al.* A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. *ACS Synth. Biol.* **5,** 172–86 (2016).

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

32. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9,** 843–54 (2008).

33. Dominski, Z. & Marzluff, W. F. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* **396,** 373–90 (2007).

34. Mackenzie, D. a, Wongwathanarat, P., Carter, a T. & Archer, D. B. 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.* **66,** 4655–61 (2000).

35. Belshaw, N. J., Haigh, N. P., Fish, N. M., Archer, D. B. & Alcocer, M. J. C. Use of a histone H4 promoter to drive the expression of homologous and heterologous proteins by Penicillium funiculosum. *Appl. Microbiol. Biotechnol.* **60,** 455–60 (2002).

36. Kelemen, Z. *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.* **11,** 69–72 (2002).

37. Bawa, Z. *et al.* Functional recombinant protein is present in the pre-induction phases of. *Microb. Cell Fact.* **13,** 127 (2014).

38. Inan, M. & Meagher, M. M. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. *J. Biosci. Bioeng.* **92,** 585–9 (2001).

39. Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V & Cregg, J. M. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186,** 37–44 (1997).

40. Hahn, S. & Young, E. T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189,** 705–36 (2011).

41. Portela, R. M. C., Vogl, T., Ebner, K., Oliveira, R. & Glieder, A. Pichia pastoris Alcohol Oxidase 1 (AOX1) Core Promoter Engineering by High Resolution Systematic Mutagenesis. *Biotechnol. J.* **13,** e1700340 (2018).

42. Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T. & Glieder, A. Synthetic Core Promoters for Pichia pastoris. *ACS Synth. Biol.* **3,** 188–91 (2014).

43. Portela, R. M. C. *et al.* Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species. *ACS Synth. Biol.* **6,** 471–484 (2017).

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

45. Lelli, K. M., Slattery, M. & Mann, R. S. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.* **46,** 43–68 (2012).

46. Miller, C. A., Martinat, M. A. & Hyman, L. E. Assessment of aryl hydrocarbon receptor complex interactions using pBEVY plasmids: expressionvectors with bi-directional promoters for use in Saccharomyces cerevisiae. *Nucleic Acids Res.* **26,** 3577–83 (1998).

47. Vickers, C. E., Bydder, S. F., Zhou, Y. & Nielsen, L. K. Dual gene expression cassette vectors with antibiotic selection markers for engineering in Saccharomyces cerevisiae. *Microb. Cell Fact.* **12,** 96 (2013).

48. Da Silva, N. A. & Srikrishnan, S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. *FEMS Yeast Res.* **12,** 197–214 (2012).

49. Li, A. *et al.* Construction and characterization of bidirectional expression vectors in Saccharomyces cerevisiae. *FEMS Yeast Res.* **8,** 6–9 (2008).

50. Partow, S., Siewers, V., Bjørn, S., Nielsen, J. & Maury, J. Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae. *Yeast* **27,** 955–64 (2010).

51. Blazeck, J., Liu, L., Redden, H. & Alper, H. Tuning gene expression in Yarrowia lipolytica by a hybrid promoter approach. *Appl. Environ. Microbiol.* **77,** 7905–14 (2011).

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

53. Rajamanickam, V., Metzger, K., Schmid, C. & Spadiut, O. A novel bi-directional promoter system allows tunable recombinant protein production in Pichia pastoris. *Microb. Cell Fact.* **16,** 152 (2017).

54. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8796–801 (2002).

55. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* **48,** 365–74 (2012).

56. Uwimana, N., Collin, P., Jeronimo, C., Haibe-Kains, B. & Robert, F. Bidirectional terminators in Saccharomyces cerevisiae prevent cryptic transcription from invading neighboring genes. *Nucleic Acids Res.* 1–10 (2017). doi:10.1093/nar/gkx242

57. Yurimoto, H., Oku, M. & Sakai, Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int. J. Microbiol.* **2011,** 101298 (2011).

58. Weinhandl, K., Winkler, M., Glieder, A. & Camattari, A. Carbon source dependent promoters in yeasts. *Microb. Cell Fact.* **13,** 5 (2014).

59. Khalil, A. S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150,** 647–58 (2012).

60. Nevozhay, D., Zal, T. & Balázsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* **4,** 1451 (2013).

61. Geier, M., Fauland, P., Vogl, T. & Glieder, A. Compact multi-enzyme pathways in P. pastoris. *Chem. Commun.* **51,** 1643–1646 (2015).

62. Näätsaari, L. *et al.* Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* **7,** e39720 (2012).

63. Krainer, F. W. *et al.* Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. *Microb Cell Fact* **11,** 22 (2012).

64. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22,** 1567–72 (2004).

65. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6,** 343–5 (2009).

66. Vogl, T., Ahmad, M., Krainer, F. W., Schwab, H. & Glieder, A. Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in Pichia pastoris. *Microb. Cell Fact.* **14,** 103 (2015).

67. Geier, M. *et al.* Double site saturation mutagenesis of the human cytochrome P450 2D6 results in regioselective steroid hydroxylation. *FEBS J.* **280,** 3094–108 (2013).

68. Abad, S. *et al.* Real-time PCR-based determination of gene copy numbers in Pichia pastoris. *Biotechnol. J.* **5,** 413–20 (2010).

69. Weis, R. *et al.* Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. *FEMS Yeast Res.* **5,** 179–89 (2004).

70. Lin-Cereghino, J. *et al.* Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques* **38,** 44, 46, 48 (2005).

71. Vogl, T., Gebbie, L., Palfreyman, R. W. & Speight, R. Effect of Plasmid Design and Type of Integration Event on Recombinant Protein Expression in Pichia pastoris. *Appl. Environ. Microbiol.* **84,** AEM.02712-17 (2018).

72. Schwarzhans, J.-P. *et al.* Integration event induced changes in recombinant protein productivity in Pichia pastoris discovered by whole genome sequencing and derived vector optimization. *Microb. Cell Fact.* **15,** 84 (2016).

73. Schwarzhans, J.-P. *et al.* Non-canonical integration events in *Pichia pastoris* encountered during standard transformation analysed with genome sequencing. *Sci. Rep.* **6,** 38952 (2016).

74. Gudiminchi, R. K., Geier, M., Glieder, A. & Camattari, A. Screening for cytochrome P450 expression in Pichia pastoris whole cells by P450-carbon monoxide complex determination. *Biotechnol. J.* **8,** 146–52 (2013).

75. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116,** 699–709 (2004).

Materials and methods

Strains and vector construction

Promoter reporter vectors

The *P. pastoris* CBS7435 wildtype strain was used for most experiments. The control strain expressing the four genes of the carotenoid pathway under control of four *AOX1* promoters was available from previous studies by Geier *et al*. [61]. For CalB expression, mut^S strains [62] were used, as higher productivity on methanol has been reported [63].

Details on the promoters and terminators used in this study (including primers for amplification) and the list of primers for generating the reporter vectors and applications (pathway assembly etc.) are provided in S 2. A subset of annotated sequences of a minimal set of BDPs covering broad regulatory profiles for dual gene expression optimization is provided in supporting file S 3 in GenBank format (and summarized in Tab. 1 in the main manuscript.

For basic characterizations, a pPpT4_S [62] based expression vector (Zeocin selection marker) bearing a single *eGFP* reporter gene previously reported was used (pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer [30]). This vector contains integration sequences near the *ARG4* locus and was linearized with *Swa*I to target insertion near the *ARG4* locus, as applied in previous promoter characterizations in *P. pastoris* [30,41,43]. Also the following vectors described below were based on this vector backbone. 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 [64]) between the targeting sequence and the stuffer fragment of pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer. The vector was assembled by digesting the single reporter vector with *Asc*I and *Avr*II. Subsequently the dTomato fused to a *P. pastoris* transcription terminator sequence was PCR amplified from a *P. pastoris* cloning vector using primers TomatoAscIBmrIFWD and AOXTTSbfIAvrIIREV1. To add an additional *Sbf*I restriction site, the obtained PCR fragment was used as template for a second PCR using primers TomatoAscIBmrIFWD and AOXTTSbfIAvrIIREV2. The newly inserted part was confirmed by Sanger sequencing. This vector was named pPpT4mutZeoMlyI-intArg4-bidi-dTOM-eGFP-BmrIstuffer. Subsequently we cloned several natural bidirectional promoters and semi synthetic fusion promoters into this vector (primers provided in S 2). The promoters were either inserted in random orientation by TA cloning or directional by Gibson assembly [65].

Bidirectional terminators reporter vector and cloning of BDTs

The reporter vector for bidirectional terminators (BDTs) contained two convergent expression cassettes each consisting of an *AOX1* promoter and an *eGFP* or *dTOM* reporter gene respectively (see illustration in Fig. 7). The 3' ends of the reporter genes are separated by a stuffer fragment that can be replaced with a BDT. The reporter vector was assembled by digesting a monodirectional control vector containing an *AOX1* promoter upstream of *eGFP* (pPpT4mutZeoMlyI-intArg4-EGFP-AOX1BglII) [30] with *Not*I and *Bam*HI. Subsequently the *AOX1* promoter fused to the dTomato gene was amplified using primers stuffer-dTom-Gib and pILV5-pAOX1-Gib (from the P_{AOX1}-dTOM side of a bidirectional vector used in this study). The stuffer fragment was amplified using primers eGFP-stuffer-Gib and dTom-stuffer-Gib from the vector pPpT4mutZeoMlyI-intARG4-eGFP-BmrIstuffer as template [30]. The primers replaced the *Bmr*I sites with *Not*I sites, as the P_{AOX1} contains a *Bmr*I site and removal of the stuffer fragment using *Bmr*I would also impair the rest of the backbone. The vector backbone and the two PCR products were combined in a Gibson assembly reaction and verified by Sanger sequencing.

For cloning of the BDTs, the reporter vector was digested with *Not*I and the backbone was gel purified. The BDTs were amplified with overhangs to the 3' ends of the reporter genes (using the primers listed in S 3) and cloned by Gibson assembly into the vector. Note that for the bidirectional fusion terminators each monodirectional terminator was amplified separately with an overhang to the other one. In this case the terminators were fused in the Gibson assembly reaction by adding three fragments (vector backbone and two PCRs of the two monodirectional terminators). The inserted terminators were sequenced using primers seqEGFP-520..543-fwd and seqTomato-517..540-fwd.

Cloning vector for dual or multi gene co-expression

The aforementioned bidirectional reporter pPpT4mutZeoMlyI-intArg4-bidi-dTOM-eGFP-BmrIstuffer vector can also be used as entry vector for the co-expression of any gene pair. Therefore, a cassette consisting of the two genes to be co-expressed 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 (general concept outlined in S 8). Alternatively, also Gibson assembly can be used. This vector contains *AOX1* terminators on both sides, hence directional cloning (even by Gibson assembly) is not possible.

To facilitate the generation of entry vectors for oriented cloning of two or more genes, we generated a cloning vector, which provides two different MDTs (T_{AOX1} and T_{DAS1}) in opposite orientation separated by a *Not*I restriction site. If two genes (dual gene co-expression) or a multiple genes (multi gene co-expression) should be co-expressed, this vector can be used for insertion. We prepared two different cloning vectors: pPpT4_S-DAS1TT-NotI-AOX1TT and pPpT4mutZeoMlyI-intArg4-DAS1TT-NotI-AOX1TT. The former is based on a the pPpT4_S vector reported by Näätsaari *et al*. [62]: Following *Not*I and *Swa*I digestion and purification of the backbone a PCR product of the T_{DAS1} bearing overhangs to the vector (primers: P_AOX1_Syn-SwaI-DAS1TT-3prime-Gib and AOX1TT-5prime-NotI-DAS1TT-5prime-Gib) was cloned by Gibson assembly and subsequently confirmed by sequencing. The latter vector contained in addition a sequence to target specific genomic integration (intArg4) and a mutated *Mly*I site in the Zeocin resistance gene (silent mutation [66]). This vector was generated by digesting the aforementioned pPpT4mutZeoMlyI-intArg4-bidi-dTOM-eGFP-BmrIstuffer with *Sbf*I and *Not*I and inserting a PCR product containing the respective overhangs (primers: intARG4-SbfI-DAS1TT-3prime-Gib and AOX1TT-5prime-NotI-DAS1TT-5prime-Gib) by Gibson assembly. Again, the vector was confirmed by sequencing.

Cloning different BDPs for CYP+CPR, CalB+PDI, GGPPS+TDS expression

Our screening strategy for the optimal BDP for a certain gene pair (S 8a-c) requires an entry vector containing the two co-expressed genes in which the promoter can be easily exchanged. A stuffer fragment in this entry vector is subsequently cut out by *Bmr*I digestion and replaced with BDPs. Note that the genes to be co-expressed must not contain *Bmr*I sites.

The vector for taxadiene co-expression was generated by ordering *P. pastoris* codon optimized GGPPS and TDS genes. The genes were ordered as synthetic double stranded fragments (gBlocks by Integrated DNA Technologies) with overhangs for Gibson assembly (gBlock-GGPPS_optTV-AOX1TT-Gib, gBlock-TDS_optTV-Part1 and gBlock-TDS_optTV-Part2-DAS1TT-Gib). A stuffer fragment with complementary overhangs was amplified using primers TDS-BmrI-stuffer-Gib and GGPPS-BmrI-stuffer-Gib. The four fragments were mixed in equimolar ratios with the *Not*I digested pPpT4mutZeoMlyI-intArg4-DAS1TT-NotI-AOX1TT backbone and joined by Gibson assembly. The entire inserted cassette was sequenced. This vector was named pPpT4mutZeoMlyI-intArg4-DAS1TT-AOX1TT-TDS_optTV-GGPPS_optTV-BmrIstuffer.

After removal of the stuffer fragment by *Bmr*I digestion and gel purification, a set of the respective differently regulated promoters was amplified, cloned into the entry vectors and verified by sequencing. See S 2 for the exact primers and overhangs used.
In a similar way, entry vectors for CYP2D6/CPR co-expression and CalB/PDI co-expression were generated. The coding sequences were available from previous studies (CYP2D6/CPR [67], CalB [30], PDI [68]). See also S 2 for the exact primers and overhangs used. For CYP2D6/CPR the monodirectional control strain (Fig. 6b) containing a single copy of a vector with each gene under control of an *AOX1* promoter was available from previous work and was generated by cloning each gene into pPpT4 and pKan vectors [62] via *Eco*RI and *Not*I sites and after the transformation a transformant with a single copy of each plasmid was selected. The monodirectional CalB/PDI control constructs shown in Fig. 6c were generated by cloning the respective promoters into the same pPpT4 vector (using the standard *AOX1TT*).

Assembly of multi gene cassettes for expression of the carotenoid pathway

Constructs with different bidirectional promoters and terminators were designed for the expression of the carotenoid pathway (four genes *Crt*E, *Crt*B, *Crt*I and *Crt*Y) in *P. pastoris* and are shown in Fig. 6d. The exact promoters, terminators and primers for amplification are provided in S 2. The bidirectional promoters and terminators were selected based on their length, function and sequence characteristics. Combinations of promoters of different strength and regulation were tested (inducible, constitutive, constitutive + inducible). Also a construct with switched positions of the BDPs was created to evaluate the effect of positioning the promoter between the first two or the last two genes.
The vector backbone pPpT4_S-DAS1TT-NotI-AOX1TT containing two monodirectional terminators T_{AOX1} and T_{DAS1} in opposite orientation with a *Not*I restriction site in between was used for insertion of the pathway. The genes, bidirectional promoters and terminators were amplified by PCR, using the primers listed in S 2. The primers for the amplification of the promoter and terminator sequences contained overhangs to the carotenoid genes. The fragments were linked by Gibson assembly. In order to increase the efficiency of the Gibson assembly, the number of fragments, which have to be combined, was reduced by a pre-assembling step via overlap extension PCR (oePCR). After combining the carotenoid genes with the adjacent promoter or terminator, the preassembled fragments were connected by Gibson assembly and used to transform *E. coli*. Plasmid DNA was isolated from transformants and the sequences were verified by sequencing.

Assays, screening and cultivation conditions for *P. pastoris* experiments

The *P. pastoris* cultivations were performed using a high throughput small scale 96 deep well plate (DWP) cultivation protocol as previously described [69]. Briefly, wells containing 250 µL BMD1 (buffered minimal dextrose medium, as reported [69]) 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. Cells were induced with 250 µL of buffered media with 1% methanol (BMM2) after 60 h of growth on glucose. After 12 h, 24 h up to 48 h 50 µL of BMM10 (with 5% methanol) was added for further induction. *P. pastoris* cells were transformed with molar equivalents to 1 µg of the empty pPpT4_S vector *Swa*I linearized plasmids [70] (1 µg of the empty pPpT4_S vector was found to yield predominantly single copy integration [42,71]. Some of the vectors used in this study are however considerably large than the empty pPpT4_S vector [*e.g.* the carotenoid pathway constructs], hence in these cases we increased the DNA amounts to have an equivalent number of vector molecules compared to the empty pPpT4_S vector). The following antibiotic concentrations were used: *E. coli*: LB-medium containing 25 µg/ml Zeocin; *P. pastoris*: 100 µg/ml Zeocin. The screening and rescreening procedures to compare single *P. pastoris* strains have previously been reported [30,42]. In brief, for each construct 42 transformants (approximately half a deep well plate) were screened to avoid clonal variation observed in *P. pastoris* [71–73]. Three representative clones from the middle of the obtained

expression landscape (to avoid outliers of multi copy integration or reduced expression because of deletions [71] or undesired integration events [72,73]) were streaked for single colonies and rescreened in biological triplicates. Finally, one representative clone was selected and a final screening of all the variants together was performed.

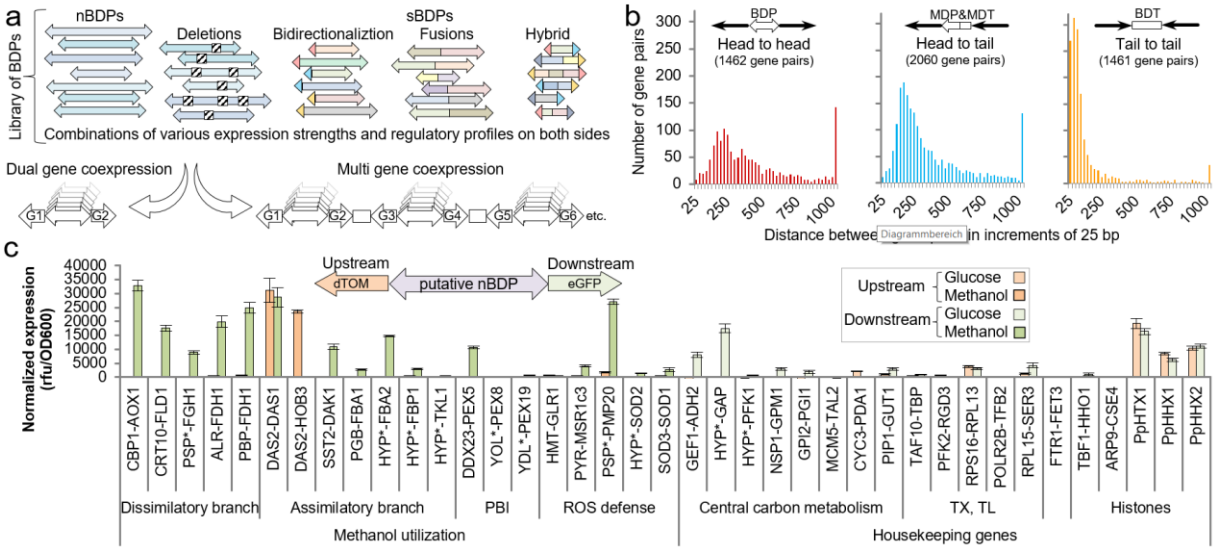
The fluorescence reporter measurements were performed using the same equipment and procedures as reported for eGfp measurements (excitation/emission wavelength: 488/507 nm) alone [30,42] but with adding a second measurement for dTomato (excitation/emission wavelength: 554/581 nm [64]). CalB activities in the supernatants were determined using p-nitrophenyl butyrate (pNPB) as substrate as previously reported [30,63]. CYP2D6 activity measurements were performed as outlined previously using 7-methoxy-4-(amino - methyl)-coumarin (MAMC) as substrate [74]. β-carotene producing strains were cultivated in shake flasks and titers determined by HPLC as described by [30]. Taxadiene producing strains were cultivated in shake flasks in BYPG media (100 mM potassium phosphate buffer pH 6.0, 1% yeast extract, 2% peptone, 1% (w/v) glycerol) with a 10% dodecane overlay and induced with methanol (final concentration of 0.5% (v/v)). Taxadiene titers were determined by GC-MS as outlined previously [3].

Data availability

All sequence data related to the *P. pastoris* promoters/terminators used in this study are available in the EMBL-EBI database (accession numbers FR839628 to FR839632) and the gene names and promoter/terminator positions are provided in S 2.

Figures

Fig. 1



<u>Fig. 1:</u> A library of bidirectional promoters (BDPs) for gene co-expression fine-tuning (a). Bidirectional histone promoters are amongst the few strong *P. pastoris* nBDPs (b,c).

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 co-expression and the assembly of multi gene co-expression cassettes (S 8).

B) The *P. pastoris* genome harbors 1462 putative nBDPs (gene pairs in divergent head to head orientation, <u>S 1</u>). 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 (MDT) 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 methanol inducible *P. pastoris* promoter [30] 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 after growth 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 <u>S 4</u>). Gene names denoted with an asterisk (*) were shortened and are provided in <u>S 2</u>. Mean values and standard deviations of biological quadruplicates are shown. PBI: peroxisome biogenesis and import; ROS: reactive oxygen species; TX,TL: transcription, translation.

Fig. 2

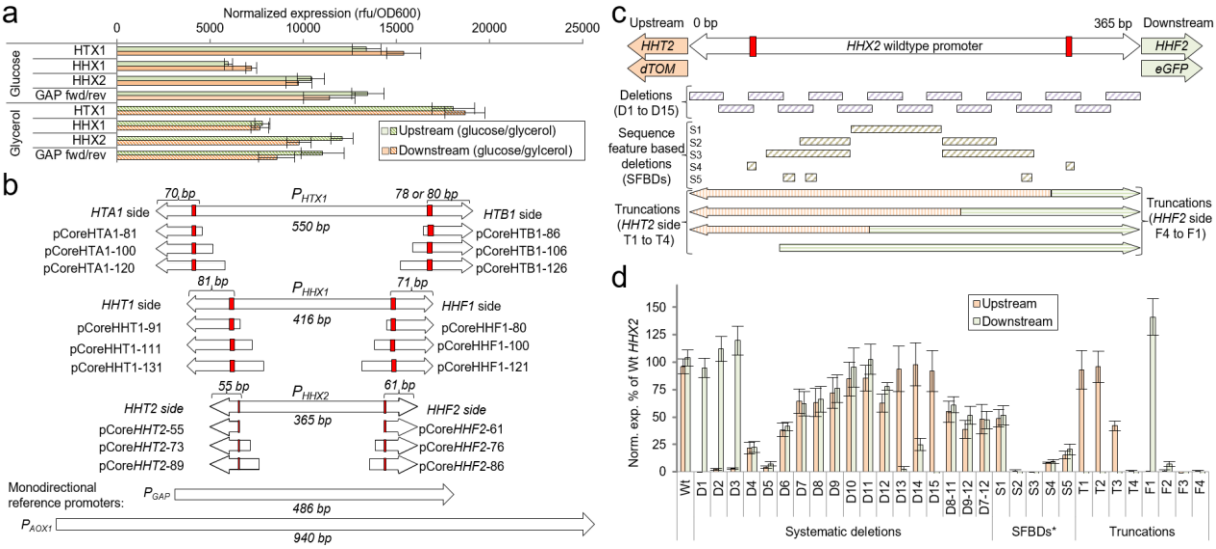


Fig. 2: Natural bidirectional histone BDPs as promoter engineering framework in *P. pastoris.* The *HTX1*, *HHX1* and *HHX2* promoters match (on glucose) or even exceed (on glycerol) the monodirectional P_{GAP} promoter (a). These three histone promoters are short compared to conventional MDPs (b) and therefore easily amenable to the generation of deletion and truncation variants as exemplified with the *HHX2* promoter (c, d).

A) Reporter protein fluorescence of the bidirectional *HTX1*, *HHX1* and *HHX2* promoters in comparison to the strong, monodirectional *GAP* reference promoter in *P. pastoris*. Cells were grown for 60 h on 1% (w/v) glucose or glycerol in 96-well plates. P_{GAP} was cloned in forward (fwd) and reverse (rev) orientation and is hence not bidirectional. 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. In panels (a) and (d) of this figure mean values and standard deviations of normalized (using the normalization factor calculated in S 4) reporter protein fluorescence measurements of biological quadruplicates grown on the respective carbon sources are shown.

B) Bidirectional histone promoters are short compared to the commonly used monodirectional *GAP* and *AOX1* promoters (all elements are drawn in the same scale). The histone promoters contain TATA boxes (red rectangle highlighting the yeast TATA box consensus sequence TATAWAWR [75]) and feature exceptionally short core promoters (pCore… & lengths indicated) useful as parts repository for promoter engineering (Fig. 3a, Fig. 4).

C,D Systematic deletions and truncations of the *P. pastoris HHX2* promoter offer shortened variants with altered cumulative expression levels and ratios. In panel (c) a schematic on the sequence variants is shown (S 2 for exact positions). TATA boxes are denoted by red rectangles. In panel (d) expression levels after growth for 60 h on glucose are shown. *SFBDs: sequence feature based deletions (*i.e.* AT/GC rich regions and TATA boxes).

Fig. 3

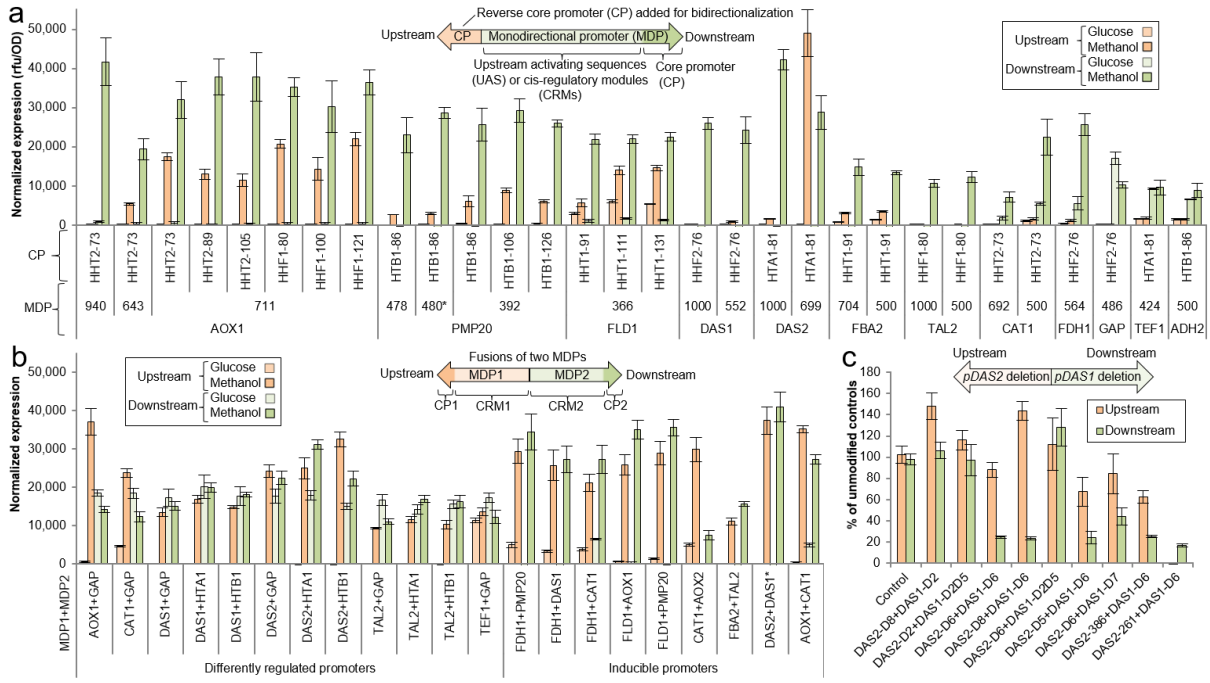


Fig. 3: Modular design strategies of synthetic bidirectional promoters (sBDPs) in *P. pastoris*: Bidirectionalization (a) and fusions of MDPs (b,c) yield sBDPs extending the repertoire of ratios and regulatory profiles.

A) Bidirectionalization of MDPs by addition of core promoters (Fig. 2b) yielded functional BDPs in most cases, but few designs gave high expression. The core promoters (CPs) indicated were fused to the indicated MDPs. The length of the MDPs is given in bp, selection criteria are outlined in S 6. *: In case of the *PMP20* promoter slightly varying sequences from the CBS7435 and the GS115 strain were tested (S 6). Strains were grown on glucose media for 60 h and subsequently induced with methanol for 48 h. In all panels of this figure mean values and standard deviations of normalized (using the normalization factor calculated in S 4) 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 co-expression of up to 10 genes without reusing any sequence (see S 6 for details on the regulatory profiles of the MDPs used). In case of P_{HTA1} and P_{HTB1} the truncated versions shown in Fig. 2b and S 7 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 ($P_{DAS2\text{-}1000}$+$P_{DAS1\text{-}1000}$).

The bidirectionalized and fusion BDPs maintained the regulatory modes of the respective MDPs[30,31]: methanol inducible and tightly glucose/glycerol repressed (P_{AOX1}, P_{PMP20}, $P_{DAS1/2}$ [and deletion variants thereof], P_{FBA2}, P_{TAL2}, P_{AOX2}), derepressed and methanol inducible (P_{CAT1}, P_{FLD1}, P_{FDH1}) and constitutive (P_{GAP}, P_{TEF1}, P_{ADH2}, $P_{HTX1\ [HTA1\text{-}HTB1]}$).

Fig. 4

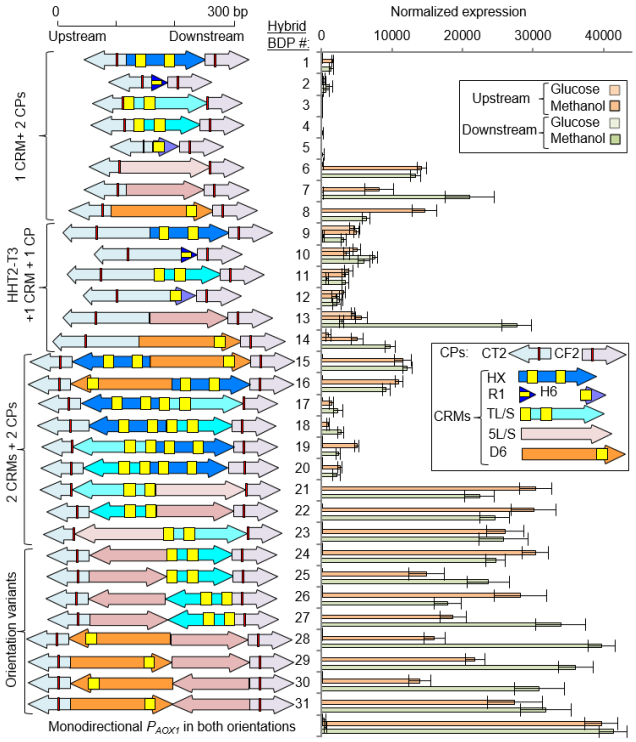
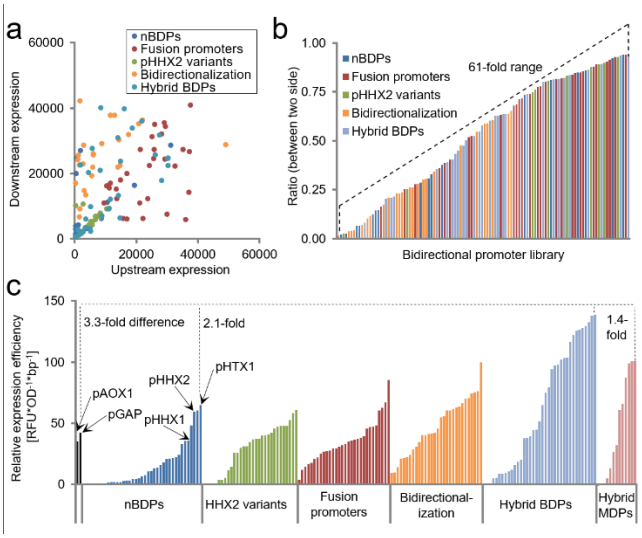


Fig. 4: Modularly designed and exceptionally short bidirectional hybrid promoters (179 to 457 bp) achieve highest expression efficiency matching the strong monodirectional *AOX1* promoter (940 bp length). The bidirectional hybrid promoters were assembled from histone core promoters (Fig. 2b) and CRMs of methanol regulated promoters (S 5, S 7). The detailed color code for the regulatory elements/abbreviations used is provided in S 7, a list of the exact designs of shBDP1-31 is provided in S 2. Yellow boxes indicate experimentally confirmed Mxr1p (methanol master regulator) binding sites in P_{AOX1} and P_{DAS2} (S 5, S 7), red boxes: TATA boxes. Additional bidirectional variants, controls and extended discussion are provided in S 7. 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 (see Fig. 2c,d) used to generate hybrid promoters with growth associated expression from one side. Strains were grown on glucose media for 60 h subsequently induced with methanol for 48 h. Mean values and standard deviations of normalized (using the normalization factor calculated in S 4) reporter protein fluorescence measurements of biological quadruplicates grown on the respective carbon sources are shown. All elements used (except for the non-regulated core promoters and constitutive *HHT2-T3*) are methanol inducible.

Fig. 5



<u>Fig. 5</u>: The library of 168 BDPs provides different absolute expression strengths (a), ratios (b) and regulatory profiles with synthetic BDPs (sBDPs) considerably surpassing the relative efficiency (c) of natural BDPs (nBDPs).

A) The library of BDPs covers the whole expression space. Normalized upstream and downstream reporter fluorescence is shown (rfu/OD as in <u>Fig. 1</u> to Fig. 4; under optimal growth conditions, by the default orientation in which the BDPs were cloned in the reporter vector).

B) 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 (under 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.

C) Relative expression efficiencies of sBDPs exceed nBDPs up to 2.1-fold and nMDPs up to 3.3-fold. 'Relative expression efficiency' is a term introduced in this study to illustrate the relationship between promoter length and promoter strength. The relative 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 (bp). Hence the expression efficiencies are relative terms and will change with different fluorescence reporter proteins used and even with different fluorospectrometers for detection. The monodirectional *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.

Fig. 6

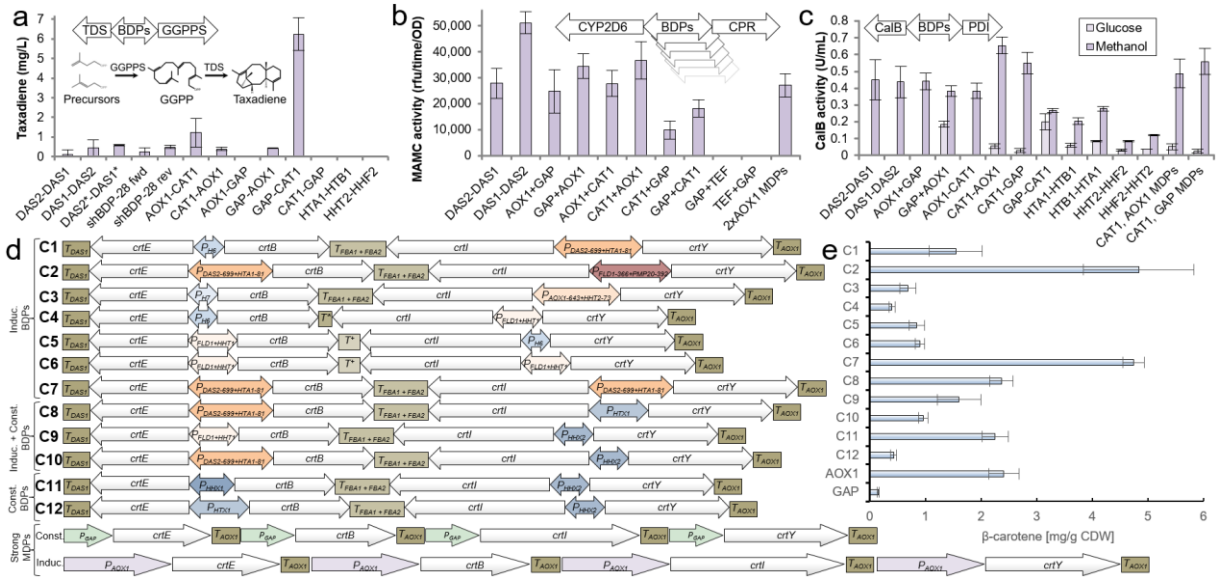


Fig. 6: Applying the library of BDPs helps to find the optimal expression condition for dual gene (a-c) and multi gene co-expression (d,e). For each pair of genes (a-c) tested, a different BDP performed best and the activity/yields for the same set of genes spanned a 5.2- to 50-fold range. The library of BDPs and BDTs (Fig. 7) facilitates the assembly and transcriptional fine-tuning of multi-gene pathways demonstrated with the four gene (*crtE, crtB, crtI, crtY*) model pathway of β-carotene biosynthesis (d,e).

A) Highest taxadiene yields were achieved using a $P_{GAP\text{-}CAT1}$ fusion promoter for GGPPS and TDS co-expression. The designs based on different BDPs span a 50-fold range in yields. DAS2*-DAS1* denotes the improved promoter variant DAS2-d8-DAS1-d2d5 (Fig. 3C, S 5). Constitutive expression of the GGPPS gene was detrimental (data not shown). Yields determined by GC-MS from shake flask cultivations (triplicates) with a dodecane overlay.

B) Highest activity for the co-expression of human CYP2D6 and its associated CPR was achieved using the natural $P_{DAS1\text{-}DAS2}$ promoter in reverse orientation. The designs based on different BDPs span a 5.2-fold activity range. '2x AOX1 MDPs' indicates a control strain expressing the two genes using two monodirectional *AOX1* promoters. The strains were pre-grown 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.

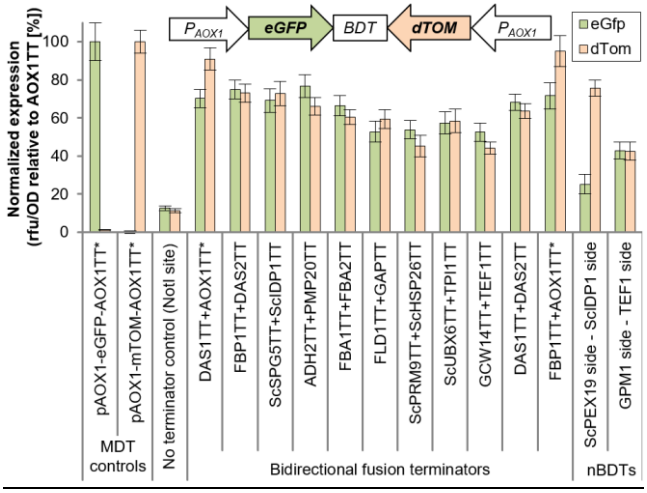
C) Bidirectional fusion promoters of P_{CAT1} to P_{AOX1} or P_{GAP} give highest volumetric activities in the co-expression of secreted CalB and the chaperone PDI. The designs based on different BDPs span a 22-fold activity range. 'CAT1, AOX1 MDPs' and 'CAT1, GAP MDPs' are control strains mimicking the best bidirectional designs with MDPs. Activities in the supernatant were measured after growth for 60 h on glucose and methanol induction for 72 h using a p-nitrophenyl butyrate (pNPB) assay.

D) Using BDPs and BDTs for pathway assembly reduces construct length and the number of parts required. Twelve bidirectional constructs were assembled by combining inducible or constitutive BDPs and combinations thereof (Induc. + const.) with a BDT and two MDTs. See S 8d for assembly strategy and supporting file S 2 (sheet 'Carotenoid pathway constructs') contains detailed information on the BDPs/BDTs used. For the BDPs, a coloring scheme similar to Fig. 5 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.

E) β-carotene titers obtained with strains based on the bidirectional constructs shown in panel e) span a 12-fold range matching or surpassing conventional P_{AOX1} and P_{GAP} based designs.

Mean values and standard deviations of triplicate cultivations in shake flasks are shown (HPLC measurements).

Fig. 7



<u>Fig. 7:</u> Bidirectional transcription terminators (BDTs) required for the assembly of bidirectional multi gene co-expression relieve expression loss associated with transcriptional collision. A reporter construct for testing bidirectional transcription termination was assembled by cloning the genes coding for eGfp and dTom in convergent orientation (small inlet). Two *AOX1* promoters were used to drive equal expression of the reporter genes. Monodirectional terminators (MDTs) were combined into bidirectional fusion terminators and two putative natural BDTs (nBDTs) were tested. A negative control lacking termination sequences and bearing solely a *Not*I restriction site was included. Additional control constructs contain only a single *AOX1* promoter, a single FP and the *AOX1** terminator. *AOX1TT** denotes the *AOX1* terminator sequence used by Vogl *et al*. [30]. Some BDTs acted also as autonomously replicating sequences (<u>S 10</u>). Mean values and standard deviations of fluorescence measurements after pre-growth on glucose followed by methanol induction of biological quadruplicates are shown.

Tables

Tab. 1 (Minimal set of diverse BDPs)

Tab. 1: Minimal set of diverse BDPs covering broad regulatory profiles for co-expression optimization. In dual gene expression applications, each BDP should be tested in forward and reverse orientation. Reporter protein fluorescences of the respective promoters are shown in Fig. 2a and Fig. 3a,b. Annotated GenBank files for these promoters are provided in supporting file S 3. For multi-gene co-expression furthermore the three histone promoters *HTX1*, *HHX1*, *HHX2* and additional methanol inducible promoters (*e.g.* shBDP23 [Fig. 4], *FLD1*+*PMP20*, *FBA2*+*TAL2* [Fig. 3b]) are useful.

BDP	Regulation	Strength	Ratio
$P_{HTX1\ (HTA1\text{-}HTB1)}$	constitutive on both sides (cell cycle/growth associated in *S. cerevisiae* [7])	strong on both sides	~ 1:1
$P_{DAS2\text{-}699\text{-}pCoreHTA1\text{-}81}$	methanol inducible (tightly glucose/glycerol repressed) on both sides	strong on both sides	~ 1:2
$P_{CAT1\text{-}FDH1}$	derepressed/methanol inducible on both sides	weak/moderate under derepression, strong on methanol	~ 1:1
$P_{AOX1\text{-}CAT1}$	methanol inducible (tightly glucose/glycerol repressed) – derepressed, methanol inducible	strong on both sides on methanol P_{CAT1} moderate under derepressed conditions	derepressed: ~ 0:1 induced: ~ 1:1
$P_{AOX1\text{-}GAP}$	methanol inducible (tightly glucose/glycerol repressed) – constitutive	strong on both sides (P_{GAP} side moderate on methanol)	glucose/derepressed: ~ 0:1 induced: ~ 1:0.5
$P_{GAP\text{-}CAT1}$	constitutive - derepressed, methanol inducible	P_{GAP} side strong constitutive (moderate on methanol), P_{CAT1} side moderate derepressed, strong on methanol	derepressed: ~ 1:0.25 induced: ~ 1:2

Supporting information

Engineered bidirectional promoters enable rapid multi-gene co-expression optimization

Thomas Vogl[1], Thomas Kickenweiz[1], Julia Pitzer[2], Lukas Sturmberger[2], Astrid Weninger[1], Bradley W. Biggs[3], Eva-Maria Köhler[1], Armin Baumschlager[1§], Jasmin Elgin Fischer[1], Patrick Hyden[1], Marlies Wagner[1], Martina Baumann[4,5], Nicole Borth[4,5], Martina Geier[2], Parayil Kumaran Ajikumar[3], Anton Glieder[1#]

[1] Institute of Molecular Biotechnology, NAWI Graz, Graz University of Technology, Petersgasse 14, Graz 8010, Austria

[2] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, Graz 8010, Austria

[3] Manus Biosynthesis, 1030 Massachusetts Avenue, Suite 300, Cambridge, MA 02138

[4] Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, Vienna 1190, Austria

[5] Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, Vienna 1190, Austria

Table of contents

S 1 (List of P. pastoris hth, htt and ttt genes)

S 1: Lists of gene pairs in the *P. pastoris* genome and extended discussion.

The genome sequence of the *P. pastoris* CBS7435 strain [76] 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.* [21] of the human genome. Genes in head to head, head to tail and tail to tail orientation are provided in the separate sheets of the excel file. 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 the excel file, gene pairs separated by gaps were omitted from the analysis.

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. Miss-annotations (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 of transcription 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 (e.g. [77,78]). 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 [9,10], 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 [79].

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 co-expression, inducible nBDPs and combinations of inducible and constitutive promoter 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 [30] and we have tested all putative nBDPs with MUT promoters on one side (Fig. 1c).

S 2 (List of BDPs & primers for cloning)

S 2: List of primers and details on sequences used in this study. Primers used for generating the vectors applied, the *P. pastoris* nBDPs tested, sBDPs generated, BDTs tested and detailed dual gene expression and carotenoid pathway assemblies are provided. The respective information is provided in different sheets of the Excel file:

- Promoters and terminators
  - Reporter vectors
    Primers for generating the reporter vectors for bidirectional promoters and terminators are provided. Also, the primers for the generation of the entry vectors for cassettes for dual or multi gene co-expression are provided (see Materials and methods for detailed descriptions).
  - nBDPs
    Detailed list on the *P. pastoris* natural BDPs (Fig. 1c) tested and primers used for amplification. Either primers for TA cloning (shorter as no overhangs are needed) or for Gibson assembly were used (overhangs denoted in different letter case). The histone promoters were cloned in both orientations, hence two primer pairs each are listed.
  - HHX2 variants
    Details on the deletion and truncation variants of the *P. pastoris HHX2* promoter (Fig. 2c,d). The deletions were achieved by either linking two PCR products up to the deletion by Gibson assembly or by ordering the promoters as gBlocks (Integrated DNA technologies).
  - Bidirectionalization
    Overview on combinations of core promoters and monodirectional promoters used to generate bidirectionalized promoters (Fig. 3a). The core promoters were ordered as long primers and fused by PCR to the monodirectional promoter and cloned via Gibson assembly into the reporter vector.
  - Fusion promoters
    Monodirectional promoters fused to each other to generate bidirectional fusion promoters combining different regulatory profiles (Fig. 3b,c). Deletions in the *DAS1/2* promoters are described in detail in the next sheet. The promoters were amplified separately with primers (with complementary overhangs between each other and to the vector) and cloned by Gibson assembly.
  - DAS1/2 deletions
    Exact deletions performed in the monodirectional *DAS1* and *DAS2* promoters (shown in S 5 and used to assemble some fusion promoters shown in Fig. 3c). The deletions were achieved by linking two PCR fragments with respective overhangs to each other using the primer combinations indicated. The olePCR products were cloned into a reporter vector via *Sbf*I and *Nhe*I sites (the *Nhe*I site is adjacent to *eGFP* reporter gene's start codon, resulting in seamless fusions).
  - Hybrid BDPs
    The exact composition of the *P. pastoris* bidirectional hybrid promoters shown in Fig. 4 is provided. Orientations of the elements are given by stylized arrows '->' or '<-', different elements are separated by '|'. The synthetic promoters were either assembled by PCR (providing short designs on a primer) or ordered as gBlocks. Fusions to the truncated HHT2-T3 variant were assembled by olePCR.
  - BDTs

Primer sequences for cloning of the bidirectional transcriptional terminators for *P. pastoris* are provided (Fig. 7, S 10).

- Dual gene applications
  Primers for generating the bidirectional dual gene co-expression vectors and cloning of the BDPs tested in *P. pastoris* for Taxadiene production, CYP2D6+CPR co-expression and CalB+PDI co-expression (Fig. 6a-c) are provided.

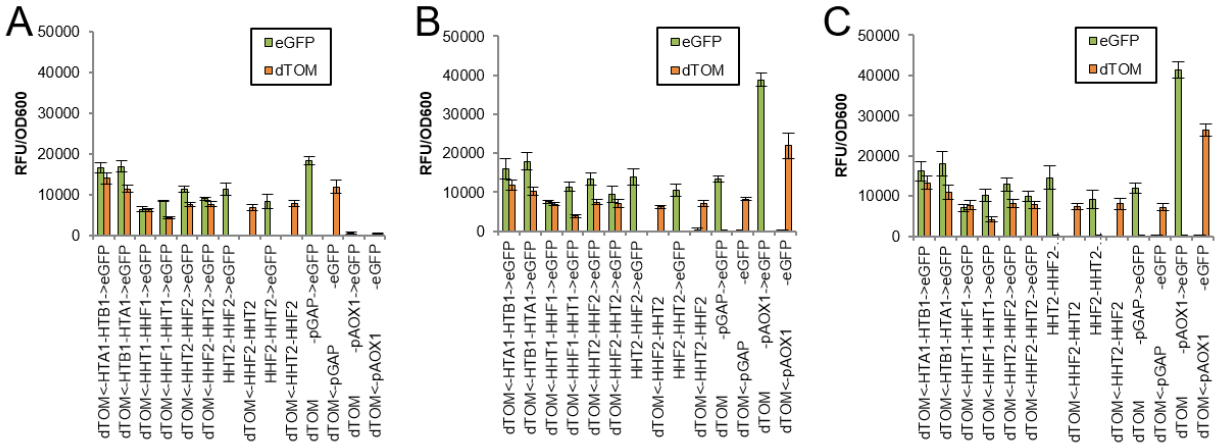
- Carotenoid pathway constructs
  Contains the exact promoters and terminators used for the pathways shown in Fig. 6d. Primers for cloning via Gibson assembly are indicated (n.a. = not applicable).

- Carotenoid pathway primers
  Primer sequences for assembling the constructs shown in the aforementioned sheet.

S 3 (Annotated sequences for minimal set of promoters)

S 3: Supporting file containing annotated sequences of a minimal set of BDPs covering broad regulatory profiles for dual gene expression optimization. Annotated sequence files in GenBank format are provided for the BDPs highlighted in Tab. 1 of the main manuscript.

S 4 (Reporter proteins normalization)



E

Values compared	Ratio eGFP/dTOM			
	60h glucose	24h MeOH	48h MeOH	72h MeOH
pHTB1-eGFP/pHTB1-dTOM	1.451	1.579	1.468	1.492
pHTA1-eGFP/pHTA1-dTOM	1.214	1.522	1.350	1.364
pHHF1-eGFP/pHHF1-dTOM	1.491	1.944	1.709	1.790
pHHT1-eGFP/pHHT1-dTOM	1.362	1.642	1.328	1.361
pHHF2-eGFP/pHHF2-mTOM	1.481	1.902	1.643	1.735
pHHT2-eGFP/pHHT2-dTOM	1.178	1.273	1.213	1.192
pHHF2-eGFP/pHHF2-dTOM (only one fluorescent protein)	1.665	2.236	1.993	1.985
pHHT2-eGFP/pHHT2-dTOM (only one fluorescent protein)	1.059	1.456	1.148	1.185
pGAP-eGFP/pGAP-dTOM	1.538	1.626	1.657	1.805
pAOX1-eGFP/pAOX1-dTOM	-	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 4: 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 *dTOM* (dTomato, an enhanced red fluorescent protein variant [64]). Reporter gene fluorescence was measured after **(a)** 60 h growth on glucose and subsequently **(b)** after 24 h, **(c)** 48 h and **(d)** 72 h methanol induction. For each construct and time point a normalization factor was calculated **(e)** by dividing the indicated eGfp value by the dTom 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. 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 control vectors with only a single FP present and cloned the *HHX2* promoter in both orientations into these vectors (the gene coding for 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.

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 on methanol for 24 hours (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 may be folding faster than dTom resulting in a higher eGfp/dTom fluorescence ratio. It appears that after 48 h and 72 h enough time has passed to allow dTom folding, resulting in a similar ratio as on glucose. This finding implies that for every measured 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. We had initially also tested alternative combinations of FPs and variants of Tomato (data not shown) but 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 devices available to us did not provide the correct filter to clearly discriminate the signals of eGfp and dTom. The high throughput characterization of the 168 *P. pastoris* BDPs was rendered possible by the availability of a monochromator based 96 well fluorescence microtiter plate reader. Notably also alternative filter based fluorescence plate readers considered did not provide by default the suitable filter sets to unambiguously discriminate eGfp and dTom fluorescence. We performed these extensive controls to ensure reliability of our plate reader measurements and we have previously shown that FACS and plate reader measurements of promoter variants are in excellent agreement in *P. pastoris* [43].

Possible interference of dTomato and OD600 measurements

In *E. coli,* it has been reported by Hecht et al. [80], that red fluorescent protein expression can bias OD600 measurements and thereby estimates for biomass. Such effects of high dTom fluorescence on OD600 measurements may theoretically also occur in *P. pastoris*. However, if there had been a major bias, our normalization experiments, where eGfp and dTom fluorescence were extensively compared, would have revealed it. We had also specifically looked at cell growth of dTom expressing cells in comparison with the wildtype strain in previous work [81] [supporting figure S6 in that paper] and had not noticed an impact of dTom expression. It is worth noting that after correction factor application, the histone promoters give nearly identical expression on both sides. Additionally, the *AOX1* promoter tested in both directions (for example shown in Fig. 4, control at the very bottom bottom) gives nearly identical values if tested with eGFP or dTOM. Thus, if there were an interference with OD600 measurements as noted by Hecht et al., it would be evident for these controls.

However why did we not experience the issues reported by Hecht et al.?

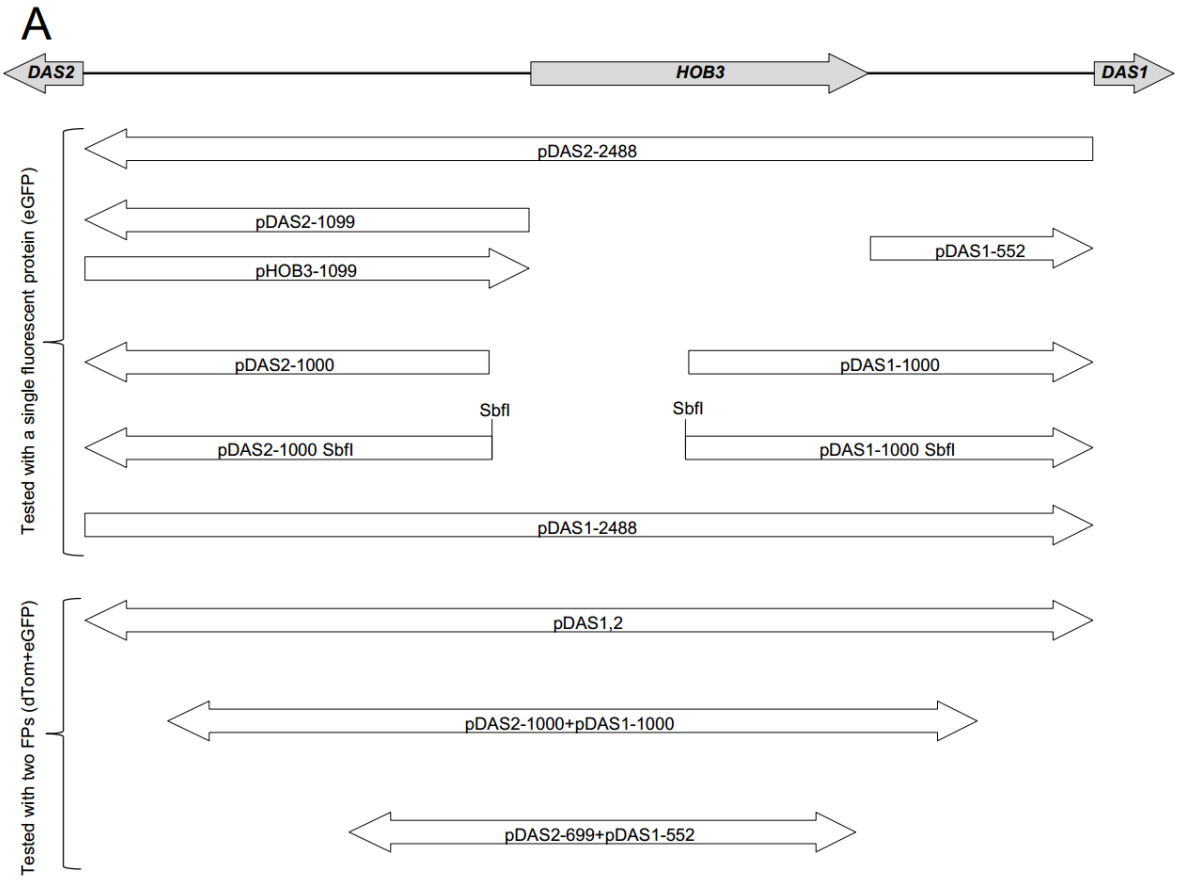
We assume that there are two main reasons:

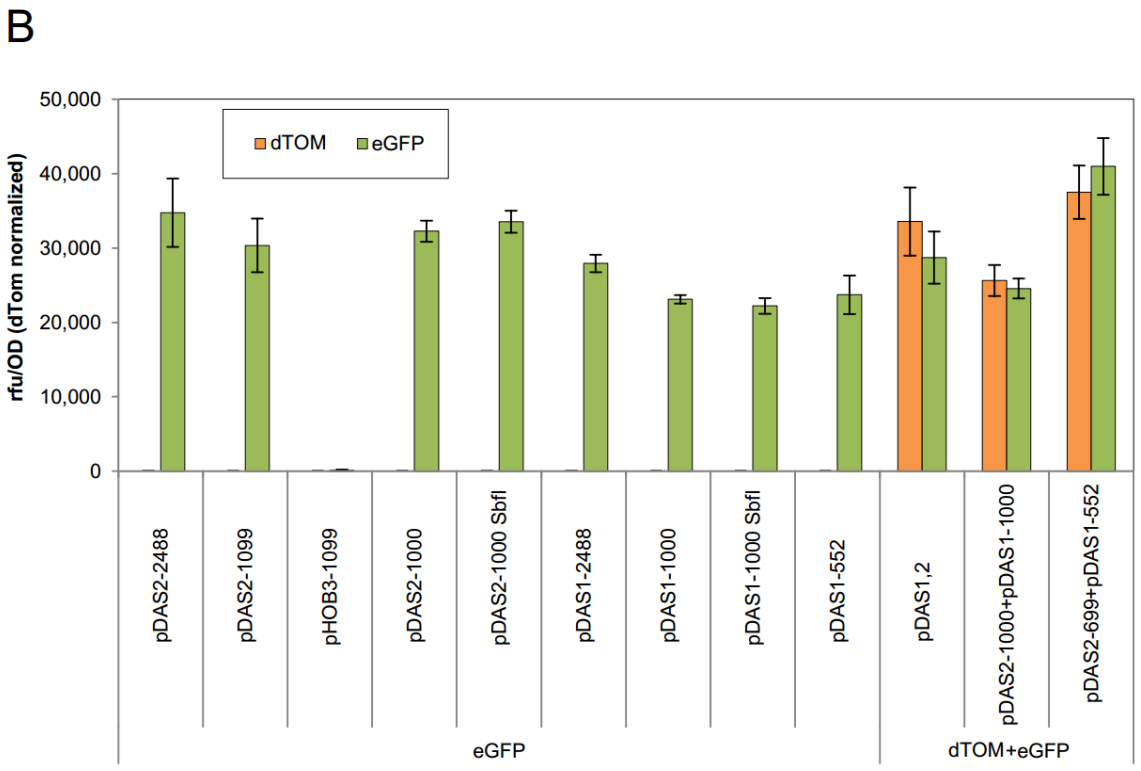
1.) We performed the experiments in the yeast *P. pastoris* and not in the bacterium *E. coli*. It appears plausible that the ratios between fluorescent proteins inside of the cells and the biomass is different between bacteria and yeasts. *P. pastori*s is known for growth to exceptionally high cell densities (up to 500 g/l cell wet weight in bioreactors and even in DWPs ODs of ~20 to 30 can be reached) whereas the fluorescent protein production does not necessarily increase the same way. Hence a 10x denser *P. pastoris* culture might actually contain relatively less protein than *E. coli* cells. Hence there could be simply relatively less fluorescent protein interfering with the biomass measurement.

2) We have used a different red fluorescent protein variant than Hecht et al. with notable differences in excitation/emission wavelengths: We have used dTomato (excitation/emission wavelength: 554/581 nm [64]) whereas Hecht et al. have used mRFP1 (excitation/emission wavelength: 584/607 nm; original mRFP1 reference: [82]). So the em/ex peaks are shifted by 30/26 nm respectively, placing mRFP1 used by Hecht et al. notably closer to 600 nm used for OD600 measurements. Hence it appears plausible that the issue noted by Hecht et al. does not occur for all RFP variants.

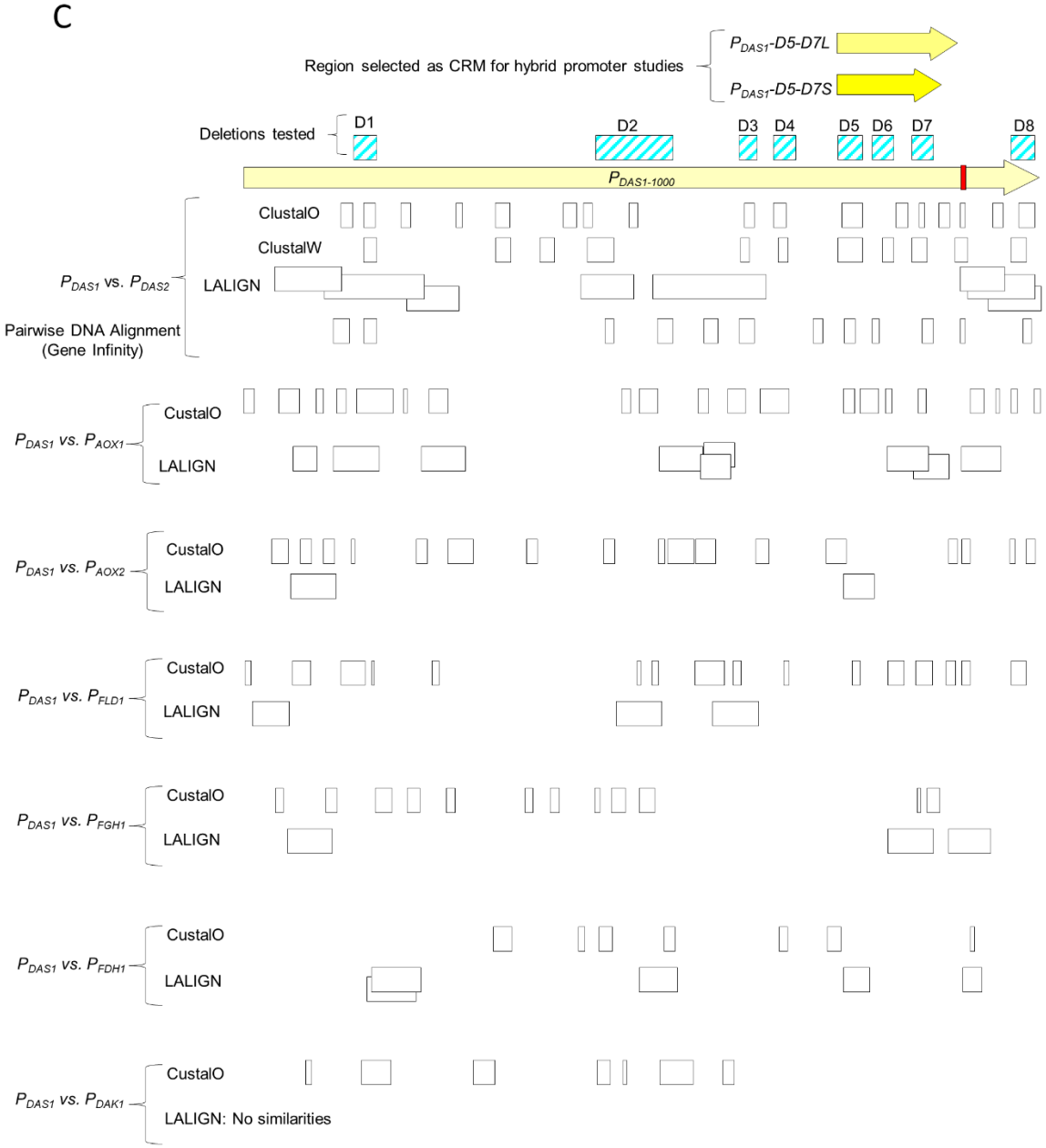
Note that while such interference effects appear to have no effect on the use of dTom in *P. pastoris*, they might be as relevant as in *E. coli* [80]in other organisms or with other fluorescent proteins and should be tested for.

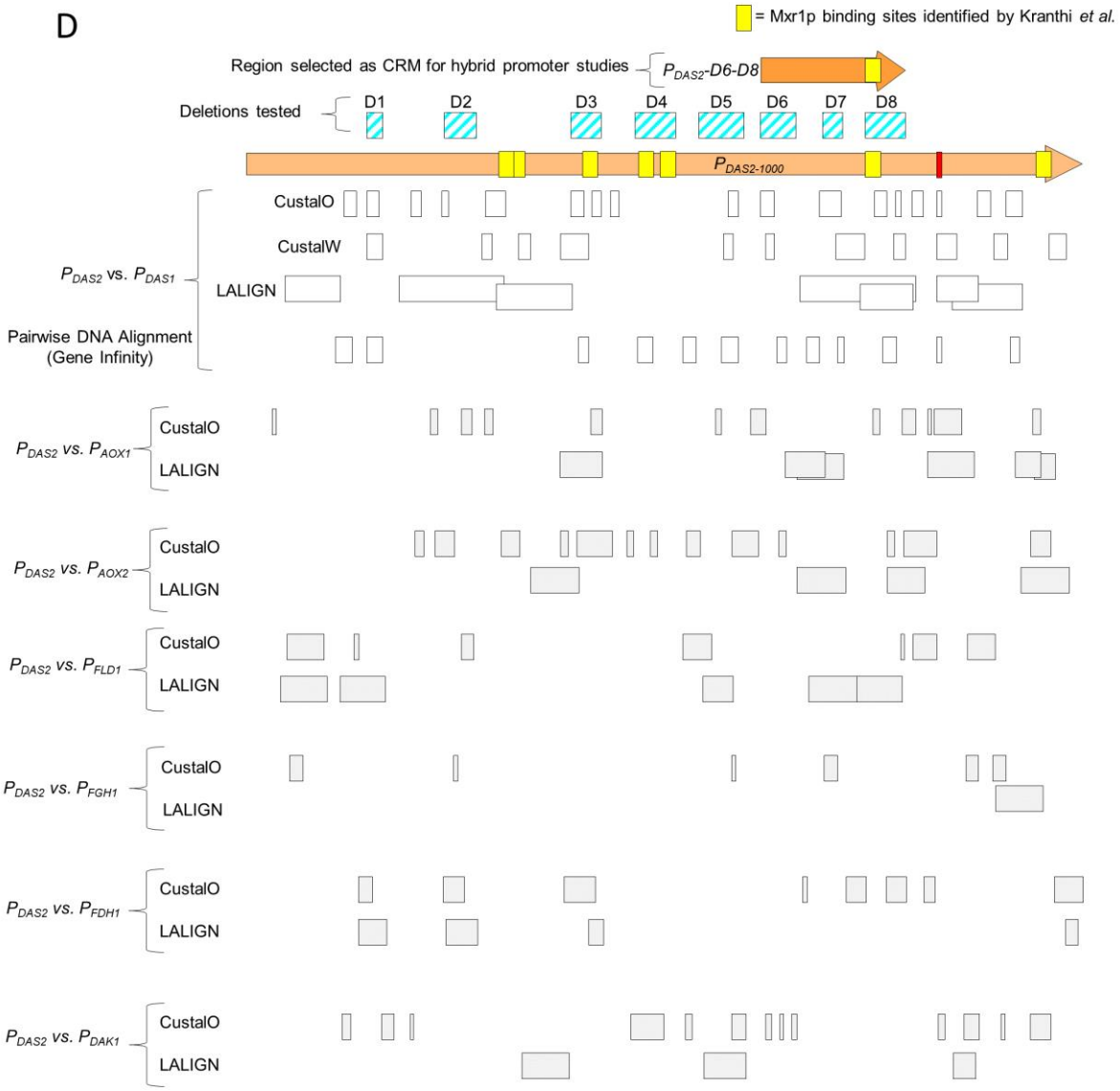
S 5 (Natural $P_{DAS1/DAS2}$, deletion variants of P_{DAS1} and P_{DAS2} + P_{AOX1} regulatory elements)





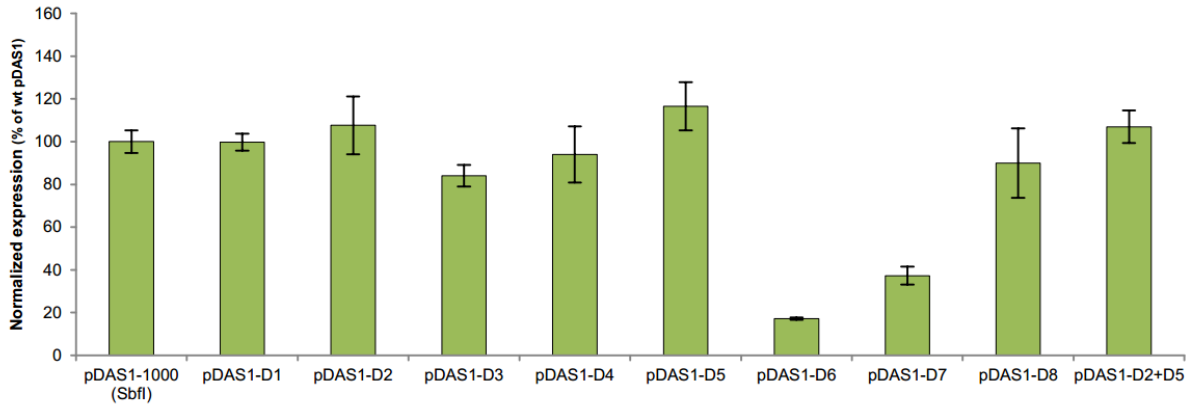
C


Region selected as CRM for hybrid promoter studies
P_{DAS1}-D5-D7L
P_{DAS1}-D5-D7S
Deletions tested
D1
D2
D3
D4
D5
D6
D7
D8
$P_{DAS1-1000}$
ClustalO
ClustalW
P_{DAS1} vs. P_{DAS2}
LALIGN
Pairwise DNA Alignment (Gene Infinity)
P_{DAS1} vs. P_{AOX1}
CustalO
LALIGN
P_{DAS1} vs. P_{AOX2}
CustalO
LALIGN
P_{DAS1} vs. P_{FLD1}
CustalO
LALIGN
P_{DAS1} vs. P_{FGH1}
CustalO
LALIGN
P_{DAS1} vs. P_{FDH1}
CustalO
LALIGN
P_{DAS1} vs. P_{DAK1}
CustalO
LALIGN: No similarities

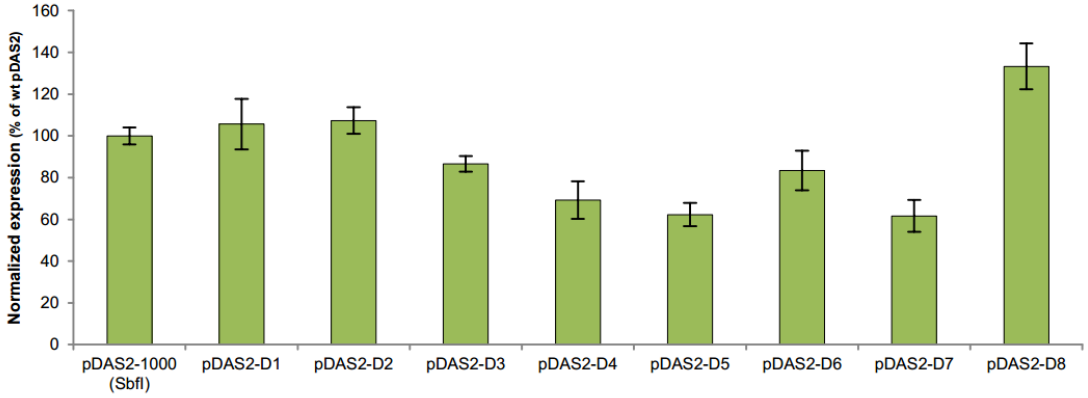

D
= Mxr1p binding sites identified by Kranthi et al.
Region selected as CRM for hybrid promoter studies
P_{DAS2}-D6-D8
Deletions tested
D1
D2
D3
D4
D5
D6
D7
D8
$P_{DAS2-1000}$
P_{DAS2} vs. P_{DAS1}
ClustalO
ClustalW
LALIGN
Pairwise DNA Alignment (Gene Infinity)
P_{DAS2} vs. P_{AOX1}
ClustalO
LALIGN
P_{DAS2} vs. P_{AOX2}
ClustalO
LALIGN
P_{DAS2} vs. P_{FLD1}
ClustalO
LALIGN
P_{DAS2} vs. P_{FGH1}
ClustalO
LALIGN
P_{DAS2} vs. P_{FDH1}
ClustalO
LALIGN
P_{DAS2} vs. P_{DAK1}
ClustalO
LALIGN

E

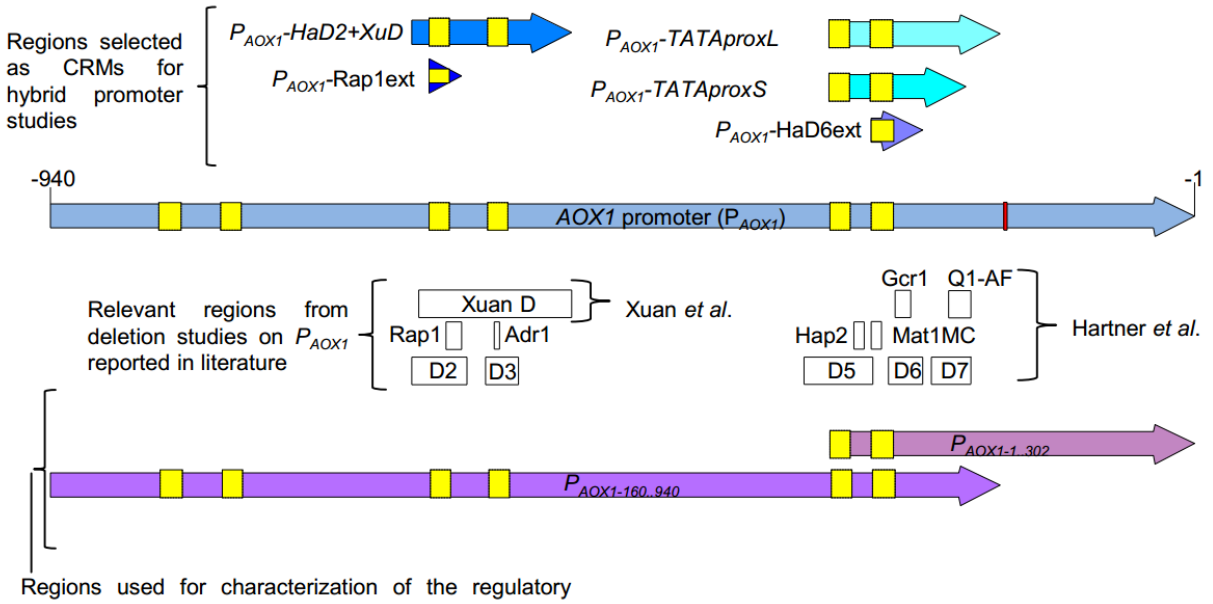
P_{DAS1} deletions



P_{DAS2} deletions



F



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 [31]) 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 *Sbf*I 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 4 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. 3c) and selected CRMs used for hybrid promoter design (Fig. 3d 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* [30]). Similar stretches from pairwise alignments using ClustalO and LALIGN are shown. For the *DAS1/DAS2* comparison Clustal W and a pairwise alignment was performed in addition. 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*. [83] are depicted.
E) Effects of single deletions depicted in panels (c) and (d) (top panel P_{DAS1}, bottom panel P_{DAS2}). eGfp reporter fluorescence (rfu/OD) of the deletion variants was normalized as percent of the unmodified wildtype promoters (pDAS1-1000, pDAS2-1000).
F) *AOX1* CRMs used shown schematically, highlighting previously deleted regions by Xuan *et al*. and Hartner *et al*. [84,85]

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 [44,84]. Either systematic deletions were performed (*i.e.* adjacent fragments [85,86]) or semi-rational considerations (such as the prediction of transcription factor binding sites [84]) were applied, as exemplified by studies on the *P. pastoris AOX1* promoter [84–86].

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 [30], 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, TFBSs may be placed at different positions between promoters. Studies on the *P. pastoris* methanol master regulator Mxr1p [83,87] showed that its binding sites are arranged pairwise over the whole *AOX1* promoter [87], whereas they are generally closer together in the *DAS2* promoter [83] (reviewed in [31] and compare the Mxr1 binding sites in S 5d and f). In addition, yeast TFBS are often short and degenerate as exemplified by the Mxr1 consensus sequence CYCCNY (N = any base, Y = C or T) [83].

Performing a multiple sequence alignment of the eight MUT promoter sequences mentioned above using Clustal Omega [88] in standard settings 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 (LALIGN is suitable to identify local sequence similarities between two sequences [89]). 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 a 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 (Bidirectionalization and fusion design considerations)

S 6: Design considerations for bidirectionalizations and fusions promoters and synergistic and antagonistic effects observed for fusions of inducible and constitutive promoters.

Design consideration for bidirectionalizations

We selected differently regulated monodirectional *P. pastoris* promoters for bidirectionalization, including methanol inducible, tightly repressed promoters of the *AOX1, PMP20, DAS1, DAS2, FBA2* and *TAL2* genes. Also methanol inducible but derepressed promoters of *CAT1, FLD1* and *FDH1* genes were tested [30]. Furthermore constitutive promoters of the *GAP, TEF* and *ADH2* genes were bidirectionalized. The short histone core promoters outlined in Fig. 2b were in reverse orientation fused to different lengths of the MDPs. Results on the bidirectionalizations are shown in Fig. 3a of the main manuscript.

Design consideration for fusions promoters

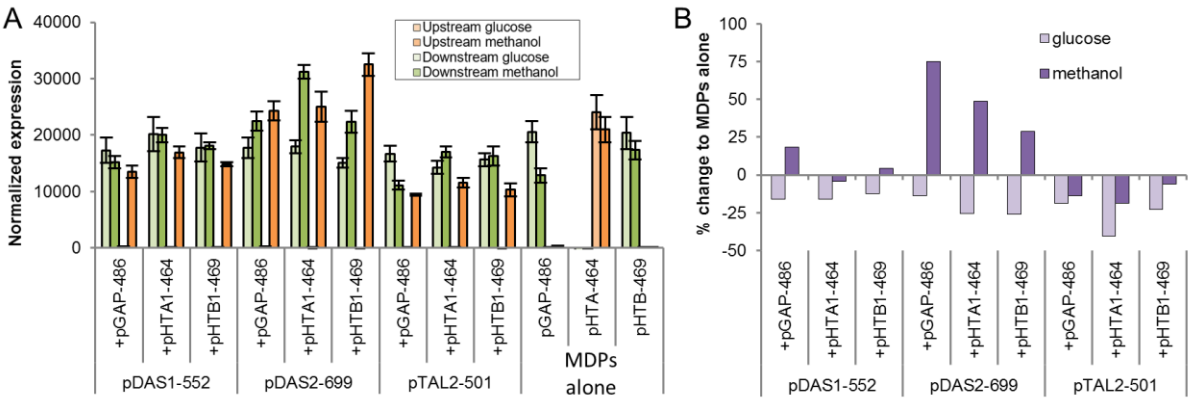
To achieve combinations of regulatory profiles, we fused differently regulated MDPs to each other in reverse orientation (see also the respective sheet in S 2). In addition, we included specifically combinations of two inducible promoters, to obtain suitable tools for regulated pathway overexpression. The properties of the combinations tested are summarized in the table on the next page (results are shown Fig. 3b of the main manuscript).

Combinations of P_{DAS1} and P_{DAS2} deletions

We generated additional methanol inducible fusion BDPs with varying expression ratios by combing different monodirectional deletion variants (Fig. 3c). 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*). Also BDPs with decreased expression were generated (*DAS2-D5+DAS1-D6*, *DAS2-386+Das1-D6*). 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).

Fusion BDP properties				Upstream promoter				Downstream promoter			
#	Functional group	Label of the fusion BDP in Fig. 3c	Length (bp)	Name	Regulation	Length (bp)	Ref.	Name	Regulation	Length (bp)	Ref.
1	Differently regulated promoters	AOX1+GAP	1426	AOX1	tightly repressed, methanol inducible	940	[30]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
2		CAT1+GAP	1178	CAT1	derepressed, methanol inducible	692	[30]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
3		DAS1+GAP	1038	DAS1	tightly repressed, methanol inducible	552	[30]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
4		DAS1+HTA1	1016	DAS1	tightly repressed, methanol inducible	552	[30]	HTA1	'constitutive' (cell cycle regulated)	464	this work
5		DAS1+HTB1	1021	DAS1	tightly repressed, methanol inducible	552	[30]	HTB1	'constitutive' (cell cycle regulated)	469	this work
6		DAS2+GAP	1185	DAS2	tightly repressed, methanol inducible	699	[30]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
7		DAS2+HTA1	1163	DAS2	tightly repressed, methanol inducible	699	[30]	HTA1	'constitutive' (cell cycle regulated)	464	this work
8		DAS2+HTB1	1168	DAS2	tightly repressed, methanol inducible	699	[30]	HTB1	'constitutive' (cell cycle regulated)	469	this work
9		TAL2+GAP	987	TAL2	tightly repressed, methanol inducible	501	[30]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
10		TAL2+HTA1	965	TAL2	tightly repressed, methanol inducible	501	[30]	HTA1	'constitutive' (cell cycle regulated)	464	this work
11		TAL2+HTB1	970	TAL2	tightly repressed, methanol inducible	501	[30]	HTB1	'constitutive' (cell cycle regulated)	469	this work
12		TEF1+GAP	1486	TEF1	constitutive (naturally involved in translation)	1000	[31]	GAP	constitutive (naturally involved in glycolysis)	486	[31]
13	Inducible promoters	FDH1+PMP20	956	FDH1	derepressed, methanol inducible	564	[30]	PMP20	tightly repressed, methanol inducible	392	[30]
14		FDH1+DAS1	1116	FDH1	derepressed, methanol inducible	564	[30]	DAS1	tightly repressed, methanol inducible	552	[30]
15		FDH1+CAT1	1064	FDH1	derepressed, methanol inducible	564	[30]	CAT1	derepressed, methanol inducible	500	[30]
16		FLD1+AOX1	1009	FLD1	derepressed, methanol inducible	366	[30]	AOX1	tightly repressed, methanol inducible	643	[30]
17		FLD1+PMP20	758	FLD1	derepressed, methanol inducible	366	[30]	PMP20	tightly repressed, methanol inducible	392	[30]
18		CAT1+AOX2	1000	CAT1	derepressed, methanol inducible	500	[30]	AOX2	tightly repressed, methanol inducible	500	[30]
19		FBA2+TAL2	1001	FBA2	tightly repressed, methanol inducible	500	[30]	TAL2	tightly repressed, methanol inducible	501	[30]
20		DAS2+DAS1*	1251	DAS2	tightly repressed, methanol inducible	699	[30]	DAS1	tightly repressed, methanol inducible	552	[30]
21		AOX1+CAT1	1632	AOX1	tightly repressed, methanol inducible	940	[30]	CAT1	derepressed, methanol inducible	692	[30]

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 MDPs alone. The same data shown in Fig. 3b and S 7c was rearranged to facilitate comparisons.
B) Changes in normalized reporter gene fluorescence of the fusions promoters compared to the MDPs alone from panel A are shown.

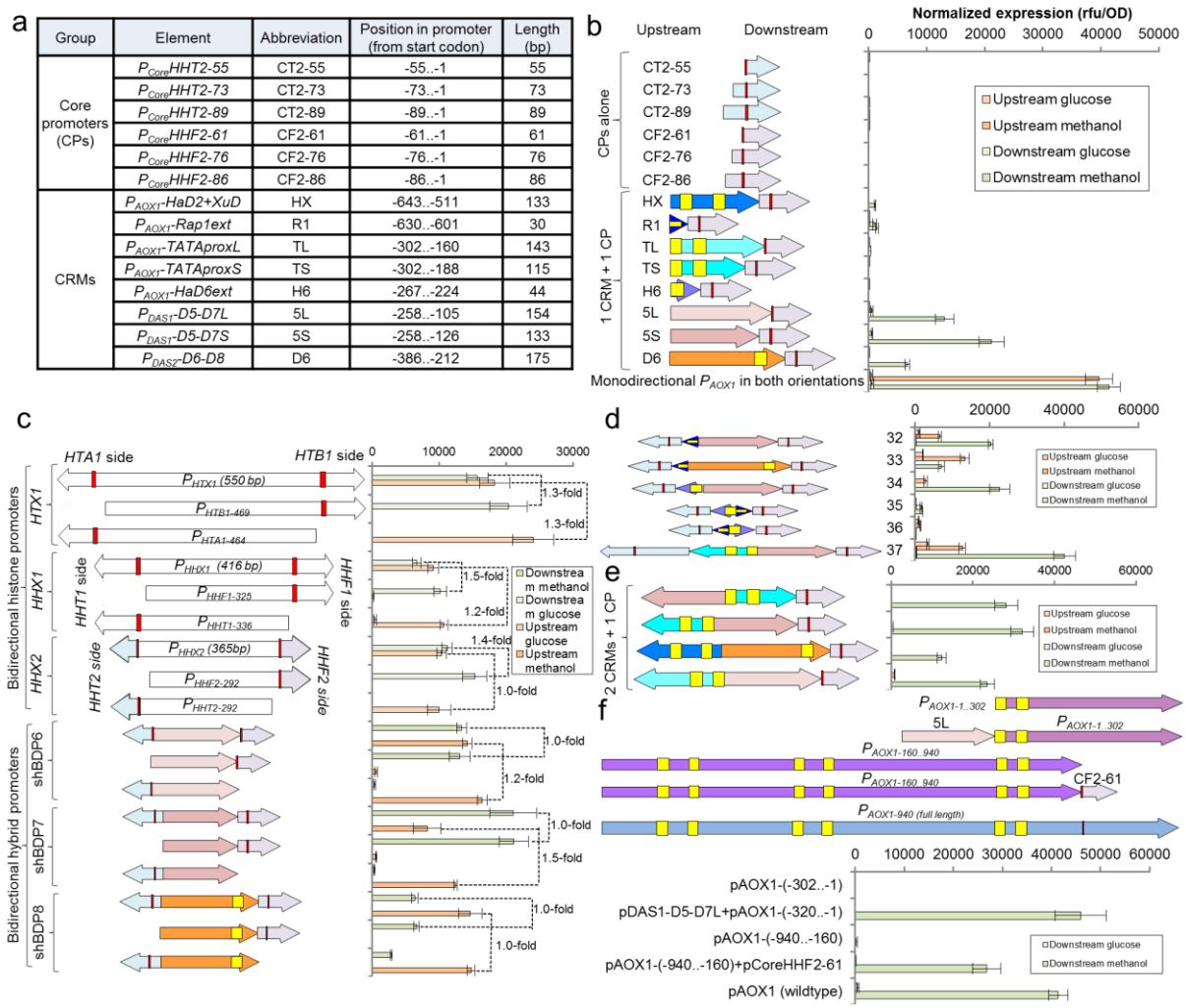
Fusions of growth-associated/constitutive *HTA1*, *HTB1* and *GAP* promoters to P_{DAS2} reached on methanol 1.3- to 1.8-fold increased expression compared to the single promoters. Notably, P_{GAP} is typically downregulated on methanol ([31], S 4), whereas fusions to P_{DAS2} showed increased expression 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. 3a) underlining the strongly activating effect on upstream fusions.
Fusions of the same growth-associated/constitutive promoters to $P_{DAS1\text{-}552}$ showed less pronounced 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 (*i.e.* 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. 4).

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 7: 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 Fig. 2b 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 monodirectional *AOX1* promoter is included 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 Fig. 2a 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. 3d.

E) Additional monodirectional hybrid promoters (combinations of 2 CRMs with 1 core promoter).
F) Using the first 302 bp of P_{AOX1} as promoter element does not elicit 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 [44,90]. We extended this strategy to BDPs, by flanking CRMs with two core promoters in opposite orientation. We used the short *HHX2* histone core promoters (Fig. 2b) successfully applied for bidirectionalization of MDPs (Fig. 3a). Six short CRMs (30 to 175 bp) from methanol regulated promoters (P_{AOX1}, P_{DAS1}, P_{DAS2}) were used. Namely, four elements from 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 reported in the literature ([84–86,91] reviewed in [31]) and binding sites reported for Mxr1p (zinc finger transcription factor and master regulator of MUT genes in *P. pastoris*) [87]. The CRMs of the *AOX1* promoter contain Mxr1p binding sites and deletions within these regions strongly affected expression [84,85,87]. P_{AOX1}-*HaD2+XuD* is a fusion of the D2 region of Hartner *et al*. [84] and region D of Xuan *et al*. [85] containing two Mxr1p binding sites [87]. 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 affected 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 Mxr1p binding sites reported for P_{DAS2} [31,83]. 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. 2c,d) 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 ($Pcore_{HTB1\text{-}86}$, Fig. 2b) resulting in a truncated P_{HTA1} promoter of 464 bp ($P_{HTA1\text{-}464}$). *Vice versa* the core promoter removal/truncated promoter pair on the other side of P_{HTX1} is $Pcore_{HTA1\text{-}81}$/$P_{HTB1\text{-}469}$. The

pairs for P_{HHX1} are $Pcore_{HHT1\text{-}91}$/$P_{HHF1\text{-}325}$ and $Pcore_{HHF1\text{-}80}$/$P_{HHT1\text{-}336}$. For P_{HHX2} the truncations F1 and T1 already shown in Fig. 1g 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 resources in the form of amino acids and capacities for protein synthesis/folding from the cell and represent a greater metabolic burden [92] 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 translational: In frame of the normalization work to compare the two FPs used (S 4), we created constructs of the full length *HHX2* promoter flanked by 1.) two FPs and 2.) 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]). Thereby the bidirectional promoter is on one side expressing the FP, but on the other side transcription is immediately stopped after the 5'UTR by the terminator. Expression of these terminator constructs was not increased compared to deletion of the core promoters (Fig. 2d). Therefore we conclude that removal of the core promoters and not just the reporter gene is required to increase expression. Hence it appears that *P. pastoris* cells have sufficient resources to produce two FPs at high levels and translation is not the limiting factor, hinting a regulatory model in which two core promoters are competing for transcription initiation by general TFs or RNAPII in a bidirectional context (possibly in line with studies in higher eukaryotes showing that antisense promoter transcription commonly relies on the sense core promoter sequence [93]).

Transcriptional 'spillover' in hybrid promoters

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 6) were observed (hybrid promoters #9-14 in Fig. 3d). If the methanol regulated CRMs were not fused to any additional sequences (S 7b) or other methanol regulated CRMs (Fig. 3d), 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 longer fusion promoters (Fig. 3b), 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. In *S. cerevisiae* DNA binding factors have been reported (Tbf1 and Mcm1), that could be used for this purpose [17].

In some respect, certain aspects of our work would rather support models from openly debated deep sequencing studies (e.g. [94–96]), in which divergent/bidirectional transcription is arising from second (cryptic) upstream core promoters (in yeast pre-initiation complexes form around both sides of nucleosome free regions associated with transcribed chromatin [97]). We have here shown that most CRMs tested were suitable to trigger upstream expression if a second core promoter was added. Hence, cryptic core promoter sequences naturally occurring upstream of CRMs/enhancers may fulfill a similar role as rationally added core promoters in our experiment (also given the fact that even random sequences can show substantial functionality as core promoters [43]). Hence we rather tend to lend credence to the notion, that widespread bidirectionality observed in deep

sequencing studies is presumably caused by the presence of cryptic upstream core promoters alongside an orientation independent recruitment of RNAPII by activators binding CRMs. Whether these cryptic/weak core promoters are evolutionary shaped and their expression serves a purpose (or is a merely an unavoidable side effect of the orientation insensitivity of the recruitment mechanisms of transcription factors) has recently been comprehensively studied [19].

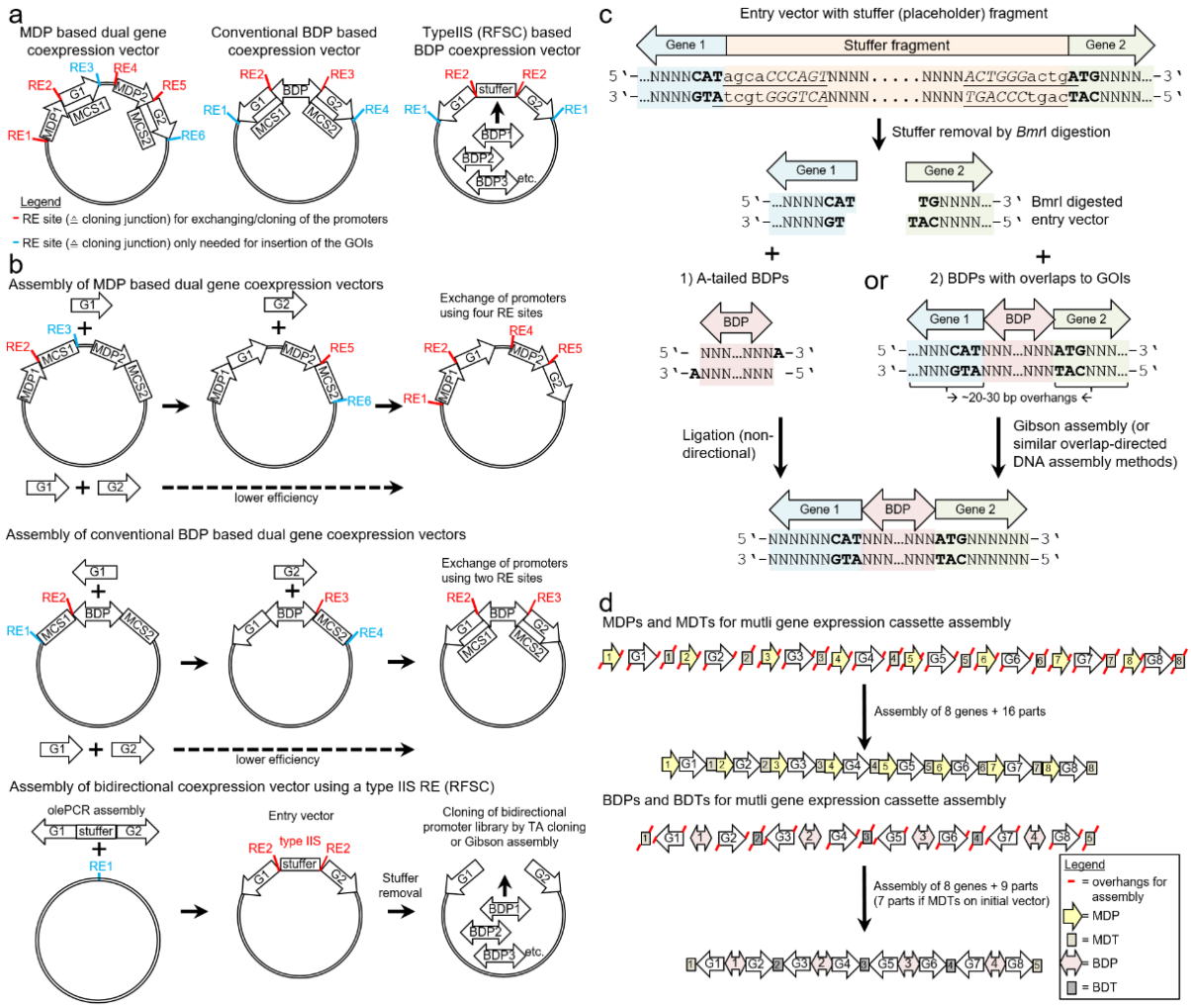
The hybrid promoter assemblies and additional controls suggest different promoter architectures for P_{AOX1} and P_{DAS1}

In all our 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 finding is surprising as a CRM from the *DAS1* promoter (P_{DAS1}-*D5-D7*) stemming from a similar 3' region close to the TATA box, did show strong activation in all contexts tested (Fig. 4 and S 7).

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 7f, 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 of the CRM to the histone core promoter. 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* [87]. 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 8: Molecular cloning of BDPs via TA cloning or Gibson assembly facilitates optimization of dual (a-c) and multi gene co-expression compared to MDPs (d).

A) Dual gene co-expression 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.* [46,49]). 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), 'RSFC': restriction site free cloning strategy [66]

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 (RSFC [66]) of BDPs by TA cloning [98] or providing junctions for Gibson assembly [65]. The start codons of the two genes are written bold, the entire *Bmr*I 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

BDPs facilitate molecular cloning for dual gene and multi gene co-expression

With a suitable library of 168 BDPs at hand, we devised a cloning strategy and optimized vectors for applying the BDPs for dual gene and multi gene co-expression (S 8). Using conventional MDPs for optimization of dual gene co-expression requires in total six unique RE sites to insert genes of interest (GOIs) and testing different promoters (S 8a,b). Currently used bidirectional vectors (*e.g.* [46,49]) rely on a fixed bidirectional promoter and sequential cloning steps using multiple cloning sites (MCSs [requiring at least four RE sites]) (S 8a,b). This strategy enables the basic use of BDPs for gene co-expression but is unfeasible for testing a library of BDPs.

Here, we use a cloning strategy based on the removal of a stuffer (placeholder) fragment via a type IIS RE [66] in combination with TA cloning [98] or Gibson assembly [65] (S 8c) allowing RE site free, seamless cloning of a large number of BDPs. 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 8b). In a subsequent cloning step the stuffer fragment is entirely cleaved out using a single type IIS RE (*Bmr*I) resulting in vector ends suitable for inserting PCR amplified BDPs S 8c). *Bmr*I recognizes a non-palindromic sequence and cleaves in a variable sequence outside of its recognition sequence, as previously applied for a resection site free cloning approach (RSFC) [66]. We positioned the variable 3' overhang generated by *Bmr*I in the beginning of the start codon of the two genes, resulting in 3' thymidine overhangs on both sides of the vector (S 8c). Adenine-tailed PCR fragments of BDPs can be directly cloned into the vector by TA cloning complementing the start codons.

Thereby, in total only two REs are needed for preparing the promoter library in 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 [99]. 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 co-expressed (extended discussion below). 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.

In addition to dual gene co-expression (S 8a-c), BDPs and BDTs (bidirectional terminators) facilitate also the assembly of multi gene expression cassettes in comparison to MDPs and MDTs (monodirectional terminators) (S 8d). Typically the efficiency of overlap-directed DNA assembly methods is decreasing with the number of fragments in the assembly [100]. 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 [100]). 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 [24,25].

Using Gibson assembly or similar overlap directed cloning methods for the cloning of BDPs relying on identical core promoters.

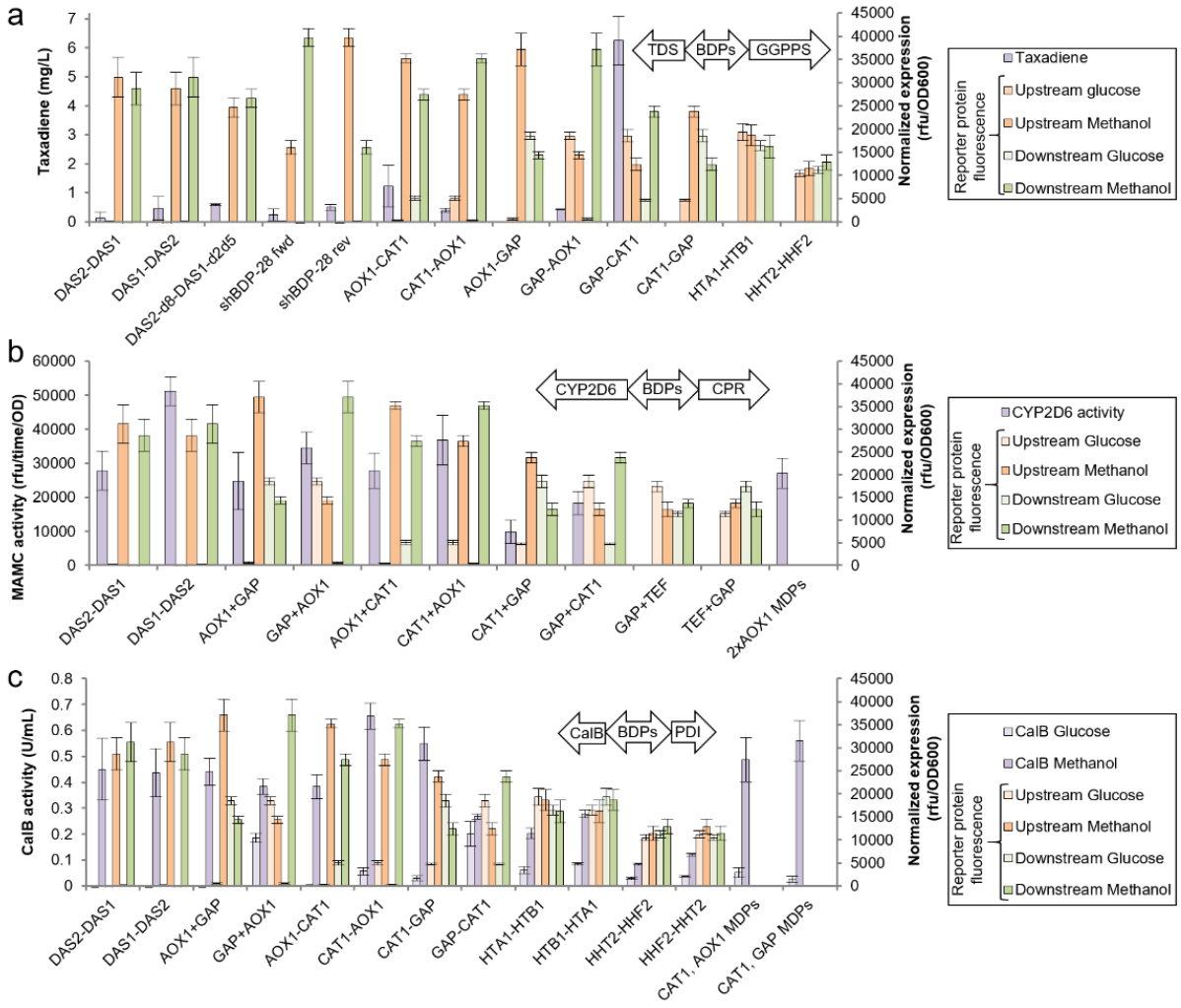
Alternatively to RE based cloning, overlap-directed DNA assembly methods such as Gibson assembly [65], CPEC [101] or SLIC/SLiCE [102,103] can be used requiring overlapping regions (here referred to as 'cloning junctions') with the vector. Thus for each gene pair a new set of primers is needed to amplify the MDPs to be tested for fine-tuning co-expression.

Cloning by overlap based cloning strategies such as Gibson assembly requires identical sequences of the ends of the PCR products of the BDPs with the two genes of interest. In the library approach described in this paper, the BDPs are PCR amplified and then cloned into the vector. For each bidirectional promoter side tested, a separate primer would be needed using this approach, which would render the library amplification of a larger number of promoters rather expensive.

However, a large set of the BDPs described in this paper can be amplified with only two primers: The natural P_{HHX2} variants (32 BDPs, Fig. 2c,d) and the synthetic hybrid BDPs (31 BDPs, Fig. 4) contain identical core promoters. Hence, their ends are matching and they can be amplified with the same primers. These more than 60 promoters cover a wide range of expression levels, ratios, regulation and are exceptionally short, providing a high relative expression efficiency (Fig. 5c).

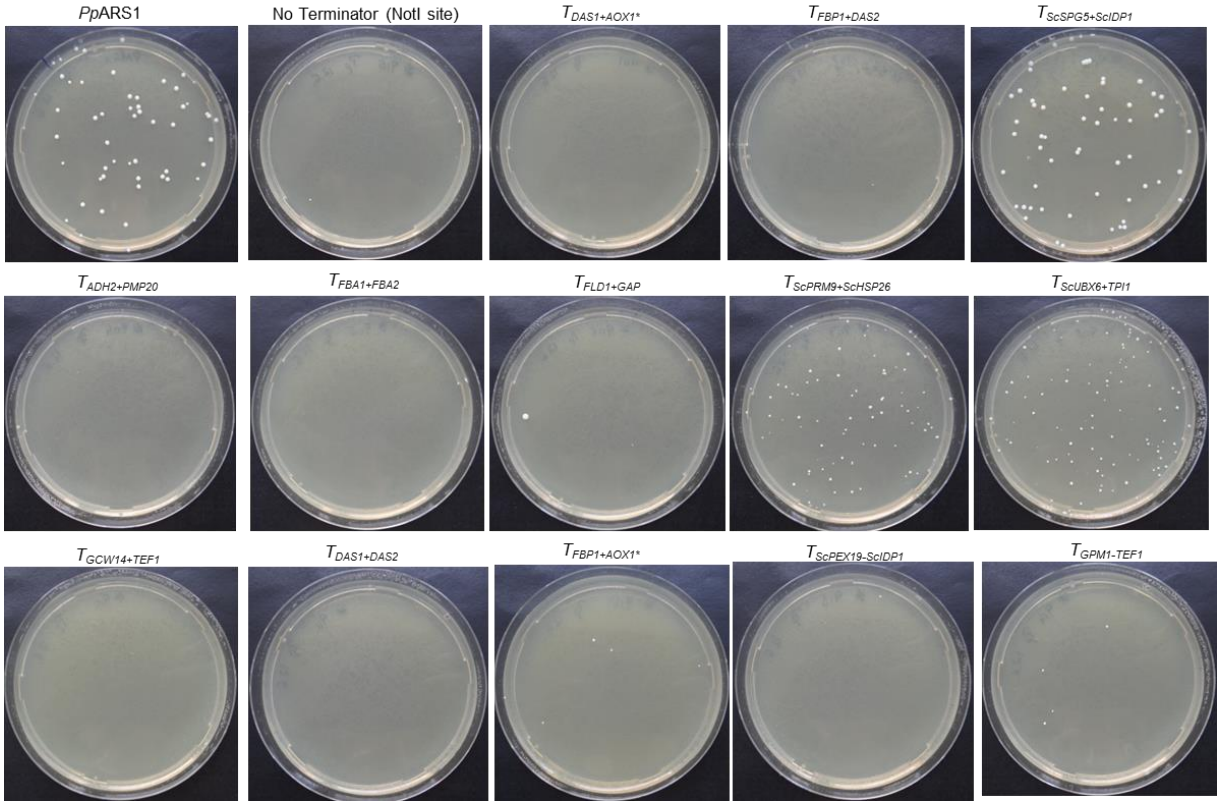
For cloning of a larger number of constructs and the availability of suitable screening systems, TA cloning is favorable. Due to higher efficiencies and less errors in the cloning junctions, we recommend Gibson assembly/overlap based cloning if a small set BDPs is to be tested.

S 9 (Reporter protein fluorescence dual gene co-expression)



S 9: Titers/activities of taxadiene (a), CYP2D6 (b) and CalB (c) do not correlate with the respective reporter protein fluorescences of the BDPs used, highlighting the need for gene pair specific co-expression optimization. The same of data of the titers/activity measurements shown in Fig. 6a-c is combined with the reporter protein measurements of the respective promoters (shown in Fig. 1 to Fig. 4).

S 10 (ARS function of bidirectional terminators, BDTs)



S 10: Autonomously replicating sequence (ARS) function of the bidirectional terminators generated.

In *S. cerevisiae*, transcription termination and autonomously replicating sequence function are associated [104]. In previous work on *P. pastoris* MDTs [30] we also noticed that some terminators show ARS function. Terminators with ARS function should be avoided as they may lead to increased background growth and strain instability of episomally replicating sequences. Therefore we tested all BDTs reported here for ARS function by transforming *P. pastoris* cells with 10 ng of the circular plasmids (see [30] for detailed information on the principle). We included a positive control of a *P. pastoris* ARS sequence [105] cloned into the vector backbone of the *P. pastoris* plasmid used (the same control was also included in previous work on MDTs [30]). All MDTs and BDTs were compared in a single experiment of which the MDT data was published [30], and its controls are also relevant, and therefore included here. 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 *Not*I 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 [106]. However circular plasmids show much lower efficiencies than linear DNA providing free ends [106] (which are typically generated for transformations of *P. pastoris* by linearizing the plasmids). Interestingly, combinations of *S. cerevisiae* terminators showed the 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 [30]. 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 previously 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.

Supporting references

76. Küberl, A. *et al.* High-quality genome sequence of Pichia pastoris CBS7435. *J Biotechnol* **154,** 312–20 (2011).

77. Liang, 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* **13,** 738 (2012).

78. Hesketh, A. R., Castrillo, J. I., Sawyer, T., Archer, D. B. & Oliver, S. G. Investigating the physiological response of Pichia (Komagataella) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. *Appl. Microbiol. Biotechnol.* **97,** 9747–9762 (2013).

79. Chang, D. T.-H., Wu, C.-Y. & Fan, C.-Y. A study on promoter characteristics of head-to-head genes in Saccharomyces cerevisiae. *BMC Genomics* **13 Suppl 1,** S11 (2012).

80. Hecht, A., Endy, D., Salit, M. & Munson, M. S. When Wavelengths Collide: Bias in Cell Abundance Measurements Due to Expressed Fluorescent Proteins. *ACS Synth. Biol.* **5,** 1024–1027 (2016).

81. Vogl, T. *et al.* Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors. *Biotechnol. Bioeng.* **115,** 1037–1050 (2018).

82. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7877–82 (2002).

83. Kranthi, B. V., Kumar, H. R. V. & Rangarajan, P. N. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris. *Yeast* **27,** 705–11 (2010).

84. Hartner, F. S. *et al.* Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Res.* **36,** e76 (2008).

85. Xuan, Y. *et al.* An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. *FEMS Yeast Res.* **9,** 1271–82 (2009).

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

87. Kranthi, B. V., Kumar, R., Kumar, N. V., Rao, D. N. & Rangarajan, P. N. 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 (2009).

88. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).

89. Huang, X. & Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12,** 337–357 (1991).

90. Perez-Pinera, P. *et al.* Synthetic biology and microbioreactor platforms for programmable production of biologics at the point-of-care. *Nat. Commun.* **7,** 12211 (2016).

91. Lin-Cereghino, G. P. *et al.* Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. *Mol. Cell. Biol.* **26,** 883–97 (2006).

92. Wu, G. *et al.* Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* **34,** 652–664 (2016).

93. van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* **35,** 145–153 (2017).

94. Duttke, S. H. C. *et al.* Perspectives on Unidirectional versus Divergent Transcription. *Mol. Cell*

60, 348–349 (2015).

95. Andersson, R. *et al.* Human Gene Promoters Are Intrinsically Bidirectional. *Mol. Cell* **60,** 346–347 (2015).

96. Duttke, S. H. C. *et al.* Human promoters are intrinsically directional. *Mol. Cell* **57,** 674–84 (2015).

97. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483,** 295–301 (2012).

98. Mead, D. A., Pey, N. K., Herrnstadt, C., Marcil, R. A. & Smith, L. M. A universal method for the direct cloning of PCR amplified nucleic acid. *Biotechnology. (N. Y).* **9,** 657–63 (1991).

99. Crook, N. C., Freeman, E. S. & Alper, H. S. Re-engineering multicloning sites for function and convenience. *Nucleic Acids Res.* **39,** e92 (2011).

100. de Kok, S. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* **3,** 97–106 (2014).

101. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6,** 242–51 (2011).

102. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4,** 251–6 (2007).

103. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40,** e55 (2012).

104. Chen, S., Reger, R., Miller, C. & Hyman, L. E. Transcriptional terminators of RNA polymerase II are associated with yeast replication origins. *Nucleic Acids Res.* **24,** 2885–93 (1996).

105. Cregg, J. M., Barringer, K. J., Hessler, A. Y. & Madden, K. R. Pichia pastoris as a host system for transformations. *Mol. Cell. Biol.* **5,** 3376–85 (1985).

106. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 6354–8 (1981).

Part II

Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase

Thomas Kickenweiz[1, 2], Anton Glieder[2*], Jin Chuan Wu [1*]

[1] Institute of Chemical and Engineering Sciences, Agency for Sciences, Technology and Research (A*STAR), 1 Pesek Road, Jurong Island, Singapore 627833;

[2] Institute of Molecular Biotechnology, Technical University of Graz, Petersgasse 14, 8010 Graz, Austria

Manuscript published in *Applied Microbiology and Biotechnology*, February 2018, Volume 102, Issue 3, pp 1297–1306

DOI: https://doi.org/10.1007/s00253-017-8656-z

Abstract

Cellulose is a highly available and renewable carbon source in nature. However, it cannot be directly metabolized by most microbes including *Komagataella phaffii* (formerly *Pichia pastoris*), which is a frequently employed host for heterologous protein expression and production of high-value compounds. A *K. phaffii* strain was engineered that constitutively co-expresses an endo-glucanase and a β-glucosidase both from *Aspergillus niger* and an exo-glucanase from *Trichoderma reesei* under the control of bidirectional promoters. This engineered strain was able to grow on cellobiose and carboxymethyl cellulose (CMC) but not on Avicel. However, the detected release of cellobiose from Avicel by using the produced mixture of endo-glucanase and exo-glucanase as well as the released glucose from Avicel by using the produced mixture of all three cellulases at 50°C indicated the production of exo-glucanase under the liquid culture conditions. The successful expression of three cellulases in *K. phaffii* demonstrated the feasibility to enable *K. phaffii* to directly use cellulose as a carbon source for producing recombinant proteins or other high-value compounds.

Keywords: *K. phaffii*, cellulose, cell-engineering, cellulases, cell growth, protein expression

Introduction

Cellulose is the most abundant organic molecule in the biosphere and the major fraction of lignocellulose. (Dashtban et al. 2009). It is a linear polymer of glucose residues which are interlinked with β 1→4 glycosidic bonds. Huge amounts of lignocellulosic biomass accumulate as waste from different industries (e.g. food, paper) and agriculture (Bayer et al. 2007; Dashtban et al. 2009; Juturu and Wu 2014). This waste is mostly burned directly or biologically degraded to greenhouse gases but is also increasingly used for biofuel production by conversion of lignocellulose to fermentable sugars and further conversion to final products. Since food, feed, energy and chemical demands are increasing worldwide while fossil fuel sources are limited, improving the efficiency of making use of renewables has high priority (Papanikolaou and Aggelis 2011; Yamada et al. 2013). Therefore, research on lignocellulosic waste especially degradation of cellulose has become increasingly important during the last decades. The current focus on improvement of bioethanol production can be summarized in three categories i) availability of lignocellulosic biomass and optimization of its pre-treatment ii), cost reduction of cellulases and iii) improvement of the enzymatic saccharification efficiency of cellulose to glucose (Dashtban et al. 2009; Klein-Marcuschamer and Blanch 2015).

Cellulases play a key role in the enzymatic degradation of cellulose to glucose by cleaving β 1→4 glycosidic bonds. β-glucosidases, endo- and exo-glucanases are three key enzymes involved in this process (Kostylev and Wilson 2012; Teeri 1997). Exo-glucanases cleave cellobiose from the ends in the crystalline region of cellulose molecules. The endo-glucanases cleave cellulose randomly in the amorphous regions and β-glucosidases cleave cellobiose further into two glucose molecules. The interaction of endo- and exo-glucanases has a synergistic effect on cellulose degradation (Kostylev and Wilson 2012; Teeri 1997). Recent discoveries like the cellulolytic activity of GH61 enzymes showed that further enzymes are needed for the complete degradation of crystalline cellulose to glucose (Morgenstern et al. 2014). Efficient cellulose-degrading fungi secrete endo- and exo-glucanases, β-glucosidases and GH61 enzymes (Mathew et al. 2008; Morgenstern et al. 2014). Most prominent fungi which are used in biotechnology as model organisms for cellulose degradation and for production of cellulases are filamentous fungi like *Trichoderma* sp., *Aspergillus* sp. and *Neurospora* sp. (Mathew et al. 2008; Tian et al. 2009). There are also some oleaginous yeasts with cellulolytic activities belonging mainly to the genera *Trichosporon* and *Cryptococcus*. These yeasts are used for production of single cell oil from

pre-treated lignocellulosic biomass which can be converted into biofuels (Dennis 1972; Liu 2012 et al.; Papanikolaou and Aggelis 2011; Stursova et al. 2012).

Beside natural cellulolytic yeasts, non-cellulolytic yeasts were engineered enabling a one-step conversion of cellulose to ethanol by heterologous protein expression of cellulases. *Saccharomyces cerevisiae* and *Kluyveromyces marxianus* were successfully engineered in which endo- and exo-glucanases and β-glucosidase were co-expressed in both strains (Chang et al. 2012; Den Haan et al. 2007; Yamada et al. 2013). The co-expression of all 3 cellulases enabled *K. marxianus* to grow on carboxymethyl cellulose (CMC) but not on phosphoric-acid swollen cellulose (PASC) or Avicel (Chang et al. 2012). It was reported that a *S. cerevisiae* strain was engineered to grow on PASC by co-expression of an endo-glucanase and a β-glucosidase (Den Haan et al. 2007). Very recently, an engineered *S. cerevisiae* strain was shown to give increased ethanol production by displaying an optimized mixture of cellulolytic enzymes (Liu et al. 2017). Moreover, Guo and co-workers demonstrated the growth of an engineered *Yarrowia lipolytica* strain on pre-treated lignocellulose (Guo et al. 2017).

The yeast *Komagataella phaffii* (formerly *Pichia pastoris*) is widely used for heterologous gene expression. One important advantage of *K. phaffii* as expression host is that only a very low amount of endogenous proteins are naturally secreted. The heterologous expressed proteins which are secreted are the vast majority of total protein in the supernatant of a *K. phaffii* culture (Cereghino and Cregg 2000; Vogl et al. 2013). It has tremendous applications in protein characterization, industrial production of proteins (e.g. pharmaceuticals) and other high value products (Cereghino and Cregg 2000; Geier et al. 2012, 2013; Vogl et al. 2013; Wriessnegger et al. 2014).

As a Crabtree negative yeast, *K. phaffii* does not produce high amounts of ethanol like Crabtree positive yeasts during cultivation conditions. Therefore, energy and carbon sources can be more efficiently used for products of interests as high-value compounds (Hagmann et al. 2014, Osawa et al. 2009; Vogl et al. 2013). Geier and co-workers demonstrated that expression and engineering of human cytochrome P450 in *K. phaffii* can be used for biocatalytic applications (Geier et al. 2013). In addition, comparison of expression levels of human cytochrome P450 in 4 different expression hosts (*Escherichia coli, S. cerevisiae, Y. lipolytica* and *K. phaffii*) showed that P450 was most efficiently expressed in *K. phaffii*. (Geier et al. 2012). Recently, *K. phaffii* was used as a whole-cell biocatalyst to produce high value aroma compound (+)-nootkatone (Wriessnegger et al. 2014). Moreover, it has also been shown to produce high value compounds β-carotene and violacein in the same *K. phaffii*

strain by co-expression of the carotenoid and violacein biosynthesis pathways (Geier et al. 2015), demonstrating new opportunities for efficient pathway design and expression in this yeast.

Therefore, a *K. phaffii* strain which is able to use natural cellulose or lignocellulosic hydrolysate as carbon source for production of high-value compounds would be a very attractive platform strain for industrial applications. Here we present the work of *K. phaffii* strains that express cellulases from filamentous fungi *Aspergillus niger* and *Trichoderma reesei* using bidirectional promoters. To the best of our knowledge, this is the first report that endo- and exo-glucanases and β-glucosidase were co-expressed in a single *K. phaffii* strain.

Material and Methods

Strains and culture conditions

K. phaffii strains BG10 (BioGrammatics, Carlsbad, CA, USA) and BSY11G1 (a Δ*gut1*, Δ*aox1* derivative of BG10) (Bisy, Hofstaetten/Raab, Austria) which is glycerol auxotrophic were used. *K. phaffii* was cultivated in YPD medium containing 2% (w/v) glucose, 2% (w/v) peptone and 1% (w/v) yeast extract (1.5% (w/v) agar YPD for agar plates). When selective markers were used, the antibiotic concentrations were 50 mg/L zeocin and 300 mg/L geneticin, respectively. The yeast cultures were incubated at 30°C and liquid cultures were shaken at 130 rpm. Buffered minimal medium (BM_ medium) was used for cultivation of *K. phaffii* in growth experiments and cellulase activity assay. BM_ medium was made as previously described (Weis et al. 2004) but with no biotin added to the medium.

E. coli Top10 and *E. coli* Top10F' (both from Thermo Fischer Scientific, Waltham, MA, USA) strains were transformed for vector amplification and cloning experiments. *E. coli* strains were incubated in LB medium. When selective markers were used, the antibiotic concentrations were 100 mg/L ampicillin, 50 mg/L kanamycin and 25 mg/L zeocin, respectively. The *E. coli* cultures were incubated at 37°C and liquid cultures were shaken at 100 rpm.

T. reesei QM9414 (ATCC 26921, CBS 392.92) and *A. niger* DSM 26641 (Ottenheim et al. 2015) were used for isolation of the genes coding for cellulases. For isolation of genomic DNA, *T. reesei* QM9414 was incubated in potato dextrose medium containing 4 g/L potato extract and 20 g/L glucose. The culture was incubated in baffled flasks at 28°C and 100 rpm for 3 days. Genomic DNA of *T. reesei* and *K. phaffii* were isolated following the method of

Namjin Chung (Balakrishnan et al. 2013). The cultivation of *A. niger* DSM 26641 was performed as previously described (Ottenheim et al. 2014).

Isolation of genes

The used genes are summarized in Table 1. All components for PCR were from Thermo Fischer Scientific (Waltham, MA, USA) if not stated otherwise. Phusion polymerase was used for PCR amplification. PCR was performed following the manufacturer's protocol. Standard overlap-extension PCR was performed as previously described (Näätsaari et al. 2012). The DNA primers were made by Integrated DNA Technologies (IDT, Coralville, IA, USA). A list of the DNA primers used is shown in Supplemental Table S1.

AnBGL1 (MF981921)**:** The gene sequence coding for β-glucosidase from *A. niger* (AM270402.1, gene <78981..>81928) was codon-optimized and synthesized by GenScript (Piscataway, NJ, USA). Its natural signal sequence was replaced with a codon-optimized version of *S. cerevisiae* alpha mating factor pre-pro signal sequence for protein secretion. A *Spe*I restriction site was added at the 3' end of the *AnBGL1* sequence. The ordered DNA sequence of the codon-optimized *AnBGL1* gene is shown in Supplemental Fig. S1. The synthesized *AnBGL1* gene was cloned by Genscript (Piscataway, NJ, USA) into pUC57-Mini plasmid.

TrBGL1 (U09580.1): The gene coding for the β-glucosidase from *T. reesei* was isolated from genomic DNA of *T. reesei* QM9414. The exons in the open reading frame of *TrBGL1* were amplified by PCR and fused in order by overlap PCR. The gene was isolated without its native signal sequence for secretion because the gene was fused with *S. cerevisiae* alpha mating factor pre-pro signal sequence instead. In addition, *Bmr*I restriction sites were removed by changing one codon of the recognition site without changing the amino acid sequence. To amplify only the exons and to remove the *Bmr*I restriction sites, *TrBGL1* was divided into four parts. These four parts were amplified by PCR with the primer pairs P1+P2, P4 + P5, P6 + P7 and P9 + P10. Part 3 of *TrBGL1* had to be amplified in a subsequent PCR step with the primer pair P7 + P8 for adding nucleotides which overlap with the sequence of part 2 for overlap extension PCR. All 4 DNA fragments were attached to each other by several overlap-extension PCR steps.

TrCBH2 **HM:** *TrCBH2*-HM is a codon-optimized gene variant coding for exo-glucanase CBH2 from *T. reesei* which had been used in a previous work (Mellitzer et al. 2014). A plasmid from this previous study was used as the template to get the sequence of *TrCBH2*-HM. *TrCBH2*-HM was amplified with the primers P11 + P12 for further work.

***TrCBH2* V09:** *TrCBH2*-V09 is a codon-optimized gene variant coding for exo-glucanase CBH2 from *T. reesei* which had been used in a previous work (Mellitzer et al. 2014). The whole gene sequence was ordered as gBlock from IDT (Coralville, IA, USA). One codon was changed to remove a *Bgl*II restriction site without changing the amino acid sequence.

AnEG-A (MF981920)**:** *AnEG-A* is coding for endo-glucanase A from *A. niger* DSM 26641. *AnEG-A* was amplified by PCR with the primers P35 + P36 using cDNA as the template. mRNA isolation from *A. niger* DSM 26641 and its transcription to cDNA was performed as previously described (Ottenheim et al. 2014). The gene sequence of *AnEG-A* is shown in Supplemental Fig. S2.

TrEG1 (M15665.1): *TrEG1* is coding for endo-glucanase 1 from *T. reesei*. Like *TrBGL1*, *TrEG1* was isolated from genomic DNA of QM9414 without the sequence coding for the native signal sequence. The exons were amplified by PCR and fused in order by overlap PCR. To amplify only the exons, *TrEG1* was divided into 2 parts. Exon 1 was amplified with the primers P18 + P19. Exon 3 was added during amplification of exon 2 by using reverse primers with overhangs (first PCR: P20 + P21 and second PCR: P20 + P22).

Vector construction

All components used for cloning were from Thermo Fischer Scientific (Waltham, MA, USA) unless otherwise specified. In parallel to the restriction enzymes from Thermo Fischer Scientific (Waltham, MA, USA), also enzymes from New England Biolabs (Ipswich, MA, USA) were used. PCR products were purified using Wizard SV Gel and PCR Clean-Up System from Promega (Fitchburg, WI, USA). All original plasmids for cloning work were from Pichia pool1 of Graz University of Technology (Näätsaari et al. 2012). For expression of cellulase genes, constitutive bidirectional promoters NB2 (synonyms: pHTX1 or natbidi 2) and NB3 (synonyms: pHHX2 or natbidi 3) were used. Naturally, NB2 and NB3 are regulating histone genes in *K. phaffii* (Vogl et al. 2015; Vogl et al. manuscript submitted). The sequence of NB2 was amplified using *K. phaffii* BG10 genomic DNA as a template. The sequences of NB3 and *eGFP* were ordered together as one gBlock from IDT (Coralville, IA, USA). The construction of the expression vectors is described in detail in supplemental data. All used vectors are listed in Supplemental Table S2.

For engineering a *K. phaffii* strain that expresses endo- and exo-glucanases and a β-glucosidase, the two expression vectors pPpKan_int2_SwaI_AOX1tt_NotI_TrCBH2-V09_NB2_AnEG-A_NotI_AOX1tt and pPpZeo_int8_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt were constructed. The maps of these

plasmids are shown in Fig.1, which were generated by using SnapGene Viewer software (GSL Biotech, Chicago, IL, USA).

For using the bidirectional promoter NB2 for mono-directional expression of only one cellulase gene (β-glucosidase, endo- or exo-glucanase) in control experiments, a stop codon was inserted 9 bases downstream of a start codon followed by a terminator sequence in its second orientation (Fig. 2).

Pichia transformation

K. phaffii transformation and making of competent cells were performed in 2 different ways. The condensed method protocol was performed as previously described (Lin-Cereghino et al. 2005). The second protocol described by Wu and Letchworth (2004) was performed with a few modifications. *K. phaffii* was inoculated in 25-50 mL of YPD to an OD_{600} of 0.4 and incubated until the culture reached an OD_{600} of 0.8-1. In the final step, the competent cells were resuspended in 1 mL of BEDS or 1 M sorbitol, respectively, and 50 to 80 µL of aliquots of the competent cell suspension was used for transformation and 500-1500 ng of linearized vectors were added to the cells. After electroporation, 1 mL of recovery medium (1:1 mixture of YPD medium with 1 M sorbitol) was added to the transformation sample. The cells were recovered at 30°C for 1-1.5 h without shaking before plating aliquots on YPD plates containing the corresponding antibiotic selection markers.

Cellulase activity assay and HPLC analysis

For detecting cellulase activity, *K. phaffii* strains which expressed cellulases were incubated in buffered minimal medium. ONCs of the test strains were used to inoculate 7 mL of BM_glycerol 1% (w/v) in 50 mL tubes (Tarsons, Kolkata, India) to an OD_{600} of 0.1. The tubes were fixed in a tilted position and the cap was loosened for aeration. The cultures were incubated at 30°C and 120 rpm for 2 days to reach the stationary phase. Subsequently, the cells were harvested and centrifuged at 3,220 × g for 5 min. The supernatant (0.3 mL) was mixed with 0.9 mL of 50 mM citrate buffer (pH5.5) containing 1% Avicel in a 1.5 mL tube. The 1.5 mL tubes were incubated horizontally at 50°C and 120 rpm for 4 h. The samples were centrifuged at 20,800 × g for 1 min and the supernatants were filtered through a 0.22 µm membrane. The measurement of cellobiose and glucose formed from Avicel using HPLC was performed as previously described (Ong et al. 2016).

Growth experiments

Growth of *K. phaffii* on cellobiose and CMC was monitored by measuring OD_{600}. The spectro-photometers Shimadzu UV-1800 (Shimadzu, Kyoto, Japan) and Eppendorf Bio Photometer plus (Eppendorf, Hamburg, Germany) were used for OD_{600} measurement. Buffered minimal liquid medium was used for the growth experiments. CMC and Avicel (crystalline cellulose) were used as model substrates for cellulose degradation to evaluate the cellulolytic activity of the engineered *K. phaffii* strains. CMC is water soluble chemically modified cellulose. It can be degraded well by endo-glucanases into shorter molecules (Teeri 1997). The concentrations of cellobiose in BM_cellobiose were 0.25% (w/v) and 0.5% (w/v), respectively. In BM_CMC, the concentration of CMC was 0.5% (w/v) and 1% (w/v), respectively. The different concentrations of the carbon sources in the media did not interfere the outcome of the experiments since these experiments were performed to see if the engineered strains were able to grow on the testing carbon sources. The cultures were inoculated to an OD_{600} of 0.05 or 0.1, respectively. The volume of the cultures was 25 mL for BM_cellobiose and 50 mL for BM_CMC, because cultivation on CMC took longer and the cell densities were much lower than in cellobiose. The cultures were incubated at 30°C and 130 rpm for a few days until stationary phase was reached.

To test for growth on Avicel, the strains were incubated in BM_Avicel 0.5% (w/v) + 0.1% (w/v) glycerol. The glycerol was used to initiate cellulase production in the starting phase. The cultures were incubated under the same conditions as stated above for BM_CMC cultures. Growth was monitored by measurement of colony forming units (cfu). For this, 1 mL of the culture was taken daily and diluted 10^{-3}, 10^{-4} and 10^{-5}. These dilutions (25 µL) were spread onto YPD plates. Another approach for detecting growth on different carbon sources was done in buffered minimal medium agar plates using the same carbon source concentration as in liquid medium described above with exception that BM_Avicel 0.5% (w/v) agar plates did not contain glycerol. The test strains were streaked onto BM_carbon source plates and incubated at 30°C for 2-3 days (cellobiose), 4-5 days (CMC) and up to 2 weeks (Avicel), respectively.

Results

Constitutive expression of β-glucosidase in *K. phaffii*

A first important step in engineering a cellulose-metabolizing *K. phaffii* strain was to confirm that constitutive expression of β-glucosidases using bidirectional promoters enables *K. phaffii* to grow on cellobiose. For this, separate expression strains were constructed for β-glucosidases AnBGL1 and TrBGL1, respectively. For expression of AnBGL1, the bidirectional promoter NB3 was used which co-expressed eGFP in its second orientation for a quick indication of expression levels (Fig. 1b). The promoter strength of this bidirectional promoter is very similar on both sides (Vogl et al. 2015; Vogl et al. manuscript submitted). This vector encoding for *AnBGL1* and *eGFP* was integrated into *K. phaffii* BG10 genome. Selected BG10 transformants which have integrated *AnBGL1* were streaked onto BM_cellobiose agar plates to check their ability to grow on cellobiose as the sole carbon source. After 2-3 days' incubation, clear differences were observed between the growth of positive clones (expressing β-glucosidase AnBGL1) and negative control (parental strain BG10) on agar plates, indicating that β-glucosidase was functionally expressed (not shown).

Engineering another β-glucosidase expressing strain was done by using vector encoding for *TrBGL1* to transform *K. phaffii* BSY11G1. TrBGL1 was expressed under control of constitutive bidirectional promoter NB2 using the construction shown in Fig. 2. Similar to BG10 transformants, the selected BSY11G1 transformants expressing TrBGL1 were also able to grow on BM_cellobiose agar plates. This indicates that it is feasible to reproduce the observed growth on cellobiose for recombinant *K. phaffii* strain expressing other β-glucosidases. Cultivation of representative clones of BG10 and BSY11G1 transformants in shake flasks confirmed that constitutive expression of β-glucosidases enabled biomass production of *K. phaffii* by utilization of cellobiose (Fig. 3).

Independent of the genetic backgrounds of the strains, the constitutive expression of AnBGL1 or TrBGL1, respectively, enabled *K. phaffii* to grow in BM_cellobiose broth medium. The respective parental strains (negative controls) did not show any relevant growth when cellobiose was used as the sole carbon source in the medium. Interestingly, the growth curve in Fig. 3b shows that the TrBGL1 expressing strain needed a long starting phase to reach the exponential growth. After 1 day's incubation, the cell density was still very low. The exponential growth just started between 30 h and 40 h of incubation, while cultures on glucose should have already reached stationary phase under such conditions. This indicates that the biomass production depends on the amount of expressed ß-glucosidase and/or its

specific activity, thereby on the release rate of glucose from cellobiose. A certain amount of β-glucosidase has to be secreted into the medium to generate enough glucose for fast cell growth.

Co-expression of endo- and exo-glucanases and β-glucosidase in a single *K. phaffii* strain

K. phaffii BG10 was co-transformed with both expression vectors (Fig. 1) in order to express all three cellulose-hydrolysing enzymes (AnBGL1 + TrCBH2 + AnEG-A). Genomic integration was evaluated by control PCR using isolated genomic DNA of transformants as a template. Selected clones were streaked onto BM_CMC agar plates to check for potential growth on CMC as the sole carbon source. A significant difference in growth between positive clones expressing all three cellulose-hydrolysing enzymes and negative control (parental strain or strains with only one integrated vector as shown in Fig. 1) was observed after 4 day's incubation (not shown). This ability to grow on CMC was confirmed by shake flask experiments (Fig. 4a).The growth curves clearly show that the parental strain BG10 and the strain expressing β-glucosidase AnBGL1 did not grow on CMC, whereas the BG10 strain expressing all three cellulases showed a clear growth (Fig. 4a), though slower than the ß-glucosidase-expressing strains on cellobiose. Interestingly, the strain expressing all three cellulases entered the stationary phase with an OD_{600} of about 1 when the CMC concentration in the medium was 1% (w/v). The experiment was stopped after 55 h's incubation when it seemed to reach the end.

With the fact that the strain expressing all three cellulases can grow on CMC as the sole carbon source, it is still unclear if all 3 expressed cellulases are required for the growth or not. Therefore, two engineered strains which express TrCBH2 + AnBGL1 and TrEG1 + AnBGL1, respectively, were investigated. As mentioned above, AnBGL1 was also constitutively expressed under the NB3 promoter, and TrCBH2 and TrEG1, respectively, were constitutively expressed using only one orientation of the NB2 promoter (Fig. 2). The strain expressing all three cellulases was used as a positive control in this experiment. The growth of engineered control strains in BM_CMC 0.5% is shown in Fig. 4b. It is seen that the strain expressing TrEG1 + AnBGL1 had a similar growth behaviour with the strain expressing all three cellulases. The strain expressing TrCBH2 + AnBGL1 did not reach the OD_{600} of the other strains on CMC. Further experiments showed that the observed initial growth within the first 8 h after inoculation (Fig. 4b) was an experimental artefact caused mainly by glucose contamination from inoculation with the pre-culture which can be reduced but not completely avoided by washing the cells of the pre-culture with H_2O before

inoculation of the main cultures (not shown). Although it remained unclear at this stage if the *TrCbh2* gene was expressed at all, the result indicates that the co-expression of β-glucosidase AnBGL1 and an endo-glucanase is sufficient for growth of *K. phaffii* in media with CMC as the sole carbon source.

Hydrolysis of Avicel by co-expressed enzymes from engineered *K. phaffii* strain

The BG10 strain expressing all three cellulases was evaluated for potential growth on Avicel. No significant growth was observed on BM_Avicel agar plates or in BM_Avicel liquid medium compared to the negative BG10 control strains (not shown). In order to analyse if this is due to a failure in co-expression of all three enzymes or just due to the insufficient enzyme activity to release enough glucose from crystalline cellulose, the supernatant was harvested from different *K. phaffii* cultures after they reached stationary phase and mixed with Avicel. These samples were analysed by HPLC to detect glucose and/or cellobiose after incubation at 50°C to check cooperative action of all three enzymes (Fig. 5).

It is seen that about 0.17 g/L of glucose released from Avicel was detected in the supernatant of the strain expressing all three cellulases. This is a proof that the cellulases mixture from the *K. phaffii* triple enzyme expression strain was able to convert Avicel into glucose. Besides, no cellobiose was detected in this sample whereas cellobiose was detected at 0.06 g/L in the supernatant of the strain expressing TrCBH2 + AnEG-A. As predicted, no glucose or cellobiose was detected in the supernatants of the BG10 strain expressing AnBGL1 and parental strain BG10.

Discussion

So far, successful expression of different β-glucosidases by *K. phaffii* was mainly focused on protein yields and characterization of expressed β-glucosidases (Chen et al. 2011; Dan et al. 2000; Hong et al. 2007; Ramani et al. 2015). It has also been demonstrated that addition of cellobiose improved growth of *K. phaffii* Mut^S (methanol utilization slow) strains in methanol-containing medium when β-glucosidases were expressed under the control of inducible *AOX1* promoter (Hong et al. 2007; Ramani et al. 2015). Here we described for the first time the growth of *K. phaffii* on cellobiose as a sole carbon source as a result of the constitutive β-glucosidase expression using histone promoters. Similar results based on constitutive expression were reported for *S. cerevisiae* and *K. marxianus* (Chang et al. 2012; Van Rensburg et al. 1998).

The engineered cellobiose-utilizing *K. phaffii* strain expressing β-glucosidase AnBGL1 was further modified to co-express an endo-glucanase (AnEG-A) and an exo-glucanase (TrCBH2). This triple hydrolase expression strain as well as the strain co-expressing β-glucosidase (AnBGL1) and endoglucanase (TrEG1) were able to grow on amorphous carboxymethylated cellulose (CMC). Although it is known that many non-cellulolytic organisms are able to hydrolyse CMC with different enzymes acting on β-glucans (Lynd et al. 2002), such an effect can be excluded in the case of *K. phaffii* because the parental strain and the strain expressing only AnBGL1 were unable to grow on CMC. Interestingly, the culture reached an OD_{600} of about 1 in BM_CMC (1%, w/v), which is unexpectedly low (Fig. 4a). This might be due to CMC degradation products which were not further converted to appropriate sugars for *K. phaffii* to digest. Medium viscosity CMC from Sigma Aldrich was used for these growth experiments. The product information describes a substitution grade of 0.65-0.95 per glucose residue in this type of CMC, which might explain the observed growth limitation. The final OD_{600} values with different CMC concentrations (0.5-1%, w/v) seem to correlate directly with the amounts of added CMC, indicating that a possible inhibitory effect from CMC degradation products can be excluded.

Co-expression of both enzymes, endo-glucanase and β-glucosidase was required to enable *K. phaffii* to grow on CMC. The growth on CMC also indicated that the endo-glucanase was functional and reasonably expressed under the control of the bidirectional histone promoter NB2. The additional co-expression of the exo-glucanase TrCBH2 did not support growth on CMC at 30C°. This was in line with previous reports. Mellitzer and co-worker described that TrCBH2 expressed in *K. phaffii* had a much lower activity on CMC than on Avicel (Mellitzer et al. 2012). Furthermore, it was reported that only very low or no activity of TrCBH2 on CMC was detected when it was expressed in its native host *T. reesei* or *Schizosaccharomyces pombe* (Okada et al. 1998). This might be a reasonable explanation why no relevant growth on CMC was observed when just TrCBH2 was co-expressed with AnBGL1.

Culture supernatant of the triple enzyme expression strain released significant amounts of glucose from Avicel when incubated at 50°C. Although the observed glucose concentration of 0.17 g/L should be sufficient to enable significant growth of the engineered strain, no obvious growth of *K. phaffii* was observed at 30°C, which might indicate a too low exo-glucanase (TrCBH2) activity at this temperature. The HPLC analysis showed only a little cellobiose was released from Avicel (Fig. 5) in the sample treated with the supernatant of the strain co-expressing TrCBH2 and AnEG-A.

Therefore, the next steps will be to improve expression and activity of TrCBH2 in this system. Mellitzer and co-workers achieved high cellobiose concentrations from Avicel by multiple *TrCBH2* gene expression in *K. phaffii* using the inducible *AOX1* promoter (Mellitzer et al. 2012), indicating the feasibility of producing higher expression levels of TrCBH2 in *K. phaffii*. Using an alternative promoter is one possibility in addition to engineering CBH2 for higher activity at lower temperatures. The research on promoters in *K. phaffii* had made much progress in the past few years, which increases the chances to find a more effective promoter for expression of TrCBH2 (Vogl et al. 2016). Geier and co-workers established a technology in *K. phaffii* making it possible to integrate multi-gene pathways into the genome and to successfully express them even just with a single alternative strong constitutive promoter (Geier et al. 2015). This method might be used for integration of multiple *TrCBH2* gene copies or for integration of different cellulase genes in the engineered *K. phaffii* strain to improve cellulose degradation to glucose. In conclusion, this work demonstrated that it is possible to use *K. phaffii* as a whole-cell biocatalyst to produce high value compounds using cellulose or even lignocellulosic hydrolysates as cheap and renewable carbon source.

Acknowledgements

This work was financially supported by the Science and Engineering Research Council (SERC) of the Agency for Science, Technology and Research (A*STAR), Singapore (SERC grant no 1526004159, ICES/15-175A02) and by the A*STAR Research Attachment Programme Scholarship from the A*STAR Graduate Academy. Thanks are also given to Dr. Christoph Ottenheim for his support in the work with *A. niger* and critical reading of the manuscript.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical statement

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Balakrishnan B, Ayyavoo J, Sadayan P, Abimannan A (2013) Evaluation of antioxidant activity of *Clitoria ternatea* and *Alternanthera sessilis* plant extracts using model system for yeast cells. Afr J Basic Appl Sci 5: 134-138. doi: 10.5829/idosi.ajbas.2013.5.3.1135

Bayer EA, Lamed R, Himmel ME (2007) The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol 18: 237–245. doi: https://doi.org/10.1016/j.copbio.2007.04.004

Chang J-J, Ho C-Y, Ho F-J, Tsai T-Y, Ke H-M, Wang C. H-T, Chen H-L, Shih M-C, Chieh-Chen Huang C-C, Li W-H (2012) PGASO: A synthetic biology tool for engineering a cellulolytic yeast. Biotechnol Biofuels 5: 53. doi: 10.1186/1754-6834-5-53

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24: 45-66. doi: https://doi.org/10.1111/j.1574-6976.2000.tb00532.x

Chen P, Fu X, Ng TB, Ye X-Y (2011) Expression of a secretory β-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. Biotechnol Lett 33: 2475–2479. doi: 10.1007/s10529-011-0724-3

Dan S, Marton I, Dekel M, Bravdo B-A, He S, Withers SG, Shoseyov O (2000) Cloning, expression, characterization, and nucleophile identification of family 3, *Aspergillus niger* β-glucosidase. J Biol Chem 275: 4973–4980. doi: 10.1074/jbc.275.7.4973

Dashtban M, Schraft H, Qin W (2009) Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci 5: 578-595. doi:10.7150/ijbs.5.578

Den Haan R, Rose SH, Lee R, Lynd LR, van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 9: 87–94. doi: https://doi.org/10.1016/j.ymben.2006.08.005

Dennis C (1972) Breakdown of cellulose by yeast species. J Gen Microbiol 71: 409-411. doi: 10.1099/00221287-71-2-409

Geier M, Braun A, Emmerstorfer A, Pichler H, Glieder A (2012) Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes – a comparative study. Biotechnol J 7: 1346–1358. doi: 10.1002/biot.201200187

Geier M, Braun A, Fladischer P, Stepniak P, Rudroff F, Hametner C, Mihovilovic MD, Glieder A (2013) Double site saturation mutagenesis of the human cytochrome P450 2D6 results in regioselective steroid hydroxylation. FEBS J 280: 3094–3108. doi: 10.1111/febs.12270

Geier M, Fauland P, Vogl T, Glieder A (2015) Compact multi-enzyme pathways in *P. pastoris*. Chem Commun 51: 1643. doi: 10.1039/C4CC08502G

Guo Z-P, Duquesne S, Bozonnet S, Cioci G, Nicaud J-M, Marty A, O'Donohue MJ (2017) Conferring cellulose-degrading ability to *Yarrowia lipolytica* to facilitate a consolidated bioprocessing approach. Biotechnol Biofuels 10: 132**.** doi: 10.1186/s13068-017-0819-8

Hagman A, Säll T, Piskur J (2014) Analysis of the yeast short-term Crabtree effect and its origin. FEBS J 281: 4805–4814. doi: 10.1111/febs.13019

Hong J, Tamaki H, Kumagai H (2007) Cloning and functional expression of thermostable β-glucosidase gene from *Thermoascus aurantiacus*. Appl Microbiol Biotechnol 73: 1331–1339. doi: 10.1007/s00253-006-0618-9

Juturu V, Wu JC (2014) Microbial cellulases: Engineering, production and applications. Renew Sust Energ Rev 33: 188–203. doi: https://doi.org/10.1016/j.rser.2014.01.077

Klein-Marcuschamer D, Blanch HW (2015) Renewable fuels from biomass: Technical hurdles and economic assessment of biological routes. AIChE J 61: 2689–2701. doi: 10.1002/aic.14755

Kostylev M, Wilson D (2012) Synergistic interactions in cellulose hydrolysis. Biofuels. 3: 61-70. doi: http://dx.doi.org/10.4155/bfs.11.150

Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, Lin-Cereghino GP (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 38: 44–48

Liu W, Wang Y, Yu Z, Bao J (2012) Simultaneous saccharification and microbial lipid fermentation of corn stover by oleaginous yeast *Trichosporon cutaneum*. Bioresour Technol 118: 13–18. doi: https://doi.org/10.1016/j.biortech.2012.05.038

Liu Z, Inokuma K, Ho S-H, den Haan R, van Zyl WH, Hasunuma T, Kondo A (2017) Improvement of ethanol production from crystalline cellulose via optimizing cellulase ratios in cellulolytic *Saccharomyces cerevisiae*. Biotechnol Bioeng 114: 1201–1207. doi: 10.1002/bit.26252

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66: 506-577 doi: 10.1128/MMBR.66.3.506-577.2002

Mathew GM, Sukumaran RK, Singhania RR, Pandey A (2008) Progress in research on fungal cellulases for lignocellulose degradation. J Sci Ind Res 67: 898-907

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

Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, Purkarthofer T, Glieder A (2014) Synergistic modular promoter and gene optimization to push cellulase secretion by *Pichia pastoris* beyond existing benchmarks. J Biotechnol 191: 187–195. doi: https://doi.org/10.1016/j.jbiotec.2014.08.035

Morgenstern I, Powlowski J, Tsang A (2014) Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. Brief Funct Genomics 13: 471-481. doi: https://doi.org/10.1093/bfgp/elu032

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7: e39720. doi: https://doi.org/10.1371/journal.pone.0039720

Okada H, Sekiya T, Yokoyama K, Tohda H, Kumagai H, Morikawa Y (1998) Efficient secretion of *Trichoderma reesei* cellobiohydrolase II in *Schizosaccharomyces pombe* and characterization of its products. Appl Microbiol Biotechnol 49: 301-308. doi: 10.1007/s002530051173

Ong SA, Ng ZJ, Wu JC (2016) Production of high concentration of L-lactic acid from cellobiose by thermophilic *Bacillus coagulans* WCP10-4. Appl Microbiol Biotechnol 100: 6501–6508. doi: 10.1007/s00253-016-7600-y

Osawa F, Fujii T, Nishida T, Tada N, Ohnishi T, Kobayashi O, Toshihiro Komeda T, Yoshida S (2009) Efficient production of L-lactic acid by Crabtree-negative yeast *Candida boidinii*. Yeast 26: 485–496. doi: 10.1002/yea.1702

Ottenheim C, Verdejo C, Zimmermann W, Wu JC (2014) Hemicellulase production by *Aspergillus niger* DSM 26641 in hydrothermal palm oil empty fruit bunch hydrolysate and transcriptome analysis. J Biosci Bioeng 118: 696-701. doi: https://doi.org/10.1016/j.jbiosc.2014.05.014

Ottenheim C, Meier K, Zimmermann W, Wu JC (2015) Isolation of filamentous fungi exhibiting high endoxylanase activity in lignocellulose hydrolysate. Appl Biochem Biotechnol 175: 2066-2074. doi: 10.1007/s12010-014-1427-8

Papanikolaou S, Aggelis G (2011) Lipids of oleaginous yeasts. Part II: Technology and potential applications. Eur J Lipid Sci Technol 113: 1052–1073. doi: 10.1002/ejlt.201100015

Ramani G, Meera B, Vanitha C, Rajendhran J, Gunasekaran P (2015) Molecular cloning and expression of thermostable glucose-tolerant β-glucosidase of *Penicillium funiculosum* NCL1 in *Pichia pastoris* and its characterization. J Ind Microbiol Biotechnol 42: 553–565. doi: 10.1007/s10295-014-1549-6

Stursova M, Zifcakova L, Leigh MB, Burgess R, Baldrian P (2012) Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. FEMS Microbiol Ecol 80: 735–746. doi: https://doi.org/10.1111/j.1574-6941.2012.01343.x

Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol 15: 160-167. doi: https://doi.org/10.1016/S0167-7799(97)01032-9

Tian C, Beeson WT, Iavarone AT, Sun J, Marletta MA, Cate JHD, Glass NL (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. Proc Natl Acad Sci 106: 22157–22162. doi: 10.1073/pnas.0906810106

Van Rensburg P, van Zyl WH, Pretorius IS (1998) Engineering yeast for efficient cellulose degradation. Yeast 14: 67-76. doi: 10.1002/(SICI)1097-0061(19980115)14:1<67::AID-YEA200>3.0.CO;2-T

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: https://doi.org/10.1016/j.copbio.2013.02.024

Vogl T, Kickenweiz T, Sturmberger L, Glieder A (2015) Bidirectional promoter. US 20150011407 A1.

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. doi: 10.1021/acssynbio.5b00199

Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A (2004) Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. FEMS Yeast Res 5: 179–189. doi: https://doi.org/10.1016/j.femsyr.2004.06.016

Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, Schürmann M, Mink D, Zellnig G, Schwab H, Pichler H (2014) Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris***.** Metab Eng 24: 18–29. doi: https://doi.org/10.1016/j.ymben.2014.04.001

Wu S, Letchworth GJ (2004) High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. Biotechniques 36:152-154.

Yamada R, Hasunuma T, Kondo A (2013) Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. Biotechnol Adv 31: 754–763. doi: https://doi.org/10.1016/j.biotechadv.2013.02.007

Table 1: List of used genes and the sources where the DNA sequences came from.

Gene	**Enzyme**	**NCBI reference**	**Source**
AnBGL1 codon-optimized	β-glucosidase	XP_001398816.1	synthetic gene (GenScript)
TrBGL1	β-glucosidase	XP_006964076.1	genomic DNA (*T. reesei* QM9414)
TrCBH2 HM codon-optimized	Exo-glucanase (CBH2)	XP_006962580.1	vector (Mellitzer et al. 2014)
TrCBH2 V09 codon-optimized	Exo-glucanase (CBH2)	XP_006962580.1	gBlocks (IDT)
AnEG-A	Endo-glucanase A	-	cDNA (*Aspergillus niger* DSM 26641)
TrEG1	Endo-glucanase 1	AAA34212.1	genomic DNA (*T. reesei* QM9414)
eGFP	Green fluorescent protein	ABC47319.1	gBlock (IDT)

Figure captions

Fig. 1: Plasmid maps of used expression vectors for engineering the strain that expresses all 3 cellulases. The vector for the expression of TrCBH2 V09 and AnEG-A contains a geneticin selection marker (kanMX cassette) (a). The vector for the expression of AnBGL1 and eGFP contains a zeocin selection marker (Zeo resistance cassette). The alpha mating factor pre-pro signal sequence (sig. seq.) upstream of *AnBGL1* is shown in white (b). CDS of genes and selection marker cassettes are shown in black, promoters (NB2, NB3) in dark grey, terminators (*AOX1* TT) in grey, origin of replication (pUC ORI) in white and the targeting site (int2, int8) containing the *Bgl*II site for yeast transformation in bright grey. Plasmid maps were generated by using SnapGene Viewer.

Fig. 2: Schematic figure of the construction for using only 1 orientation of the bidirectional promoter NB2 to express the gene coding for a β-glucosidase (*TrBGL1*), an exo-glucanase (*TrCBH2*) or an endo-glucanase (*TrEG1*), respectively.

Fig. 3: Growth of β-glucosidase expressing *K. phaffii* on cellobiose in shake flasks. 3 different time measurements of AnBGL1 expressing strain culture (grey bar) in BM_cellobiose 0.25%. Parental strain BG10 was the negative control (black bar) (**a**). Growth curve of TrBGL1 expressing strain (circle) in BM_cellobiose 0.5%. The negative control (triangle) was parental strain BSY11G1 (**b**).

Fig. 4: Comparison of growth of different *K. phaffii* BG10 strains on CMC: Growth curve of parental strain BG10 (filled triangle), AnBGL1 expressing strain (filled circle) and strain expressing all 3 cellulases (empty circle) in BM_CMC 1% **(a).** Growth curve of the strains expressing TrCBH2+AnBGL1 (empty triangle), TrEG1+AnBGL1 (filled square) and strain expressing all three cellulases (empty circle) in BM_CMC 0.5% (**b**).

Fig. 5: HPLC measurements of samples containing supernatants of *K. phaffii* strains and Avicel. Glucose measurement (black bar) and cellobiose measurement (grey bar) of the samples containing the supernatant of the strains expressing AnBGL1, TrCBH2 V04+AnEG-A, all three cellulases and supernatant of the non transformed parental strain BG10 as negative control. ND means “not detectable”.

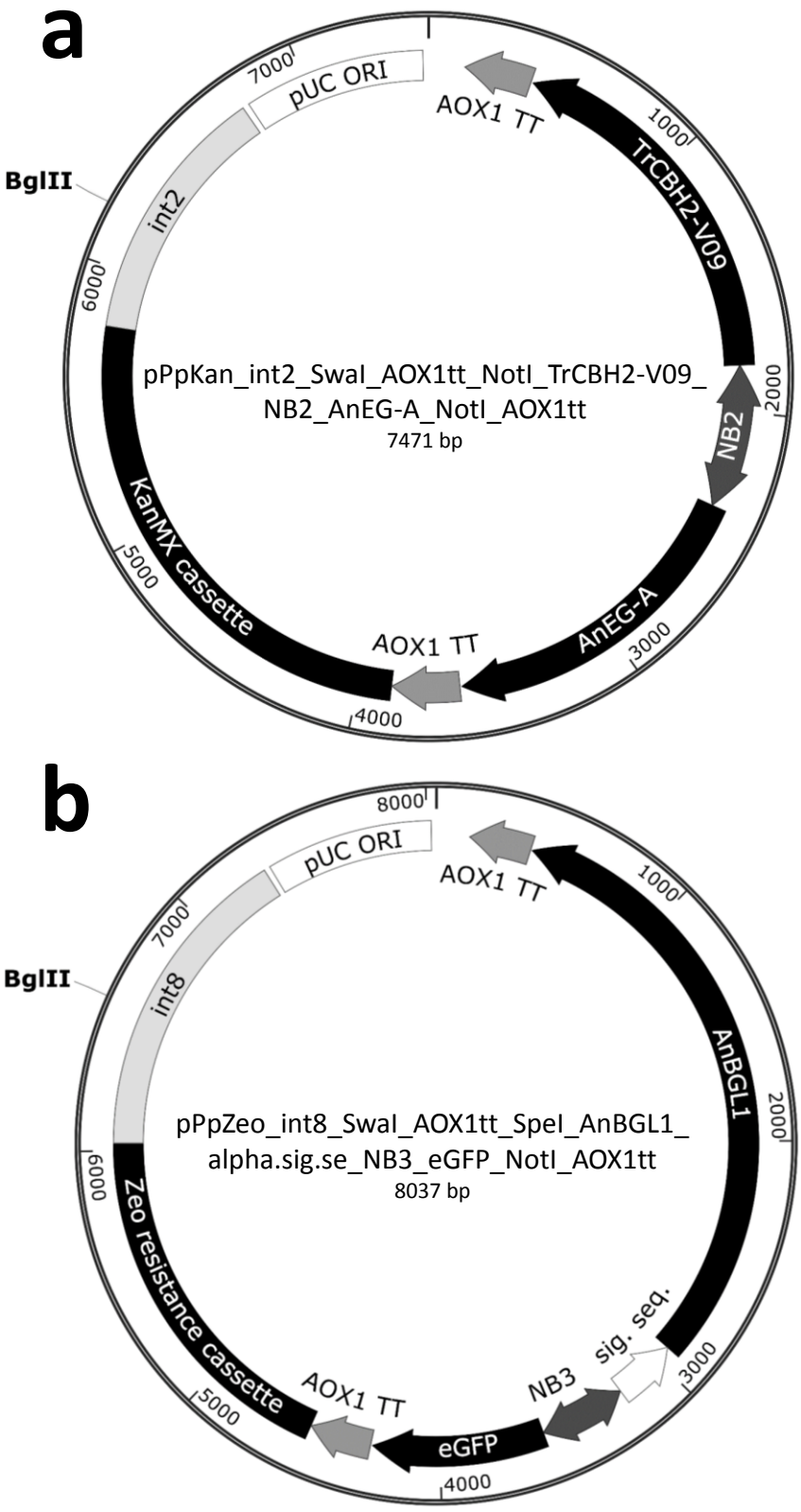

a
pPpKan_int2_SwaI_AOX1tt_NotI_TrCBH2-V09_
NB2_AnEG-A_NotI_AOX1tt
7471 bp
BglII
pUC ORI
AOX1 TT
TrCBH2-V09
NB2
AnEG-A
AOX1 TT
KanMX cassette
int2
1000
2000
3000
4000
5000
6000
7000
b
pPpZeo_int8_SwaI_AOX1tt_SpeI_AnBGL1_
alpha.sig.se_NB3_eGFP_NotI_AOX1tt
8037 bp
BglII
pUC ORI
AOX1 TT
AnBGL1
sig. seq.
NB3
eGFP
AOX1 TT
Zeo resistance cassette
int8
1000
2000
3000
4000
5000
6000
7000
8000

Fig.1

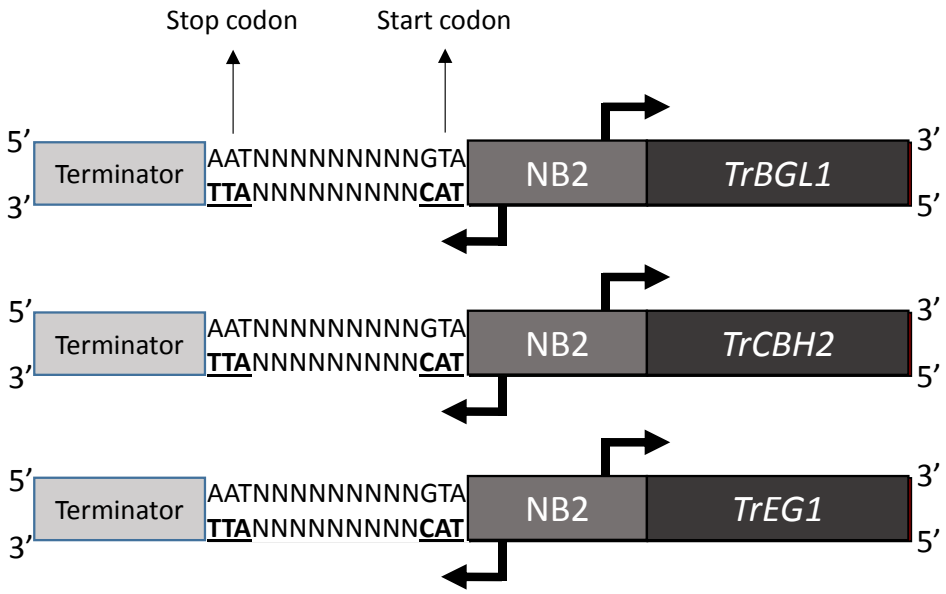

Stop codon
Start codon
5’
3’
Terminator
AATNNNNNNNNNGTA
TTANNNNNNNNNCAT
NB2
TrBGL1
3’
5’
5’
3’
Terminator
AATNNNNNNNNNGTA
TTANNNNNNNNNCAT
NB2
TrCBH2
3’
5’
5’
3’
Terminator
AATNNNNNNNNNGTA
TTANNNNNNNNNCAT
NB2
TrEG1
3’
5’

Fig.2

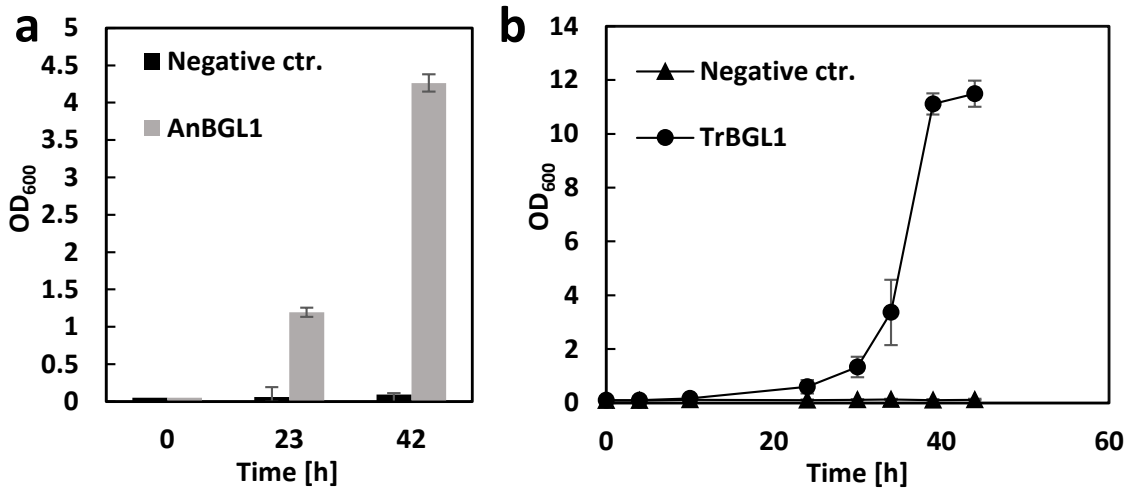

a
Negative ctr.
AnBGL1
OD600
Time [h]
0
23
42
b
Negative ctr.
TrBGL1
OD600
Time [h]
0
20
40
60

Fig.3

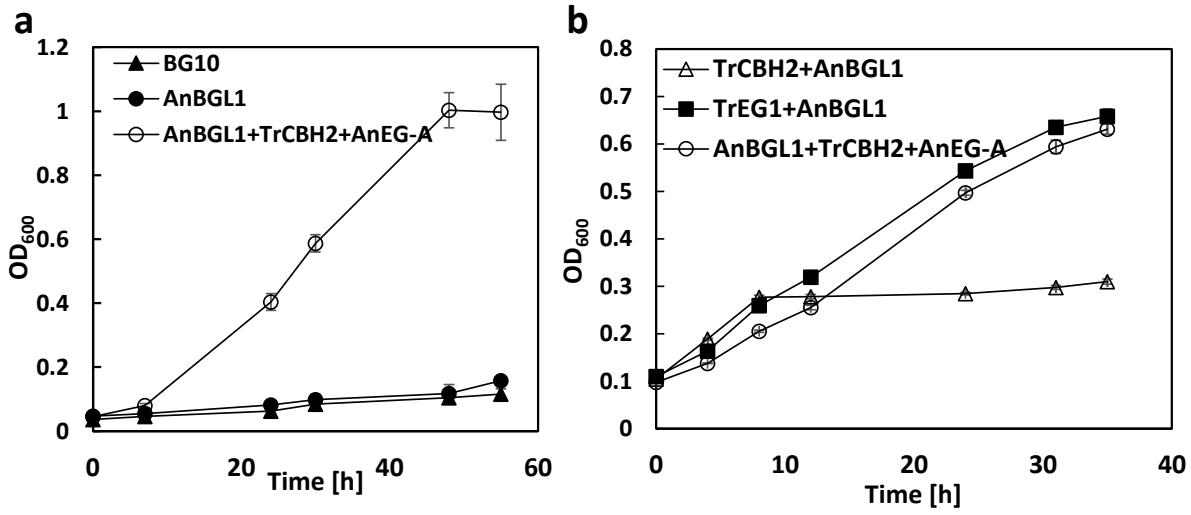

a
BG10
AnBGL1
AnBGL1+TrCBH2+AnEG-A
OD600
Time [h]
b
TrCBH2+AnBGL1
TrEG1+AnBGL1
AnBGL1+TrCBH2+AnEG-A
OD600
Time [h]

Fig. 4

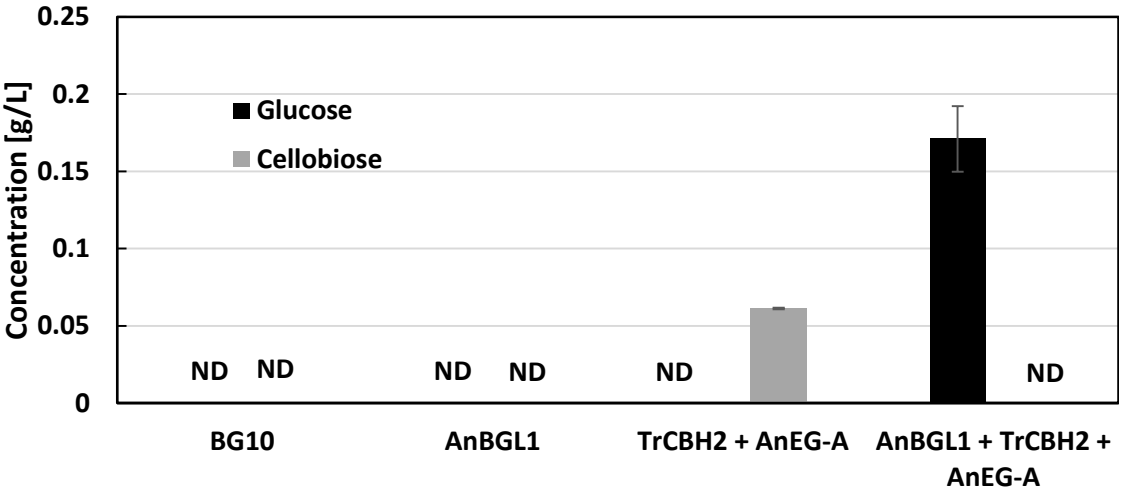

0.25
0.2
0.15
0.1
0.05
0
Concentration [g/L]
Glucose
Cellobiose
ND
ND
ND
ND
ND
ND
BG10
AnBGL1
TrCBH2 + AnEG-A
AnBGL1 + TrCBH2 + AnEG-A

Fig. 5

Supplemental data

Vector construction

All components for the cloning work were from Thermo Fischer Scientific (Waltham, MA, USA) unless otherwise specified. The restriction enzymes were either from Thermo Fischer Scientific (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). PCR products were purified using Wizard SV Gel and PCR Clean-Up System from Promega Fitchburg, WI, USA). All origin plasmids were from Pichia Pool1 of Graz University of Technology (Näätsaari et al. 2012). Primers were ordered from IDT (Coralville, IA, USA). All primers and constructed vectors used for expression of cellulases are listed in Supplemental Tables S1 and S2. DNA fragments were cloned into vectors by using T4 ligase unless otherwise specified. When DNA fragments were cloned into vectors using Gibson cloning, it was performed as previously described (Vogl et al. 2016)**.**

Cloning of vectors for expression of the genes *TrBGL1*, *TrCBH2-HM* and *TrEG1*, respectively, under NB2 control. First, the successfully isolated *TrBGL1* gene without introns was cloned into pPpT4_Zeo_pGAP_alpha as entry vector for storage. (Zeo stands for zeocin selection marker.) For this, *TrBGL1* was amplified with the primers P3 + P10 to add the *Xho*I restriction site which was located within alpha mating factor pre-pro signal sequence. Afterwards, pPpT4_Zeo_pGAP_alpha and the isolated *TrBGL1* DNA fragment were cut with *Xho*I and *Not*I and ligated with each other to create pPpT4_Zeo_pGAP_alpha_TrBGL1.

For another study, isolated *TrBGL1* and *TrCBH2*-HM were used to create a vector for co-expressing both genes (not shown). In this vector *TrBGL1* was fused with *S. cerevisiae* alpha mating factor pre-pro signal sequence. The vector was designed for being able to clone bidirectional promoters into this vector to control the co-expression of both genes. NB2 was cloned into this vector after PCR amplification with the primers P13 + P14 by Gibson cloning**.** This vector containing TrCBH2-HM_NB2_alpha.sig.seq._TrBGL1 was used as template to amplify NB2_TrCBH2-HM with the primers P16 + P17 and NB2_alpha.sig.seq._TrBGL1 with the primers P15 + P10, respectively.

pPpB1GAP_S vector was modified by removing P_{GAP} by restriction with *Swa*I and *Not*I and insertion of *ADH*tt terminator sequence in minus orientation afterwards. *ADH*tt terminator was amplified using the primers P25 + P26 and cloned into this vector after restriction with *Swa*I and *Not*I forming B1_S_Zeo_ADHtt_BmrI_NotI_AOX1tt. This vector and the PCR products NB2_TrCBH2-HM and NB2_alpha.sig.seq._TrBGL1, respectively, were digested with *Bmr*I and *Not*I and ligated to create B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt and B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt. The zeocin resistance gene cassette was replaced by *GUT1* marker for *K. phaffii* and ampicillin as prokaryotic selection marker for *E. coli* in B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt to create pPpGUT1_Amp_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt.

TrEG1 contains a *Xho*I restriction site. The isolated *TrEG1* gene without introns was amplified by PCR using the primers P23 + P24 to add ends which are homologous to B1_S_ ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt that had been cut with *Xho*I and *Not*I. Gibson cloning was performed to ligate both fragments to create B1_S_Zeo_ ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt.

Vectors for expression of the gene *AnBGL1*. *AOX1* promoter was removed from B1_S_Kan_ by restriction with *Swa*I and *Not*I. (Kan stands for kanamycin selection marker that can also be used as geneticin selection marker in *K. phaffii*.) A gBlock was ordered from IDT (Coralville, IA, USA) coding for alpha mating factor pre-pro signal sequence, NB3 and *eGFP* with a *Swa*I and a *Spe*I restriction sites at the end with alpha mating factor pre-pro signal sequence and a *Not*I restriction site at the downstream end of *eGFP*. This gBlock was amplified with the primers P27 + P28, digested with *Swa*I and *Not*I and ligated with digested B1_S_Kan_ to create pPpKan_SwaI_SpeI_alpha.sig.seq_NB3_eGFP_AOX1tt. *AOX1*tt terminator sequence was amplified with the primers P29 + P30 to add a *Spe*I restriction site at the upstream end of the terminator sequence and a *Swa*I restriction site at the downstream end. The PCR product and pPpKan_SwaI_SpeI_alpha.sig.seq_NB3_eGFP_AOX1tt were digested with *Spe*I and *Swa*I and ligated with each other to create pPpKan_SwaI_AOX1tt_SpeI_alpha.sig.seq_NB3_eGFP_AOX1tt.

The codon-optimized version of *AnBGL1* which was fused with alpha mating factor pre-pro signal sequence from *S. cerevisiae* was cloned into this vector after restriction with *Xho*I and *Spe*I to form pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_AOX1tt.

Another gBlock was ordered from IDT (Coralville, IA, USA) coding for the zeocin resistance gene cassette containing P_{ILV5}, P_{EM72}, zeocin resistance gene and *AOD*tt terminator sequence. The sequence of this gBlock was identical with the zeocin resistance cassette of T4_S plasmid. This gBlock was amplified with the primers P45 + P46. There were a *Bam*HI restriction site at upstream end and a *Pst*I restriction site at the downstream end of the zeocin resistance cassette. By restriction with *Bam*HI and *Pst*I, the selection marker in the expression vectors containing geneticin resistance cassette was exchanged.
The targeting sites int2, int3 and int8 were amplified from *K. phaffii* BG10 genomic DNA with the primers P39 + P40 (int2), P41 + P42 (int3) and P43 + P44 (int8) to add *Pst*I restriction sites at both ends. The restriction sites were inserted into pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt and pPpZeo_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt to create pPpKan_int3_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt and pPpZeo_int8_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt which were used for expression of β-glucosidase AnBGL1.

Vector for co-expression of the genes *TrCBH2*-V09 and *AnEG-A*. First, *TrCBH2*-V09 was cloned into pPpKan_S as entry vector. For this, gBlock coding for *TrCBH2*-V09 was amplified using the primers P31 + P33. Then, pPpKan_S and *TrCBH2*-V09 were cut with *Eco*RI and *Not*I and ligated with each other to create pPpKan_S_pAOX1_TrCBH2-V09_AOX1tt. TrCBH2_NotI_AOX1tt was amplified with the primers P32 + P34 using this vector as template. AnEG-A was amplified with the primers P35 + P36 adding a *Not*I restriction site at its downstream end. NB2 was amplified with the primers P37 + P38 to add homologous sequences to *TrCBH2*-V09 and *AnEG*-A at each end for overlap extension PCR. NotI_AnEG-A, NB2 and TrCBH2-V09_NotI_AOX1tt were fused with each other by overlap extension PCR. For the next cloning step, a vector pPpKan_int2_SwaI_AOX1tt_NotI_gene-x_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt was used. This vector and the PCR product were cut with *Not*I. The vector backbone and NotI_TrCBH2-V09_NB2_AnEG-A_NotI were ligated with each other to create pPpKan_int2_SwaI_AOX1tt_NotI_TrCBH2-V09_NB2_AnEG-A_NotI_AOX1tt, which was used for co-expression of endo-glucanase (AnEG-A) and exo-glucanase (TrCBH2).

Primers and Vectors used

Table S1: List of used primers.

Name	Primer name	Secquence 5' → 3'
P1	Bgl1Exon1_alpha_1st step_fwd	gaagagagaggccgaagctGTTGTACCTCCTGCAGGGACTCCATG GGGAAC
P2	Bgl1Exon1part1mutB mrI rev	GCCCGTGAGATATGGATCGACACCGAAGCCCTCCCAATT GCG
P3	Bgl1Exon1_alpha_ 2nd step_fwd	gtctctctcgagaagagagaggccgaagctGTTGTAC
P4	Bgl1Exon1part2 fwd	AGGGCTTCGGTGTCGATCCATATCTCACGG
P5	Bgl1Exon1part2 rev	GTAGTTGAACTTGGTGTAAGACAGTCCATAGCCGAACTC G
P6	Bgl1Exon2part1mutB mrI fwd	CTTACACCAAGTTCAACTACTCACGCCTCTCCGTCTTGTC GACCGCCAAGTCTGGTCCTGCGACTGGCGCCGTTGTG
P7	Bgl1Exon2part1mutB mrI rev	TGCGAAGCCGTGTCCCAATAGCTGAGATCTCGTCGTCGG ATGTTG
P8	Bgl1Exon2part1ole fwd	CGAGTTCGGCTATGGACTGTCTTACACCAAGTTCAACTA C
P9	Bgl1Exon2part2 fwd	GGGACACGGCTTCGCAGAAATGGGTG
P10	Bgl1Exon2part2_NotI rev	gactaGCGGCCGCCTACGCTACCGACAGAGTGCTCGTCAGC CTGATATC
P11	CBH2_stuf-_ole_1st-step fwd	ACTGGGtgctatgattgtcggtatcttgactaccttg
P12	AOXtt_3'end_ApaI_ole_1st-step rev	GGGCCCgcacaaacgaaggtctcacttaatc
P13	alpha_NB2_ort1 rev	cagcggtgaaaatagatgggaatctcattttgatttgtttaggtaacttgaactggatg
P14	Cbh2-HM_NB2_ort1 fwd	caaggtagtcaagataccgacaatcatTGTTGTAGTTTTAATATAGTTTG AGTATGAG
P15	Nb2ort2_ADHtt_ole rev	CCTAGGTTAGACTGGGTACATtgttgtagttttaatatagtttgagtatgagat ggaac
P16	Nb2ort1_ADHtt_ole rev	CCTAGGTTAGACTGGGTACATtttgatttgtttaggtaacttgaactggatg
P17	Cbh2-HM_NotI rev	gactaGCGGCCGCTTACAAGAAAGATGGATTGGCGTTAGTC AACAG
P18	TrEg1 alpha-all fwd	GTCTCTCTCGAGAAGAGAGAGGCCGAAGCTCAGCAACC GGGTACCAGCACCCCCGAG
P19	TrEg1_Exon1ole rev	GCCGTAGTAGCTTTTGTAGCCGCTGCCATAGGGGTTG
P20	TrEg1_Exon2ole fwd	CAGCGGCTACAAAAGCTACTACGGCCCCGGAGATACCG TTG
P21	TrEg1_Exon2ole rev	CATTGCGAGTAGTAGTCGTTGCTATACTGGCACGTAGTG
P22	TrEg1_Exon3_NotI rev	gactaGCGGCCGCCTAAAGGCATTGCGAGTAGTAGTCGTTG
P23	TrEg1_alpha_gibs fwd	CTGCTAAGGAAGAGGGTGTCTCTCTCGAGAAGAGAGAG GCCGAAGCTC
P24	TrEg1_AOX1tt_gibs	CTCTTGAGCGGCCGCCTAAAGGCATTG

	rev	
P25	ADHtt_NotI fwd	agtcaGCGGCCGCATGTACCCAGTCTAACCTAGGGCGAATTTC
P26	ADHtt_SwaI rev	gactaATTTAAATTTGTCCTCTGAGGACATAAAATACACACCGAG
P27	alpha_3end_SwaI fwd	TCAGATTTAAATTCAGACTAGTCTAAGCTTCG
P28	eGFP_3end_NotI rev	ttgagcggccgcttacttgtac
P29	AOX1tt_5end_SpeI fwd	TCAGACTAGTTCAAGAGGATGTCAGAATGCCATTTG
P30	AOX1tt_3end_SwaI rev	TCAGATTTAAATGCACAAACGAAGGTCTCACTTAATCTTC
P31	CBH2-V09_EcoRI fwd	TCAGAATTCGAAACGATGATAGTCGGAATC
P32	CBH2-V09 fwd	ATGATAGTCGGAATCTTAACTACGCTGG
P33	CBH2-V09_NotI rev	ACTTGCGGCCGCTTATAAGAATGAGGGATTTGCATTAGTTAGCAACTG
P34	AOX1tt_3end_SwaI rev	TCAGATTTAAATGCACAAACGAAGGTCTCACTTAATCTTC
P35	EG-A fwd	ATGAAGACTCTCTCCCTTGC
P36	EG-A_NotI rev	TTGAGCGGCCGCCTACAAACACTGCGAATACCAC
P37	NB2_EG-A fwd	GCAAGGGAGAGAGTCTTCATtgttgtagttttaatatagtttgagtatgagatggaac
P38	NB2_CBH2-V09 rev	CGTAGTTAAGATTCCGACTATCATtttgatttgtttaggtaacttgaactggatg
P39	integrsite2 fwd	ACTGCTGCAGGAAAACCGGTGTCCAGAGAGATTC
P40	integrsite2 rev	ACTGCTGCAGAGTGACGACGAATCTATTGCTGAC
P41	integrsite3 fwd	ACTGCTGCAGGACTTGATTGTCGAGATTGGTGAG
P42	integrsite3 rev	ACTGCTGCAGTCTGCTAGCCAGATGGTGATTATG
P43	integrsite8 fwd	ACTGCTGCAGCAATCATCCACTCTGCCTCTCTG
P44	integrsite8 rev	ACTGCTGCAGAGGATAGTGCTTTCATGCCCAAG
P45	PILV5_BamHI fwd	tgtgcggatccttcagtaatgtc
P46	AODtt_PstI rev	agtggtacctgcagctaaggtaatc

Table S2: List of all expression vectors. zeocin (Zeo) and geneticin (Gen) were antibiotic markers used. Glycerol is an auxotrophic marker and used in *K. phaffii* BSY11G1 genetic background

Vector name	**Eukaryotic selection marker**	**Restriction site for linearization in yeast transformation**
pPpKan_int2_SwaI_AOX1tt_NotI_TrCBH2-V09_NB2_AnEG-A_NotI_AOX1tt	Gen	*Bgl*II
pPpZeo_int8_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt	Zeo	*Bgl*II
pPpKan_int3_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt	Gen	*Bgl*II
pPpGUT1_Amp_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt	glycerol	*Swa*I
B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt	Zeo	*Swa*I
B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt	Zeo	*Swa*I

Alpha mating factor pre-pro signal sequence_AnBGL1 sequence made as synthetic DNA:

ACACAGAATTCCGAAACGATGAGATTCCCATCCATTTTTACCGCAGTCTTGTTTGCCGCCAGTTCCGCACTTGC
CGCCCCAGTCAATACCACAACCGAGGACGAAACTGCTCAAATTCCAGCTGAAGCCGTTATCGGTTACTCTGAT
TTGGAGGGAGATTTTGACGTTGCTGTCTTGCCTTTCTCTAACTCCACAAACAACGGTTTGCTTTTTATTAACACT
ACAATCGCTTCAATCGCTGCCAAGGAAGAGGGAGTTAGTCTCGAGAAGAGAGAAGCAGAGGCTGATGAGTT
GGCCTACTCTCCACCTTACTATCCATCCCCTTGGGCAAACGGACAGGGTGACTGGGCCGAGGCATATCAAAGA
GCAGTTGATATTGTCTCACAGATGACCTTGGCTGAAAAGGTTAATCTTACCACTGGAACTGGTTGGGAATTGG
AGCTTTGTGTCGGTCAAACAGGTGGAGTTCCAAGATTGGGTATCCCTGGAATGTGCGCTCAGGACTCCCCATT
GGGTGTTAGAGATTCAGACTACAATAGTGCCTTTCCAGCTGGTGTTAACGTCGCAGCTACTTGGGATAAGAAT
TTGGCCTATCTTAGAGGTCAAGCAATGGGACAGGAATTTTCTGACAAAGGTGCTGATATTCAATTGGGACCAG
CCGCAGGTCCTCTTGGAAGAAGTCCAGACGGTGGAAGAAACTGGGAGGGTTTTTCACCAGATCCTGCATTGA
GTGGAGTTCTTTTCGCTGAAACAATTAAGGGTATCCAAGATGCTGGAGTTGTCGCTACCGCCAAACATTATATT
GCCTACGAACAAGAGCACTTTAGACAGGCTCCAGAGGCCCAAGGATACGGTTTCAACATCACCGAATCAGGT
AGTGCTAATTTGGATGACAAGACTATGCATGAATTGTACCTTTGGCCTTTTGCAGATGCTATTAGAGCCGGAG
CAGGTGCTGTTATGTGTTCTTACAACCAAATCAACAACTCTTACGGTTGCCAAAACTCTTACACTTTGAACAAG
TTGCTTAAAGCTGAACTTGGTTTTCAAGGATTCGTCATGTCCGACTGGGCTGCCCATCACGCTGGTGTTTCTGG
AGCCTTGGCAGGTCTTGATATGTCCATGCCAGGAGATGTTGACTACGATTCAGGTACCAGTTATTGGGGAACA
AATTTGACCATTTCTGTTCTTAACGGTACTGTCCCTCAATGGAGAGTTGATGACATGGCTGTCAGAATCATGGC
AGCTTACTATAAAGTTGGTAGAGACAGATTGTGGACCCCACCTAACTTTTCTTCCTGGACTAGAGATGAATAC
GGTTTTAAATACTACTACGTCTCCGAAGGACCATACGAGAAGGTTAACCAGTTCGTTAACGTCCAAAGAAACC
ATTCTGAATTGATCAGAAGAATCGGTGCAGACTCCACTGTTTTGCTTAAGAATGATGGTGCTTTGCCACTTACA
GGAAAAGAAAGATTGGTTGCTCTTATTGGAGAGGATGCCGGATCAAATCCTTACGGTGCTAACGGATGTAGT
GACAGAGGTTGCGATAACGGAACTTTGGCCATGGGATGGGGTTCTGGAACAGCAAATTTTCCATATTTGGTTA
CCCCTGAACAAGCTATTTCCAACGAGGTCTTGAAGAACAAAAATGGTGTTTTCACCGCCACTGACAACTGGGC
AATTGATCAGATCGAGGCTTTGGCCAAGACTGCTTCTGTCTCCTTGGTTTTCGTCAATGCCGATTCTGGTGAAG
GATACATTAATGTTGACGGTAACTTGGGAGATAGAAGAAATTTGACACTTTGGAGAAACGGAGATAATGTTA
TTAAAGCCGCAGCTTCAAACTGTAACAACACCATCGTTATCATCCACAGTGTCGGTCCAGTTTTGGTCAATGAG
TGGTATGACAACCCTAATGTTACTGCTATTTTGTGGGGTGGACTTCCTGGTCAAGAATCTGGAAACTCCTTGGC
CGATGTTCTTTACGGTAGAGTCAATCCAGGAGCAAAGTCTCCTTTTACTTGGGGTAAAACAAGAGAAGCTTAC
CAAGACTATTTGTACACTGAGCCAAACAATGGTAACGGAGCCCCTCAGGAAGATTTTGTTGAGGGTGTCTTCA
TTGACTACAGAGGATTCGATAAGAGAAATGAAACTCCAATCTATGAGTTTGGTTACGGATTGTCATATACAAC
CTTCAACTACAGTAATTTGCAAGTTGAAGTCCTTTCTGCACCAGCTTATGAGCCTGCTTCCGGTGAAACAGAGG
CCGCACCTACCTTTGGTGAAGTTGGAAACGCTTCCGACTATTTGTACCCAGATGGTCTTCAGAGAATCACTAAA

TTCATCTATCCTTGGTTGAACTCTACTGACCTTGAGGCTTCAAGTGGAGATGCCTCATACGGACAAGACGCTAG
TGATTATTTGCCAGAAGGTGCAACTGATGGATCTGCTCAGCCAATTTTGCCAGCTGGTGGTGGTGCTGGTGGT
AATCCTAGATTGTACGACGAGCTTATTAGAGTTTCAGTCACCATCAAGAACACTGGTAAAGTTGCTGGAGATG
AAGTCCCACAATTGTATGTTTCTCTTGGAGGTCCAAACGAACCTAAGATTGTTTTGAGACAGTTTGAGAGAATC
ACCTTGCAGCCTTCTAAAGAAACTCAATGGTCCACTACATTGACAAGAAGAGACCTTGCTAACTGGAATGTCG
AAACTCAAGATTGGGAGATTACATCTTACCCAAAAATGGTTTTCGCAGGTTCCAGTAGTAGAAAACTTCCACTT
AGAGCTTCACTTCCAACAGTCCATTAGACTAGTCTGAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTG

Fig. S1: DNA sequence of codon-optimized *AnBGL1*. Restriction sites (yellow background), Kozak sequence (green background), codon optimized alpha mating factor pre-pro signal sequence (blue background) and codon optimized *AnBGL1* cds (olive background).

Used gene sequence of *Aspergillus niger* DSM 26641 endo-glucanase (AnEG-A):

ATGAAGACTCTCTCCCTTGCCGCTGTCTTGCTCGTCCAAGCCTGGACTGCGAACAGCAGTCCGAGGGCTAAGC
GGTCGTCTGCTTTCCAATGGTTCGGCAGCAACGAATCCGGCGCCGAATTTGGCTCCGGGAATATCCCCGGCAC
TTTGGGCACGGATTACATCTGGCCTGATGCGACGGCTATCGCGACGCTGCATAGTGCCGGGATGAATATCTTT
CGCGTGCCGTTTCTGATGGAGCGGTTGGTGCCGGATGAGTTGACTGGTGCGGCGGATGCGACGTATATGGCT
GATCTGAAGGCTACGGTGCAAACCATCACAGACCTTGGTGCTTATGCGGTGGTGGATCCGCATAATTTTGGAA
GATACTACGGCAACATCATCACCTCAACCAGCGACTTTGCAGCCTTCTGGACAACAGTCGCCAGCGAGTTCAC
CGACAATGACCTTGTCATCTTTGATACCAACAACGAATTCCACGACGAAGACCAAACTCTCGTGCTCGATCTCA
ATCAAGCCGCCATCACCGCCATCCGCGCCACCGGTGCCACAACGCAATATATCTTCGTCGAAGGCAATTCCTAC
ACCGGCGCCTGGACCTGGACCACCACCAACACGAACCTCGTCAACCTGACCGACCCCAACGACCATCTCATCT
ACGAAATGCACCAGTACCTGGACAGCGATGGCTCCGGGACGAGTTCGACGTGCGTCAGCTCGACAATTGGTG
CGGAGCGCATTGCTGATGCCACAACCTGGTTGAAAGAGAATGGCAAGCAGGCGGTGCTGGGGGAGTTCGCG
GGCGGCGCGAATAGTGTATGTGAGGAGGCGGTTACAGGCATGTTGGATGCGATGGTGGATGCTAGTGATGT
TTGGCTGGGGGGCATGTGGTGGTCTGCTGGGCCTTGGTGGGGGGATTATATCTTCTCGTTGGAGCCCGATGA
TGGAATTGCTTATGAGTATTATTTGGGGGTGTTGGAGGGGTATACGCCTGGTTCATCTTCGGGATCATCTGGA
GGGAGTACAACAAGTGCCGTGGCGAGTGCTAGTAGTACGTCGTCTGCTGGGAGTACTACGAGCAAAGTGACT
ACTACTACTGCTGCGACTGCTACATCGACGGCTGTAGCTGCGCACTATGCTCAGTGTGGTGGGGAGGGCTGG
ACTGGGGCGACGACTTGTGCGGATGGATATACCTGCACAGTAGAAAATGAGTGGTATTCGCAGTGTTTGTAG

Fig. S2: Sequenced DNA sequence of *AnEG-A* isolated from *A. niger* DSM 26641. Start codon (yellow background), Stop codon (blue background).

References

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression

and synthetic biology. *PLoS One* 7: e39720. doi: https://doi.org/10.1371/journal.pone.0039720

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. **doi:** 10.1021/acssynbio.5b00199

Part III: Improving protein secretion by *K. phaffii*

Abstract

Many different microorganisms have been implemented as industrial hosts for the production of enzymes and high-value compounds. The methylotrophic yeast *Komagataella phaffii* is such a broadly used and efficient eukaryotic host employed for multiple industrial purposes. Due to the simplicity of downstream processing (DSP) protein secretion is often favoured compared to intracellular expression. In addition some proteins need posttranslational processing which occurs along the secretory pathway. However, bottlenecks in the secretory pathway of eukaryotic expression hosts limit the specific productivity and titers of heterologous proteins in the culture media. Therefore, improving the expression system and the elimination or reduction of such bottlenecks is a key to commercial success. In general applied approaches to improve protein secretion include mutagenesis experiments which require convenient high throughput screening methods. However, most screening methods allow limited throughput as the cultivation of the clones has to be done individually in liquid medium. In this study, a simple screening method is presented which enables mutant screening for better protein secretion based on faster growth on agar plates. For this purpose, a *K. phaffii* strain with two integrated copies of a gene coding for a β-glucosidase was used as screening strain. This strain showed no faster growth on cellobiose containing media compared to a single copy strain while growth was still lower than on pure glucose. After random mutagenesis experiments, mutants were screened for faster growth on minimal medium plates with cellobiose as sole carbon source. Similarly, another strain expressing the three key cellulases which are required for cellulose saccharification was tested. This latter platform strain was used in screenings for better growth on carboxymethyl cellulose.

Introduction:

1. Secretion: General overview

Protein secretion in eukaryotes has been part of extensive research with the purpose to reveal its major mechanisms. Many different model organisms like plant and mammalian cells, filamentous fungi and yeasts are used, to understand the secretory pathway better [Aguiar et al. 2014; Delic et al. 2013; Jürgens and Geldens 2002; Nickel 2005; Richter et al. 2009], revealing specific as well as general effects. The motivation for doing research on protein secretion differs. For medicine, it is important to understand how protein secretion works in human cells because defects in secretory pathway can trigger diseases in human [Wang and Kaufmann 2012; Yoshida 2007]. There is another medical interest to study the secretory pathway in pathogenic fungi because of the linkage of their virulence to protein secretion [Schaller et al. 2005; Sorgo et al. 2013]. On the other side, there are also industrial interests to understand the secretory pathway of fungi which are used for production of secreted proteins or other high value compounds [Rebnegger et al. 2014; Saloheimo and Pakula 2012] in order to make commercial processes more efficient.

The secretory pathway contains several steps which include embracing translocation and protein transport to different cell compartments, the folding and glycosylation of the proteins and their secretion. The secretory pathway in yeasts was comprehensively reviewed by Delic et al. 2013. It starts at the membrane of Endoplasmic Reticulum (ER) with the translocation of the synthesized peptide chain from the cytosolic side into the ER. Two types of translocations are known. In co-translational translocation, the translocation of the proteins is combined with their translation involving ribosomes. Whereas in post-translational translocation, no ribosomes are involved and the proteins are fully translated before they are translocated. The translocation of proteins depends on their hydrophobicity and amino acid constellation [Delic et al. 2013].

Differences in cotranslational and posttranslational export of proteins by *S. cerevisiae* were studied in more detail by the Glick lab [Fitzgerald and Glick 2014].

In the ER, chaperons support the folding of the translocated peptide chains into the right structure. Additionally, disulfide bonds are shuffled and optimized by protein disulfide isomerases (PDI) which facilitates correct protein folding. These foldases are able to recognize misfolded proteins and reduce the amount of incorrect disulphide bonds [Delic et al. 2013; Labunskyy et al. 2007; Parakh and Atkin 2015]. The formation of disulfide bonds

seems to be one cause for the formation of reactive oxygen species (ROS). H_2O_2 is formed during re-oxidation of PDI by transferring electrons to O_2. This can lead to ER stress (ROS/Protein folding stress) when the protein expression machinery reaches its limits [Delic et al. 2013; Parakh and Atkin 2015]. For example, it has been reported that expression from multiple gene copies or very strong promoters in *K. phaffii* can cause such ER stress [Hohenblum et al. 2003; Hyka et al. 2010; Liu et al. 2014].

Peptidyl-prolyl isomerases (PPiase) seem to be another type of enzymes which are also involved in protein folding in the ER. It has been reported that expression of PPiases was upregulated under stress in *S. cerevisiae*. It might be that prolyl isomerisation supports the folding of proteins and stabilizes folded proteins [Delic et al. 2013; Kim et al. 2010; Schmidpeter et al. 2015]. Another step of secretory pathway which is located in ER is the core glycosylation of proteins. N-glycosylation of proteins is a very important step in protein maturation. Furthermore, it is also an important part of quality control by marking the proteins which are not correctly folded [Delic et al. 2013; Moremen et al. 2012; Roth et al. 2010].The sugar residues are linked to the proteins by glycosyl- and oligosaccharyl-transferases. N- and O-glycosylation are the two different types of protein glycosylation. In N-glycosylation, the glycans are linked to asparagine residues of the protein whereas glycans are linked to serine and threonine residues in O-glycosylation. [Delic et al. 2013; Goto 2007; Moremen et al. 2012; Tannous et al. 2015]. Further modification of the linked oligosaccharide chains is performed by glucosidases and mannosidases. Misfolded proteins are differently glycosylated than correctly folded proteins. This leads either to the degradation of the misfolded protein or to the correction of its protein folding. Only proteins which are correctly folded and passed the "quality control" in ER can be expected to be transported further to Golgi [Delic et al. 2013; Moremen et al. 2012; Tannous et al. 2015].

The intracellular transport of proteins between different cell compartments and from the ER to cellular membranes occurs by vesicles. These vesicles are split off from membranes of the cell compartments (ER, Golgi) by budding. There are different types of vesicles involved in vesicular protein transport. COPII vesicles transport the proteins from the ER to the Golgi (anterograde transport). COPI vesicles are the vesicles that transport the proteins from cis-Golgi back to ER (retrograde transport) and the secretory vesicles (regulated by exocyst) mediate the protein transport from trans-Golgi to the cell membrane. The regulation of protein transport between the cell compartments is very complex, involving many different factors and protein complexes. This is to ensure that the proteins reach the right destinations

[Delic et al. 2013]. Details about the regulation of protein transport in vesicles were summarized by Delic et al. 2013.

Once the proteins have reached the Golgi, they undergo further glycosylation. As mentioned previously, protein glycosylation is not only crucial for quality control of proteins but also influences the properties of proteins (e.g. osmo-tolerance, solubility and stability) and their final destination (e.g. secretion to extracellular space) [Delic et al. 2013; Goto 2007; Gunasekaran et al. 1996]. The type and extend of glycomodifications in the Golgi is highly variable among fungal species and performed by mannosyl-transferases and mannan polymerases [Delic et al. 2013; Goto 2007]. Secreted proteins are transported from cis- to trans-Golgi (over mid-Golgi) and then, in exocyst regulated vesicles to the cell membrane resulting in secretion of the proteins. This step is called Exocytosis [Delic et al. 2013; Geva et al. 2017].

Genes coding for secreted proteins, contain a signal sequence. The signal sequence is required for the transport of the expressed proteins to the correct final destination. The most commonly used signal sequence for heterologous gene expression in *K. phaffii*, is the *S. cerevisiae* mating factor alpha pre-pro-signal sequence. The pre-sequence might be responsible for the posttranslational translocation to the ER of yeasts and the pro-sequence for the transport from the ER to Golgi [Fitzgerald and Glick 2014]. In the Golgi, the signal sequence is cleaved at the C-terminal end of the pro-sequence by the protease Kex2p and the mature protein is then further transported to the cell membrane [Ahmad et al. 2014] where it is either integrated by additional hydrophobic signals or exported to the extracellular space.

2 Relevance of secretion for heterologous gene expression in biotechnology

Many different organisms are used in industrial biotechnology to produce enzymes and other proteins or high-value small compounds. The spectrum of expression hosts ranges from prokaryotes (bacteria) to eukaryotes as filamentous fungi, yeast, and insect and mammal cells [Demain and Vaishnav 2009].

Eukaryotes, especially mammalian cells are the preferred expression hosts for production of therapeutic glycoproteins [Gutierrez and Lewis 2015]. In fact, eukaryotic expression hosts are especially useful for the production of complex eukaryotic proteins since those often are post-translationally modified. This includes the formation of disulphide bonds and glycosylation

which (both) cannot be performed naturally by bacteria [Demain and Vaishnav 2009]. The methylotrophic yeast *K. phaffii* is one of these eukaryotic expression hosts which is also used for production of therapeutics and proteins. A milestone to increase the chance to use yeasts as an alternative to mammalian cells was the glyco-engineering of *K. phaffii* to humanize its glycosylation of proteins [Hamilton et al. 2006]. Another more economic reason to increase the efficiency of protein secretion is to increase the yields of secreted proteins in industry. The secretion of heterologous proteins is a common used strategy in industry because the amount of host proteins is low in the supernatant of yeasts and some other eukaryotic hosts. This simplifies the purification of heterologous proteins and lowers the costs of the downstream processing. An increase of protein secretion of the heterologous proteins in yeasts and other eukaryotic hosts, which are used for production of secreted proteins, would increase the product yields and facilitate the downstream processing [Porro et al. 2011].

The aim is to produce the highest yields of the demanded product without compromising its quality while keeping the amounts of secreted endogenous host proteins such as proteases low. Therefore, there is the motivation to improve the productivity of expression hosts by genetic modification and optimize their expression efficiency to produce demanded proteins [Gutierrez and Lewis 2015; Zheng et al. 2016]. The secretory pathway has been revealed as major bottleneck in the expression of secreted proteins in *K. phaffii* and other eukaryotic hosts [Delic et al. 2014; Idiris et al. 2010a; Routenberg Love et al. 2012]. Routenberg Love and co-workers have located a rate-limiting step in the secretory pathway, in the capacity to transport proteins from the ER to the extracellular space [Routenberg Love et al. 2012].
Since the secretory pathway is a major bottleneck in eukaryotic protein production, there is a strong focus on improving the efficiency of protein secretion in expression hosts.

3 Approaches to improve protein secretion

Many different methods have been established to increase protein production by heterologous expression hosts. In general, it can be differentiated between a) mutagenesis of the expression host or genes involved in protein expression and folding; b) overexpression or knock-out of specific genes affecting protein production and folding; and c) sequence optimization of heterologous genes and vectors (copy number, choice of promoter and signal sequence, codon-optimization of gene sequence) [Liu et al. 2014; Xiao et al. 2014; Zheng et al. 2016]. As the secretory pathway is very complex, many genes are involved in this multi step

procedure and its regulation which all might be putative targets in mutagenesis experiments to improve the efficiency of protein secretion. In addition the number of possible combinatorial approaches is seemingly unlimited and therefore often knowledge based approaches are studied [Delic et al. 2014; Idiris et al. 2010a]. On the other hand more random based approaches such as random integration, random deletion and mutagenesis provide opportunities to learn about new so far undiscovered mechanisms as it was reported for example in Agaphonov et al. 2005 and Huang et al. 2015.

3.1 Mutagenesis

In general mutagenesis experiments causing point mutations can be categorized into rationally and randomly introduced mutations.

Random mutagenesis is the classical method to improve expression host strains [Singh et al. 2017]. The secretion of heterologously expressed human urokinase-type plasminogen activator in the methylotrophic yeast *Hansenula polymorpha* could be improved by chemical mutagenesis. The isolated mutation which was responsible for the increased protein secretion, inactivated a gene coding for an *O*-mannosyltransferase [Agaphonov et al. 2005]. Huang and co-worker showed that protein secretion in *S. cerevisiae* was improved after random mutagenesis with UV-light [Huang et al. 2015].

The *S. cerevisiae* alpha mating factor pre-pro-signal sequence was also targeted by random mutagenesis approaches. It was possible to improve the secretion of single chain antibodies in *S. cerevisiae* using signal sequences which had been modified by error prone PCR. Analysis of the mutated *S. cerevisiae* alpha mating factor pre-pro-signal sequence showed that mutations of the LLFI motif of the leader to amino acids with more polar side chains were responsible for the better secretion of the single chain antibodies. This positive effect was enhanced by a further mutation at position 22 of the leader (valine to alanine). This *S. cerevisiae* alpha mating factor pre-pro-signal sequence variant also improved the secretion of other proteins as horseradish peroxidase and interleukin-2. Evidence could be found by knock-out studies that the mutations in the leader increased the secretion of the single chain antibodies by decreasing the vacuolar mis-trafficking rate of the secreted protein [Rakestraw et al. 2009]. Random mutagenesis approaches to improve protein secretion were also performed with filamentous fungi. For example, it has been reported that secretion of heterologous expressed proteins in the filamentous fungi

Ashbya gossypii could be increased by chemical treatment with ethyl methane sulfonate [Ribeiro et al. 2013].

Random mutagenesis was also used for improving the secretion of homologous expressed proteins. The production of homologous expressed cellulases in different filamentous fungi like *Trichoderma* sp., *Aspergillus* sp., and *Humicola* sp. has been improved by random mutagenesis. For example the *Trichoderma reesei* mutant strain RUT-C30 which was created by random mutagenesis, showed a 20 times higher secretion of cellulases than the parental strain, [Singh et al. 2017].

In **rational mutagenesis**, the focus is on mutation of a specific gene. In this approach, it is required to have information about the DNA sequence of the gene and its gene product (e. g. protein structure and reaction mechanism) [Ali et al. 2016]. The mutation(s) is introduced into the gene in a systematic way by modifying usually only one selected codon. As a consequence, only one amino acid is changed in the protein. The modified protein is then tested if its properties have been improved (e.g. enzymatic activity, stability) [Ali et al. 2016]. Rational mutagenesis has been performed to modify the signal sequence of expressed heterologous proteins in *K. phaffii* and thus improving their secretion. Lin-Cereghino and co-workers reported that the deletion of amino acids forming the 2nd and 3rd helix of the *S. cerevisiae* alpha mating factor pre-pro-signal sequence, increased the secretion of a lipase and a horseradish peroxidase in *K. phaffii* [Lin-Cereghino et al. 2013]. Quite recently, Zheng and co-worker were able to improve the secretion of an alpha-galactosidase in *K. phaffii* by changing one amino acid of the Kex2 protease recognition site in the *S. cerevisiae* alpha mating factor pre-pro-signal sequence [Zheng et al. 2016].

It has been also reported that rational mutagenesis was done on a *Gaussia princeps* signal sequence to improve the protein secretion in the yeast *Kluyveromyces marxianus*. Thereby it was possible to increase the secretion of *G. princeps* luciferase by *K. marxianus* by using synthetic signal sequences which were based on the native signal sequence of *G. princeps* luciferase. The deletion of the 16th amino acid of the signal sequence (a Glu residue) had an enhancing effect on the secretion of the luciferase. Furthermore, it was shown that insertion of stretches of the same hydrophobic amino acid (either Phe, Leu, Ile, or Met) into the *Gaussia princeps* luciferase signal sequence had also an enhancing effect on luciferase secretion. The best secretion of the luciferase was reached with a synthetic signal sequence consisting a stretch of 16 methionine residues. The enhancing effect on

protein secretion of this signal sequence was confirmed by also being able to increase the secretion of the human leukemia inhibitory factor protein expressed in *K. marxianus* [Yarimizu et al. 2015]. Since rational mutagenesis approaches are strongly limited due to the missing information about the targeted protein, random mutagenesis is more extensively performed in protein engineering of other proteins than proteins involved in the secretory pathway [Ali et al. 2016].

3.2 Overexpression and knock-out of genes involved in protein secretion

The understanding of the secretory pathway especially in *S. cerevisiae* has been increased during the last decades. Many genes of this yeast which are involved in the secretory pathway have been identified and characterized. This knowledge has been used for finding putative homologues of these identified genes in other expression hosts [Delic et al. 2013; Gasser et al. 2007]. Another strategy to get evidence if genes are involved in secretory pathway or not is to compare the transcriptome of strains which express secreted (heterologous) proteins with the transcriptome of their parental strain. In this strategy, upregulated and downregulated genes of the heterologous protein expressing strain are studied equally [Gasser et al. 2007; Sims et al. 2005]. For example, Gasser and co-worker analysed the transcriptome of an antibody fragment overexpressing *K. phaffii* strain by chip hybridisation and compare it with that of its parental strain. Since the *K. phaffii* genome was not published yet at this time, microarrays of *S. cerevisiae* were used in this study. They were able to identify 13 genes with a potential function in protein secretion which were upregulated in the antibody fragment expressing strain. Their *S. cerevisiae* homologues were co-expressed in a *K. phaffii* strain expressing human antibody Fab fragment. The co-expression of 6 out of these 13 genes showed an enhancing effect on secretion of the antibody fragment. These 6 genes are coding for chaperons (Ssa4, Sse1), proteins involved in protein transport (Bfr2, Bmh2) and subunits (Cup5, Kin2) of a vacuolar ATPase that is linked to exocytosis [Gasser et al. 2007].

Strain engineering by overexpression and knock-out of genes involved in protein secretion in order to improve protein secretion, was reviewed by Idiris et al. 2010a. In general, the targets in this approach can be divided in four areas:

1) Translocation, protein folding and glycosylation modification in ER,
2) Vesicular transport of proteins,

3) Glycosylation modification in Golgi and
4) Components of the protein degradation system [Idiris et al. 2010a].

It should be mentioned that mutagenesis of glycosylation modifications in Golgi are mainly performed to do glyco-engineering as it was performed for example in *K. phaffii* to humanize the glycosylation of proteins which increased the production efficiency of pharmaceutical proteins in *K. phaffii* [Hamilton et al. 2006; Idiris et al. 2010a].

Mutagenesis of components involved in translocation. The translocation of proteins, the first step of the secretory pathway, could be improved by overexpression of genes involved in this step. For example, overexpression of *S. cerevisiae SSA1* in *K. phaffii* increased the secretion of expressed granulocyte colony stimulating factor [Zhang et al. 2006]. The homologous overexpression of co-translational translocation components Srp14p and Srp54p improved secretion of different heterologous proteins in *S. cerevisiae*. By overexpressing Srp14p and Srp54p, it was possible to increase the secretion of a β-glucosidase, an endo-glucanase, and an α-amylase in *S. cerevisiae*. In this study, the three secreted enzymes were expressed with their native signal sequences and also with *S. cerevisiae* Yap3 and Suc2 signal sequences. The increasing effect of Srp14p and Srp54p overexpression on the secretion of the three enzymes could be seen in each signal sequence variant [Tang et al. 2015].

Mutagenesis of components involved in protein folding. Engineering of protein folding in ER is a very common approach to increase protein secretion and has been performed in many different expression hosts.

The overexpression of the chaperons PDI and BiP, respectively, increased the secretion of different heterologous proteins in *S. cerevisiae* [Idiris et al. 2010a; Shusta et al. 1998]. It was also shown that the additional knock-out of the Mg^{2+} ATPase *PMR1* enhanced the improving effect of BiP overexpression on bovin prochymosin secretion in *S. cerevisiae*. However, this synergistically enhancing effect was not observed in the secretion of plant thaumatin in *S. cerevisiae* [Harmsen et al. 1996]. On the other hand, *S. cerevisiae* strains carrying a null allele of PMR1 showed glycosylation defects and secreted unprocessed mating factor alpha [Rudolph et al. 1989]. In *K. phaffii*, deletion of PMR1 caused a significant reduced growth which could partially be restored by addition of $Ca2^{+}$ [Dux and Inan 2006]. Overexpression of PDI and BiP was also performed in *K. phaffii*. It has been reported that secretion of single chain antibody in *K. phaffii* could be increased by

homologous overexpression of BiP but not by homologous overexpression of PDI [Damasceno et al. 2005]. Interestingly, heterologous overexpression of PDI from *S. cerevisiae* increased secretion of a Fab antibody fragment in *K. phaffii* [Gasser et al. 2007]. Overexpression of PDI had also positive effects on protein secretion in other eukaryotic expression hosts. It has been reported that homologous overexpression of genes coding for protein disulfide isomerase and of polyubiquitin in *K. lactis* increased the secretion of heterologous expressed proteins [Bao and Fukuhara 2001; Yun et al. 2016]. Furthermore, homologous overexpression of genes identified as PDI in fission yeast *Schizosaccharomyces pombe* increased the secretion of human serum transferrin [Mukaiyama et al. 2010].

Further studies showed that also overexpression of other chaperons (Sec63, Ssa1, Ydj1) from *S. cerevisiae* increased the secretion of *Candida antarctica* lipase B in *K. phaffii* [Samuel et al. 2013]. Koskela and co-workers were able to show that antibody secretion in *S. cerevisiae* was improved by heterologous overexpression of folding factors which were found to be overexpressed in mammalian plasma cells during antibody production [Koskela et al. 2017]. Also overexpression of the membrane protein calnexin which is located in the ER and part of the protein quality control system, increased the secretion of different heterologous expressed proteins in *H. polymorpha* [Klabunde et al. 2007].

Mutagenesis of components involved in vesicular trafficking. There is also big potential for improving protein secretion by engineering vesicular protein trafficking. It has been reported that overexpression of factors involved in vesicular protein transport like the syntaxins Sso1 and Sso2 improved protein secretion in *S. cerevisiae* [Ruohonen et al. 1997]. It was also able to show that knock-out of *MON2* which is coding for a scaffold protein responsible for vesicular formation, increased the secretion of heterologous expressed luciferase in *S. cerevisiae* [Idiris et al. 2010a; Kanjou et al. 2007].

Mutagenesis of components involved in protein degradation. Another approach to increase protein secretion is the knocking-out of genes coding for proteases to influence the protein secretion. The proteolytic degradation of the heterologous expressed proteins can be decreased by knocking-out proteases. Idiris and co-workers reported a 30 fold improved secretion of a heterologous expressed human growth hormone in *S. pombe* after knocking-out seven genes coding for proteases [Idiris et al. 2010b].

Beside overexpression or knock-out of genes which are directly involved in the secretory pathway or protein degradation, the approach to influence the secretory pathway indirectly by overexpression of genes coding for transcription factors (TF) was also successful. Several independent studies demonstrated that overexpression of HAC1 (TF for unfolded protein response), Nrg1 and Aft1 increased secretion of different heterologous expressed proteins in *K. phaffii* [Guerfal et al. 2010; Ruth et al. 2014; Stadlmayr et al. 2010].

4 Screening method are a limiting factor in mutagenesis experiments

Targeted approaches for improving protein secretion have been discussed above. A big challenge and limiting factor for improvements by random mutagenesis experiments is the screening of the generated mutants. Usually, (many) thousands of clones have to be screened for having a chance to get a positive one. Therefore, the choice of the screening method or establishing a novel convenient screening method is crucial for such approaches. There are different levels of phenotype screening: i) population level, ii) single/multiple cells level and iii) single cell level. Screenings on population level are the most commonly applied screening methods. They provide a good insight on the phenotype but are inconvenient for high throughput screenings due to the big scale (mL/µL amount) that is needed (e.g. usage of 96 deep-well plates) [Huang et al. 2015; Vervoort et al. 2017].

Although screening on population level made progress in the last decades using -omics technologies, screening on single and multiple cells levels are getting more suitable for screening of mutants [Vervoort et al. 2017]. Huang and co-workers used droplet microfluidics-based cell sorting as high throughput screening. They screened *S. cerevisiae* libraries for improved α-amylase secretion. In this method, single cells from the *S. cerevisiae* libraries were encapsulated with oil. A fluorogenic α-amylase substrate had been added to the liquid culture before encapsulation step. It was used to detect cells with the highest α-amylase concentrations by fluorescence activated droplet sorting [Huang et al. 2015].

Since there is a trend to reduce the volumes of the sample size and to increase in parallel the throughput, many new screening methods are getting established. However, such high throughput screening methods require special equipment (e.g. for cell encapsulation and cell sorting) which might not be available in each lab [Huang et al. 2015; Vervoort et al. 2017]. Therefore, it is very useful to create novel screening methods which are simple and convenient to perform without the need of such special equipment. Here, a simple screening

method is presented with which it is possible to screen *K. phaffii* mutants for better protein secretion for example by simply observing colony growth on agar plates. Therefore new *K. phaffi* protein expression strains were developed as platform strains for further development to efficient secreter strain.

5 Introns in genes involved in secretory pathway as interesting targets for secretion experiments

Very recently, it has been reported that *K. phaffii* genome has around 5100 open reading frames. Only about 11% of the 5100 open reading frames are coding for genes which are spliced [Sturmberger et al. 2016]. It is known that introns are involved in the regulation of eukaryotic gene expression. They have effects on transcription, polyadenylation, mRNA export, translation, and even mRNA decay [Nott et al. 2003]. It has been reported that introns can have enhancing effects on gene expression named "intron mediated enhancement" (IME) [Gallegos and Rose 2015]. On the other side, introns can also have inhibiting effects on gene expression. For example, it has been reported that expression of *HAC1* is inhibited by an intron in *K. phaffii*. The *HAC1* mRNA is spliced under stress conditions (ER folding stress) which just leads to its translation. *HAC1* is a key transcription factor which activates expression of genes involved in unfolded protein response (UPR). Heterologous gene expression can lead to folding stress in ER which triggers UPR pathway [Puxbaum et al. 2015]. This example shows that an intron (in *HAC1*) can have strong influence on the secretory pathway and thereby, on heterologous gene expression. However, it is unknown what the purpose of the identified introns in the genes is or if they have an influence on protein secretion in *K. phaffii*?

Since it was reported that introns are involved in the regulation of the secretory pathway in *K. phaffii*, the approach was done in this work to search for known genes which are involved in secretory pathway of *K. phaffii* and to check if they contain introns. In a next step, the introns of selected genes containing introns were removed to see if protein secretion can be improved in *K. phaffii* when the transcribed mRNA of these genes has not to undergo splicing.

6 ARTP (Artificial Room-Temperature Plasma) as a novel whole-cell mutagenesis method

Beside genome shuffling [Luhe et al. 2011] and removing introns of genes which are involved in protein secretion, ARTP was also used for mutagenesis experiments in this work. ARTP is a novel method for whole-cell mutagenesis experiments which has been successfully used for mutating bacteria, fungi, yeasts, dinoflagellates and plant cells. It has been reported that following improvements by ARTP mutagenesis experiments in tested strains can be classified in

i) increased productivity of the producer strain,

ii) improved growth of the strain and higher tolerance against chemical compounds and other environmental conditions during cultivation and

iii) improved activity of expressed enzymes [Ottenheim et al. 2018].

ARTP requires Helium as working gas and does not produce hazardous radiation or toxic waste. Although the cause for mutations by physical plasma is not fully understood, there is strong evidence that reactive chemical species are generated by ARTP. It is thought that these reactive chemical species interact either directly on DNA or indirectly on other biomolecules to introduce mutations. Comparison of different mutagenesis methods showed that ARTP caused more DNA damage with higher mutation rate than UV radiation and two selected chemical mutagenesis methods [Ottenheim et al. 2018]. The principles of the method and the state of the art of ARTP was summarized very recently by Ottenheim et al. 2018. Interestingly, no reported study using ARTP mutagenesis on *K. phaffii* was found. However, since there are reports about successful mutagenesis experiments on other yeasts [Ottenheim et al. 2018], we found this work to be a good opportunity to perform mutagenesis experiments on *K. phaffii* using ARTP to improve the protein secretion.

Material and Methods

Strains and Medium

K. phaffii BG10 (BioGrammatics, Carlsbad, CA, USA) was used for creating the screening strain strain expressing two gene copies of AnBGL1. The *K. phaffii* strain expressing all three cellulases which had been described in the published manuscript Kickenweiz et al. 2018 (see

Part II of this thesis), was used as second screening strain in the screening for mutants with improved protein secretion.

In this study, *K. phaffii* was incubated in YPD complex medium and minimal medium (BM_ medium). YPD medium and BM_ medium were made as described in Kickenweiz et al. 2018 (see Part II of the thesis). For BM_ medium, cellobiose 0.5% (w/v), carboxymethyl-cellulose (CMC) 1% (w/v) and Avicel 1% (w/v) were used as carbon source. For making agar plates, 2% (w/v) agar was added to the medium before autoclavation.

Vectors and genes

The construction of the used vectors containing the genes *AnBGL1* (β-glucosidase from *Aspergillus niger*) and *eGFP* (enhanced Green Fluorescent Protein) were previously described in Kickenweiz et al. 2018 (see Part II of the thesis). For creating multi *AnBGL1* gene copy strains, *K. phaffii* BG10 was transformed with pPpKan_intX_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt vectors containing different targeting sites (see Appendix A1.4 and A4 of the Thesis).

Yeast transformation

K. phaffii transformation was performed as previously described in Kickenweiz et al. 2018 (see Part II of the thesis) using protocols based on Lin-Cereghino et al. 2005 and Wu and Letchworth 2004.

Ethanol precipitation

PCR products were purified and concentrated by ethanol precipitation. For this, the sample containing the PCR product was filled up to 100 µL with H_2O. 4 µL glycogen (50 µg/mL) solution and 10 µL 3M sodium-acetate were added to the sample. Then, the sample was mixed by inverting the tube. 240 µL of ice-cold absolute ethanol were added and the mix was incubated at -80°C for at least 15 min. After incubation, the sample was centrifuged at maximum speed 4°C for 20 min. The supernatant was removed and the pellet was dried. At the end, the pellet was dissolved in 10-20 µL H_2O.

Mutagenesis and Screening for mutants with improved protein secretion:

Genome shuffling

Genome shuffling for whole genome mutagenesis was performed as previously described by Luhe et al. 2011. Taq polymerase was used for error Prone PCR. 5 ng of isolated *K. phaffii*

BG10 genomic DNA (without RNA) were used as template. Random primers were made by Integrated DNA Technologies (IDT, Coralville, IA, USA). For introducing mutations, the PCR sample contained 0.25 mM $MnCl_2$ and the mixture of dNTPs was 0.6 mM each dCTP and dTTP, 0.2 mM each dATP and dGTP. The PCR mix content was as follows:

PCR mix:
DNA…….....x µL (5 ng isolated genomic DNA)
10x Taq Buffer.....5 µL
MgCl2 (25 mM)…1.5 µL
MnCl2 (15 mM)....0.8 µL
dNTPs (10 mM)…1 µL
dCTP (0.6 mM)….4 µL
dTTP (0.6 mM).....4 µL
random primer…...0.25 µL (Primer solution: 100 pmol/µL)
Taq Pol..……....1 µL
H_2O……..…..38-x µL
Total = 50 µL

The PCR program was set as follows:

PCR program:
95°C……. 5 min
92°C……..1 min
37°C…….2 min x50 cycles
55°C……...4 min
55°C……..5 min

The PCR products in the PCR samples were purified by ethanol precipitation. Freshly prepared competent *K. phaffii* cells were transformed with the purified PCR products. The competent cells for transformation were prepared using protocols based on Lin-Cereghino et al. 2005 and Wu and Letchworth 2004 as previously described in Kickenweiz et al. 2018 (see Part II of the Thesis). However, the resulting resuspension of competent cells at the end of competent cell preparation was diluted 1:1000 with 1M sorbitol. 50 µL of this dilution were used for the yeast transformation. 3-4 µg of the purified error prone products were added to the 50 µL competent cells resuspension. After the electric shock, 900 µL of recovery medium were added to the transformation samples. The recovery medium for screening on BM_CMC 1% agar plates was a 1:1 mixture of BM_CMC 1% (w/v) and 1M Sorbitol. The recovery medium for screening on BM_cellobiose 0.5% (w/v) agar plates was a 1:1 mixture of

BM_cellobiose 0.5% (w/v) and 1M Sorbitol. The transformation samples were incubated for recovery at 30°C for 2 h.

100 µL aliquots of the transformation samples were plated onto BM_CMC 1% (w/v) agar plates and BM_cellobiose 0.5% (w/v) agar plates, respectively. The plates were incubated for a few days at 30°C.

Mutagenesis experiments by removing introns of the genes involved in secretory pathway

All components for PCR were from Thermo Fischer Scientific (Waltham, MA, USA) if not stated otherwise. Phusion polymerase was used for PCR amplification. PCR was performed following the manufacturer's protocol. Standard overlap-extension PCR was performed as previously described by Näätsaari et al. 2012. PCR products were purified using Wizard SV Gel and PCR Clean-Up System from Promega (Fitchburg, WI, USA). The DNA primers were made by Integrated DNA Technologies (IDT, Coralville, IA, USA).

The introns of the genes which are involved in secretory pathway, were removed by overlapping extension PCR (olePCR). For this, the exons of the genes were amplified by PCR without the introns using *K. phaffii* BG10 genomic DNA as template. The primers added overlapping sequences to the exons which could be used for the olePCR. A list of the DNA primers used is shown in Supplemental Table S3.2. The resulted PCR products from the olePCR contained the intron-free gene variants. A part of the intron-free gene variants were mixed together and concentrated by ethanol precipitation. This purified DNA sample containing the intron-free gene variants of all isolated genes, was used for transformation of the screening strains. The screening strains were transformed with the purified PCR products as described above for the transformation with the error Prone PCR products.

ARTP (Artificial Room Temperature Plasma) mutagenesis

For ARTP mutagenesis experiments, an overnight culture (ONC) was inoculated with the screening strain and incubated at 30°C and 150rpm overnight. On the next day, a main culture of 25mL BM_glycerol 1% (w/v) was inoculated with the ONC to an OD_{600} of 0.4. This main culture was incubated at 30°C and 150rpm for 4h. 1-1.5 mL culture was taken for further work when the culture reached an OD_{600} of about 1. The token culture was centrifuged at maximum speed for 1 min. The supernatant was discarded and the pelleted cells were washed in 1 mL H_2O. After another centrifugation step the supernatant was

discarded and the cells were resuspended in 1 mL of 1M sorbitol. 10 µL of the prepared cell resuspension were pipetted onto each metal plate which was used for mutagenesis experiment. Each metal plate with the sample on it was used for a certain stress exposure time. For introducing the stress, the ARTP mutation breeding system from Beijing Siqingyuan (BEST)BIOTECH.Co.,Ltd. (Beijing, China) was used. The stress exposure time was set as recommended for yeast in the manufacturer's protocol (0sec, 30sec, 60sec, 90sec, 120sec and 150sec). After triggering the stress, the metal plates with the samples on it were transferred to a 1.5 mL tube containing 990 µL of recovery medium. The recovery medium for screening on BM_CMC 1% (w/v) agar plates and BM_Avicel 1% (w/v) agar plates was a 1:1 mixture of BM_CMC 1% (w/v) and 1M Sorbitol. The recovery medium for screening on BM_cellobiose 0.5% (w/v) agar plates was a 1:1 mixture of BM_cellobiose 0.5% (w/v) and 1M Sorbitol. The samples were vortexed for a few seconds and diluted 1:5 by transferring 200 µL of the vortexed samples to new tubes which contained 800 µL of fresh recovery medium. The samples were incubated for recovery at 30°C for 2 h. 100 µL aliquots were plated onto screening agar plates and incubated for few days at 30°C.

Results

Generation of multi-copy *AnBGL1* expressing *K. phaffii* strains as screening platform strains

It was possible to demonstrate that constitutive expression of β-glucosidases enables *K. phaffii* to grow on cellobiose [Kickenweiz et al. 2018]. Here, the engineered *K. phaffii* strain expressing AnBGL1 was used as a platform strain for mutagenesis experiments to improve protein secretion. Since only one enzyme is needed for growth on cellobiose, growth effects on cellobiose by expressing different gene copy numbers of *AnBGL1* could be easily investigated. As described in Kickenweiz et al. 2018, *K. phaffii* BG10 was transformed with vectors for constitutive co-expression of the genes *AnBGL1* (β-glucosidase from *Aspergillus niger*) and *eGFP* (enhanced Green Fluorescent Protein) but with different integration sites. The constitutive co-expression of *AnBGL1* and *eGFP* in this construct (AnBGL1_alph.sigseq_NB3_eGFP) was controlled by the NB3 promoter. The mating factor alpha pre-pro-signal sequence from *S. cerevisiae* was used for the secretion of AnBGL1 [Kickenweiz et al. 2018]. Positive transformants were incubated in 96 deep-well plates and fluorescent signal of expressed eGFP was measured. The eGFP signal which was described to

be directly correlated to the gene copy number [Abad et al. 2010] indicated the copy number which had been integrated into the *K. phaffii* genome. Additional, the clones were all tested for their growth on cellobiose by incubating them in BM_cellobiose. The growth on cellobiose as sole carbon source indicated that *AnBGL1* was functionally expressed in the tested clones.

For further work, a clone with two integrated copies of AnBGL1_alph.sigseq_NB3_eGFP, was selected (2 copies strain). Recultivation of the 2 copies strain on BM_glycerol in Deep well plates confirmed that both copies of the vectors containing AnBGL1_alph.sigseq_NB3_eGFP are stable integrated in the genome (see Fig. 3.1).

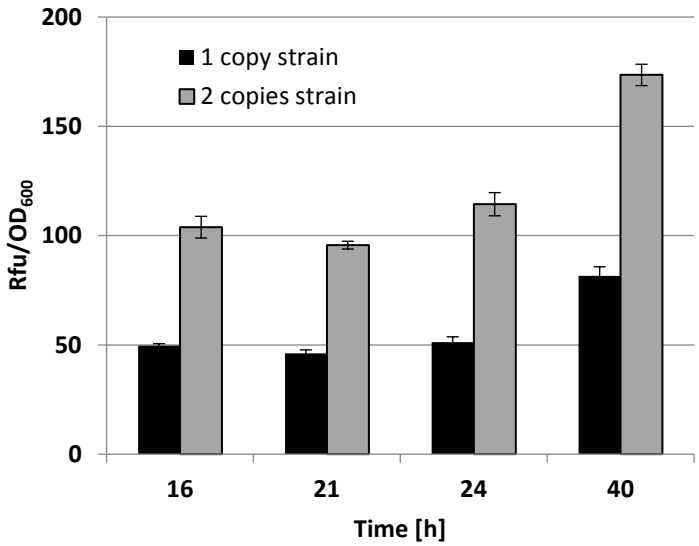


Fig. 3.1: Cultivation of a selected clones expressing either one *AnBGL1* gene copy or two gene copies on glycerol. Measured fluorescent signal of expressed eGFP was normalized to OD_{600}. 1 copy strain (black) and 2 copy strains (grey) were batch cultivated on BM_glycerol 0.25% in Deep-well plates without any additional glycerol feeds. Rfu is for random fluorescent units. Error bars show the standard deviation of biological triplicates.

It was able to see that the measured eGFP signals normalized to OD_{600} of 1 copy strain and 2 copies strain correlated well with the gene copy number during cultivation on glycerol. Since the strains were not cultivated under selective pressure (presence of geneticin or growth on cellobiose), the measured eGFP signals indicated that the integration of the two *eGFP* gene copies were stable or at least the expressed GFP reporter was not degraded. The increase in fluorescence at the end of the cultivation might indicate some potential cell lysis. In a next step, the growth of the 2 copies strain on cellobiose was compared with the representative

single copy strain in flask experiments. The growth curves of both strains are shown in Fig. 3.2.

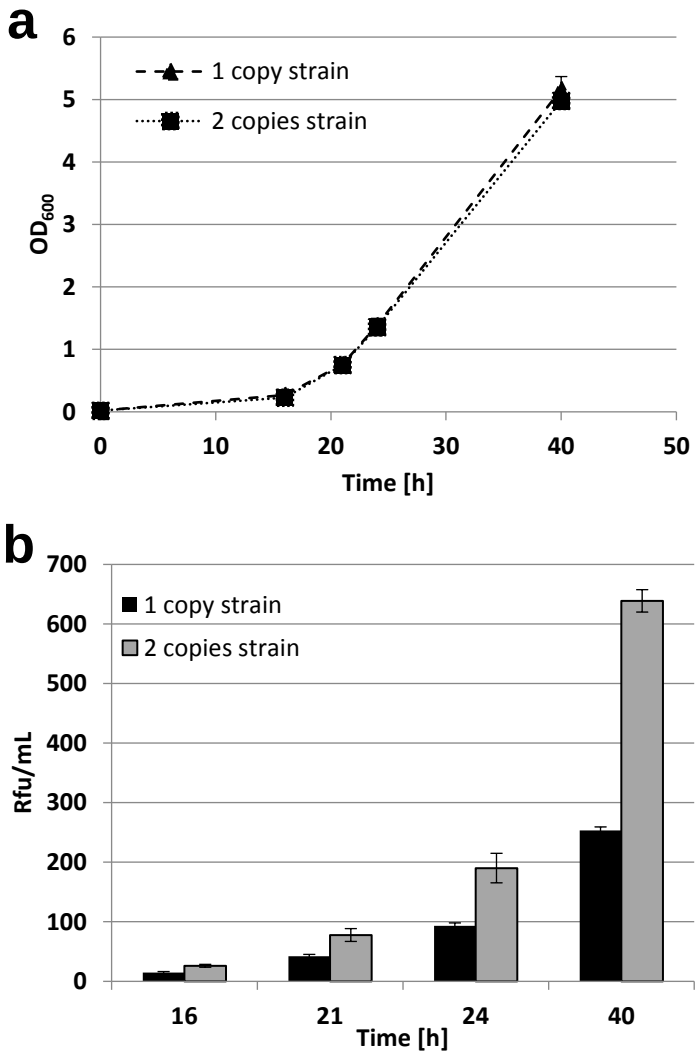


Fig. 3.2: Comparison of *K. phaffii* strains expressing different gene copies of *AnBGL1* during incubation on cellobiose in flasks. Growth curves of *K phaffii* strains expressing one gene copy of *AnBGL1*/*eGFP* (1 copy strain, triangle) and two gene copies of *AnBGL1*/*eGFP* (2 copies strain, square) on batch BM_cellobiose 0.25% in flasks (**a**). Measured fluorescent signal of expressed eGFP in taken samples. 1 copy strain (black) and 2 copy strains (grey) were cultivated on BM_cellobiose 0.25% in flasks. Rfu is for random fluorescent units (**b**). Error bars show the standard deviation of biological triplicates.

Compared to the 1 copy strain, no better growth on cellobiose was observed in the cultures containing the 2 copies strain. Parallel measurement of the eGFP signal per sample during incubation on cellobiose showed that the fluorescent signals in both strains correlated well with the gene copy number. It can be clearly seen that at 16 h, 21 h and 24 h, the measured OD600 was similar in both strains whereas the amount of expressed eGFP (fluorescence

signal per sample) was double as high in the 2 copies strain (see Fig. 3.2 a) and b)). After 40 h of incubation, however, the eGFP concentration in the cultures with the 2 copies strain was more than double as high as the measured eGFP signals per sample in the 1 copy strain cultures (see Fig. 3.2 b). This effect was less significant during incubation on glycerol and the measured eGFP signal also correlated well with the copy number at 40 h of incubation (see Fig. 3.1). A growth delay or slower growth due to *AnBGL1* expression compared to original strain (BG10) was not observed during incubation on glycerol (not shown).

To reproduce the result that multiple gene copy expression of *AnBGL1* did not lead to better growth on cellobiose, *K. phaffii* BG10 strains expressing two and three *AnBGL1* gene copies were redone. Strains which had according to the eGFP measurement two and three copies of *AnBGL1* integrated, were selected and tested for growth on cellobiose with 1 copy strain as reference strain. The growth curves of the three strains on cellobiose in flasks are shown in Fig. 3.3a.

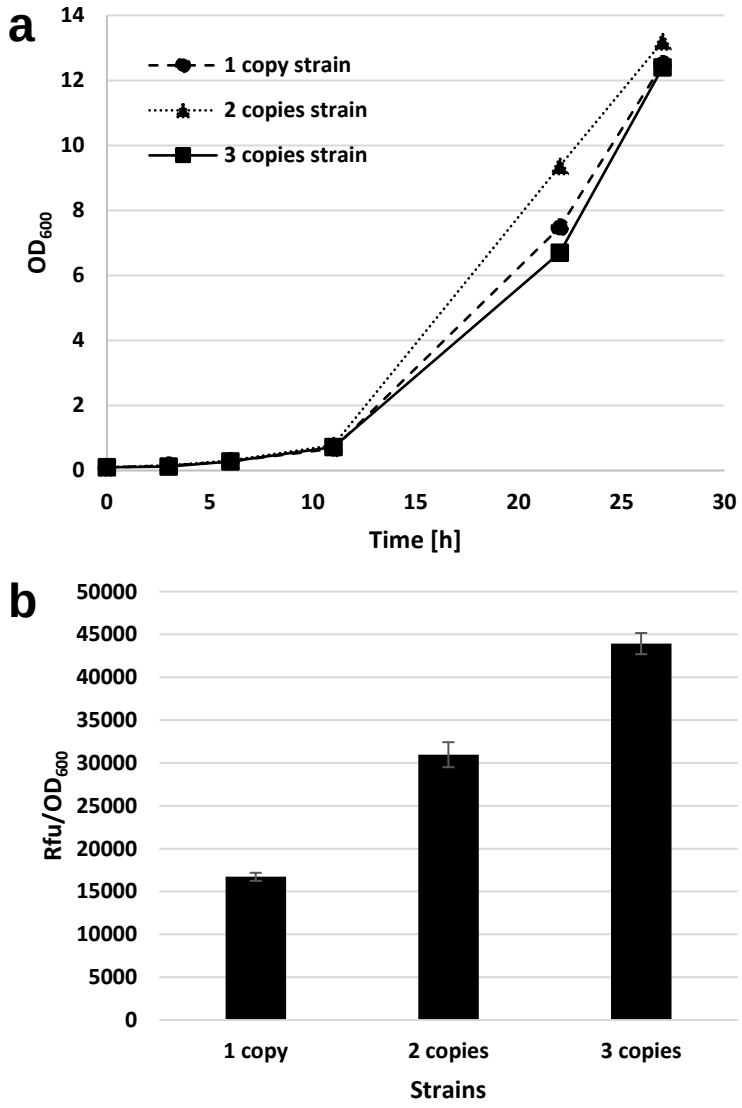


Fig. 3.3: Cultivation of *K. phaffii* BG10 strains expressing one, two and three gene copies of AnBGL1 on cellobiose in flasks. Growth curve of 1 copy (circle), 2 copies (triangle) and 3 copies strain (square) cultivated in BM_cellobiose 0.5% batch cultures in flasks (**a**). Fluorescence measurement of these incubated cultures after 27 h of incubation (**b**). Error bars show the standard deviation of technical triplicates

After 22 h of incubation, the 2 copies strain showed the highest cell density but the 1 copy strain had a higher OD_{600} than the 3 copies strain. Although these results have to be interpreted with caution due to the missing of biological triplicates in this experiment, it is interesting that with exception of the measured OD_{600} values after 22 h of incubation, there were no relevant differences in the growth curve between 1 copy, 2 copies and 3 copies strains on cellobiose. Even after 11 h of incubation, the measured OD_{600} value of the 2 copies strain (0.783) was only slightly higher than the OD_{600} of the 1 copy strain (0.675) and 3 copies strain (0.723). After 22 h of incubation, the 2 copies strain showed the highest biomass

based on OD measurements (OD600 = 9.37) followed by 1 copy strain (OD600 = 7.48) and 3 copies strain (OD600 = 6.7).

The eGFP signal measurement of all strains after 27 h of incubation (Fig 3.3 b) showed well that the fluorescence signals of expressed eGFP was increased according to the copy gene numbers. This indicated that the gene copies integrated in the genome were stable. As in the case of the other strains, no growth delay or slower growth compared to BG10 strain was observed during incubation on glycerol (not shown).

Removing introns from genes involved in secretory pathway

Literature was searched for putative targets for mutagenesis experiments to increase protein secretion in *K. phaffii*. In literature, most of found genes which were supposed to be involved in secretory pathway of *K. phaffii* were named after their homologues in *S. cerevisiae*. The found genes which were supposed to be involved in secretory pathway, was mainly based on the work of Delic et al. 2013, Delic et al. 2014 and Rebnegger et al. 2014. To get the sequence of the reported genes, blastp was performed with the amino acid sequence of the *S. cerevisiae* homologues and *P. pastoris* CBS7435 genome as reference. The list with found genes which are supposed to be involved in secretory pathway as putative targets for mutagenesis experiments are shown in Supplemental Table S3.1. NCBI nucleotide entries of identified genes in CBS7435 genome were checked for predicted introns. Genes which contained predicted introns were further checked by using transcriptomic data of *K. phaffii* BG10 [also indicated in Sturmberger et al. 2016] to confirm if these genes are spliced. Genes which were confirmed to be spliced, were amplified from BG10 genome without their introns to create intron-free gene versions of them for mutation experiments on *K. phaffii* improving protein secretion. The list of intron-free modified genes is in Table 3.1.

Table 3.1: List of genes involved in secretory pathway which have been intron-free modified in this work for mutagenesis experiments.

	Gene	NCBI annotation	Step involved
1	CPR1-1	CCA36430.1	Cytosol
2	SBH1	CCA37917.1	Translocation of synthesized polypeptides to the ER lumen
3	GTT1	CCA37819.1	Protein folding: Chaperones in ER
4	SBA1	CCA36956.1	Protein folding: Chaperones in ER
5	KRE5-2	CCA36284.1	Glycosylation process in ER
6	SEC22	CCA39894.1	COPII vesicle - Anterograde transport from ER to Golgi
7	ERV41-1	CCA39272.1	COPII vesicle - Anterograde transport from ER to Golgi
8	YIP3	CCA38494.1	COPII vesicle - Anterograde transport from ER to Golgi
9	SEC17	CCA38323.1	COPII vesicle - Anterograde transport from ER to Golgi

10	SAR1	CCA36641.1	COPII vesicle - Anterograde transport from ER to Golgi
11	SEC23-1	CCA36506.1	COPII vesicle - Anterograde transport from ER to Golgi
12	SEC13	CCA36223.1	COPII vesicle - Anterograde transport from ER to Golgi
13	BET3	CCA36244.1	COPII vesicle - Anterograde transport from ER to Golgi
14	BET5	CCA37932.1	COPII vesicle - Anterograde transport from ER to Golgi
15	SLY1	CCA40908.1	COPII vesicle - Anterograde transport from ER to Golgi
16	BET1	CCA36330.1	COPII vesicle - Anterograde transport from ER to Golgi
17	YPT6	CCA41004.1	Golgi trafficking
18	SEC28	CCA39470.1	COPI vesicle - Retrograde transport from Golgi to ER
19	SEC20	CCA37383.1	COPI vesicle - Retrograde transport from Golgi to ER
20	APS2-1	CCA36308.1	Exocytosis
21	APS2-3	CCA40057.1	Exocytosis
22	TPM2	CCA41053.1	Exocytosis
23	RHO3	CCA38545.1	Exocytosis
24	MYO2-3	CCA39331.1	Exocytosis

Method for screening mutants for improved protein secretion based on growth

The *K. phaffii* strain expressing two gene copies of *AnBGL1* from Fig. 3.1 (*AnBGL1* 2 copies strain) and the *K. phaffii* strain expressing all three cellulases which was described in Kickenweiz et al. 2018 (see Part II of this thesis) were used as screening strains in this screening. One approach on how to mutate the platform strains, was to remove introns in genes which are supposed to be involved in secretory pathway. For this, the platform strain was transformed with the intron-free gene variants mentioned above. The introns should be removed by integration of the intron-free gene versions via homologous recombination. Beside this approach, genome shuffling and ARTP (Artificial Room Temperature Plasma) were chosen as methods for introducing mutations into *K. phaffii* genome.

After the mutagenesis step, the samples containing the mutated cells were plated onto the screening plates. The *AnBGL1* 2 copies strain was screened using BM_cellobiose agar plates and the strain expressing all three cellulases was screened employing BM_CMC and BM_Avicel agar plates, respectively.

Small grown colonies became visible after 2-3 days of incubation on BM_cellobiose screening plates and after 3-4 days of incubation on BM_CMC screening plates. For being able to see small grown colonies on Avicel plates, the plates had to be incubated for 5-7 days. In the chosen setting, the number of colonies per BM_cellobiose and BM_CMC agar plates varied from 100-300 colonies when cells were mutated by using ARTP and genome shuffling.

Interestingly, the numbers of grown colonies on BM_Avicel screening plates had the tendency to be lower (less than 50 colonies) in the same screening setup. Lower numbers of

grown colonies on screening plates were also observed after transforming the screening strains with the intron-free gene variants. Less colonies were grown on the plates with the mutants compared to negative control where the competent cells were transformed with water instead.

In ARTP mutagenesis experiments, the pulse was set as suggested by the supplier for mutagenesis experiments on yeast. It should be mentioned here that longer stress exposure on the sample seemed not to have a clear effect on the number of grown colonies on screening plates. The number of grown colonies on the screening plates did not become less when the samples were exposed to higher the stress level was.

Some colonies which were bigger in size than the average on the same screening plate, were isolated for further tests to confirm the results by cultivation in liquid medium containing the specific carbon source. However, further cultivations in liquid medium containing either cellobiose or CMC as sole carbon source did not confirm a faster growth of these selected clones compared to the non-mutated reference strain.

Discussion

AnBGL1 multi-copy strains seemed to be suitable screening strains for mutagenesis experiments

Although the gene copy numbers in the tested *AnBGL1* expressing strains were indicated by measuring the eGFP signal only and not confirmed by further methods, strong evidence for the reliability of the results was made by redoing specific *AnBGL1* multi-copy strains by targeted multi site integration using separate markers and reproduction of their growth behaviour on cellobiose (see Fig. 3.3). It was possible to show that multiple gene copy expression of *AnBGL1* in this system did not enable a better growth of *K. phaffii* on cellobiose in liquid medium compared to the 1 copy strain (see Fig. 3.2 and 3.3). Although the results of Fig. 3.3 have to be interpreted with caution due of the missing biological triplicates, it is remarkable that the growth curve of all 3 tested strains were similar in the first 11 h of incubation. After 22 h of incubation, the 2 copies strain showed the highest cell density strains but its growth and that of the 3 gene copies strain on cellobiose are far below the theoretical expected growth compared to the 1 copy strain. In this experiment, the cultures were inoculated to an OD_{600} of 0.1 which was higher than in other experiments. By starting

with a higher OD_{600} value in the cultures, a growth effect should be detected earlier during cultivation. A higher cell number should produce higher AnBGL1 titers from begin onwards which consequently should lead to faster glucose liberation from cellobiose and therefore a faster increase of cell densities. In Kickenweiz et al. 2018 (see Part II of this thesis), it was shown that no significant increase in cell density was observed when the parental *K. phaffii* BG10 strain was incubated in BM_cellobiose liquid medium. This underlined the importance of β-glucosidases for *K. phaffii* to grow on cellobiose. In the experiment shown in Fig. 3.3, the OD_{600} value of the cultures was increased after 11 h of incubation from 0.1 in the beginning to about 0.7 which means that β-glucosidase was already produced and secreted in the cultures. Therefore, a higher measured OD_{600} value of the 2 copies strain and 3 copies strain had been expected to be visible after 6 h and 11 h of incubation, respectively.

Based on the results of the growth experiments (see Fig. 3.2 and 3.3), it seemed that the limit of efficient *AnBGL1* expression or secretion, respectively is (almost) reached with the expression of one *AnBGL1* gene copy because the integration of more gene copies of *AnBGL1* in *K. phaffii* had no relevant effects on growth on cellobiose. This strongly indicates that this method is suitable as screening method for detecting *K. phaffii* mutants with better protein secretion. However, final conclusion can only be made after measuring the gene copy number of *AnBGL1* on genomic level, the expression of *AnBGL1* on transcription level (qPCR) and the secretion of AnBGL1 on protein level (SDS-PAGE) in the *AnBGL1* 2 copies strain. Just the comparison with the 1 copy strain will show then if there is a bottleneck and on which level it is. Furthermore, another control experiment can be to add β-glucosidase to the medium to see if this leads to a faster growth of 1 copy and the 2 copies strains on cellobiose containing medium.

The non-correlation between better growth on cellobiose and *AnBGL1* gene copy number points to a bottleneck in the secretory pathway. The question is where the bottleneck in the secretory pathway is? As described in the introduction, secretory pathway is very complex. Intracellular proteins and extracellular proteins have some steps of the secretory pathway in common. Compared to eGFP, the protein structure of AnBGL1 is much bigger and more complex than eGFP. It can be that the expression of more complex intracellular proteins would also be negatively affected in the multi-copy strains. It has been reported that heterologous overexpression in *K. phaffii* can lead to cell stress (folding stress in ER) [Gasser et al. 2008; Liu et al. 2014]. The cell stress depends on the target protein [Hyka et al. 2010]. The measured eGFP signals still correlated with the gene copy numbers (see Fig. 3.2) meaning that expression of “simple” intracellular eGFP was not affected in multi-copy

strains. Maybe, it is still possible for cells to express simple proteins on normal level in the presence of folding stress whereas expression of proteins with more complex structures is negatively affected. Although such a stress cannot be excluded in tested multi-copy strains, no evidence for a bottleneck in earlier steps of the secretory pathway (e.g. protein folding) or even in earlier stages (e.g. transcription) was found, since no cell growth delay on glycerol and glucose was observed in these strains. It is likely that a bottleneck is located in the later steps of secretory pathway affecting primarily the secretion of extracellular proteins in the tested multi-copy strains. However, as mentioned above, it is required to perform a qPCR and a SDS-PAGE for being able to make a conclusion.

A bottleneck in the secretory pathway which prevented better growth of multi copy strains on cellobiose, made the *AnBGL1* multi-copy strains to interesting platform strains for mutagenesis experiments to improve efficiency of protein secretion in *K. phaffii*.

Introns in genes involved in secretory pathway as interesting targets for secretion experiments

It was surprising that so many genes which are supposed to be involved in secretory pathway contain introns. Especially many introns were detected in genes which are involved in COPII vesicle transport and exocytosis. 56 genes were found in literature to be involved in COPII vesicle transport and 28 genes which are involved in exocytosis. From these genes, 19 out of the 56 genes which are involved in COPII vesicular transport, were annotated in NCBI to contain introns. It could be confirmed by analysing transcriptomic data that 11 genes of these 19 genes are spliced. From 28 genes involved in exocytosis, 8 genes were annotated on NCBI to contain introns and it could be confirmed by using transcriptomic data that 6 genes are spliced. This means that about 20% of the found genes which are involved in these two key steps of protein secretion, are spliced. This is an intron frequency which is double as high as the intron frequency of the total open reading frames on the *K. phaffii* genome. The finding that the introns frequency is very high in genes which are involved in secretory pathway, might be evidence for a possible regulative function of introns in secretory pathway. Since no clone with improved protein secretion was isolated in this work and therefore no further tests could not be done, it remains unclear what the purpose of the identified introns in the genes is or if they have an influence on protein secretion in *K. phaffii*.

The observation that the number of grown colonies on the screening plate was low after yeast transformation with the intron-free gene variants, might be evidence for the unviability of

some mutants. Since all intron-free gene variants were mixed for yeast transformation, it is not possible to know which gene mutants are not viable. The reason for mixing all intron-free gene variants was to have also the chance to test the effect of possible mutation combinations on protein secretion. Although no mutant showing better growth on cellobiose/CMC was isolated during screening, it is not clear if removing introns from these genes can have a positive effect on protein secretion in *K. phaffii*. Further experiments which are focused on removing the intron of only one gene at once, could give clearance if removing introns have an effect on protein secretion in *K. phaffii*. In addition, expression from weaker or differently regulated promoters might counteract the possible growth retarding effect of some transformants or the observed reduction of the transformation rate.

Screening based on growth might be a simple method for identification of mutants with improved protein secretion

Screening methods based on cell growth are a powerful approach. Changes in growth behaviour can be easily detected without requiring expensive equipment. To the best of our knowledge, the screening presented here is the first method based on growth to screen for *K. phaffii* mutants with improved protein secretion.

Besides screening on cellobiose and CMC agar plates on which the strains were able to grow (the two *AnBGL1* gene copy strain was able to grow on cellobiose and the strain expressing all three cellulases on CMC), the strain expressing all three cellulases was also screened onto BM_Avicel plates after mutagenesis using ARTP. In Kickenweiz et al. 2018 (see Part II of the thesis), it was shown that no growth of this strain was observed on Avicel. The idea to screen for improved protein secretion on Avicel plates was that improving the protein secretion in this screening strain might enable it to grow on Avicel. Unfortunately, no clone showing growth on Avicel could be isolated. However, it was interesting that the number of visible colonies on Avicel plates was lower than on CMC plates although the same samples were used for plating onto the different screening plates. Since background growth was observed on minimal medium agar plates, it was expected that the numbers of colonies on Avicel screening plates would be about the same as on CMC screening plates. An explanation for this observation might be that in contrast to the samples on CMC, not enough glucose was liberated from Avicel for the cells to fully recover from stress after mutagenesis.

Another observation during the screening was the effect of cross-feeding. The secreted cellulases/*AnBGL1* liberated glucose from CMC/cellobiose which could also be used by other cells. This effect was clearly seen when strains expressing cellulases were streaked onto the same CMC and cellobiose agar plates with negative controls. The distance between the colonies has a strong influence on this effect. The closer the negative clone was to the cellulase expressing clone on the agar plate, the better the better the negative clone grew on CMC and cellobiose. However, colonies of negative clones were not observed to reach the colony size of cellulase expressing cells on agar plates.
This effect is relevant for this screening since it could influence the screening sensitivity, especially when the screening strains already had the ability to grow on CMC and cellobiose, respectively. This problem could be solved by isolation of all bigger colonies in a cluster on the screening plate. The clone with improved protein secretion which was responsible for the better growth of the colonies in the cluster, can be then identified in further quantitative tests.

However, a better growth on cellobiose/CMC of the few clones which had been picked for further tests in the screening, could not be reproduced. One reason might be that the mutation which might have affected protein secretion, was not stable. However, it is more likely that the better cell growth of these clones was caused by a better growth conditions as for example a faster recovery after mutagenesis or better localisation on the screening plate. It has been shown that after few days of incubation, colonies which were located in parts of the screening plate with lower colony densities formed bigger colonies than colonies located in parts with higher colony densities. The reason for this seems to be that the resources are limited when colony densities are high. This gives isolated colonies which were more distant from other colonies of the plate, an advantage in later phases of incubation. The effect mentioned above that the clones of the same colony cluster grow better due to more secreted enzymes is valid as long the nutrients are not restrictive. Another growth effect which was observed during screening was that colonies located at the edge of the agar plate tended to grow faster. It might be that colonies located at the edge of the plate benefited from that the enzymes and glucose from cellobiose/CMC cannot diffuse further at the edge side. Therefore, the amount of available glucose at the edge area might have been higher than in other parts of the plate.

Although no clone with improved protein secretion was isolated in this study, this screening technique can be suitable for screening mutants with higher protein secretion efficiency. The method can be easily performed and is primarily limited by the number of prepared agar

plates. It can be useful as first line screening to test a high number of mutants and to isolate promising candidates for further (quantitative) tests. To increase the chance to find a positive clone in the screening, it is also possible to try out further mutagenesis methods as e.g. UV-rays, X-rays or chemical mutagenesis. Furthermore, it would be good to introduce an already known mutation which increases protein secretion in the platform strains and test this screening method with this positive control. In that way, it would be possible to get final proof that this screening method is working.

Acknowledgements

This work was financially supported by the Science and Engineering Research Council (SERC) of the Agency for Science, Technology and Research (A*STAR), Singapore (SERC grant no 1526004159, ICES/15-175A02) and by the A*STAR Research Attachment Programme Scholarship from the A*STAR Graduate Academy.

References

Abad S, Kitz K, Schreiner U, Hoermann A, Hartner F, Glieder A (2010) Real time PCR based determination of gene copy numbers in *Pichia pastoris*. Biotechnol. J 5: 413-420. doi: https://doi.org/10.1002/biot.200900233

Agaphonov MO, Sokolov SS, Romanova NV, Sohn J-H, Kim S-Y, Kalebina TS, Choi E-S, Ter-Avanesyan MD (2005) Mutation of the protein-*O*-mannosyltransferase enhances secretion of the human urokinase-type plasminogen activator in *Hansenula polymorpha*. Yeast **22**: 1037–1047. doi: https://doi.org/10.1002/yea.1297

Aguiar TQ, Ribeiro O, Arvas M, Wiebe MG, Penttilä M, Domingues L (2014) Investigation of protein secretion and secretion stress in *Ashbya gossypii*. BMC Genomics 15:1137. doi: https://doi.org/10.1186/1471-2164-15-1137

Ahmad M, Hirz M, Pichler H, Schwab H (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98:5301–5317. doi: 10.1007/s00253-014-5732-5

Ali Q, Sultan A, Azhar A, Kanwal N, Ali F, Rasheed A, Adnan M, Ahmad S, Naseem MU, Nasir IA, Husnain T (2016) Protein Engineering: A brief overview methodologies and applications. Life Sci J 13:12.

Bao W-G, Fukuhara H (2001) Secretion of human proteins from yeast: stimulation by duplication of polyubiquitin and protein disulfide isomerase genes in *Kluyveromyces lactis*. Gene 272:103-110. doi: https://doi.org/10.1016/S0378-1119(01)00564-9

Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CB (2005) Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. Appl Microbiol Biotechnol 74:381–389. doi: https://doi.org/10.1007/s00253-006-0652-7

Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D, Brigitte Gasser B (2013) The secretory pathway: exploring yeast diversity. FEMS Microbiol Rev 37:872–914. doi: https://doi.org/10.1111/1574-6976.12020

Delic M, Göngrich R, Mattanovich D, Gasser B (2014) Engineering of Protein Folding and Secretion - Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins. Antioxid Redox Signal 21: 414–437. doi: 10.1089/ars.2014.5844

Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27:297–306. doi: https://doi.org/10.1016/j.biotechadv.2009.01.008

Dux MP, Inan M (2006) Identification and characterization of calcium and manganese transporting ATPase (*PMR1*) gene of *Pichia pastoris*. *Yeast* 2006 23:613–621. doi: https://doi.org/10.1002/yea.1379

Fitzgerald I, Glick BS (2014) Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Microb Cell Fact 13:125. doi: https://doi.org/10.1186/s12934-014-0125-0

Gallegos JE, Rose AB (2015) The enduring mystery of intron-mediated enhancement. Plant Science 237:8–15. doi: https://doi.org/10.1016/j.plantsci.2015.04.017

Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D (2007) Transcriptomics-Based Identification of Novel Factors EnhancingHeterologous Protein Secretion in Yeasts Appl Environ Microbiol 73:6499–6507. doi: 10.1128/AEM.01196-07

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact **7**:11. doi: 10.1186/1475-2859-7-11

Geva Y, Crissman J, Arakel EC, Gómez-Navarro N, Chuartzman SG, Stahmer KR, Schwappach B, Miller EA, Schuldiner M (2017) Two novel effectors of trafficking and maturation of the yeast plasma membrane H+-ATPase. Traffic 18:672–682. doi: 10.1111/tra.12503

Goto M (2007) Protein O-Glycosylation in Fungi: Diverse Structures and Multiple Functions. Biosci Biotechnol Biochem 71:1415-1427. doi: http://dx.doi.org/10.1271/bbb.70080

Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, Callewaert N (2010) 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 **9**:49. doi: https://doi.org/10.1186/1475-2859-9-49

Gunasekaran P, Augur C, Viniegra-González G, Favela-Torres E (1996) Approaches Towards Improvement of Microbial Strains Used in Solid State Fermentation. In: Galindo E (ed) Fronteras en biotecnología y bioingeniería, Sociedad Mexicana de Biotecnología y Bioingenieria, Mexico, pp 359-364

Gutierrez JM, Lewis NE (2015) Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modelling. Biotechnol J 10:939–949. doi: 10.1002/biot.201400647

Hamilton SR. Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU (2006) Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins. Science 313:1441-1443. doi: 10.1126/science.1130256

Harmsen MM, Bruyne MI, Raué HA, Maat J (1996) Overexpression of binding protein and disruption of the *PMR1* gene synergistically stimulate secretion of bovine prochymosin but not plant Thaumatin in yeast. Appl Microbiol Biotechnol 46:365–370. doi: https://doi.org/10.1007/BF00166231

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D (2003) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotechnol Bioeng 85:367-375. doi: https://doi.org/10.1002/bit.10904

Huang M, Bai Y, Sjostrom SL, Hallström BM, Liu Z, Petranovic D, Uhlén M, Joensson HN, Andersson-Svahn H, Nielsen J (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. PNAS 112: E4689–E4696. doi: https://doi.org/10.1073/pnas.1506460112

Hyka P, Züllig T, Ruth C, Looser V, Meier C, Klein J, Melzoch K, Meyer H-P, Glieder A, Kova K (2010) Combined use of fluorescent dyes and flow cytometry to quantify the physiological state of *Pichia pastoris* during the production of heterologous proteins in high-cell-density fed-batch cultures. Appl Environ Microbiol 76:4486–4496. doi: 10.1128/AEM.02475-09

Idiris A, Tohda H, Kumagai H, Takegawa K (2010a) Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86:403–417. doi: 10.1007/s00253-010-2447-0

Idiris A, Tohda H, Sasaki M, Okada K, Kumagai H, Giga-Hama Y, Takegawa K (2010b) Enhanced protein secretion from multiprotease-deficient fission yeast by modification of its vacuolar protein sorting pathway. Appl Microbiol Biotechnol 85:667–677. doi: 10.1007/s00253-009-2151-0

Jürgens G, Geldner N (2002) Protein secretion in plants: from the *trans*-Golgi network to the outer space. Traffic 3:605–613. doi: https://doi.org/10.1034/j.1600-0854.2002.30902.x

Kanjou N, Nagao A, Ohmiya Y, Ohgiya S (2007) Yeast mutant with efficient secretion identified by a novel secretory reporter, Cluc. Biochem Biophys Res Commun 358:429–434. doi: https://doi.org/10.1016/j.bbrc.2007.04.140

Kickenweiz T, Glieder A, Wu JC (2018) Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase. Appl Microbiol Biotechnol 102: 1297–1306. doi: https://doi.org/10.1007/s00253-017-8656-z

Kim I-S, Kim H-Y, Shin S-Y, Kim Y-S, Lee DH, Park KM, Yoon H-Y (2010) A cyclophilin A CPR1 overexpression enhances stress acquisition in *Saccharomyces cerevisiae*. Mol Cells 29:567-574. doi: https://doi.org/10.1007/s10059-010-0071-6

Klabunde J, Kleebank S, Piontek M, Hollenberg CP, Hellwig S, Degelmann A (2007) Increase of calnexin gene dosage boosts the secretion of heterologous proteins by *Hansenula polymorpha*. FEMS Yeast Res 7:1168–1180. Doi:10.1111/j.1567-1364.2007.00271.x

Koskela EV, de Ruijter JC, Frey AD (2017) Following nature's roadmap: folding factors from plasma cells led to improvements in antibody secretion in *S. cerevisiae*. Biotechnol. J. 2017, 12:1600631. Doi: ttps://doi.org/10.1002/biot.201600631

Labunskyy VM, Hatfield DL, Gladyshev VN (2007) The Sep15 Protein Family: Roles in Disulfide Bond Formation and Quality Control in the Endoplasmic Reticulum. IUBMB life 59:1-5. doi: https://doi.org/10.1080/15216540601126694

Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, Lin-Cereghino GP (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 38: 44–48

Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chang JW, Chavan A, Tsai JW, Lin-Cereghino J (2013) The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene 2013 519: 311–317. doi:10.1016/j.gene.2013.01.062

Liu H, Qin Y, Huang Y, Chen Y, Cong P, He Z (2014) Direct evaluation of the effect of gene dosage on secretion of protein from yeast *Pichia pastoris* by expressing eGFP. *J Microbiol Biotechnol* (2014), 24:144–151. doi: http://dx.doi.org/10.4014/jmb.1308.08010

Luhe AL, Tan L, Wu J, Zhao H (2011) Increase of ethanol tolerance of *Saccharomyces cerevisiae* by error-prone whole genome amplification. Biotechnol Lett 33:1007–1011. doi: 10.1007/s10529-011-0518-7

Moremen KW, Tiemeyer M, Nairn AV (2012) Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 13:448–462. doi:10.1038/nrm3383

Mukaiyama H, Tohda H, Takegawa K (2010) Overexpression of protein disulfide isomerases enhances secretion of recombinant human transferrin in *Schizosaccharomyces pombe*. Appl Microbiol Biotechnol 86:1135–1143. doi: 10.1007/s00253-009-2393-x

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* 7: e39720. doi: https://doi.org/10.1371/journal.pone.0039720

Nickel W (2005) Unconventional secretory routes: direct protein export across the plasma membrane of mammalian cells. Traffic 6:607–614. doi: 10.1111/j.1600-0854.2005.00302.x

Nott A, Meislin SH, Moore MJ (2003) A quantitative analysis of intron effects on mammalian gene expression. RNA 9:607–617. doi: http://www.rnajournal.org/cgi/doi/10.1261/rna.5250403

Ottenheim C, Nawrath M, Wu JC (2018) Microbial mutagenesis by atmospheric and room-temperature plasma (ARTP): the latest development. Bioresour Bioprocess 5:12. doi: https://doi.org/10.1186/s40643-018-0200-1

Parakh S, Atkin JD (2015) Novel roles for protein disulphide isomerase in disease states: a double edged sword? Front Cell Dev Biol 3:30. doi: 10.3389/fcell.2015.00030

Porro D, Gasser B, Fossati T, Maurer M, Branduardi P, Sauer M, Mattanovich D (2011) Production of recombinant proteins and metabolites in yeasts. Appl Microbiol Biotechnol 89:939–948. doi: 10.1007/s00253-010-3019-z

Puxbaum V, Mattanovich D, Gasser B (2015) Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl Microbiol Biotechnol 99:2925–2938. doi: 10.1007/s00253-015-6470-z

Rakestraw JA, Sazinsky SL, Piatesi A, Antipov, Wittrup KD (2009) Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *S. cerevisiae*. Biotechnol Bioeng 103:1192–1201. doi:10.1002/bit.22338.

Rebnegger C, Graf AB, Valli M, Steiger MG, Gasser B, Maurer M, Mattanovich D (2014) In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. Biotechnol J 9:511-525. doi: https://doi.org/10.1002/biot.201300334

Ribeiro O, Magalhães F, Aguiar TQ, Wiebe MG, Penttilä M, Domingues L (2013) Random and direct mutagenesis to enhance protein secretion in *Ashbya gossypii*. Bioengineered 4:322–331. doi: https://doi.org/10.4161/bioe.24653

Richter S, Voß U, Jürgens G (2009) Post-Golgi Traffic in Plants. Traffic 10: 819–828. doi: 10.1111/j.1600-0854.2009.00916.x

Roth J, Zuber C, Park S, Jang I, Lee Y, Kysela KG, Le Fourn V, Santimaria R, Bruno Guhl B, Cho JW (2010) Protein *N*-glycosylation, protein folding, and protein quality control. Mol Cells 30: 497–506. doi: https://doi.org/10.1007/s10059-010-0159-z

Routenberg Love K, Politano TJ, Panagiotou V, Jiang B, Stadheim TA, Love JC (2012) Systematic Single-Cell Analysis of *Pichia pastoris* Reveals Secretory Capacity Limits Productivity. PLoS ONE 7: e37915. doi:10.1371/journal.pone.0037915

Rudolph HK, Antebi A, Fink GR, Buckley CM, Dorman TE, LeVitre JA, Davidow LS, Mao J-i, Moir DT (1989) The yeast secretory pathway is perturbed by mutations in *PMR1*, a member of a Ca^{2+} ATPase family. Cell 58:133–145. doi: https://doi.org/10.1016/0092-8674(89)90410-8

Ruohonen L, Toikkanen J, Tieaho V, Outola M, Soderlund H, Keranen S (1997) Enhancement of protein secretion in *Saccharomyces cerevisiae* by overproduction of Sso protein, a late-acting component of the secretory machinery. Yeast 13:337–351. doi: https://doi.org/10.1002/(SICI)1097-0061(19970330)13:4<337::AID-YEA98>3.0.CO;2-K

Ruth C, Buchetics M, Vidimce V, Kotz D, Naschberger S, Mattanovich D, Pichler H, Gasser B (2014) *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion. Microb Cell Fact 13:120. doi: https://doi.org/10.1186/s12934-014-0120-5

Saloheimo M, Pakula TM (2012) The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*). Microbiology 158:46–57. doi: 10.1099/mic.0.053132-0

Samuel P, Vadhana AKP, Kamatchi R, Antony A, Meenakshisundaram S (2013) Effect of molecular chaperones on the expression of *Candida antarctica* lipase B in *Pichia pastoris.* Microbiol Res 168: 615-620. doi: https://doi.org/10.1016/j.micres.2013.06.007

Schaller M, Borelli C, Korting HC, Hube B (2005) Hydrolytic enzymes as virulence factors of *Candida albicans*. Mycoses 48:365–377. doi: https://doi.org/10.1111/j.1439-0507.2005.01165.

Schmidpeter PAM, Schmid FX (2015) Prolyl isomerization and its catalysis in protein folding and protein function. J Mol Biol 427:1609-1631. doi: https://doi.org/10.1016/j.jmb.2015.01.023

Shusta EV, Raines RT, Plückthun A, Wittrup KD (1998) Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. Nat Biotechnol 16:773–777.

Sims AH, Gent ME, Lanthaler K, Dunn-Coleman NS, Oliver SG, Robson GD (2005) Transcriptome analysis of recombinant protein secretion by *Aspergillus nidulans* and the unfolded-protein response in vivo. Appl Environ Microbiol 71:2737–2747. doi: 10.1128/AEM.71.5.2737-2747.2005

Singh A, Patel AK, Adsul M, Singhania RR (2017) Genetic modification: a tool for enhancing cellulase secretion. Biofuel Res 4:600-610. doi: 10.18331/BRJ2017.4.2.5

Sorgo AG, Heilmann CJ, Brul S, de Koster CG, Klis FM (2013) Beyond the wall: *Candida albicans* secret(e)s to survive. FEMS Microbiol Lett 338:10–17. doi: https://doi.org/10.1111/1574-6968.12049

Stadlmayr G, Benakovitsch K, Gasser B, Mattanovich D, Sauer M (2010) Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*. Biotechnol Bioeng 105:543-555. doi: https://doi.org/10.1002/bit.22573

Sturmberger L, Chappelle T, Geier M, Krainer F, Day KJ, Vide U, Trstenjak S, Schiefer A, Richardson T, Soriaga L, Darnhofer B, Birner-Gruenberger R, Glick BS, Tolstorukov I, Cregg J, Madden K, Glieder A (2016) Refined *Pichia pastoris* reference genome sequence. J Biotechnol 235:121-131. doi: http://dx.doi.org/10.1016/j.jbiotec.2016.04.023

Tang H, Bao X, Shen Y, Song M, Wang S, Wang C, Hou J (2015) Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. Biotechnol Bioeng 112:1872-1882. doi: https://doi.org/10.1002/bit.25596

Tannous A, Pisoni GB, Hebert DN, Molinari M (2015) N-linked sugar-regulated protein folding and quality control in the ER. Semin Cell Dev Biol 41:79–89. doi: 10.1016/j.semcdb.2014.12.001

Vervoort Y, Linares AG, Roncoroni M, Liu C, Steensels J, Verstrepen KJ (2017) High-throughput system-wide engineering and screening for microbial biotechnology. Curr Opin Biotechnol 46: 120-125. doi: https://doi.org/10.1016/j.copbio.2017.02.011

Wang S, Kaufmann RJ (2012) The impact of the unfolded protein response on human disease. J Cell Biol 197:857–867. doi: 10.1083/jcb.201110131

Wu S, Letchworth GJ (2004) High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *Biotechniques* 36:152-154.

Xiao S, Shiloach J, Betenbaugh MJ (2014) Engineering Cells to Improve Protein Expression. Curr Opin Struct Biol 26:32-38. doi: https://doi.org/10.1016/j.sbi.2014.03.005

Yarimizu T, Nakamura M, Hoshida H, Akada R (2015) Synthetic signal sequences that enable efficient secretory protein production in the yeast *Kluyveromyces marxianus*. Microb Cell Fact 14:20. doi: https://doi.org/10.1186/s12934-015-0203-y

Yoshida H (2007) ER stress and diseases. *FEBS J 274:*630–658. doi:10.1111/j.1742-4658.2007.05639.x

Yun C-R, Kong J-N, Chung J-H, Kim M-C, Kong K-H (2016) Improved secretory production of the sweet-tasting protein, Brazzein, in *Kluyveromyces lactis*. J Agric Food Chem 64:6312−6316. doi: 10.1021/acs.jafc.6b02446

Zhang W, Zhao H-L, Xue C, Xiong X-H, Yao X-Q, Li X-Y, Chen H-P, Liu Z-M (2006) Enhanced secretion of heterologous proteins in *Pichia pastoris* following overexpression of *Saccharomyces cerevisiae* chaperone proteins. Biotechnol Prog 22:1090-1095. doi: https://doi.org/10.1021/bp060019r

Zheng X, Fang B, Han D, Yang W, Qi F, Chen H, Li S (2016) Improving the Secretory Expression of an α-Galactosidase from *Aspergillus niger* in *Pichia pastoris*. doi: https://doi.org/10.1371/journal.pone.0161529

Supplemental data

Table S3.1: List of genes which are supposed to be involved in secretory pathway.

Gene name	NCBI annotation	Introns found in ORF according to NCBI annotation	Introns confirmed by transcriptomic data
Cytosol			
FES1	CCA38940.1	No	-
CNS1	CCA38509.1	No	-
SIS1	CCA37853.1	No	-
SSE1	CCA37673.1	No	-
YDJ1	CCA37768.1	No	-
SSA1-1	CCA40589.1	No	-
SSA1-2	CCA39802.1	No	-
HCH1-1	CCA40005.1	Yes	No
HCH1-2	CCA36980.1	No	-
SSB2-1	CCA39425.1	No	-
SSB2-2	CCA39764.1	No	-
CBR6	CCA39596.1	No	-
JJJ1	CCA37438.1	Yes	No
EGD2	CCA38206.1	No	-
HSP104-1	CCA36271.1	No	-
HSP104-2	CCA38657.1	No	-
SNL1	CCA36999.1	No	-
ZUO1	CCA38702.1	No	-
CPR7	CCA39789.1	No	-
EGD1	CCA36340.1	Yes	No
CPR1-1	CCA36430.1	Yes	Yes
CPR1-2	CCA37613.1	No	-
CPR1-3	CCA40596.1	No	-
SSA3	CCA39942.1	No	-
ER			
Tanslocation of synthesized polypeptides to the ER lumen			
SPC2	CCA38383.1	No	-
SPC3	CCA40254.1	No	-
SEC11	CCA37097.1	No	-
SEC61	CCA36368.1	No	-
SEC62	CCA39128.1	No	-
SEC66	CCA38548.1	No	-
SSH1	CCA37554.1	No	-
SRP101	CCA38600.1	No	-
YET3	CCA38944.1	No	-
HRI1	CCA38532.1	No	-
4_0650-1	CCA40378.1	No	-
4_0650-2	CCA38360.1	No	-
SGT2	CCA37018.1	No	-
SRP72	CCA37121.1	No	-
SEC65	CCA39628.1	No	-
SBH1	CCA37917.1	Yes	Yes
SEC63	CCA40752.1	No	-
SEC72	CCA38530.1	No	-
SRP54	CCA40829.1	No	-
SLS1	CCA36429.1	No	-
Protein folding: Chaperones in ER			

ERJ5	CCA36303.1	No	-
4_0765	CCA40369.1	No	-
SLP1	CCA37358.1	No	-
CHS7	CCA36929.1	Yes	No
CPR5	CCA36730.1	No	-
ERO1	CCA36464.1	No	-
KAR2-1	CCA38843.1	No	-
KAR2-2	CCA36494.1	No	-
LHS1	CCA36228.1	No	-
MPD1	CCA36295.1	No	-
SCJ1-1	CCA36342.1	No	-
SCJ1-2	CCA39683.1	No	-
SCJ1-3	CCA38037.1	No	-
SIL1	CCA36700.1	No	-
YNL181W	CCA36278.1	No	-
1-1_0160	CCA36329.1	No	-
GTT1	CCA37819.1	Yes	Yes
4_0991	CCA40358.1	No	-
DJP1	CCA37789.1	No	-
SWA1-1	CCA40389.1	No	-
SWA1-2	CCA40730.1	No	-
SWA1-3	CCA38830.1	No	-
SWA1-4	CCA37780.1	No	-
SWA1-5	CCA39906.1	No	-
JEM1	CCA37709.1	No	-
HLJ1	CCA36545.1	No	-
SBA1	CCA36956.1	Yes	Yes
STI1	CCA38459.1	No	-
Initial steps of O- and N-glycosylation			
PMT1-1	CCA38772.1	No	-
PMT1-2	CCA36754.1	No	-
PMT1-3	CCA40355.1	No	-
PMT1-4	CCA38726.1	No	-
PMT1-5	CCA36946.1	No	-
SWP1	CCA36416.1	No	-
CNE1	CCA38660.1	No	-
MNS1	CCA38210.1	No	-
MNL1-1	CCA39253.1	No	-
MNL1-2	CCA39254.1	No	-
Glycosylation process in ER			
OST1	CCA39413.1	No	-
OST3	CCA40528.1	No	-
OST6	CCA38994.1	No	-
WBP1	CCA38558.1	No	-
STT3	CCA37419.1	No	-
SEC53	CCA37752.1	No	-
RFT1	CCA40701.1	No	-
KRE5	CCA39213.1	No	-
KRE5-2	CCA36284.1	Yes	Yes
Vesicular transport			
Anterograde transport from ER to Golgi			
PHO86	CCA40569.1	No	-
SEC22	CCA39894.1	Yes	Yes

BST1	CCA38267.1	No	-
KRE6	CCA40635.1	No	-
SEY1	CCA39889.1	Yes	No
YOP1	CCA38438.1	No	-
EMP24	CCA38334.1	No	-
EMP47	CCA40714.1	No	-
ERV41-1	CCA39272.1	Yes	Yes
ERV41-2	CCA38162.1	No	-
MRS6	CCA39578.1	No	-
GYP7	CCA36422.1	No	-
TRS33-1	CCA39036.1	Yes	No
TRS33-2	CCA36257.1	Yes	No
YIP3	CCA38494.1	Yes	Yes
MDR1	CCA39353.1	No	-
TRX2-1	CCA40871.1	No	-
TRX2-2	CCA40413.1	No	-
TRX2-3	CCA39488.1	No	-
SEC16	CCA41070.1	No	-
HRR25-1	CCA38763.1	No	-
HRR25-2	CCA40571.1	No	-
HRR25-3	CCA38624.1	No	-
SEC17	CCA38323.1	Yes	Yes
COG4	CCA36712.1	No	-
COG6	CCA41094.1	Yes	No
COG8	CCA37935.1	No	-
SEC12	CCA40532.1	No	-
SAR1	CCA36641.1	Yes	Yes
SEC23-1	CCA36506.1	Yes	Yes
SEC23-2	CCA40539.1	No	-
SEC24	CCA39547.1	No	-
SEC13	CCA36223.1	Yes	Yes
SEC31	CCA37142.1	No	-
SFB2	CCA40658.1	No	-
ERV29	CCA38697.1	Yes	No
ERV25	CCA39000.1	No	-
ERP1	CCA38410.1	Yes	No
ERP2	CCA39474.1	Yes	No
BET3	CCA36244.1	Yes	Yes
BET5	CCA37932.1	Yes	Yes
TRS23	CCA37886.1	No	-
TRS20	CCA38501.1	No	-
TRS85	CCA38185.1	No	-
YPT1-1	CCA39941.1	No	-
YPT1-2	CCA40036.1	No	-
YIP1	CCA38583.1	No	-
YIF1	CCA36246.1	No	-
SLY1	CCA40908.1	Yes	Yes
VPS33	CCA37971.1	No	-
VPS45	CCA36612.1	No	-
SED5	CCA36939.1	No	-
BOS1	CCA37749.1	Yes	No
BET1	CCA36330.1	Yes	Yes
YKT6	CCA37598.1	No	-
SEC18	CCA39825.1	No	-
Golgi			
Golgi trafficking			

VTI1	CCA39049.1	No	-
YPT6	CCA41004.1	Yes	Yes
TVP18	CCA38743.1	No	-
COY1	CCA37224.1	No	-
COG3	CCA41107.1	No	-
COG5	CCA39545.1	Yes	No
COG7	CCA36573.1	No	-
TLG1	CCA40135.1	Yes	No
TLG2	CCA39225.1	No	-
Glycosylation process			
GDA1	CCA41147.1	No	-
KRE2-1	CCA39958.1	Yes	No
KRE2-2	CCA39959.1	No	-
KRE2-3	CCA36307.1	No	-
KRE2-4	CCA37807.1	No	-
KRE2-5	CCA36963.1	No	-
KRE2-6	CCA37906.1	Yes	No
MNN9-1	CCA41064.1	No	-
MNN9-2	CCA39647.1	No	-
MNN9-3	CCA38187.1	No	-
MNN10	CCA37891.1	No	-
MNN11	CCA37828.1	No	-
VRG4	CCA39228.1	No	-
MNN4-1-1	CCA37327.1	No	-
MNN4-1-2	CCA37328.1	No	-
MNN4-1-3	CCA38247.1	No	-
MNN4-2	CCA38260.1	No	-
MNN2-3	CCA39797.1	No	-
MNN2-1	CCA39368.1	No	-
OCH1	CCA36419.1	No	-
HOC1	CCA39540.1	Yes	No
MNN5	CCA36950.1	No	-
CWH41	CCA36680.1	No	-
ROT2	CCA38181.1	No	-
GTB1	CCA39996.1	No	-
Vesicular transport			
COPI vesicle - Retrograde transport from Golgi to ER			
GOS1	CCA38826.1	No	-
VPS74	CCA40071.1	No	-
YOL107	CCA37827.1	No	-
SEC28	CCA39470.1	Yes	Yes
GLO3	CCA39604.1	No	-
SEC20	CCA37383.1	Yes	Yes
USE1	CCA37014.1	No	-
GET3	CCA40742.1	No	-
SEC39	CCA38538.1	No	-
Exocytosis			
APS2-1	CCA36308.1	Yes	Yes
APS2-2	CCA39093.1	Yes	No
APS2-3	CCA40057.1	Yes	Yes
SEC9	CCA37836.1	No	-
TPM2	CCA41053.1	Yes	Yes
SEC15	CCA38686.1	No	-

SEC5	CCA39291.1	No	-
SEC10	CCA37250.1	No	-
SFH5-1	CCA37781.1	No	-
SFH5-2	CCA39504.1	Yes	Yes
SFH5-3	CCA40575.1	No	-
BUD7-1	CCA38842.1	No	-
BUD7-2	CCA40599.1	No	-
SEC1	CCA41033.1	No	-
SSO2	CCA37207.1	No	-
RHO3	CCA38545.1	Yes	Yes
KIN2	CCA39819.1	No	-
SRO77	CCA36975.1	No	-
SEC3	CCA37472.1	No	-
SEC6	CCA37371.1	No	-
SEC8	CCA38943.1	Yes	No
EXO70	CCA40443.1	No	-
EXO84	CCA41090.1	No	-
YPT31	CCA37447.1	No	-
MYO2-1	CCA37604.1	No	-
MYO2-2	CCA40521.1	No	-
MYO2-3	CCA39331.1	Yes	Yes
SNC1	CCA37603.1	No	-

Table S3.2: Primers used in this study

Gene	Primer name	Primer sequence 5' → 3'
CPR1	CPR1_pt1 fwd	CTTCACAGAAATCGAACAGAGGAG
	CPR1_pt1_ole rev	ACGGCCCAATGGTTGGTCGTTGGAAGAAACGTCGAAGAATG
	CPR1_pt2_ole fwd	CGACGTTTCTTCCAACGACCAACCATTGGGCCGTATTGTC
	CPR1_pt2 rev	CTAGTGGCGGTTGATAACAATTGATG
SBH1	SBH1_pt1 fwd	GACGAAGAAATGGCATTAACTCTGTTC
	SBH1_pt1_ole rev	CTTTAACAACATCCATAATGTCTACAGCAATTCCAGGAGGACAG
	SBH1_pt2_ole fwd	GAATTGCTGTAGACATTATGGATGTTGTTAAAGCTTCGGTC
	SBH1_pt2 rev	CTTCGTCCGTTGATGTGAAGTTTG
GTT1	GTT1_pt1 fwd	CTCACTCTTGGGATGATGGCTAG
	GTT1_pt1_ole rev	CAAAGATTATAGTACACTGGCTCAATCAATCTAGAGCAAATTACATTATCTG
	GTT1_pt2_ole fwd	CTCTAGATTGATTGAGCCAGTGTACTATAATCTTTGGCGACGATG
	GTT1_pt2 rev	GTTGTTTCACTTACTAGCGGCATTG
SBA1	SBA1_pt1 fwd	GAAATGTAATGACCGTGTTGTCCAC
	SBA1_pt1_ole rev	GTCAATTCCTCCAGAGGTTCTGTGGGCACAACGCTCTAAC
	SBA1_pt2_ole fwd	GTTGTGCCCACAGAACCTCTGGAGGAATTGACATTGTATATTCGAAATATTG
	SBA1_pt2 rev	GCTACAACATCAATATAGCAAAGCGTG
SEC 22	SEC22_pt1 fwd	CATCGCTTTGTGACAATGGAGTTAAG
	SEC22_pt1_ole rev	GTGGCTGACAAAGGAAGGCCATCGTTTCTAAAGATCAAGGTGGAC
	SEC22_pt2_ole fwd	GATCTTTAGAAACGATGGCCTTCCTTTGTCAGCCACGGTG

	SEC22_pt2_ole rev	GAGGTAAGTTCTCGTGTTTTATCAGAGCCTCCAAATTGATTC
	SEC22_pt3_ole fwd	GAGGCTCTGATAAAACACGAGAACTTACCTCAAAATTTATTCTTATCCTCAATCAC
	SEC22_pt3 rev	CTATTCAAACAGTATGGTCCTTCAGACC
SEC 17	SEC17_pt1 fwd	CATTGAACAATGTATTGACAATCGTTGACGAG
	SEC17_pt1_ole rev	GAGCAACTTATAGCAAAGGCTGACAAGAAATGCGCCCCGGTTAG
	SEC17_pt2_ole fwd	GCATTTCTTGTCAGCCTTTGCTATAAGTTGCTCTGCCTCAG
	SEC17_pt2 rev	GTTCTCCGCTAAGTCATACCCTTCG
SAR1	SAR1_pt1 fwd	GAAGGCAGTAGCGTTGCTAGAAAAG
	SAR1_pt1_ole rev	GTGGGTACTAAACTGGTTCCAGGATGTGTTGGCATCTTTAG
	SAR1_pt2_ole fwd	CCAACACATCCTGGAACCAGTTTAGTACCCACATTATTGATGGTTTG
	SAR1_pt2 rev	GATCCCAATTCCATAACTTTATTTTACTGTCATCTG
BET3	BET3_pt1 fwd	GTTGCTGAGTCTGGCTCTAG
	BET3_pt1_ole rev	ATGTCAAAGTGAACAGTTCAGAGTTAATCTTTTCTGTCCTATTCTTATAAACATCCTC
	BET3_pt2_ole fwd	AATAGGACAGAAAAGATTAACTCTGAACTGTTCACTTTGACATATG
	BET3_pt2 rev	AAGAACTCCGTGCAAACGTAC
SLY1	SLY1_pt1 fwd	CTTCCTCCCCTCTGTGTTCAG
	SLY1_pt1_ole rev	GTGATAGGCAGATAGCCACTCTGGAGAAAATGCTTCATTTG
	SLY1_pt2_ole fwd	CATTTTCTCCAGAGTGGCTATCTGCCTATCACGAAGACTGG
	SLY1_pt2 rev	CACAGTTTCGACATGTGGGAC
BET1	BET1_pt1 fwd	GTATCCCAATGGCAGGCGTAC
	BET1_pt1_ole rev	CTTCGATCATGTCAAGTCGCTATTCGTCAAACTTACATCAAAGAG
	BET1_pt2_ole fwd	GTTTGACGAATAGCGACTTGACATGATCGAAGACGTGTCCTG
	BET1_pt2 rev	TCTCAGCTGGGAAACGACCATTAC
APS2	APS2_1_pt1 fwd	TGCAAGGGACCGCCAAATACTC
	APS2_1_pt1_ole rev	CTATAAACAATTTTGTGATCTTTATATTCCAAGACATTGCACATCTTTGCCTTTC
	APS2_1_pt2_ole fwd	GTCTTGGAATATAAAGATCACAAAATTGTTTATAGAAGGTATGCTTCTTTATTTTTC
	APS2_1_pt2 rev	AGTATGTTAGTGAATGACGGTAAGTGGTG
SEC20	SEC20_pt1 fwd	TCTTCTTCGTGAGTCACTTCGTC
	SEC20_pt1 rev	CTTGTTGAAAATAATCCTCATCCATGATATGCCAGTTTTGC
	SEC20_pt2 fwd	CATATCATGGATGAGGATTATTTTCAACAAGACTATCAGCCTATAG
	SEC20_pt2 rev	CTATCTGACAACGTTTGTAGATCCCTG
TPM2	TPM2_pt1 fwd	GTACATACGAGCAGTAGCCAATG
	TPM2_pt1 rev	GATAAACTTAAGGAAAAAATCAACAAGCTCAAATTGGACG
	TPM2_pt2 fwd	GAGCTTGTTGATTTTTTCCTTAAGTTTATCCATCTTTTACTGTTG
	TPM2_pt2 rev	CGAACGACATGGAAGTTTGATC
RHO3	RHO3_pt1 fwd	GTTTGGAGCCATTATGAAGTGAGATG
	RHO3_pt1 rev	CTATTTTCCTTTGAATAGGTTTATGAGATCCACTACAAAACAC
	RHO3_pt2 fwd	GTGGATCTCATAAACCTATTCAAAGGAAAATAGTCATTTTGGGTG
	RHO3_pt2 rev	TTACATTATTGTACAGCTGGTTTTATTGTC

APS2_3	APS2_3_pt1 fwd	CTTAAATGCAAAGAGACCCTTTGATGC
	APS2_3_pt1 rev	CTGTGCTGATATTCAACAATGATGGGCTTCCTCG
	APS2_3_pt2 fwd	CCATCATTGTTGAATATCAGCACAGAATGGATCATGATTG
	APS2_3_pt2 rev	ATAGAGATCCTCTCAATAGAGTGCAAC
KRE5-2	KRE5-2 fwd	CCAAGCCATGTCCTGAATGTC
	KRE5-2_pt1ole rev	GGAAGCTTTCAGTGGTTTGCTTATGCAGATCATCAAACTTCC
	KRE5-2_pt2ole fwd	GATCTGCATAAGCAAACCACTGAAAGCTTCCCAACC
	KRE5-2 rev	GAGTATGAATGGTACTGGCGTGTG
ERV41	ERV41 fwd	GTTCTGGCTGTGGATGATCCTC
	ERV41_pt1ole rev	GACAAAGGTAAACGCATCCATCGCCTACGACGAAATTG
	ERV41_pt2ole fwd	GTCGTAGGCGATGGATGCGTTTACCTTTGTCAATATATCTGG
	ERV41 rev	GATCTATCTCTGTAGCCATAGCCTTTCC
YIP3	YIP3 fwd	CAATCTGCTGCTGGATATTTGGAG
	YIP3_pt1ole rev	TGTTCTCTGAGAGTTGAGAAAATTGAAAATTTGCAAATTGGGACATTGAC
	YIP3_pt2ole fwd	CAATTTGCAAATTTTCAATTTTCTCAACTCTCAGAGAACATGTCAG
	YIP3 rev	CTAGACTGCTTCTTCAAACACGGTTTC
SEC23	SEC23 fwd	CTCCCACAGACTCACTAGCATC
	SEC23_pt1ole rev	GATCAGAATCTACGATGGACCAAGACGCGTTTGAGACC
	SEC23_pt2ole fwd	CAAACGCGTCTTGGTCCATCGTAGATTCTGATCCACTTGGGAG
	SEC23 rev	CTTTCCTGGGAATCCAATTACCGAG
SEC13	SEC13 fwd	GTTCTACACAGTACCATGGCTCTG
	SEC13_pt1ole rev	CGTTTCCAATTGTAACCAGCAAGGGACTTGTACGGAAAACTATAC
	SEC13_pt2ole fwd	GTACAAGTCCCTTGCTGGTTACAATTGGAAACGCACATGATGAC
	SEC13 rev	CATGCCACATCTCTGACCCAG
BET5	BET5 fwd	CTTTCGTCAAGGGCGGTCTAC
	BET5_pt1ole rev	CCTATTGTAAATACAATTACAGTGTCTATCAAATATCCAAAACGAGTAAATC
	BET5_pt2ole fwd	GGATATTTGATAGACACTGTAATTGTATTTACAATAGGGATTTTGCTCAAC
	BET5 rev	GTTTGAGGTCAGCTGATGTCATC
SEC28	SEC28 fwd	CTTGTAAATCCTACCACCGGTCAG
	SEC28_pt1ole rev	CTAATTTGATAGAGCTCACCGCTGTCAGAGAATGAATCC
	SEC28_pt2ole fwd	CTCTGACAGCGGTGAGCTCTATCAAATTAGACAGCAGTTTTTC
	SEC28 rev	CAAGTGTGAGGTTGACCCAAC
YPT6	YPT6 fwd	TGAGTGATGCACGTCCTACTAC
	YPT6_pt1ole rev	GGAAGTTTTTCCCACACCTTGATCGCCTAAGAACAC
	YPT6_pt2ole fwd	CTTAGGCGATCAAGGTGTGGGAAAAACTTCCCTAATTACAAGATTC

	YPT6 rev	CACATTTGCCTCTGCTAGATCGATC
MYO2	MYO2 fwd	GCCTGTTCCTCACTAAGCTG
	MYO2_pt1ole rev	CAAAGCTTAACGAGTCTAATTTAAGTTTTGAAAAAGCCTCTGCTGAG
	MYO2_pt2ole fwd	GCTTTTTCAAAACTTAAATTAGACTCGTTAAGCTTTGACTCCAAC
	MYO2 rev	GACGAACTCCCTAGAATCCGAAC

Part IV – Using *K. phaffii* as platform strain for screening for novel cellulases

Abstract

Enzymes became an important factor in the industrial world as well as for pharmaceutical applications. It was estimated that cellulases make over 20% of the global industrial enzyme market. Cellulases are used in different industries where they have to show high activity and stability under harsh process conditions. Moreover, the harsh process conditions differ from industry to industry thus requiring different cellulases. There is high potential in improving the process efficiency by finding novel or engineered cellulases which show better activity and stability under respective specifically applied process conditions. Therefore, many diverse approaches were followed to isolate novel cellulases. In many cases, the used screening methods in these approaches were limited by laborious cultivation steps in liquid medium for the expression of the different enzyme variants. Here, a screening method is presented which used *K. phaffii* as screening strain for screening a cDNA library for cellulases. This screening method is very simple. It screens for *K. phaffii* transformants relying on cellulase expression which enables growth on cellobiose and carboxymethyl cellulose (CMC) as sole carbon sources. It was possible to establish a screening allowing to identify transformants expressing active β-glucosidases and thereby enabling growth on cellobiose. Furthermore cDNA libraries based on a *K. phaffii* platform strain already expressing a β-glucosidase, were screened for active expression of putative endo-glucanases on CMC containing media.

Introduction

Enzymes in Industry

Microorganisms have been used by mankind for many years. Archaeological evidence indicates that microorganisms were already used for production of fermented drinks since 7000 BC in China and 6000 BC by the Sumerians [Legras et al. 2007; Singh et al. 2016]. Although microorganisms where used for that long time by mankind, it was only in the 19th century that enzymes were discovered first. A milestone was the demonstration of the conversion from glucose to ethanol by cell free yeast extract performed by Eduard Buchner in 1897. Just in 1926, however, it was possible to isolate a first enzyme in its pure form. In 1960, a bacterial protease alcalase was produced by NOVO Industry on commercial scale using *Bacillus licheniformis* as producer [Gurung et al. 2013; Rao et al. 1998].

Since that time, research has made huge progress and the use of enzymes in cell-free systems has become standardized as e.g. for DNA cloning, PCR, or in vitro protein expression [Endo and Sawasaki 2006; Gibson et al. 2009]. Reactions catalysed by enzymes are from high interest in research and industry because enzymes are non-toxic and an environmental friendly alternative to chemical catalysts, needing only a low energy input [Singh et al. 2016]. Nowadays, enzymes of all 6 enzyme classes are used for industrial purposes but hydrolases are by far the mostly used enzymes in the industries. There is a diverse range of applications for enzymes in the industries and the market was growing significantly in the past decade. It was estimated that the global market of industrial enzymes was 3.3 billion USD in 2010. In 2014, it was already 4.2 billion USD and a further annual growth was expected of 7% from 2015 to 2020 reaching nearly 6.2 billion USD was expected [Gurung et al. 2013; Singh et al. 2016]. Examples for industrial enzymes of all enzyme classes are listed in Table 4.1.

Table 4.1: Examples for industrial enzymes and the different industries which are using them. EC number stands for Enzyme Commission number [Gurung et al. 2013; Singh et al. 2016].

EC number	Class of enzymes	Examples for Industrial enzymes	Examples for Industries using these enzymes
EC1	Oxidoreductases	Catalases Glucose oxidases Laccases	Dairy Baking, Beverage, Polymer Pulp & Paper, Cosmetics, Textile, Waste*
EC2	Transferases	Acyltransferases	Organic synthesis
EC3	Hydrolases	Amylases Cellulases Lipases Mannanases Pectinases Phytases Proteases	Baking, Beverage, Paper and Pulp, Detergent Beverage, Pulp & Paper, Detergent Dairy, Baking, Pulp & Paper, Polymer, Cosmetics Detergent Beverage Animal feed Pulp and Paper, Detergent, Cosmetics, Waste*

		Pullulanases Xylanases	Beverage Pulp & Paper, Baking, Animal feed
EC4	Lyases	Pectate lyases Alpha-acetolactate decarboxylases	Textile Beverage
EC5	Isomerases	Glucose isomerases	Organic synthesis
EC6	Ligases	DNA ligases	Molecular biology

*Waste stands for Waste management

Usage of cellulases in industry

Among the industrially important sector of hydrolases, proteases are dominating (e.g. in detergent and dairy industries). Amylases and cellulases are the second biggest group and those are used e.g. in textile, detergent and bakery industries [Gurung et al. 2013]. Estimates showed that cellulases cover about 20% of the total enzyme market [Singh et al. 2016; Srivastava et al. 2015]. A list of the industries which use cellulases is shown in Table 4.2. As mentioned in Kickenweiz et al. 2018 (see Part II of this thesis), cellulases are mainly represented by the key enzymes endo-glucanases, exo-glucanases and β-glucosidases which cleave β 1,4-glycosidic linkages of cellulose and cellobiose molecules [Kostylev and Wilson 2012; Teeri 1997].

Table 4.2: List of industries using cellulases and their function [Chen et al. 2007; Singh et al. 2016; Srivastava et al. 2015].

Industry	**Function**
Beverage	Fruit liquefaction
Pulp & Paper	Deinking, drainage improvement
Detergent	Colour clarification
Textile	Denim washing
Waste treatment	Degradation of cellulose
Energy	Saccharification of cellulose for biofuel production

Although cellulases are used in many different industries, their most prominent application in industrial biotechnology is the saccharification of (ligno-)cellulose for biofuel production. Since fossil energy sources are strongly limited, it is considered that these problems can be partially solved by changing to biofuels [Lynd et al. 2017]. There was an increase of total bioethanol production of 17% to 105 billion litres from 2009 to 2010. However, economical barriers in biofuel production became more challenging in recent years. One reason contributing to that was the introduction of new oil extraction technologies which led to an oversupply of oil and a decrease of the oil price. As a consequence, the costs for cellulosic biofuels production have to be reduced further before it can become an economical

alternative to fossil oil [Lynd et al. 2017; Srivastava et al. 2015]. Therefore, there is a vital interest of industry to reduce the costs of saccharification process of (ligno-) cellulosic biomass. One big potential for cost reduction is in cellulase cost reduction and increasing the efficiency of (ligno-)cellulose degradation [Klein-Marcuschamer and Blanch 2015; Kuhad et al. 2016, Saini et al. 2016]. Although there are many approaches to improve the (ligno-) cellulose pre-treatment [Mood et al. 2013; Poornejad et al. 2013; Ren et al. 2016], a main focus is still on improving the cellulase efficiency for these processes [Kuhad et al. 2016].

Approaches for improving cellulase efficiency

The specific mix of cellulases has a huge influence on the (ligno-)cellulosic biomass degradation efficiency. The filamentous fungus *Trichoderma reesei* is one of the main producers of commercialized cellulases. It secretes a high amount of endo- and exo-glucanases but the concentration of produced β-glucosidase is low. As consequence, β-glucosidase has to be added to the cellulases mix from another source to improve the enzymatic saccharification of (ligno-)cellulosic biomass. On the other hand *Aspergillus niger*, another filamentous fungus that is used for cellulase production, seems to have a deficiency in exo-glucanases compared to other cellulase producers [Srivastava et al. 2015; Zhao et al. 2016]. Therefore, there is the desire to find alternative cellulase producers whose cellulase mixes degrade cellulose more efficiently. For example, the filamentous fungi *Penicillium oxalicum* might be a potential alternative to *T. reesei*, since it produces an efficient cellulase mix with a higher β-glucosidase concentration than *T. reesei*. However, its secretion efficiency has still to be improved to become an economically efficient alternative option for industrial cellulase production [Zhao et al. 2016]. Another major microorganism for lignocellulase degrading enzyme cocktails is *Myceliophthora thermophile*, nicknamed C1 and recently acquired from Dyadic by Dupont [Dyadic].

The importance of improving the cellulase efficiency to lower the process costs in total is not only limited to biofuel production. As mentioned before, cellulases are used in different industries with differing specific process conditions. For example the energy industry (biofuel production) needs cellulases which show a high heat stability at low pH, while the detergent industry uses thermostable cellulases which are active in a broad pH range and halo-tolerant. Furthermore, waste industry prefers types of thermostable cellulases applied at high pH. This

shows the importance of finding novel cellulases to reduce the process costs by fulfilling these requirements better than commonly used cellulases [Kuhad et al. 2016].

Optimization of known cellulases

One approach to improve the properties of cellulases which are required for industrial purposes, is to perform mutagenesis experiments on already known cellulases. It has been reported that different endo-, exo-glucanases and β-glucosidases could be engineered by rational and random mutagenesis [Kuhad et al. 2016; Liu et al. 2009]. Since goal of this thesis was to screen for novel cellulases, mutagenesis experiments on known cellulases were not performed.

Isolation of novel cellulases

In an alternative and complementary approach, novel cellulases are isolated from natural sources. Since the mentioned process conditions in industry are extreme (e.g. temperature, pH and ionic concentration), many studies focus on screening of extremophile organisms for novel cellulases. Extremophile organisms live under harsh conditions and the chances are high that enzymes of interest have been adapted to such conditions during evolution. It has been reported that cellulases which were adapted to a harsh environment (e.g. hyper-thermostable and thermo-alkali stable cellulases) were successfully isolated by screening extremophiles [Ando et al. 2002; Jang and Chen 2002; Kuhad et al. 2016]. It should be kept in mind, however, that enzymes adapted to harsh conditions can also be found in ambient environment [Duan and Feng 2010]. For example, a halo-tolerant cellulase was isolated from soil samples containing much lower salt concentration [Voget et al. 2006]. Another example showing that the properties of enzymes have not to correlate with the extreme environment is that Pang and co-workers isolated a cellulase from high-temperature compost samples which showed activity at low temperature [Pang et al. 2009]. These examples indicate that it is also worthy to search for cellulases in mesophilic environments.

Isolation of cellulases from culturable organisms. The classic way to isolate novel cellulases is to screen strain collections and natural samples for microorganisms which are able to grow on cellulose as sole carbon source. It was possible to isolate cellulolytic bacteria and fungi by such functional screening [Houfani et al. 2017; Kim et al. 2014; Monga and Chadha 2014; Yan et al. 2001]. However, this approach to find novel cellulases requires

culturable organisms. Since it was estimated that only 1 % of all microorganisms are culturable in labs [Colin et al. 2015], this approach is strongly limited.

Isolation of cellulases from unculturable organisms. Since only such a little percentage of microorganisms are culturable, the vast majority of microorganisms cannot be tested for their ability to grow on cellulose. Therefore, alternative molecular approaches were required for being able to screen unculturable microorganisms for cellulases [Colin et al. 2015; Duan and Feng 2010; Wang et al. 2008]. In 1995, a first cellulase was isolated from an unculturable extremophile bacteria by screening metagenomic libraries [Duan and Feng 2010]. The "metagenomic" DNA was directly isolated from (environmental) samples and cloned into suitable expression systems. Plasmids, fosmids, cosmids, and λ phages were used as library types for metagenomic libraries in bacterial hosts (mainly *E. coli*). Expression hosts are transformed with the built metagenomic libraries and the clones are tested for cellulolytic activity. Screening of metagenomic libraries for cellulases became very popular in the last 15 years and many cellulases from unculturable microorganisms were isolated and characterized. However, metagenomic libraries seems not to be suitable for discovering eukaryotic genes since eukaryotic promoters which are also contained in the metagenomic DNA samples, cannot be used in the common prokaryotic expression hosts for expressing the eukaryotic gene [Duan and Feng 2010; Guazzaroni et al. 2015; Uchiyama and Miyazaki 2009; Zhou et al. 2016]. No report about using an eukaryotic expression host for a metagenomic approach to discover cellulases was found. The reason for that might be that there is no suitable eukaryotic expression host and expression system for such an approach.

Another approach to screen of unculturable microorganisms for cellulases is the screening of metatranscriptomic libraries. In this approach, mRNA is directly isolated from the (environmental) samples and transcribed into cDNA. The cDNA is then used to create a cDNA library. This process is very tricky because mRNA is not as stable as DNA. However, it has been reported that cellulases were identified by screening of metatranscriptomic libraries. With this approach it became also possible to screen for novel eukaryotic cellulases [Duan and Feng 2010; Grant et al. 2006; Findley et al. 2011]. For example, Findley and co-worker reported that they were able to discover protozoan cellulases from cow rumen. In this study, *E. coli* clones expressing the protozoan cDNA library were screened for cellulolytic activity by functional screening with plates containing dye-linked insoluble polysaccharide substrates [Duan and Feng 2010; Findley et al. 2011].

Screening methods for cellulases

So far, the importance of cellulases in industry and the reasons why it is important to search for novel cellulases have been discussed. However, the choice of the screening method is highly relevant for the isolation of cellulases and influences the respective findings. As in each screening, the efficiency and outcome highly depends on the chosen screening method. The aim is to combine high throughput and high sensitivity in one method [Vervoort et al. 2017; Zhang et al. 2006].

In general, it has to be differentiated between facilitated screening and random screening. In facilitated screening, a qualitative screening is performed first to reduce the number of clones by selecting promising ones. In random screening, the clones do not undergo a qualitative pre-screening and the tested clones are chosen randomly. In cluster screenings, several clones are pre-screened as groups to increase the throughput initially before re-screenings are done to identify the respective specific hits.

There are three main screening methods, i) screening based on growth on cellulose or cellobiose, respectively ii) colorimetric assays, and iii) fluorometric assays [Zhang et al. 2006]. Such assays usually provide the opportunity for higher throughput than classical cellulose degradation assays combined with enzymatic or chemical sugar analysis.

Screening based on growth on cellulose/cellobiose. This approach is mainly used for the enrichment and identification of culturable cellulolytic microorganisms. Environmental samples are first incubated in cellulose containing liquid medium to enrich cellulolytic microorganisms. Otherwise, diluted aliquots of the environmental samples can be directly spread onto agar plates containing cellulose as sole carbon source. This approach has been used to screen cellulolytic bacteria and fungi [Soares Jr et al. 2012; Zhang et al. 2014]. During their work on generation of thermostable β-glucosidases, Liu and co-workers used a screening based on growth on cellobiose for the qualitative pre-screening of a library of mutated β-glucosidase in *E. coli* to identify clones expressing functional β-glucosidase variants [Liu et al. 2009].

Colorimetric screening methods. The most commonly used qualitative colorimetric assay to isolate cellulolytic microorganisms is Congo red dyeing. The testing samples are spread onto agar plates containing crystalline cellulose or CMC. The plates are dyed with Congo red after a few days of incubation. Halos appear around cellulolytic organisms indicating cellulose degradation by cellulases [Yoon et al. 2007; Zhang et al. 2006]. Congo red shows strong

interaction with cellulose. Therefore, only areas of the agar plate are dyed where CMC is not degraded. Degraded CMC is either metabolized or diffused. As consequence, "non-dyed" halos appear around cellulolytic organisms [Theater and Wood 1982]. Although CMC is mainly degraded by endo-glucanases only, Congo red dyeing using CMC as substrate is still broadly used to get first evidence for cellulolytic activity in testing microorganisms [Yoon et al. 2007; Zhang et al. 2006]. Beside Congo-red, it is also possible to use Phenol red and Remazol brilliant blue as dyes in qualitative plate screening. However, it has been reported that Congo red is more sensitive than the other dyes [Yoon et al. 2007].

Another strategy for qualitative agar-plate assays is the usage of dyed cellulose/CMC. Unlike the first method in which incubated agar plates are stained with dyes, dyed cellulose/CMC is already added to the medium during medium preparation. The cellulose/CMC molecules are covalently linked with the dye [Biely et al. 1985; Zhang et al. 2006]. Insoluble dyed cellulose can be synthesized by using the dyes Remazol Brilliant Blue, Reactive orange and Reactive blue 19. An example for soluble dyed cellulose is CMC dyed with Remazol Brilliant Blue [Zhang et al. 2006]. In general, dyed cellulose/CMC is added to the agar medium. Cellulolytic activity of growing microorganisms leads to a release of the chromophores or of smaller diffusible dyed saccharide molecules. These small molecules diffuse which becomes visible as formed decoloured halos around the grown colonies after few days of incubation [Ten et al. 2004]. Usage of dyed cellulose/CMC is not only limited to plate assays, it is also used for quantitative cellulase activity assays. The grade of hydrolysed Remazol Brilliant Blue dyed cellulose can be measured by measuring the optical density at 595 nm of the supernatant in a liquid sample [Holtzapple et al. 1984].

Another type of substrates for colorimetric cellulase activity assays are trinitrophenyl-carboxymethyl cellulose (TNP-CMC). Due to the high sensitivity TNP-CMC is widely used for quantitative cellulase activity assays on the basis of enzymatic release of TNP. Interestingly, since no reports were found about using TNP-CMC in agar plates, it seems that this chromophore is not used in plate assays [Carder 1986; Robson and Chambliss 1986; Zhang et al. 2006]. It is also possible to measure cellulase activity by measurement of reducing sugar ends. Reducing sugar ends are generated during cellulose degradation with each cleavage by a cellulase. Therefore, the increase of reducing sugar ends in the sample correlates with cellulose degradation and gives information about the cellulase activity. The most common colorimetric reducing sugar end assays are the DNS (dinitrosalicyclic acid) and pHBAH (4-hydroxybenzoylhydrazine) assays. As TNP-CMC, DNS and pHBAH assays are usually used for quantitative measurement of cellulase activity in liquid samples [Carder

1986; Mellitzer et al. 2012a; Zhang et al. 2006]. Since the tested clones have to be cultivated separately in liquid cultures, the throughput of these assays is limited by the number of clones which can be cultivated in the lab. Furthermore, it should be mentioned that TNP, DNS and pHBAH are phenolic compounds. As consequence, hazardous waste is generated in these assays.

Fluorometric screening methods. Very common used fluorogenic substrates for cellulases in screenings are analogues of cellulose which are substituted with methylumbelliferyl groups. 4-methylumbelliferone is a coumarin derivative. The degradation of fluorogenic substrates (by cellulolytic enzymes) can be measured by measuring the fluorescence signal of liberated methylumbelliferyl groups. It is possible to use fluorogenic substrates in both, agar plate assays and liquid samples [Coniglio et al. 2017; Kim et al. 2007; Ko et al. 2013a; Li et al. 2011]. However, the agar plates might quench the fluorescence signal, making detection difficult when the colonies were not plotted on membranes or filters prior to the assay. It has been reported that 14 fungal strains were tested for β-glucosidase activity on agar plates containing 4-methylumbelliferyl glucoside. Furthermore, they were also tested for exo-glucanase activity on agar plates containing 4-methyl umbelliferyl cellobioside. Positive strains showed fluorescence under UV light after a few days of incubation. However, it seemed that the clones had to be incubated separately on the agar plates containing 4-methyl umbelliferyl cellobioside for screening in this study which required one plate per tested clone [Coniglio et al. 2017]. Kim and co-workers screened metagenomic libraries for β-glucosidases by using 4-methylumbelliferyl β-D-cellobioside (MUC; a fluorogenic analog of cellulose) as substrate in liquid cultures incubated in 96 well plates. Positive clones were detected under UV light [Kim et al. 2007]. Ko and co-workers established a robotic high throughput screening system for exo-glucanases using the same strategy as reported by Kim et al. 2007 [Ko et al. 2013a, Ko et al. 2013b]. Recently, it has been suggested to use halogenated fluorogenic substrates as 6-chloro-4-methylumbelliferone for screening of metagenomic libraries for glycosidases because the halogenated fluorogenic substrates might be used in a wider pH spectrum in screenings compared to their non-halogenated equivalents [Chen et al. 2016]. Beside 4-methylumbelliferone, also other coumarins are used for screening cellulases [Ostafe et al. 2013].

Another fluorophore which is used for measurement of cellulolytic activity is aminofluorescein (fluram). As with coumarins, cellulose can also be labelled with

aminofluorescein. An assay for the quantitative measurement of cellulolytic activity was described by Helbert and co-workers [Helbert et al. 2003].

As already mentioned in Part III of this thesis, there is a trend to increase throughput while reducing the volume of the samples in screenings. This trend can be also observed in screening of DNA libraries to isolate cellulases. Ostafe and co-worker described an ultrahigh-thoughput screening strategy for cellulases based on flow cytometry. *S. cerevisiae* cells expressing an endo-glucanase from *T. reesei* were sorted from non-expressing *S. cerevisiae* cells by using a FACS. The cells were compartmentalized using double emulsion technology. Cellulase activity of positive cells was detected by measuring fluorescence of oxidized 3-carboxy-7-(40-aminophenoxy) coumarin [Ostafe et al. 2013]. Körfer and co-worker reported a flow cytometry based ultrahigh throughput screening for directed cellulase evolution. Interestingly, in that case the cellulase variants were expressed in vitro (cell free enzyme production) using in vitro compartmentalization. The randomly mutated cellulase library was screened by measuring fluorescence of liberated fluorescein from fluorescein-di-β-D-cellobioside using FACS [Körfer et al. 2016].
With both screening methods described in Körfer et al. 2016 and Ostafe et al. 2013, it was possible to screen a huge number of clones. However, equipment in both approaches which is not available in each lab (e.g. FACS) was required. On the other side, the colorimetric and fluorometric assays mentioned above which require the separate cultivation of test clones, are not very suitable for ultra high throughput screening. The number of separately cultivated clones which can be tested, depends on the availability of space, shakers, flasks, Deep well plates etc. in a lab. Therefore, there was a high motivation to establish novel screening methods which are simple and suitable for high throughput. Here, a method for screening DNA-libraries for cellulases is presented which is based on the findings in Part II of this thesis. The yeast *Komagataella phaffii*, which is naturally not able to metabolize either cellulose or cellobiose efficiently, was used as screening host to identify cellulases by detecting growth on cellulose and cellobiose agarose plates.

Material and Methods

Strains

Two screening strains were used for the selection method described here. The first one was BSY11G1 (a Δ*gut1*, Δ*aox1* derivative of BSYBG10) (Bisy, Hofstaetten/Raab, Austria). The second one was a *K. phaffii* BSY11G1 strain expressing *TrBGL1* under NB2 control which was described in Part II of the thesis. Making of YPD medium and cultivation of *K. phaffii* were performed as described in Material and Methods of Part II. BSM medium was made as previously described in Mellitzer at al. 2012b with the adaption that agarose 1% (w/w) was added to make BSM_agarose plates. In Mellitzer et al. 2012b, BSM medium contained EDTA. Here, BSM_agarose plates were made with and without EDTA for screening. Cellobiose 0.5% (w/v), CMC 1% (w/v) and Avicel 0.5% (w/v) were used as sole carbon sources in the screening. The zeocin concentrations were 50 mg/L zeocin in YPD medium and 100 mg/L in BSM_agarose plates.

E. coli Top10F' (Thermo Fischer Scientific, Waltham, MA, USA) strains were transformed for vector amplification and cloning experiments. *E. coli* strains were incubated in LB medium. When selective markers were used, the antibiotic concentration was 25 mg/L zeocin. The *E. coli* cultures were incubated at 37°C and liquid cultures were shaken at 100 rpm.

Construction of *Pteridium aquilinum* cDNA library and vectors

All components for the cloning work were from Thermo Fischer Scientific (Waltham, MA, USA) unless otherwise specified. The restriction enzymes were either from Thermo Fischer Scientific (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). PCR products were purified using Wizard SV Gel and PCR Clean-Up System from Promega Fitchburg, WI, USA). Primers were ordered from IDT (Coralville, IA, USA). List of used primers is shown in Table 4.1. cDNA library was kindly provided by Lukas Sturmberger and contained cDNA isolated from *Pteridium aquilinum*. cDNA library was created by using template-switching method as described by Fischer, 2016.

The vector pPpT4_S_SpARS_pCAT1-500_sTom_Zeo (see plasmid map in Fig. 4.1) which was for the expression of sTomato under pCAT promoter, was used as a negative control in the screening method. The sTomato gene is coding for a red fluorescent protein which was previously described by Shaner et al. 2004. The cDNA was cloned into the same vector backbone. The vector pPpT4_S_SpARS_pCAT1-500_sTom_Zeo and a prepared sample of

the linearized vector backbone without sTomato gene for Gibson cloning (for cloning another gene) were kindly provided by Lukas Sturmberger.

The prepared vector backbone was open at the 3'end of the *CAT1*-500 promoter which made it possible to clone genes into this vector by using Gibson cloning. This was used to create positive controls by cloning the genes coding for the cellulases TrBGL1, TrCBH2-HM and TrEG1 into this vector backbone. These genes were amplified by PCR using the plasmids mentioned in Supplemental data of Part II of the thesis [Kickenweiz et al. 2018] and B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt as template. *TrCBH2*-HM was amplified using the primers TrCBH2ole fwd and TrCBH2ole rev and B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt as template. *TrBGL1* was amplified using the primers alphole fwd and TrBGL1ole rev and pPpGUT1_Amp_ADHtt_NB2_alpha.sig.seq_TrBGL1_AOX1tt as template. *TrEG1* was amplified using the primers alphole fwd and TrEG1ole rev. The PCR products were cloned into the vector backbone by using Gibson cloning to create the vectors pPpT4_S_SpARS_pCAT1-500_TrCBH2-HM_Zeo,
pPpT4_S_SpARS_pCAT1-500_alpha.sig.seq_TrBGL1_Zeo and
pPpT4_S_SpARS_pCAT1-500_alpha.sig.seq_TrEG1_Zeo.

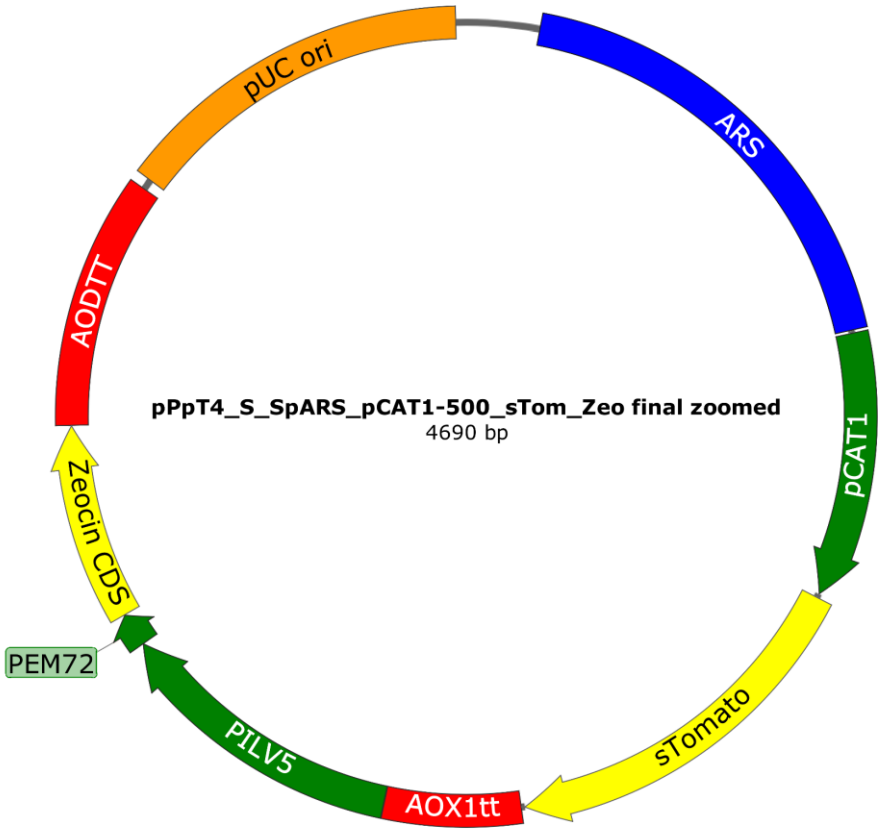


Fig. 4.1: Plasmid map of pPpT4_S_SpARS_pCAT1-500_sTom_Zeo. The expression of the gene coding for sTomato is under control of P_{CAT1}. The other used plasmids had the same vector backbone but the genes *TrBGL1*, *TrCBH2*-HM, *TrEG1* and cDNA were downstream of P_{CAT1} instead of *sTomato*. ARS sequence is shown in blue, CDS in yellow, terminator sequences in red and promoters in green. Plasmid maps were generated by using SnapGene Viewer.

Table 4.1: List of used primers

Primer name	**Sequence 5' → 3'**
TrCBH2ole fwd	CTTACAATTAAAACTAGACTAGTTCGAAACGATGATTGTCGGTATCTTGACTACCTTG
TrCBH2ole rev	CATTCTGACATCCTCTTGAGCGGCCGCGCTTACAAGAAAGATGGATTGGCGTTAG
alphole fwd	CTTACAATTAAAACTAGACTAGTTCGAAACGATGAGATTCCCATCTATTTTCACCGCTG
TrEG1ole rev	CATTCTGACATCCTCTTGAGCGGCCGCTAAAGGCATTGCGAGTAGTAGTCG
TrBGL1ole rev	CATTCTGACATCCTCTTGAGCGGCCGCTACGCTACCGACAGAGTGCTC
transformation_ casette_fwd	ACAATGAGTGTCAGATAAGTCACTATGTCC
transformation_ casette_rev	TCAAGAAGATCCTTTGATCTTTTCTACGGG

Screening

As mentioned before, BSM_agarose medium with and without EDTA was used for making screening plates. The screening plates contained 0.5% (w/v) cellobiose (for β-glucosidase screening), 1% (w/v) CMC (for endo-glucanase screening) or 0.5% (w/v) Avicel (for exo-glucanase screening) as sole carbon source. There were two approaches for expressing the cDNA library and cellulases in *K. phaffii*, episomal expression and chromosomal integration of the vectors into *K. phaffii* genome.

For episomal expression, 20-30 ng DNA of control ARS plasmids containing the genes *TrBGL1*, *TrCBH2*, *TrEG1*, or *sTomato* were used for the yeast transformation. For transforming *K. phaffii* with the cDNA library, 300-400 ng of ARS plasmids containing cDNA were used for the transformation to get comparable number of growing transformants on the screening plates. The difference of growing transformants between cDNA library and control plasmids might depend on the quality of the cDNA library or the expression of proteins which are toxic for *K. phaffii*. Yeast transformation was performed as previously described in Lin-Cereghino et al. 2005. The gained yeast resuspension of competent cells in the last step of the competent cell preparation was diluted 1:10 to reduce background growth on the screening plates. 50 µL of this 1:10 dilution was used for transformation and mixed with the DNA containing the ARS plasmids. After electric shock, 900 µL of YPD/1M sorbitol was added to the transformation samples and incubated at 28°C for 1 h. 100 µL aliquots of the transformation samples were plated onto the screening BSM_agarose plates containing the specific carbon source. The plates were incubated for a few days at 28°C.

For chromosomal integration, the ARS plasmids had to be linearized without the ARS sequence. For this, the vectors containing *sTomato*, *TrBGL1*, *TrCBH2-HM*, *TrEG1* and also cDNA were amplified by PCR using the primers transformation_casette_fwd and transformation_casette_rev. These primers bound on the vectors outside of the ARS sequence amplifying them without the ARS sequence. After the PCR step, DpnI digestion was performed with the PCR samples to degrade methylated DNA templates (ARS plasmids, prepared from *E. coli*). The samples were purified by running an agarose gel and gel purification afterwards. Yeast transformation was performed as previously described in Lin-Cereghino et al. 2005. 50 µL of the undiluted yeast resuspension containing the competent cells were used for the transformation. 100 ng of the purified PCR products were added to the competent cells. After electric shock, 900 µL of YPD/1M sorbitol was added to the transformation samples and incubated at 28°C for 2 h. Then, the samples were centrifuged for 30 sec at maximum speed and washed with 0.5M sorbitol. After another centrifugation step at

maximum speed for 30 sec, the pellet was resuspended in 1 mL 0.5M sorbitol. 100 µL aliquots of the transformation samples were plated onto the screening BSM_agarose plates containing the specific carbon source. The plates were incubated for a few days at 28°C.

Results and Discussion

Usage of BSM-agarose plates reduced background growth on screening plates

In the screening described in Part III of this thesis, minimal medium agar plates (BM agar plates) were used. This minimal medium contained yeast nitrogen base, potassium buffer, agar and a carbon source. In preliminary test experiments it has been shown that *K. phaffii* strains which were streaked-out onto BM agar plates lacking a carbon source, were able to form small colonies after few days of incubation. The size of the colonies depended on the amount of inoculum which had been used on a plate. The more cells were inoculated on the same agar plate the smaller were the grown colonies (not shown). It was even possible to show that colonies were formed on agar plates containing only agar after 5 days or more of incubation. These colonies were very tiny but visible with the naked eye. All these observations were evidence for that BM medium was not optimal for a screening based on growth. In the screening described in Part III of this thesis, background growth did not interfere with the screening, since the used screening strains were already able to grow on the used carbon source. The biggest grown colonies were isolated for further tests. Another important difference between both screenings was that an antibiotic selective marker could not be used for the screening of Part III of the thesis. In the screening method described in Part III, the strains were transformed with mutated DNA sequences but not with an antibiotic selection marker. Therefore, the transformed strains could not be selected by a gained antibiotic resistance after transformation but by better growth on the specific carbon source.

Here, the screening for novel cellulases used zeocin as antibiotic selection marker meaning that all positive transformants can grow on zeocin. Here, it should be mentioned that zeocin is an antibiotic which is pH sensitive [Invivogen]. Maybe, minimal medium is not buffered as good as complex medium at a quite neutral pH value. Although zeocin in minimal medium agar plates was by far not as effective as in e.g. YPD plates, it still strongly slowed down the growth of cells which did not express the zeocin resistance gene.

In a first step, it was necessary to decrease the background growth of positive transformants which are not expressing cellulases. This was important for increasing the sensitivity of the

screening and to differentiate more clearly all colonies expressing cellulases from those colonies expressing no cellulases.

For reducing the background growth, basal salt medium (BSM_agarose plates) was used. BSM medium is a common used defined broth medium used in fermentation experiments with *K. phaffii* [Mellitzer et al. 2012b]. It contains even less compounds which might be used as carbon source by *K. phaffii*, than BM medium. Therefore, BSM medium was slightly adapted to create BSM_agarose plates. Agarose was used instead of agar because comparative growth experiments with *K. phaffii* on plates containing either only agar or agarose showed that the tiny grown colonies on the agar plates mentioned before were not observed on the agarose plates. In total, it was possible to decrease the background growth on plates by using BSM medium instead of BM medium and agarose instead of agar.

De-repressed *CAT1* promoter was chosen for expression of the cellulase genes and cDNA library

Since the screening for novel cellulases was based on growth, the choice of promoter was highly relevant. The constitutive histone promoter NB2 which has already been used in Part II of the thesis, seemed to be a good choice. β-glucosidases and endo-glucanases were successfully expressed in *K. phaffii* under control of histone promoters enabling growth on cellobiose and CMC. However, Part II of the thesis also showed that the exo-glucanase TrCBH2 was insufficiently expressed under NB2 promoter in *K. phaffii* for an observable growth on Avicel.

In this study *CAT1* promoter (P_{CAT1}) was chosen as promoter for the screening. It is a de-repressed promoter. That means that it is inactive at the presence of glucose but becomes strongly active when the glucose is depleted. Its activity can even be strongly enhanced by addition of methanol [Vogl et al. 2016]. It has been reported that different cellulases were successfully expressed in high yields using inducible promoters in *K. phaffii* [Chen et al. 2011; Mellitzer et al. 2012b; Quay et al. 2011]. It was also shown that higher yields of TrCBH2 were reached during heterologous expression in *K. phaffii* by using the inducible *AOX1* promoter than with the constitutive GAP promoter [Mellitzer et al. 2012b]. Since P_{CAT1} is not a constitutive promoter but a de-repressed one, it was an interesting option for expression of the cDNA library to screen for novel cellulases.

Episomal expression of the DNA library and cellulases was not suitable due of lacking the required stability

An important question for the screening was if the DNA library, should be chromosomal integrated or episomal expressed in *K. phaffii*. Episomal expression has the advantage that the transformation efficiency is much higher. Episomal expression in *K. phaffii* was not used for heterologous gene expression due to the instability of the plasmids. However, research on ARS plasmids for episomal expression in *K. phaffii* made certain progress in the last years. The plasmids became more stable which made them more considerable for usage in experiments with *K. phaffii* [Vogl et al. 2016].

For testing the suitability of ARS plasmids for the screening, cellulases for control experiments were cloned into ARS plasmids (see plasmid map Fig. 4.1). The expression of the cellulases on the ARS plasmids were under control of P_{CAT1}. *K. phaffii* BSY11G1 was transformed with an ARS plasmid for expression of a β-glucosidase (TrBGL1) and an ARS plasmid containing *sTomato* instead of *TrBGL1* as negative control mixed in the same transformation sample. The transformation samples were plated onto BSM_cellobiose agarose plates. In parallel, a *K. phaffii BGL1* strain which had been chromosomally integrated *TrBGL1*, was transformed with an ARS plasmid for expression of an endo-glucanase (TrEG1) and plated onto BSM_CMC agarose plates. The ARS plasmid containing the gene stomato was used here as negative control on the same plate too.

After few days of incubation, the grown colonies on the plates were still very small. There was no clear difference of colony size between transformants expressing cellulase and negative control expressing stomato. This indicated that the episomal expression was not very efficient in this screening systems.

Grown transformants were isolated and streaked out onto YPD plates containing zeocin. Interestingly, the inoculum contained only a little number of cells which were able to grow on YPD-Zeo plates. Only a few colonies were formed in the streak-out on the plate and the vast majority of the inoculum stayed as a mass of non-growing cells. This was strong evidence that many cells of the isolated colony had already lost the ARS plasmid. Since, this problem was not observed on YPD plates containing zeocin, it seemed that the selective pressure of zeocin was not strong enough in BSM agarose plates to keep the ARS plasmid in *K. phaffii* cells. Furthermore, the longer required incubation time >5 days increased the chance of losing the plasmid in the cells. Since the applied ARS plasmids based on the ARS elements of *S. pombe* which had been reported by Clyne and Kelly 1995, were not stable enough for

episomal expression of the cDNA library in the screening, the strategy was changed to chromosomal integration of the cDNA library.

Control experiments for cellulase screening using genomic integration

The advantage of chromosomal integration is that the integration of genes into *K. phaffii* genome is stable even after a long incubation time. The control experiments which had been done before with the ARS plasmids (see above), were repeated but with chromosomal integration of the expression vectors into the *K. phaffii* genome. For this, the ARS plasmids were amplified by PCR but without the ARS sequence. After transformation of the screening strains, the samples were plated onto BSM agarose plates containing either cellobiose or CMC as sole carbon source. After few days of incubation, it was possible to differentiate positive clones from negative ones on BSM_cellobiose plates. Longer incubation of the cellobiose plates showed that the screening for β-glucosidases even worked when the background growth was strong (see Fig. 4.2).

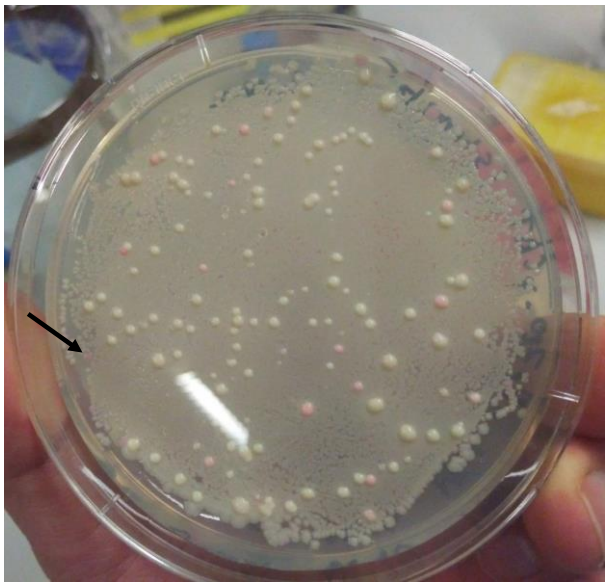


Fig. 4.2: BSM_cellobiose agarose plate with grown transformants expressing TrBGL1 after 14 days of incubations. *K. phaffii* was co-transformed with a mixture of amplified expression cassettes for the expression of stomato and TrBGL1, respectively. The transformants expressing TrBGL1 are able to grow well on cellobiose and form big colonies whereas transformants expressing only stomato are not able to form big colonies. For comparison, the black error in the figure points to a tiny red colony of a negative control expressing stomato only under P_{CAT1}. Big grown colonies with red colour are due to co-expression of both expression cassettes (stomato and TrBGL1).

The plate shown in Fig. 4.2 demonstrated very impressive that β-glucosidase was needed to form big colonies. It was already possible to see that positive clones form bigger colonies after 5-6 days of incubation. At this time the background growth was visible as tiny colonies and not as a background lawn as shown in Fig. 4.2 after 14 days of incubation. This plate also showed the occurrence of co-transformation in *K. phaffii*. While clones expressing only stomato did not show growth on this plate (see black error), the red big colonies were tested positive to have integrated both linearized expression vectors (TrBGL1 and stomato). It has to be mentioned here that there would be not that many positive clones expressing a β-glucosidase on an agarose plate during screening. The high number of transformants expressing β-glucosidase released a high amount of β-glucosidase to the medium which had an increasing effect on background growth by cleaving cellobiose to glucose molecules.

On the BSM_CMC plates, the grown colonies were much smaller than on the BSM_cellobiose plates. The reason for that seemed to be that in contrast to cellobiose, only a small part of added CMC can be used as carbon source. As mentioned in Part II of the thesis, it is likely that only the unsubstituted glucose residues in CMC can be used as carbon source by *K. phaffii* expressing a β-glucosidase and an endo-glucanase. A background of tiny colonies was also observed on the BSM_CMC plates. Further reduction of the background was necessary for being able to screen for endo-glucanases on CMC and even for exo-glucanase on Avicel.

The absence of EDTA in BSM_agarose plates seems to strongly reduce unintended background growth

While first charges of made BSM_agarose plates still contained EDTA, EDTA was not added in later charges to the BSM medium any more. The reason for that was that EDTA contains a complex carbon structure which can be used as carbon source by microorganisms [Nörtemann 1992]. It was very interesting that the colour of the grown *K. phaffii* colonies were not white but brown when incubated on BSM_cellobiose agarose medium without EDTA. The colonies were checked under microscopy to be sure that there were no contaminations. Microscopy proved that the tested colonies are budding yeasts. Streak-out of brown colonies on YPD and BM_glucose agar plates containing zeocin showed that the grown colonies became white again as it is usual for *K. phaffii* colonies.

The effect of forming brown colonies could be reproduced by redoing BSM medium with and without EDTA. This is strong evidence that *K. phaffii* strains form brown colonies on

BSM_agarose plates when EDTA was missing. This effect was first observed in BSM agar plates containing zeocin as selective marker. To be sure that the effect did not occur in the presence of zeocin only, *K. phaffii* strains were also streaked out on plates without zeocin. The brown colour of grown colonies on these BSM_agarose plates without zeocin proved that the effect was not caused by zeocin. In a next step, it was tested if this effect was limited to glucose as carbon source by inoculating BSM_glucose, BSM_glycerol, BSM_cellobiose and BSM_methanol agarose plates with *K. phaffii* strains. The tested strains are listed in Table 4.2.

Table 4.2: List of strains tested on BSM_agarose plates with different carbon sources

Nr.	Tested strain	Growth on carbon source
1	*K. phaffii* BSY11G1 Δgut1 pPpGUT1_NB2_TrBGL1; pPpT4_Zeo_pCAT1_stomato	Glucose, glycerol, cellobiose, methanol*
2	*K. phaffii* BSY11G1 Δgut1 pPpT4_Zeo_pCAT1_TrBGL1	Glucose, cellobiose, methanol*
3	*K. phaffii* BSY11G1 Δgut1	Glucose, methanol*
4	*K. phaffii* CBS7435	Glucose, methanol

* AOX1 is only knocked out in *K. phaffii* BSY11G1 Δgut1 strain which strongly slows down growth on methanol but does not fully inhibit it.

All strains which showed growth on the certain carbon sources, formed brown colonies. This indicated that the effect was not limited to a certain strain. As example, the BSM_glucose agarose plate is shown in Fig. 4.3.

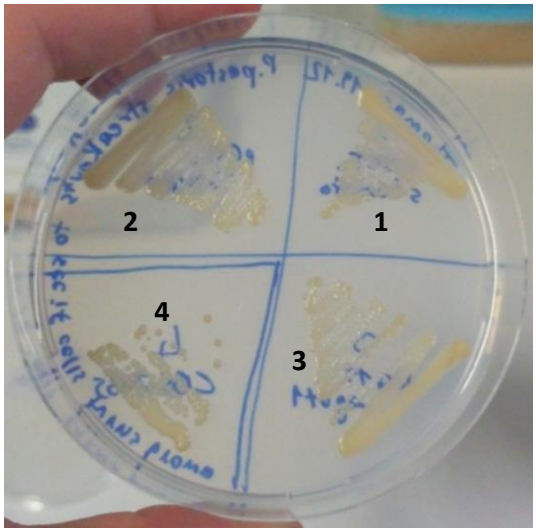


Fig. 4.3: Different *K. phaffii* strains incubated on a BSM_glucose agarose plate. Strain labelling is explained in Table 4.2.

Furthermore, the effect was not limited when sugar was the carbon source since the grown strains also formed brown colonies on methanol and glycerol. It is unknown what the reason for the brown colour of the colonies is. EDTA is a chelator for 2+ charged ions. It is likely that the 2+ charged ions which normally would be chelated by EDTA, are directly or indirectly responsible for this effect [Oviedo and Rodriguez 2003].

The finding itself that *K. phaffii* formed brown colonies on these BSM_agarose plates without EDTA was not really relevant for the screening for cellulases. It only improved the contrast between the white grown *K. phaffii* colonies and the white murky screening plates. However, it became highly relevant for such screening after the observation that only growing colonies seemed to be brown. It seemed that the brown colonies changed to white colour when the carbon source was used up. It was not able to observe real further growth of a colony on BSM agarose screening plates when the colony stayed white.
In the first days of incubation, all grown colonies on a screening plate were brown. In the next phase, first colonies changed to white colour while other colonies located e.g. at the edge of the plate stayed still brown (see Fig. 4.4).

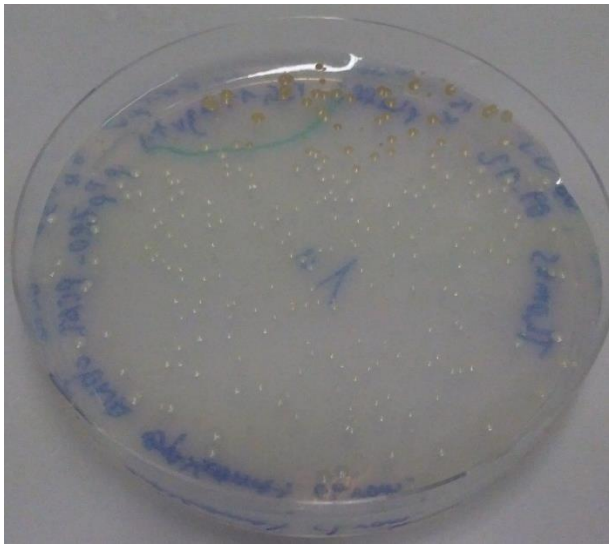


Fig. 4.4: β-glucosidase screening on a BSM_cellobiose agarose plate (without EDTA) with *K. phaffii* which was transformed with *P. aquilinum* cDNA library.

Interestingly, all colonies even the positive control changed to white colour after this phase but positive clones turned back to brown colour after growth on the certain carbon source was initiated due to cellulase expression (see Fig. 4.5). After further incubation of the agarose

plate, the brown colonies (positive clones) became bigger indicating that this clones are able to grow on that carbon source (see Fig. 4.6).

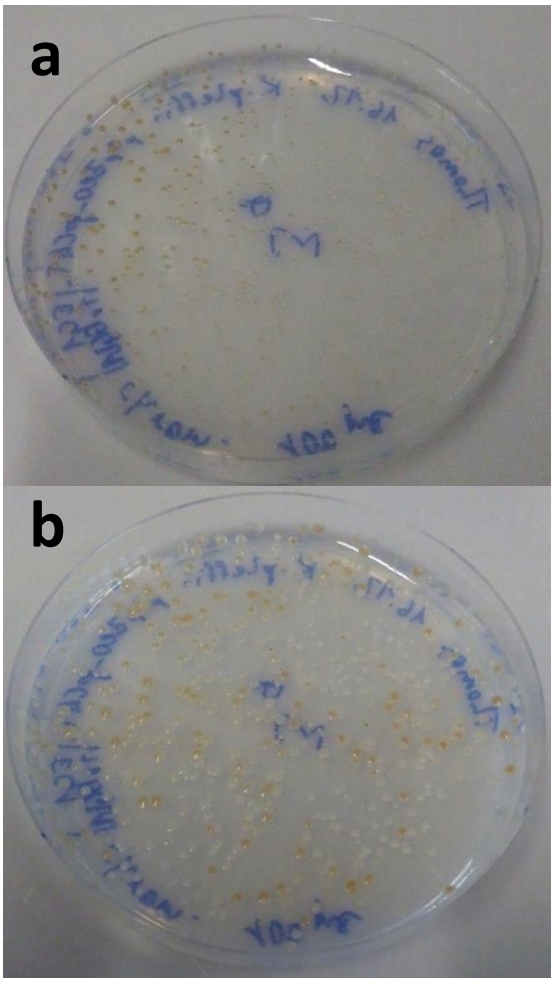


Fig. 4.5: Control experiment on BSM_cellobiose agarose plates by transforming *K. phaffii* with vectors for TrBGL1 expression as positive control and TrEG1 expression as negative control. *K. phaffii* was co-transformed with a mixture of both vectors. The vectors were linearized for chromosomal integration. The cellulases were expressed under P_{CAT1}.The plate is shown after 4 days of incubation **(a)** and after 7 days of incubation **(b)**. After 4 days of incubation, all colonies are still turning white. It can still be seen brown colonies in the left upper area of the agarose plate. After 7 days of incubation some colonies become fully brown again and form larger colonies indicating that these clones are expressing TrBGL1.

The effect of cross feeding which has been previously discussed in Part III of the thesis, also occurred in this screening. Negative clones which profited from positive clones in the neighbourhood, had a light brown colour. However, these clones could be easily differentiated from positive clones because they lost the brown colour in the outer rim of the colony. Only the centre of the colony stayed brown for a while until this part became also white (see Fig. 4.6).

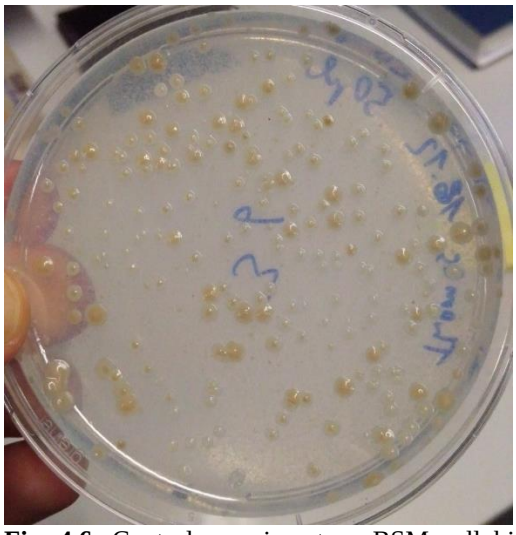


Fig. 4.6: Control experiment on BSM_cellobiose agarose plates by transforming *K. phaffii* with vectors for TrBGL1 expression as positive control and TrEG1 expression as negative control. *K. phaffii* was co-transformed with a mixture of both vectors. The vectors were linearized for chromosomal integration. The plate is shown after 10 days of incubation. It can be clearly seen that some colonies are fully brown and have formed large colonies indicating that these clones are expressing TrBGL1.

Furthermore, these colonies were much smaller in size than the positive clones. When no positive clone was on the screening plate, all colonies stayed white and no real further growth of the colonies was observed on this plate (see Fig. 4.4). To get certainty, all colonies were isolated which became completely brown again after >7 days of incubation in the control experiments (see Fig. 4.5 and 4.6). The isolated clones were tested for their ability to grow on cellobiose. The isolates of the completely brown colonies were able to grow on cellobiose whereas streak-outs of isolated white colonies as negative control did not show growth on cellobiose (not shown).

The finding that growing *K. phaffii* cells on BSM agarose plates without EDTA were brown and turned white when they stopped growing, increased the sensitivity of this screening. By selecting brown colonies, it was much easier in the control experiments to identify real positive clones and to filter out clones which formed larger colonies due to their "good" location on the plate.

Another interesting effect of BSM_agarose plates without EDTA was that the selective pressure of zeocin seemed to be stronger in this plates than in plates containing EDTA. On the screening plates with EDTA, there was still a weak background growth even after increasing the zeocin concentration to 100 µg/mL. Whereas no background growth was

observed on the screening plates without EDTA which also contained 100 µg/mL zeocin (compare Fig. 4.4 with Fig. 4.7). It is unknown if the absence of EDTA has an indirect effect on *K. phaffii* sensitivity to zeocin or if it had an effect on Zeocin.

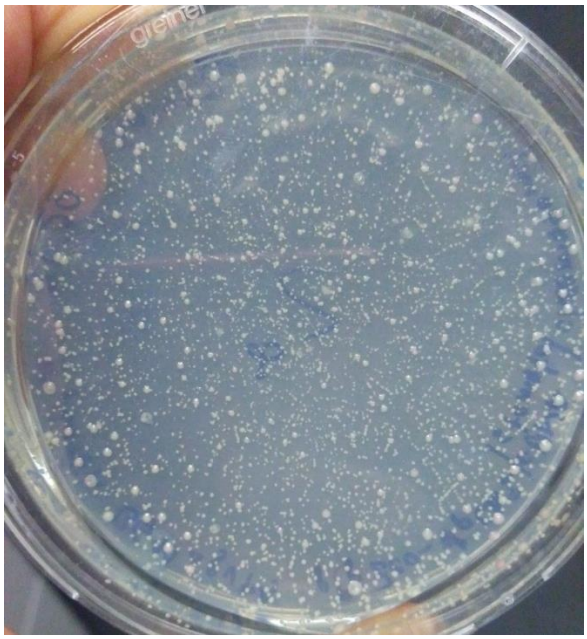


Fig. 4.7: β-glucosidase screening on a BSM_cellobiose agarose plate (with EDTA) with *K. phaffii* which was transformed with *P. aquilinum* cDNA library. Although the plate contained 100 µg/mL zeocin, a background growth of non-transformants could be observed.

Furthermore, a screening by selection of brown colonies on BSM agarose plates without EDTA might be supported by using the de-repressed P_{CAT1} for cellulase expression. It is supposed that P_{CAT1} became active when the cell growth in the colonies on the screening plates had stopped which supports the colour change back to brown in positive clones.

No positive clone was isolated in the screening for cellulases β-glucosidases and endo-glucanases

Since *K. phaffii* needs only one enzyme for being able to grow on cellobiose, the screening for β-glucosidases based on growth could be performed quite simply. The control experiments (see Fig. 4.2, 4.5 and 4.6) showed that it was possible to screen for β-glucosidases by using the described method. The *P. aquilinum* cDNA library were screened for β-glucosidases on BSM_cellobiose plates (see Fig. 4.4 and 4.7). However, no β-glucosidase expressing clone was detected.

The screening of the *P. aquilinum* cDNA library for endo-glucanases was not that simple as the screening for β-glucosidases. The co-expression of endo-glucanases and also of a β-glucosidase was required for enabling *K. phaffii* to grow on CMC. A *K. phaffii* BSY11G1 strain expressing TrBGL1 under NB2 promoter control was used in the screening for endo-glucanases. As mentioned before, the grown colonies of *K. phaffii* co-expressing TrBGL1 and TrEG1 under were very small on BSM_CMC agarose plates with EDTA which made a screening based on growth more challenging. To increase the size of growing colonies on CMC plates, the CMC concentration was increased from 0.5% to 1%. Furthermore, the screening was improved by using BSM_CMC agarose plates without EDTA. By using plates without EDTA, the interfering background growth of non-transformants was eliminated. Furthermore, positive clones were able to be selected not only by size but also by brown colour as demonstrated in control experiments with β-glucosidase on cellobiose (Fig. 4.5 and 4.6). The cDNA library was also screened for endo-glucanases. However, no positive clone was detected.

The described screening method is not suitable for screening for exo-glucanases on Avicel

Even more challenging is the creation of a high throughput screening method for exo-glucanases. Ko and co-worker reported a high throughput screening method for exo-glucanases in liquid samples using robotics [Ko et al. 2013b]. However to the best of my knowledge, no functional high throughput plate screening method has been reported yet. In this work, it was tried to screen for exo-glucanases according to growth on Avicel. Here, *TrCBH2* was expressed under P_{CAT1} control to see if the co-expression of *TrCBH2* with *TrBGL1* (under NB2 control) enabled *K. phaffii* to grow on Avicel. However, no active growth of *K. phaffii* co-expressing TrBGL1 and TrCBH2 was observed on BSM_Avicel agarose plates with EDTA. The grown colonies of the transformants expressing TrCBH2 or sTomato did not differ in size on the same BSM_Avicel agarose plate (not shown).

Conclusion and Outlook

Here, a new growth based screening method on agarose plates for isolation of cellulases was presented. With the usage of the yeast *K. phaffii* as expression host, it was also possible to screen for eukaryotic cellulases. As eukaryotic expression host, *K. phaffii* is able to do typical eukaryotic posttranslational modifications on proteins [Cereghino and Cregg 2000] which is required for functional expression of eukaryotic proteins. To the best of my knowledge, there has not been any report about a high throughput plate screening for cellulases based on growth using *K. phaffii* as a host for screening or selection.

Although it was not possible to find a novel cellulase by screening the *P. aquilinum* cDNA library, it was clearly demonstrated that a screening for β-glucosidases was possible according to growth on cellobiose. This screening method was also used for screening cDNA library for endo-glucanases. However, a screening for endo-glucanases only according to cell growth might be too troublesome. The clones co-expressing TrBGL1 and TrEG1 needed a longer incubation time to grow on CMC and the grown colonies were very small compared to grown colonies on cellobiose plates. At the beginning, background growth was an additional problem during the screening for endo-glucanases. However, the change to BSM_CMC agarose plates without EDTA made the described screening method also suitable for screening cDNA libraries for endo-glucanases. The background growth was eliminated and the brown colour of growing cells made the screening of the small colonies easier. The change to BSM_cellobiose agarose plates without EDTA also improved the sensitivity of the screening for β-glucosidases.

It was not possible to establish a functional screening for exo-glucanases due to the lack of growth on Avicel. It seemed that TrCBH2 is not only insufficiently expressed under NB2 promoter but also under P_{CAT1}. Since TrCBH2 was successfully expressed in high concentration under inducible *AOX1* promoter in *K. phaffii* [Mellitzer et al. 2012b], it should be considered to induce P_{CAT1} with methanol during the screening. As mentioned before, activity of de-repressed P_{CAT1} when glucose is absent, can be strongly enhanced by induction with methanol [Vogl et al. 2016]. *K. phaffii* BSY11G1 which was used in this work, is a Mut^S strain. Growth on methanol is strongly slowed down in *K. phaffii* Mut^S strains by the knock-out of *AOX1* [Krainer et al. 2012]. Therefore, it might be worth to repeat the approach for screening for exo-glucanases which was done in this study. With the change that the screening plates (BSM_Avicel agarose) should also contain 0.5% (w/v) methanol for

inducing TrCBH2 expression under P_{CAT1} control. Theoretically, expression of TrCBH2 under P_{CAT1} control is strongly induced by methanol when glucose is absent and is repressed when enough glucose is liberated from Avicel. That means that positive clones co-expressing exo-glucanases might be able to grow on Avicel forming larger colonies than negative clones. This strategy was already shown to be working in the screening for β-glucosidases and endo-glucanases where the cellulases were also expressed under P_{CAT1} control. Such a screening would strongly facilitate the search for new exo-glucanases.

References

Ando S, Ishida H, Kosugi Y, Ishikawa K (2002) Hyperthermostable Endoglucanase from *Pyrococcus horikoshii*. Appl Environ Microbiol 68: 430–433. doi: 10.1128/AEM.68.1.430–433.2002

Biely P, Markovič O, Mislovičová D (1985) Sensitive detection of endo-1,4-β-glucanases and endo-1,4-β-xylanases in gels. Anal Biochem 144: 147-151. doi: https://doi.org/10.1016/0003-2697(85)90096-X

Carder JH (1986) Detection and quantitation of cellulase by Congo red staining of substrates in a cup-plate diffusion assay. Anal Biochem 153: 75-79. doi: https://doi.org/10.1016/0003-2697(86)90063-1

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24: 45-66. doi: https://doi.org/10.1111/j.1574-6976.2000.tb00532.x

Chen J, Wang Q, Hua Z, Du G (2007) Research and application of biotechnology in textile industries in China. Enzyme Microb Technol 40:1651-1655. doi: https://doi.org/10.1016/j.enzmictec.2006.07.040

Chen P, Fu X, Ng TB, Ye X-Y (2011) Expression of a secretory β-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. Biotechnol Lett 33: 2475–2479. doi: 10.1007/s10529-011-0724-3

Chen H-M, Armstrong Z, Hallam SJ, Withers SG (2016) Synthesis and evaluation of a series of 6-chloro-4-methylumbelliferyl glycosides as fluorogenic reagents for screening metagenomic libraries for glycosidase activity. Carbohydr Res 421:33-39. doi: https://doi.org/10.1016/j.carres.2015.12.010

Clyne RK, Kelly TJ (1995) Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. The EMBO Journal 14: 6348-6357. doi: https://doi.org/10.1002/j.1460-2075.1995.tb00326.x

Colin P-Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M, Morgavi DP, Janssen DB, Hollfelder F (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nat Commun 6:10008. doiI: 10.1038/ncomms10008

Coniglio RO, Fonseca MI, Villalba LL, Zapata PD (2017) Screening of new secretory cellulases from different supernatants of white rot fungi from Misiones, Argentina. Mycology 8:1-10. doi: 10.1080/21501203.2016.1267047

Duan C-J, Feng J-X (2010) Mining metagenomes for novel cellulase genes. Biotechnol Lett 32:1765–1775. doi: 10.1007/s10529-010-0356-z

Dyadic: https://www.dyadic.com/wp-content/uploads/2014/11/Next-Generation-Enzymes.pdf (24.09.2018)

Endo Y, Sawasaki T (2006) Cell-free expression systems for eukaryotic protein production. Curr Opin Biotechnol 17:373–380.

Findley SD, Mormile MR, Sommer-Hurley A, Zhang X-C, Tipton P, Arnett K, Porter JH, Kerley M, Gary Stacey G (2011) Activity-Based Metagenomic Screening and Biochemical Characterization of Bovine Ruminal Protozoan Glycoside Hydrolases. Appl Environ Microbiol 77:8106–8113. doi: 10.1128/AEM.05925-11

Fischer JE (2016) Master's Thesis: Development of new Methods for reliable cDNA Library Generation and Expression. Graz University of Technology, Faculty of Technical Chemistry, Chemical and Process Engineering, Biotechnology

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA III, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods 6:*343–345. doi:10.1038/NMETH.1318

Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, Heaphy S (2006) Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Appl Environ Microbiol 72:135–143. doi:10.1128/AEM.72.1.135–143.2006

Guazzaroni M-E, Silva-Rocha, Ward RJ (2015) Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. Microb Biotechnol 8:52-64 doi: https://doi.org/10.1111/1751-7915.12146

Gurung N, Ray S, Bose S, Rai V (2013) A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. Biomed Res Int vol. 2013, Article ID 329121, 18 pages. doi: http://dx.doi.org/10.1155/2013/329121

Jang H-D, Chen K-S (2002) Production and characterization of thermostable cellulases from *Streptomyces* transformant T3-1. World J Microbiol Biotechnol 19: 263–268. doi: https://doi.org/10.1023/A:1023641806194

Helbert W, Chanzy H, Husum TL, Schülein M, Ernst S (2003) Fluorescent Cellulose Microfibrils As Substrate for the Detection of Cellulase Activity. *Biomacromolecules* 4:481–487**.** doi**:** 10.1021/bm020076i

Holtzapple MT, Caram HS, Humphrey AE (1984) Determining the inhibition constants in the HCH-1 model of cellulose hydrolysis. Biotechnol Bioeng 26:753-757. doi: https://doi.org/10.1002/bit.260260719

Houfani AA, Větrovský T, Baldrian P, Benallaoua S (2017), Efficient screening of potential cellulases and hemicellulases produced by *Bosea* sp. FBZP-16 using the combination of enzyme assays and genome analysis. World J Microbiol Biotechnol 33:29. doi: 10.1007/s11274-016-2198-x

Invivogen: www.invivogen.com/PDF/Zeocin_powder_TDS.pdf *(10.01.2018)*

Kickenweiz T, Glieder A, Wu JC (2018) Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase. Appl Microbiol Biotechnol 102: 1297–1306. doi: https://doi.org/10.1007/s00253-017-8656-z

Kim S-J, Lee C-M, Kim M-Y, Yeo Y-S, Yoon S-H, Kang H-C, Koo B-S (2007) Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA. J Microbiol Biotechnol 17: 905–912.

Kim JJ, Kwon Y-K, Kim JH, Heo S-J, Lee Y, Lee S-J, Shim W-B, Jung W-K, Hyun J-H, Kwon KK, Kang D-H, Oh C (2014) Effective microwell plate-based screening method for microbes producing cellulase and xylanase and its application. J Microbiol Biotechnol 24:1559–1565. doi: http://dx.doi.org/10.4014/jmb.1405.05052

Klein-Marcuschamer D, Blanch HW (2015) Renewable fuels from biomass: Technical hurdles and economic assessment of biological routes. AIChE J 61: 2689–2701. doi: 10.1002/aic.14755

Ko K-C, Lee JH, Han Y, Choi JH, Song JJ (2013a) A novel multifunctional cellulolytic enzyme screened from metagenomics resources representing ruminal bacteria. Biochem Biophys Res Commun 441:567–572. doi: http://dx.doi.org/10.1016/j.bbrc.2013.10.120

Ko K-C, Han Y, Cheong D-E, Choi JH, Song JJ (2013b) Strategy for screening metagenomic resources for exocellulase activity using a robotic, high-throughput screening system. J Microbiol Methods 94:311–316. doi: http://dx.doi.org/10.1016/j.mimet.2013.07.010

Körfer G, Pitzler C, Vojcic L, Martinez R, Schwaneberg U (2016) *In vitro* flow cytometry-based screening platform for cellulase engineering. Sci Rep 6:26128. doi: 10.1038/srep26128

Kostylev M, Wilson D (2012) Synergistic interactions in cellulose hydrolysis. Biofuels. 3: 61-70. doi: http://dx.doi.org/10.4155/bfs.11.150

Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A (2012) Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact 11:22. doi: https://doi.org/10.1186/1475-2859-11-22

Kuhad RC, Deswal D, Sharma S, Bhattacharya A, Jain KK, Kaur A, Pletschke BI, Singh A, Karp M (2016) Revisiting cellulase production and redefining current strategies based on major challenges. Renew Sust Energ Rev 55: 249–272. doi: http://dx.doi.org/10.1016/j.rser.2015.10.132

Legras J-L, Merdinoglu D, Cornuet J-M, FRANCIS Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol 16: 2091–2102. doi: https://doi.org/10.1111/j.1365-294X.2007.03266.x

Li X, Gruber MY, Hegedus DD, Lydiate DJ, Gao M-J (2011) Effects of a Coumarin derivative, 4-Methylumbelliferone,on seed germination and seedling establishment in *Arabidopsis*. J Chem Ecol (2011) 37:880–890. doi: 10.1007/s10886-011-9987-3

Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, Lin-Cereghino GP (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 38: 44–48

Liu W, Hong J, Bevan DR, Zhang Y-HP (2009) Fast identification of thermostable beta-glucosidase mutants on cellobiose by a novel combinatorial selection/screening approach. Biotechnol Bioeng 103:1087-1094. doi: https://doi.org/10.1002/bit.22340

Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE (2017) Cellulosic ethanol: status and innovation. Curr Opin Biotechnol 45:202–211. doi: http://dx.doi.org/10.1016/j.copbio.2017.03.008

Mellitzer A, Glieder A, Weis R, Reisinger C, Flicker K (2012a) Sensitive high-throughput screening for the detection of reducing sugars. Biotechnol. J 7:155–162. doi: 10.1002/biot.201100001

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

Monga A, Chadha BS (2014) Characterization and evaluation of cellulases and hemicellulases produced by diverse fungal strains. Curr Biotechnol 3:257-265

Mood SH, Golfeshan AH, Tabatabaei M, Jouzani GS, Najafi GH, Gholami M, Ardjmand M (2013) Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. Renewable and Sustainable Energy Reviews 27: 77–93. doi: https://doi.org/10.1016/j.rser.2013.06.033

Nörtemann B (1992) Total degradation of EDTA by mixed cultures and a bacterial isolate. *Appl Environ Microbiol 58:*671-676

Ostafe R, Prodanovic R, Commandeur U, Fischer R (2013) Flow cytometry-based ultra-high-throughput screening assay for cellulase activity. Anal Biochem 435: 93–98. Doi: http://dx.doi.org/10.1016/j.ab.2012.10.043

Oviedo C and Rodriguez J (2003) EDTA: the chelating agent under environmental scrutiny. Quim. Nova 26:901-905.doi: http://dx.doi.org/10.1590/S0100-40422003000600020

Pang H, Zhang P, Duan C-J, Mo X-C, Tang J-L, Feng J-X (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Curr Microbiol 58:404–408. doi: 10.1007/s00284-008-9346-y

Poornejad N, Karimi K, Behzad T (2013) Improvement of saccharification and ethanol production from rice straw by NMMO and [BMIM][OAc] pretreatments. Ind Crops Prod 41:408– 413. doi: http://dx.doi.org/10.1016/j.indcrop.2012.04.059

Quay DHX, Bakar FDA, Rabu A, Said M, Illias RM, Mahadi NM, Hassan O, Murad AMA (2011) Overexpression, purification and characterization of the *Aspergillus niger* endoglucanase, *EglA*, in *Pichia pastoris*. Afr J Biotechnol 10:2101-2111. doi: 10.5897/AJB10.1046

Rao MB, Tanksale AM, Ghatge MS, Deshpande VV (1998) Molecular and Biotechnological Aspects of Microbial Proteases. *Microbiol Mol Biol Rev* 62:597–635.

Ren H, Zong MH, Wu H, Li N (2016) Efficient pretreatment of wheat straw using novel renewable cholinium ionic liquids to improve enzymatic saccharification. Ind Eng Chem Res 55:1788−1795. doi: 10.1021/acs.iecr.5b03729

Robson LM, Chambliss GH (1986) Cloning of the *Bacillus subtilis* DLG beta-1, 4-glucanase gene and its expression in *Escherichia coli* and *B. subtilis*. *J Bacteriol 165:*612-619. doi: 10.1128/jb.165.2.612-619.1986

Saini JK, Patel AK, Adsul M, Singhania RR (2016) Cellulase adsorption on lignin: A roadblock for economic hydrolysis of biomass Renew Energ 98:29-42. doi: https://doi.org/10.1016/j.renene.2016.03.089

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat Biotechnol 22: 1567–1572

Singh R, Kumar M, Mittal A, Mehta PK (2016) Microbial enzymes: industrial progress in 21st century. 3 Biotech 6: 174. doi: https://doi.org/10.1007/s13205-016-0485-8

Soares FL, Melo IS, Dias ACF, Andreote FD (2012) Cellulolytic bacteria from soils in harsh environments. World J Microbiol Biotechnol 28:2195–2203. doi: https://doi.org/10.1007/s11274-012-1025-2

Srivastava N, Srivastava M, Mishra PK, Singh P, Ramteke PW (2015) Application of Cellulases in Biofuels Industries: An Overview. J Biofuels Bioenergy 1: 55-63. doi: 10.5958/2454-8618.2015.00007.3

Theater RM and Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol 43:*777-780.

Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol 15: 160-167. doi: https://doi.org/10.1016/S0167-7799(97)01032-9

Ten LN, Im WT, Kim MK, Kang MS, Lee ST (2004) Development of a plate technique for screening of polysaccharide-degrading microorganisms by using a mixture of insoluble chromogenic substrates. J Microbiol Methods 56:375– 382. doi: https://doi.org/10.1016/j.mimet.2003.11.008

Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol 20:616–622. doi: 10.1016/j.copbio.2009.09.010

Vervoort Y, Linares AG, Roncoroni M, Liu C, Steensels J, Verstrepen KJ (2017) High-throughput system-wide engineering and screening for microbial biotechnology. Curr Opin Biotechnol 46: 120-125. doi: https://doi.org/10.1016/j.copbio.2017.02.011

Voget S, Steele HL, Streit WR (2006) Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol 126:26–36. doi:10.1016/j.jbiotec.2006.02.011

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. doi: 10.1021/acssynbio.5b00199

Wang F, Li F, Chen G, Liu W (2008) Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res 164: 650—657. doi:10.1016/j.micres.2008.12.002

Yan Y, Smant G, Davis E (2001) Functional Screening Yields a New β-1,4-Endoglucanase Gene from *Heterodera glycines* that May be the Product of Recent Gene Duplication. Mol *Plant Microbe Interact* 14:63–71. doi: https://doi.org/10.1094/MPMI.2001.14.1.63

Yoon JH, Park JE, Suh DY, Hong SB, Ko SJ, Kim SH (2007) Comparison of dyes for easy detection of extracellular cellulases in fungi. Mycobiology 35:21-24. doi: https://doi.org/10.4489/MYCO.2007.35.1.021

Zhang YHP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. Biotechnol Adv 24: 452-481. doi: https://doi.org/10.1016/j.biotechadv.2006.03.003

Zhang Z, Liu J-L, Lan J-Y, Duan C-J, Ma Q-S, Feng J-X (2014) Predominance of *Trichoderma* and *Penicillium* in cellulolytic aerobic filamentous fungi from subtropical and tropical forests in China, and their use in finding highly efficient β-glucosidase. Biotechnol Biofuels 7:107

Zhao S, Yan Y-S, He Q-P, Yang L, Yin X, Li C-X, Mao L-C, Liao L-S, Huang J-Q, Xie S-B, Nong Q-D, Zhang Z, Jing L, Xiong Y-R, Duan C-J, Liu J-L, Feng J-X (2016) Comparative genomic, transcriptomic and secretomic profiling of *Penicillium oxalicum* HP7-1 and its cellulase and xylanase hyper-producing mutant EU2106, and identification of two novel regulatory genes of cellulase and xylanase gene expression. Biotechnol Biofuels **9**:203. doi: https://doi.org/10.1186/s13068-016-0616-9

Zhou Y, Wang X, Wei W, Xu J, Wang W, Xie Z, Zhang Z, Jiang H, Wang Q, Wei C (2016) A novel efficient β-glucanase from a paddy soil microbial metagenome with versatile activities. Biotechnol Biofuels 9:36. doi: 10.1186/s13068-016-0449-6

Summary and Conclusion

In this Thesis, bidirectional promoters (NB2 and NB3) of *K. phaffii* were identified (see Part I) and used as tool for co-expression of cellulases (see Part II). This study demonstrated the suitability and simplicity of controlling the co-expression of a few genes. In Part II, the non-cellulolytic yeast *Komagataella phaffii* was successfully engineered to metabolize cellobiose and modified cellulose. A *K. phaffii* strain, co-expressing a β-glucosidase, an endo-glucanase, an exo-glucanase and eGFP was generated which was able to grow on cellobiose and carboxymethyl cellulose (CMC) as sole carbon source (see Fig. 4a of Part II). Although no growth of this strain was observed on Avicel, evidence for the functional expression of the exo-glucanase (TrCBH2) in this system was found in the experiments which are shown in Fig. 5 of Part II and Fig. A2 of Appendix part. It was possible to show that the enzyme mix in the supernatant of the engineered strain was able to release glucose from Avicel when it was incubated together with Avicel at 50°C indicating a functional expression of TrCBH2. It seemed that the expression of the exo-glucanase (TrCBH2) was too insufficient for an observable growth on Avicel. Nevertheless, this was the first time that all three key cellulases which are required for cellulose saccharification, were co-expressed in one *K. phaffii* strain.

Although the expression of the exo-glucanase might have been insufficient in this system, the results of this study indicate the possibility to enable *K. phaffii* to grow on Avicel by cell-engineering. Due to the ongoing progress in heterologous gene expression in *K. phaffii* (e.g. promoter research and polycistronic gene expression) [Geier et al. 2015, Vogl et al. 2016], it is likely that exo-glucanases can be sufficiently expressed or alternative exo-glucanases will allow to use Avicel as carbon source in the future. A *K. phaffii* strain that is able to use (ligno-) cellulose as cheap non-food carbon source for production of high value compounds would be highly interesting for industrial applications. The presentation of an engineered *K. phaffii* strain that can metabolize CMC, might be a first important step towards this direction.

Part III of this thesis describes a method for screening *K. phaffii* mutants with improved protein secretion. Two screening strains were used whose secretion of the expressed cellulases might be restricted by a bottleneck in the secretory pathway. The first screening strain has integrated two copies of a gene coding for a β-glucosidase (AnBGL1). This strain did not grow faster on cellobiose than a strain expressing only one gene copy of *AnBGL1* (see

Fig. 3.2 and 3.3 of Part III) although the strain was grown on media where the amount of available glucose determined the specific growth rate. Measuring the fluorescence of co-expressed eGFP was used to indicate the gene copy numbers. Since the measurement of the fluorescence signals showed that the expression of eGFP was seemingly not affected in multi gene copy strains, it might be that there was a bottleneck in the secretory pathway of the two gene copy strain. Although further research has to be done for getting more evidence for a bottleneck in the secretory pathway in this strain, the two gene copy strain was used as screening strain to improve its growth on cellobiose by mutagenesis experiments.

The second screening strain was the strain expressing all three cellulases described in Part II. Previously it has been reported that a bottleneck in the secretory pathway limits the protein secretion in multi-copy *K. phaffii* strains expressing secreted proteins [Routenberg Love et al. 2012]. In the second screening strain, three genes were heterologously expressed which coded for secreted proteins (AnBGL1, TrCBH2, AnEG-A). It might either be that a bottleneck in the secretory pathway also restricted the secretion of the expressed cellulases in this strain or alternatively TrCBH2 simply was not the choice for these studies. Nevertheless this strain might be a convenient screening strain for mutagenesis experiments to improve the protein secretion in *K. phaffii* or for studies including overexpression of further enzyme components of typical natural cellulase cocktails.

The described screening method is very simple and does not require special equipment or toxic dyes. In this screening, the mutated *K. phaffii* cells were plated onto minimal medium agar plates containing cellobiose, CMC or Avicel as sole carbon source. Clones which showed a better growth on the screening plates than the average, could be isolated for further tests to check if the protein secretion was increased in these clones. To the best of my knowledge, it was the first time that a method was presented for screening *K. phaffii* mutants with improved protein secretion according to faster growth on cellobiose or CMC.

Although no clone with improved protein secretion was isolated in this study, this method might be very useful for facilitated screening after the mutagenesis step to select promising clones.

In Part IV, a screening method for identifying novel cellulases was described. Based on the results of Part II, it was possible to create a screening for β-glucosidases and endo-glucanases using *K. phaffii* as screening strain. A fern cDNA library was screened for β-glucosidases by detecting *K. phaffii* clones which can grow on cellobiose as sole carbon source. Endo-

glucanases might be discovered by using a *K. phaffii* strain that expresses a β-glucosidase (TrBGL1) as screening strain and finally allows growth on CMC as sole carbon source.
For this screening, a novel solid minimal medium was introduced. This medium was adapted from BSM (Basal Salt Medium) broth medium which is usually used for *K. phaffii* fermentation experiments [Mellitzer et al. 2012]. Agarose was used for the plates instead of agar. By introducing BSM_agarose plates, it was possible to reduce unintended background growth to a minimum.

The making of BSM_agarose plates without EDTA seemed to have two beneficial effects for the screening. The first effect was that growing *K. phaffii* cells formed brown colonies (see Fig. 4.5 of Part IV). The brown colonies became white (usual colour of *K. phaffii* colonies) after the available carbon source in the plates was used up. Furthermore, it seemed that the selective pressure of zeocin was stronger in BSM_agarose plates without EDTA than in plates containing EDTA. This had a positive effect on the screening by elimination of the unintended background growth of non-transformants. Another important advantage was that the clones were not only screened according to their growth but also according to the brown colour of growing colonies. Colonies which were still completely brown after >7 days of incubation can be interpreted to be positive clones expressing cellulases. This increased the sensitivity of the screening for both, β-glucosidases and endo-glucanases.
It was not possible in this study to create a functional screening for exo-glucanases. Even the construct made as a positive control did not allow for the planned growth on Avicel. As concluded in Part II, it seemed that expression of the exo-glucanase TrCBH2 was insufficient for a growth on Avicel or this enzyme is not efficient enough by itself for the targeted release of glucose or cellobiose.
Although no novel cellulase was isolated in the screening of a cDNA library, the described screening method might be highly convenient for screening for β-glucosidases and endo-glucanases. To the best of my knowledge, it was the first time that a screening method was presented using *K. phaffii* to screen for cellulases according to cell growth on cellobiose and CMC. Furthermore, to the best of my knowledge, it has not been described in literature before that growing *K. phaffii* cells form brown colonies on BSM_agarose plates in the absence of EDTA. EDTA is a chelator for 2+ charged metal ions. It has been reported that EDTA can be used as carbon source by microorganisms [Nörtemann 1992, Oviedo and Rodriguez 2003] but this cannot explain the observed effect of the brown coloured *K. phaffii* colonies. It was not possible to find a reference which clearly explains the role of EDTA in *K.*

phaffii medium. Since EDTA is a chelator for 2+ charged metal ions, the reason for the brown coloured *K. phaffii* colonies might be a higher concentration of free metal ions in the medium which would be chelated in the presence of EDTA. It would be interesting to find out which 2+ charged metal ion (for example Fe^{2+} or Cu^{2+}) is responsible for this effect. There is a possibility that a higher intracellular concentration of such metal ions in growing cells might lead to a brown colony-colour. Is that the case, an question would be if the effect in the cells is caused by an increased concentration of formed protein complexes containing these metal ions (e.g. Fe^{2+} in heme containing proteins) or by free metal ions. Research to reveal the cause for the brown colonies might be topic in further studies.

References

Geier M, Fauland P, Vogl T, Glieder A (2015) Compact multi-enzyme pathways in *P. pastoris*. Chem Commun 51: 1643. doi: 10.1039/C4CC08502G

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

Nörtemann B (1992) Total degradation of EDTA by mixed cultures and a bacterial isolate. *Appl Environ Microbiol 58:*671-676

Routenberg Love K, Politano TJ, Panagiotou V, Jiang B, Stadheim TA, Love JC (2012) Systematic Single-Cell Analysis of Pichia pastoris Reveals Secretory Capacity Limits Productivity. PLoS ONE 7: e37915. doi:10.1371/journal.pone.0037915

Vogl T, Kickenweiz T, Sturmberger L, Glieder A (2015) Bidirectional promoter. US 20150011407 A1.

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. **doi:** 10.1021/acssynbio.5b00199

Appendix

As mentioned in the general introduction at the beginning of the thesis, the Appendix part contains the experiments done in the *K. phaffii* engineering project which are not part of the paper published by Kickenweiz et al. 2018 (Part II of the thesis). All material and methods are used and performed as described in Kickenweiz et al. 2018 (Part II of the thesis) if not stated otherwise.

A1. Experiments

A1.1 pHBAH assays revealed that TrEG1 was functionally expressed under control of NB2 promoter in *K. phaffii*:

To see if *TrCBH2* and *TrEG1* were functionally expressed in *K. phaffii* under control of NB2 promoter, *K. phaffii* BSY11G1 strains were created which have integrated the vectors B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt and B1_S_Zeo_ ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt in their genomes. Furthermore, *K. phaffii* BSY11G1 was also transformed with the vector B1_S_Zeo_ADHtt_NB2_TrEG1_TrBGL1_alpha.sig.seq_NB2_TrCBH2-HM to test the cellulolytic activity when *TrCBH2* was co-expressed with *TrEG1* in *K. phaffii*. Positive transformants grown on YPD-Zeo plates were isolated for further cultivation in 96 deep-well plates. *K. phaffii* cultures were incubated in 96 deep-well plates in 250 µL of BMD 1% at 28°C and 320 rpm for 60 h. Then, 50 µL of BMD 3.5% were added to the cultures and further incubated for 12 h. After reaching stationary phase, the supernatant was harvested for testing the expressed cellulases. Cellulolytic activity of expressed cellulases was detected by measuring the reducing sugar ends of degraded cellulose. For this, pHBAH assay was performed as described in A2 of the Appendix part. The results from the pHBAH assay are shown in Fig. A1.

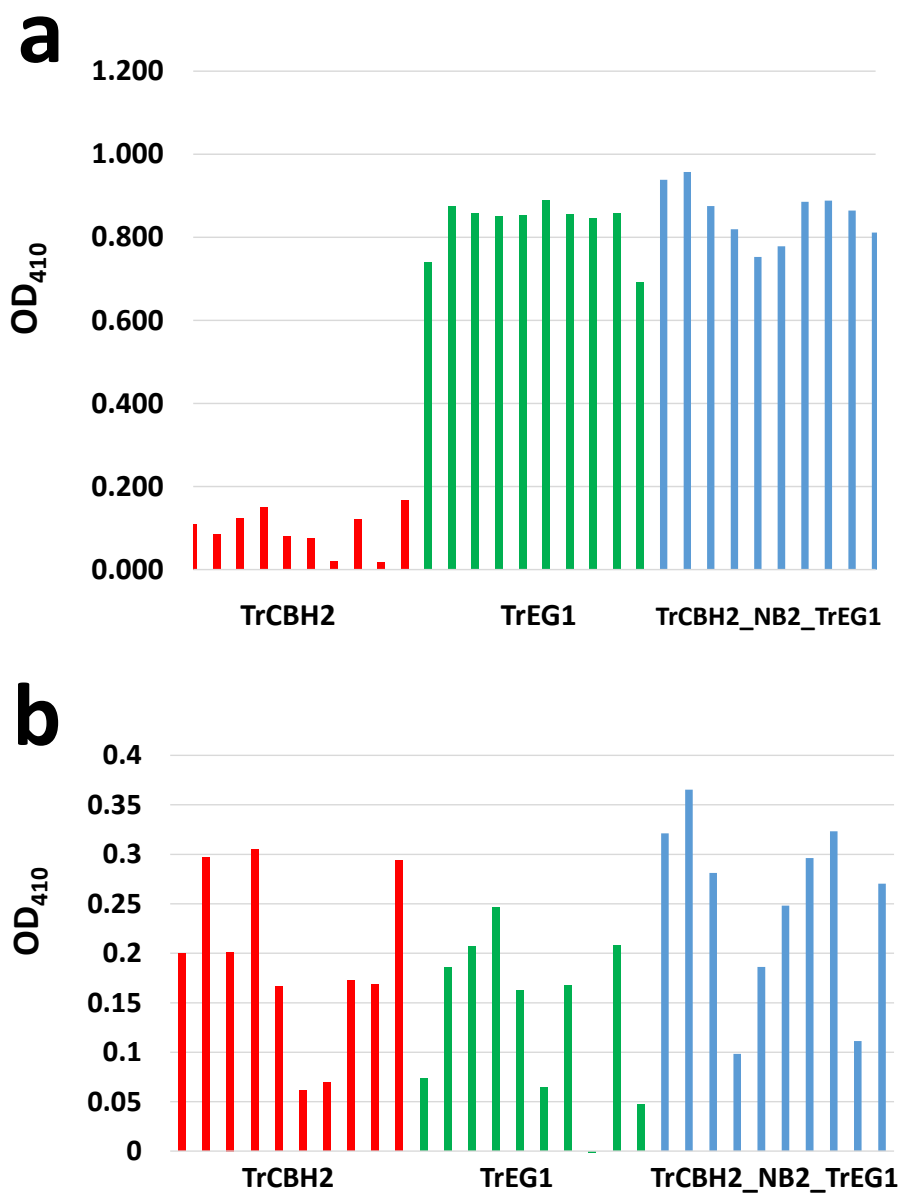


Fig. A1: pHBAH assays with the supernatant of engineered *K. phaffii* strains. pHBAH assay was performed with Avicel as substrate (**a**) and with CMC (**b**) after the samples were incubated at 50°C for 2 h. Values are shown after subtraction of BSY11G1 reference strain values. Values of the clones expressing TrCBH2 are shown in red, expressing TrEG1 in green and co-expressing both genes in blue bars.

The pHBAH assay showed very clearly that TrEG1 was functionally expressed under NB2 promoter control in *K phaffii*. High OD_{410} values were measured in the samples containing TrEG1 only and the combination TrCBH2 with TrEG1 whereas the OD_{410} values were very low in the samples with TrCBH2 only (see Fig. A1a). It is known from the literature that endo-glucanases are active on amorphous cellulose (e.g. CMC) cutting cellulose randomly in the amorphous region. On the crystalline cellulose (e.g. Avicel), however, the activities of endo-glucanases are much lower [Henrissat et al. 1985, Lynd et al. 2002]. This fits very well

to the results of the pHBAH assay. The samples containing TrEG1 showed high OD_{410} values with CMC whereas the measured OD_{410} values were very low in the samples containing TrEG1 with Avicel.

The OD_{410} values of the samples containing only TrCBH2 were low when it was incubated with, Avicel and CMC as substrate. However, TrCBH2 has been reported to have high activity when expressed in *K. phaffii* [Mellitzer et al. 2012a, Mellitzer et al. 2014]. Since TrCBH2 was expressed successfully before in *K. phaffii*, it was expected the activity of TrCBH2 to be higher in the samples with Avicel in the pHBAH assay. This might be evidence for insufficient expression of *TrCBH2* under control of NB2 promoter in *K. phaffii*. However, it should also be mentioned that the OD_{410} values had a tendency to be higher in the samples with Avicel containing the combination TrCBH2 with TrEG1 than in the samples containing TrEG1 only. This might suggest an additive effect of endo-glucanase and exo-glucanase on Avicel degradation although not sufficient to allow growth on this cellulose material as sole carbon source.

A1.2 TrCBH2 with AnBGL1 co-expressed in *K. phaffii* liberate glucose from Avicel when incubated at 50°C:

One high priority was to find evidence for the liberation of glucose from Avicel by co-expression of either *TrCBH2* or *TrEG1* with *AnBGL1*. Therefore, *K. phaffii* BSY11G1 strains were created which co-expressed either *TrCBH2* with *AnBGL1* or *TrEG1* with *AnBGL1*. The list of the used strains in this experiment are shown in Table A1. In this approach all used genes were expressed under control of constitutive histone promoters.

Table A1: List of the strains used in this experiment

Used strains in this experiment	**Cellulases expressed**
K. phaffii BSY11G1 pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_AOX1tt	AnBGL1
K. phaffii BSY11G1 pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_AOX1tt B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt	AnBGL1 + TrCBH2
K. phaffii BSY11G1 pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_AOX1tt B1_S_Zeo_ ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt	AnBGL1 + TrEG1

Positive transformants grown on YPD plates containing zeocin and geneticin, were isolated for cultivation in 96 deep-well plates. In 96 deep-well plates, *K. phaffii* cultures were incubated in 250 µL of BMD 1% at 28°C and 320 rpm for 60 h. Then, 50 µL of BMD 3.5% was added to the cultures and further incubated. After reaching stationary phase, the supernatant of the cultures was harvested to test the expressed cellulases located in the supernatant for their ability to form glucose from Avicel. The procedure was performed as described in A2 of the Appendix part but with the exception that the incubation time of the supernatants with Avicel at 50°C was 4 h. The results are shown in Fig. A2.

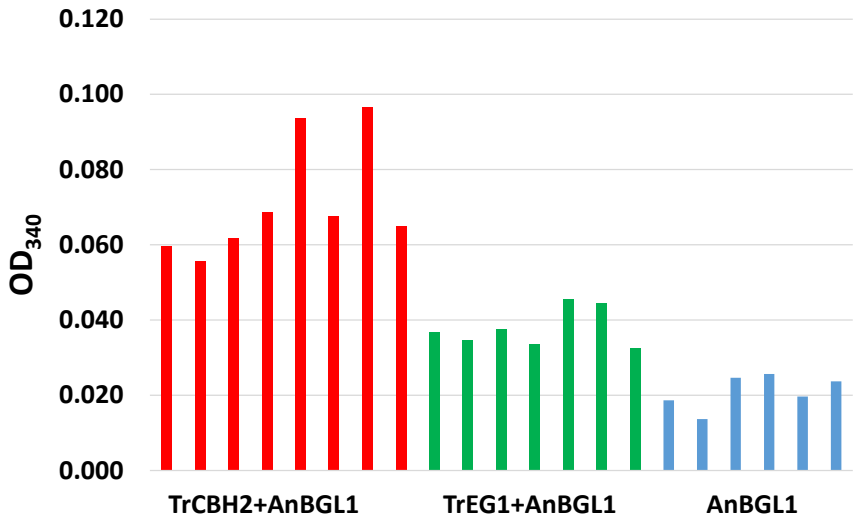


Fig. A2: Measurement of liberated glucose from Avicel by the enzymes in the supernatant of engineered *K. phaffii* strains. Glucose measurement was performed after the samples were incubated at 50°C for 4 h. Values are shown after subtraction of BSY11G1 reference strain values. Values of the clones co-expressing TrCBH2+AnBGL1 are shown in red, co-expressing TrEG1+AnBGL in green and expressing only AnBGL1 in blue bars.

K. phaffii BSY11G1 was cultivated as reference strain in this experiment. The strain expressing only AnBGL1 was also tested to see the additive effects on glucose liberation when TrCBH2 or TrEG1 was co-expressed with AnBGL1. Interestingly, an OD_{340} signal was also detected in the samples containing AnBGL1 only. The measured OD_{340} value was a bit higher in the samples containing AnBGL1 and TrEG1. The highest measured OD_{340} signal was detected in the samples containing AnBGL1 and TrCBH2. It should be mentioned here that the measured OD_{340} values in all samples were very low which indicates that only tracers of glucose were detected. However, these results also indicate that the combination of AnBGL1 with TrCBH2 was more efficient in liberating glucose from Avicel than the combination of AnBGL1 with TrEG1. This outcome was expected since it has been reported in literature that TrCBH2 is more active on crystalline cellulose (Avicel) than TrEG1 [Henrissat et al. 1985]. This experiment gave strong evidence that TrCBH2 was functionally expressed. However, its expression or activity on Avicel seemed to be insufficient for enabling *K. phaffii* to grow on Avicel in this system.

A1.3 Secreted cellulases whose genes were expressed in an integrated whole-in-one vector containing all three cellulases in *K. phaffii*, were able to liberate glucose from Avicel:

Another approach for engineering a *K. phaffii* strain that is able to grow on cellulose was the construction of whole-in-one plasmid containing the genes *TrBGL1*, *TrCBH2*-HM, *TrEG1*. All genes are coding for cellulases from *Trichoderma reesei*. With such a plasmid, the genes coding for the three cellulases could be integrated at once into *K. phaffii* genome for heterologous protein expression. The construction of this vector (B1_S_Zeo_ADHtt_NB2_TrEG1_TrBGL1_alpha.sig.seq_NB2_TrCBH2-HM) is described in A.3 of the appendix part.

For this approach, *Pichia pastoris* (*K. phaffii*) CBS 7435 (NRRL-Y11430, ATCC6273) was transformed with this plasmid after linearizing it with *Swa*I. Positive transformants grown on YPD-Zeo plates were isolated for further cultivation in 96 deep-well plates. In 96 deep-well plates, *K. phaffii* cultures were incubated in 250 µL of BMD 1% at 28°C and 320 rpm for 60 h. Then, 50 µL of BMD 3.5% was added to the cultures and further incubated. After reaching stationary phase (after 12 h of further incubation) the supernatant of the cultures was harvested to test the expressed cellulases located in the supernatant for their ability to form glucose from Avicel. The procedure for glucose measurement is described in A2 of the appendix part.

Interesting clones were isolated for further tests. It was possible to isolate one clone whose expressed cellulases in the supernatant seemed to be able to form glucose from Avicel (see Fig. A3).

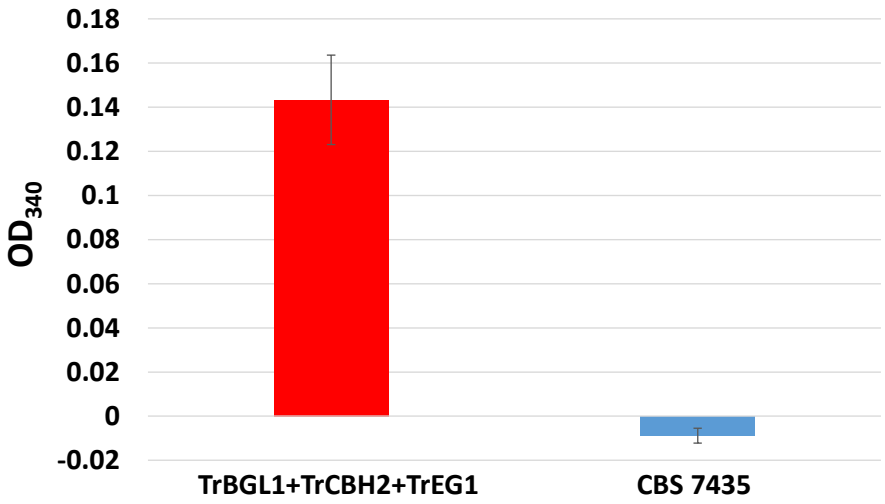


Fig. A3: Comparison of measured liberated glucose amount from Avicel by the enzymes in the supernatant of engineered the *K. phaffii* strain and refrence strain. Glucose measurement was performed after the samples were incubated at 30°C overnight. The strain co-expressing TrBGL1+TrCBH2+TrEG1 is shown in red, the reference strain CBS7435 in blue.

The measured OD_{340} signal in the sample with the supernatant of the clone expressing all three cellulases indicated that glucose was liberated from Avicel. However, the very low measured OD_{340} value also indicated that only traces of glucose were detected in this sample. This experiment showed that it was possible to express cellulases for cellulose saccharification in *K. phaffii* by integration of only one plasmid. However, it also demonstrated the limitation of this approach. By using only one plasmid, the gene copy number ratio of the genes TrBGL1, TrEG1 and TrCBH2 for protein expression was fixed to 1:1:1. Increasing the gene copy number of one gene to improve cellulolytic activity would also increase the gene copy number of the other two genes. This might not be efficient and is a strong limitation for engineering a *K. phaffii* strain that is able to grow on Avicel. Therefore, the approach to use two different vectors, one for expression of a β-glucosidase and a second one for co-expression of endo- and exo-glucanase seemed to be more flexible and more promising.

A.1.4 Establishment of targeting sites in expression vectors for chromosomal integration into *K. phaffii* genome:

For the approach to use two vectors for the co-expression of all three cellulases in *K. phaffii*, one vector was constructed for expressing *AnBGL1* and a second one for co-expressing *AnEG-A* and *TrCBH2-V09*. Both vectors had the same vector backbone (see vector construction in supplemental data of Part II of the thesis, Kickenweiz et al. 2018). Furthermore, both vectors were linearized for *K. phaffii* transformation by using the same restriction enzyme *Swa*I. As consequence, both linearized plasmids had homologous sequences at each end.

This could cause problems in creating stable strains. When both vectors were integrated by two yeast transformations in sequence, the first integrated vector could be replaced by the second vector during the second transformation. Even when both vectors were integrated in parallel during the same yeast transformation, it could happen that their integration was not stable. It happened that positive clones which had been isolated, lost the genes of one vector after re-cultivation steps. Since both cases were observed in first test experiments (not shown), this problem should be solved by establishing targeting sites. It has been reported that homologous sequences facilitate homologous recombination in *K. phaffii* the longer they are [Näätsaari et al 2012]. The targeting sites should facilitate the integration of both vectors into different loci creating stable strains. *Bgl*II was chosen as restriction enzyme for linearization of the vectors.

Isolation of targeting sites and testing their suitability. For finding suitable targeting sequences, the *K. phaffii* genome was browsed for natural *Bgl*II recognition sites. Loci containing a *Bgl*II restriction site which were not located within an annotated ORF, were amplified by PCR. With this PCR step, *Pst*I recognition sites were added at each end of the isolated targeting sites. The sequences of the targeting sites are listed in A4 of the Appendix part. Targeting sites were inserted into the vectors as described in supplemental data of Part II of the thesis.

K. phaffii BG10 was transformed with the vectors containing different targeting sites (pPpKan_intX_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt) to check their suitability. Positive grown transformants on YPD-Gen plates were isolated and incubated in 250 µL BMD 1% medium at 30°C and 180 rpm for 48 h in a 96 deep-well plate. After reaching stationary phase, samples for eGFP fluorescence measurement were taken

after vortexing the 96 deep-well plate. The eGFP fluorescence measurement was performed as previously described by Vogl et al. 2016. The results are shown in Fig. A4.

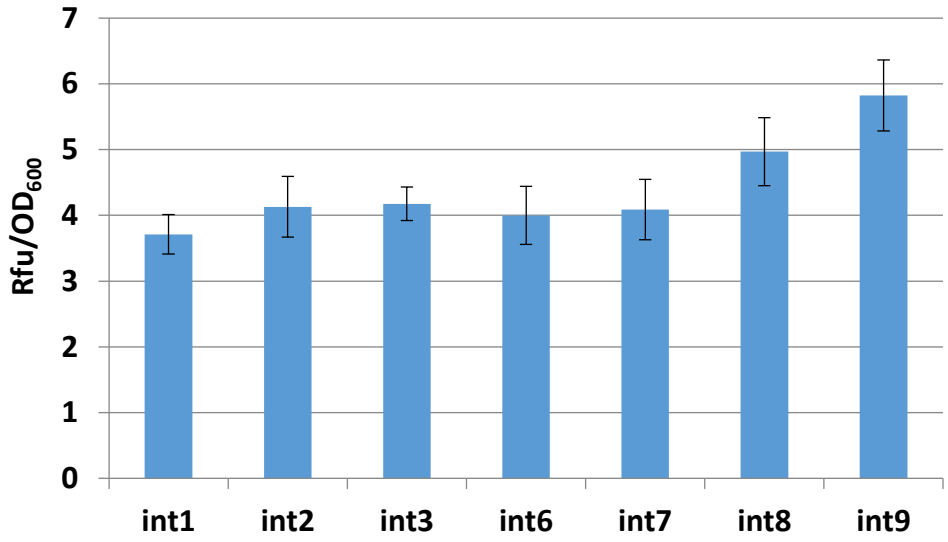


Fig. A4: eGFP signal measurement of *K. phaffii* clones which have integrated vectors with different targeting sites. The samples for eGFP measurement were taken after 48 h of incubation in BMD 1%. The fluorescence signal was measured at 488/507 nm (excitation/emission) and normalized to OD_{600}. A control strain with ectopic integration was not performed in this experiment since the direct comparison with the targeting sites was not required. This experiment primarily showed that the used targeting sites are suitable for gene expression since eGFP fluorescence signals could be measured indicating that eGFP was functionally expressed in all clones.

The yeast transformation with the vectors containing the different targeting sites (int1, int2, int3, int6, int7, int8 and int9) worked well. Furthermore, the eGFP signal measurement showed that eGFP was successfully expressed in all tested *K. phaffii* clones. This indicated the suitability of the isolated targeting sites for the approach to engineer a strain expressing all three cellulases. However, it should be mentioned that it was not checked if the vectors were integrated into the intended loci.

A.2 Methods:

pHBAH (*para*-hydroxybenzoic acid hydrazide) assay for measurement of reducing sugar ends:

After the cultures have reached stationary phase, the cells in the 96 deep-well plates cultures were pelleted at maximum speed (3,220 × g) for 5 min. 20 µL of the supernatant was used for the pHBAH assay. The pHBAH assay was performed as described by Mellitzer et al. 2012b with the only exception that 100 mM potassium-phosphate buffer (pH 6) was used instead of 50 mM sodium-citrate buffer (pH 4.8). The reason for the change to potassium-phosphate buffer was to adjust the experiment conditions more to the growth conditions of *K. phaffii*. 1% Avicel (w/w) and 1% CMC (w/w) were used as substrates.

Glucose UV hexokinase kit for measurement of glucose concentration:

After reaching stationary phase, the cultures in 96 deep-well plates were centrifuged at maximum speed (3,220 × g) for 5 min to pellet the cells. 20 µL of the supernatant was added to 150 µL of 100 mM Potassium-Phosphate buffer (pH 6) with 1% Avicel (w/v) in 96 microtiter plates. In parallel, 20 µL supernatant were added to 150 µL of 100 mM potassium-phosphate buffer (pH 6) without Avicel as control on the same microtiter plate. The microtiter plate was incubated at 30°C at 300 rpm overnight. After incubation, the microtiter plate was centrifuged at maximum speed (3,220 × g) for 10 min to centrifuge down the Avicel particle. The supernatant was used for measurement of the glucose concentration by using Glucose UV hexokinase kit from DIPROmed (Vienna, Austria). Glucose measurement was performed in small scale as described by Sturmberger 2013. Due to the small amounts of liberated glucose, 50 µL of the supernatant was added to 200 µL of reaction solution in a new microtiter plate. The samples were incubated for 10 min and the absorption of the samples was measured at a wavelength of 340 nm afterwards.

A.3 Vector construction

All components for the cloning work were from Thermo Fischer Scientific (Waltham, MA, USA) unless otherwise specified. The restriction enzymes were either from Thermo Fischer Scientific (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). PCR products were purified using Wizard SV Gel and PCR Clean-Up System from Promega

Fitchburg, WI, USA). All origin plasmids were from Pichia Pool1 of Graz University of Technology (Näätsaari et al. 2012) if not stated otherwise. Primers were ordered from IDT (Coralville, IA, USA). All primers and constructed vectors used for expression of cellulases which had not been described in supplemental data of Part II of the thesis, are listed in Table A2 in the Appendix part. DNA fragments were cloned into vectors by using T4 ligase unless otherwise specified. When DNA fragments were cloned into vectors using Gibson cloning, it was performed as previously described (Vogl et al. 2016).

1. B1_S_Zeo_ADHtt_NB2_TrEG1_TrBGL1_alpha.sig.seq_NB2_TrCBH2-HM:

Two plasmids were used for the construction of the plasmid expressing all three cellulases, B1_S_Zeo_ADHtt_NB2_TrEG1_TrBGL1_alpha.sig.seq_NB2_TrCBH2-HM:

1) T4_S_Zeo_intARG4_AOX1tt_TrCBH2-HM_NB2_alpha.sig.seq_TrBGL1_AOX1tt
2) B1_S_Zeo_ADHtt_NB2_TrEG1_PstIBmrIstuffer

B1_S_Zeo_ADHtt_NB2_TrEG1_PstIBmrIstuffer and T4_S_Zeo_intARG4_AOX1tt_TrCBH2-HM_NB2_alpha.sig.seq_TrBGL1_AOX1tt were digested with *Apa*I and *Spe*I. The B1_S_Zeo_ADHtt_NB2_TrEG1_ backbone was ligated with the AOX1tt_TrCBH2-HM_NB2_TrBGL1_AOX1tt fragment to create the final vector B1_S_Zeo_ADHtt_NB2_TrEG1_TrBGL1_ alpha.sig.seq _NB2_TrCBH2-HM.

1) Construction of T4_S_Zeo_intARG4_AOX1tt_TrCBH2-HM_NB2_alpha.sig.seq_TrBGL1_AOX1tt:

First, T4_S_intARG4_AOX1tt_TrCBH2-HM_BmrIstuffer_ alpha.sig.seq_TrBGL1_AOX1tt was created by ligation of 4 DNA fragments using Gibson cloning:

Fragment 1: The DNA fragment coding for TrBGL1_AOX1tt was amplified by PCR using the primers P13532 and P13533 and pPpT4_Zeo_pGAP_alpha_TrBGL1 (this vector was mentioned in supplemental data of Part II of the thesis, Kickenweiz et al. 2018) as template. With this PCR step, a *Spe*I recognition site was inserted between *AOX1*tt and the *Bam*HI recognition site at the beginning of the P_{ILV5}. Furthermore, overlapping parts were added in this PCR step for Gibson cloning.

Fragment 2: In a first PCR step, *TrCBH2*-HM was amplified from a plasmid used in Mellitzer et al. 2014 using the primers P13581 and P13582. With this step, an *Apa*I recognition site was inserted at the 3' end of *TrCBH2*-HM. The obtained PCR product was used for a second PCR step with the primers P13530 and P13531 to add overlapping parts for Gibson cloning.

Fragment 3: BmrIstuffer was amplified by PCR with P13543 and P13535 using pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-BmrIstuffer (TV0159 made by Vogl 2015) as template.

Fragment 4: The vector pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-BmrIstuffer was also used as backbone for the new vector. This vector was cut with BamHI and SbfI to remove the BmrIstuffer and the genes coding for stomato and eGFP.

After preparing all fragments, they were ligated with each other by using Gibson cloning to create the plasmid T4_S_intARG4_AOX1tt_TrCBH2-HM_BmrIstuffer_ alpha.sig.seq_TrBGL1_AOX1tt.

The BmrIstuffer in T4_S_intARG4_AOX1tt_TrCBH2-HM_BmrIstuffer_ alpha.sig.seq_TrBGL1_AOX1tt was removed by digesting the vector with *Bmr*I. NB2 promoter was amplified by PCR with the primers P13681 and P13682 using B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt (this vector was mentioned in supplemental data of Part II of the thesis, Kickenweiz et al. 2018) as template. In a next step, NB2 promoter was cloned into the cut vector backbone by Gibson cloning to create T4_S_intARG4_AOX1tt_TrCBH2-HM_NB2_ alpha.sig.seq_TrBGL1_AOX1tt.

2) Construction of B1_S_Zeo_ADHtt_NB2_TrEG1_PstIBmrIstuffer:

BmrIstuffer was amplified by PCR with the primers P13593 and P13594 using pPpT4mutZeoMlyI-intArg4-bidi-sTomato-eGFP-BmrIstuffer (TV0159 mentioned before) as template. With this PCR step, PstI recognition sites were added at both ends of the BmrIstuffer creating PstIBmrIstuffer. The PstIBmrIstuffer and B1_S_Zeo_ADHtt_BmrI_NotI_AOX1tt (the construction of this vector has already been described in supplemental data of Part II of the thesis, Kickenweiz et al. 2018) were cut with *Pst*I. After this step, B1_S_Zeo_ADHtt_BmrI_NotI_AOX1tt and PstIBmrIstuffer were ligated with each other to form B1_S_Zeo_ ADHtt_PstIBmrIstuffer.

This vector and B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrEG1_AOX1tt (the creation of this vector has already been described in supplemental data of the Part II of the thesis, Kickenweiz et al. 2018) were digested with *Swa*I and *Not*I. The obtained

B1_S_Zeo_PstIBmrIstuffer backbone was ligated with ADHtt_NB2_alpha.sig.seq._TrEG1 fragment to create B1_S_Zeo_ADHtt_NB2_TrEG1_PstIBmrIstuffer.

2. B1_S_Zeo_ADHtt_TrEG1_alph.sig.seq._NB2_TrCBH2-HM_AOX1tt:

For the construction of the vector co-expressing *TrEG*1 and *TrCBH2*-HM, the vector B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt (the construction of this vector has already been described in supplemental data of the Part II of the thesis, Kickenweiz et al. 2018) was used as background. This vector was linearized by cutting with *Bmr*I. In parallel, alpha.sig.sec._TrEG1 was amplified by PCR with the primers alpha_fwd and BmrI_rev using B1_S_Zeo_ADHtt_NB2_alpha.sig.seq._TrBGL1_AOX1tt (the construction of this vector has already been described in supplemental data of the Part II of the thesis, Kickenweiz et al. 2018) as template to add overlapping parts for Gibson cloning. For this PCR, Q5 polymerase from New England Biolabs (Ipswich, MA, USA) was used. The obtained PCR product was ligated with linearized B1_S_Zeo_ADHtt_NB2_TrCBH2-HM_AOX1tt by Gibson cloning to create B1_S_Zeo_ADHtt_TrEG1_alph.sig.seq._NB2_TrCBH2-HM_AOX1tt.

Table A2: Used primers for constructing the vector which were mentioned in Appendix part

Primer no.	Primer name	Sequence 5' → 3'
P13532	Alpha-sig_ BmrIstuffer-3'end_ole fwd	CTTCCAGCTTAAgagctcACTGGGactgatgagattcccatctattttcaccgctg
P13533	AOXtt_PIVL5-5'end_SpeI_ole rev	gaaacaagacattactgaaggatccACTAGTgcacaaacgaaggtctcacttaatc
P13581	CBH2_BmrIstuffer-5'end_ole_1st-step fwd	ACTGGGtgctatgattgtcggtatcttgactaccttg
P13582	AOXtt_intARG-3'end_ApaI_ole_1st-step rev	GGGCCCgcacaaacgaaggtctcacttaatc
P13530	CBH2_BmrIstuffer-5'end_ole fwd	CAGAGGTCggcgcgccACTGGGtgctatgattgtcggtatcttgactaccttg
P13531	AOXtt_intARG-3'end_ApaI_ole rev	cattctgcgagaaggtccCCTGCAGGGGGCCCgcacaaacgaaggtctcacttaatc
P13543	T.reeseiBgl1new rev	AACCGCCTTCACCCTCCTC
P13535	Bmrstuffer-3'end _ Alpha-sig 5'end_ole	cagcggtgaaaatagatgggaatctcatcagtCCCAGTgagctcTTAAGCTGGAAG
P13681	Cbh2_Natbidi2_ort2 fwd	caaggtagtcaagataccgacaatcattttgatttgtttaggtaacttgaactggatg
P13682	Bgl1(alpha-sig.seq.)_Natbidi2_ort2 rev	cagcggtgaaaatagatgggaatctcatTGTTGTAGTTTTAATATAGTTTGAGTATGAG
P13593	Bmrstuffer-5'end_ApaI_PstI fwd	gtctctCTGCAGACTAGTcagtCCCAGTgagctcTTAAGCTGGAAG

P13594	Bmrstuffer-3'end_SpeI_PstI rev	gtctctCTGCAGACTAGTcagtCCCAGTgagctcTTAAGCTGGAAG
-	alpha_fwd	gttcaagttacctaaacaaatcaaaATGAGATTCCCATC
-	BmrI_rev	CTAGGTTAGACTGGGTACATCTAAAGGCATTGCGAGTAGTAGTC

A.4 Sequences of targeting sites

Sequences of all targeting sites were isolated from *K. phaffii* BG10 genome and cloned into the vector pPpKan_SwaI_AOX1tt_SpeI_AnBGL1_alpha.sig.seq_NB3_eGFP_NotI_AOX1tt. *Bgl*II restriction sites are highlighted in red. Primers for amplification of the targeting sites are listed in Table A3.

int1, 1761 bp

GAAGTCGGTGTAAGATGTTCGTTGTATGAAAGTATCGGTTAGTACATTCTGTTTCACACCTCTTTCCTTATTTTG
ATAACATTCTCTTACGCAATCTTCATTCTTCAAGTATTAAATGATGCTCCCCTGATGCGCGTGCAAGTGCTACTA
GTCGCAACACGCTGTAACACCTAGGGGAGGGAAGTTTTAACTATTTGAGTCATGTTTTGGATGAATGCTGTTT
GGATAGGATAAACGGAAGCAGTTGTAAGTTGTAATGCGACCTGCTATAATCAAGGAGAGAAAAAGTTATATT
GAGACATAGTCTTGAACCTAAGAATATGAATCAAAAGTGAGGGCATAATTATTGTTTCCCCTCTGAACTTGGT
GAGTAACAGAGTAGGCATAGGATCTCCTCAAAATTATTAAATCTATTGGGGAATCTTGAACGAATCTCAAAAA
TACATTGGAAAGAATGTATACTTTGGTTTGTATTAAGTTATCTATAGTAGAGTTGGATTGCGAGGAAAGGATG
CAGTTTTTCCCTGAGCAGAGAGGTAGTCAACAAATGTCAAAAATTGACCAACAAAAAAAGATCTTCAAGGTGG
TTCTACTGACAATGTGCAGATCAATTGTTACTTTACTAACTATATTGGATCTAGTCGAAGGTTTTTAACAAGAAT
AGACTCCCATGAGACCGATAAAAGCCATAGTACGGCCGAAAATAAACAAAAGAATTACCAGATTTCTGAACC
GGAAATTAATGAACCACAGGCATGAAGTCATGCATATGCATGAGTATGCATAATTTGACCCTTGACAGATCCA
CCAGAAATCAAAACTCCATGAAGCCCCTGAAAAAAATTCGGCAATGTAACTAATGAATACTCCCACAATTACG
AAGTATACCGATTTTAGACCACATTTTTATCCTTGGTTAACATTACTTTACTGTAATCTTGAGGAAATCCACGAT
CCAAACTATTAAAACTATGTCTCGTTCTGCTATCAGACGTGTATGTCCTGCTTGCTTTGCTGTGTCTAAAAATGA
TTGAAGGAAGAGATGCACAATTTTTTGGCTACAACGCTATTTCACTAATAGCTTTGCCAACATTTGGTAACTTT
AGTTGAGTAGTCAGTTTTCTTGCTTCATTATAAATGGTATAGTTATATGAGACCCGCTTTATTTCTGGGAGTCCA
CCTGAATATCTTTAAGTGAATTAAACCCTCATCATACGTAGTTCAAATATCAAAAATTCTGTCTCTCAACCTGTA
AGCAGTACAAAATACATCACACCAACATATACTTAACATTATGTATTACGTGGACATGAAAATGGGACTTTAAA
CAATTCTTCTAGTGATGTTGAGCTCTAACTCATACGATGAGACGACCACTTCAGTTTCAGCGTCCCGCGCTTCA
AACAACAAGTGTGTGTCAGGATTTTGGGGGCTCTTGTCAACACGTAACTTGGGTTTGCCAGATTGCAGCTTAA
ACCATTGCTCAAATCTTGTGCCATCTTCTTGAGTGACATTTTAGTTCTGTTACCGAACTGGTTGAAAATGCAGA

GTAGAATGGGCTAGTTGGTACGGCATCCCTTGCTTAATTACACTACTCAGTCAACTTCAATCACCTTTTTTTATT
TCACTATATACAACTAAACATCCTAAGACCAGTTTATCTCACGGGATTGGATGTTTGTGTAAAAATATAGTATT
GATTAGGGAGCAGTAATTACGATCCAAAAGCATAACAGACTCTTACCCAATCTCTGAGGCAAG

int2, 964 bp

GAAAACCGGTGTCCAGAGAGATTCCTTACTGTTTTCTTAATGCTTGTGCGCAGAATCCTTTTACGAGGTAAAAG
AAATCTCTTTTGTTGGCTAAAATTCCAGAGGGAGCAGATAGAAAGGTAGAAGAAGAAGAAAGAGTTATTTCA
AAAATAGTGTACCAAGAAATAACAACGAAAAATATTCAACACACTACCTTTATGGAGTGCGTCACTTTTTGATT
GGAATATACTGAAATAAAGAATAAAGAATACAATAAAGCATATCCATTTGTTGTACGAATAATACGGTATAAA
CAACATTAGCGGTGTGGTATATTCCAACTGTTCTTACCTGTGTGAATGTATCTCAATTCAAATTTGCCTGATAAT
GGAAAAGAAACAATCTGGGTGTTATTCTGCGGTTGTTCAAAGATCTATTGCTCACCGCTCTCGGAATTACACGT
AACTCTGGAACAGAGTTCATCGAAGGCAGTAACTTCACAAAATTCAAAGCATCTTAGTTGACCTGGGTACAGA
TATGTTCTCGTAATCTACCGCAGTCTTCTAAGAATTTTGTACAAATCGAATGGGTATGTCCTGCTCCAAATTAG
GAACACAGTCACAGGAGACGCGATTTTTTTTTACCAGTCAGGAACTTGGGGTGAAGTCAACTCGGTATTGAGA
GAAACTCTAACTGCGTCTTTTCGAGCGGCCAATTGAATGACTTTCGTTCCAAACATTTTTGCATCCCGCTTTCAT
TACTAAAAATTAGCATTGTATTCAGATGCCTGTCCAAAATAGGGCGACTGAGCTGCAAAACTTTACACTTTTCT
ATTCTCTATTCTCATTTTCTGATTACATCAACTAAGCTCTTCTATTTTTGATTGCATGCATTGTTTGTCGGGCTTC
AACTTTATCCCGAAGCATTGGGACCCCCCGACCAAAGGTTGCTGAGATTCAGCGTCAGCAATAGATTCGTCGT
CACT

int3, 1277 bp

GACTTGATTGTCGAGATTGGTGAGTGGCTCTGACCTAAATGTTATGGGCAAAATCGCAGACCTTAGAATAAAG
CACCACAATCGCAATTATTAAATATACTTCTCTTTATACTTGGTGTCTCTGCTTATAAAGTAATTGATTGCAATG
TCTTTTTGACTTCGGCTTCCTGAATGATGTTTCAATACATTTTACTATTGAATTACCGCTCAACACTGGATCTCAT
TGGGGGAACATATAAACAAAAGAGTTAAATTAGAATAGTACTGTACTTTAGCTCGTATAAGTTTTAGTCAAAC
CTATGAAGGTCTATAAGCGAGGAATACATGATGACAAGAGGCTCTGATATACTAAAGAACTAATGAATGGGT
TCTTATCAGCAGCTTTTGCAAAAGAATCGAACCCAAATCATCTGAAATACAGATGTGATTCAATTGCGTACACT
TACTTACCTACGATGAAAATCAAAGGTGAACAAAGTTCTTGGGTAGAGTGAATGAATCCCCATCTTACATCCTT
CCAACCAGTAAAGACAGACGACGGTCGCGAATTCCAAGTTTACCGATAAAAGAGCAGATGCACAGAAGTCTA
TTCATCGGGACGTCATCTCTTGCAACCATGAACCTAAATTCAATTATGTGGCAAATAATTGAGGCATAAATTCA
AAGATGCAGTATTCAACATTTTTGAAGCTAAAATTGTCAGAAGATCTCCAATCTTGCTAATTGGTTTTTAGCTG
CAAACTTACCTATAGACCTGGTTGATGCAACAAATATATTATTAGCAAATATTTTAGGGTTTTGTGGAGTTAAG
AAAGTTATCAGATTTATGAGAACGTCGAGCATGATCACCTCCGGATGGTGTGAAGTCATAAATTAACCAGGCA
TTATAACCGATAAACAAACCCTTGCAGAAGTTTCTATAATGTTCGAAAATTACTGGTGTTCTTGGCATAAACCT
ATATCATACCAACCTTTTGATACTAATGCAAACTCACTTGTTTCCAAATGGTAAGTCTTGAAGTCTTTCAGAGAG
GAGCAATGCAGAAGCTTAGCTCCCAATGTCAAGTACTTAACTTTTGTCTCTGAAACCAAGTTTCCTCAAAGATA
AACTTGTTATTCAAGCTAGTAGGTACAACAAGTGTAAGCTGAAGCCGAAATCAAATACAAGGCAAAACGATCT
TTGAAATACCGTGGAAAAAGCAAAATCCCCCTGGAAGTTACAAACCACTTGCTAATCTCCCAATATAGCTGTG
CATAATCACCATCTGGCTAGCAGA

int6, 947 bp

TTCTCTGAGTCTGTTGCTTGGATCTGTGTTGATTTTTCAACAACTCGTCATCGAATAGTTGCGGGAACAGAACG
ACGAAACGTTCTCTGTCCCAGAACTTATTACAATTTTGAAAAATGAAAAGGAAAAAGAATGACAAAACTGTGT
CCATGGTTACCATGGCTGGCGAAATGACCCAACATAGTGCCTACATCAAGACGCAACTATCACAACAATAATC
ACTAACCACGAAACTTAATCAGCCAAAGAGAAAAGCATACCACTCTGGGCCCACAAAAAGAAACAACATTACA
ATTCACCGCAGGTACGTCGAACTCACAATAACGAAAACACCGAGGAAAAAATTAATCATGTTAAATTTGTTCC
ATGCAGATCTGTCAGGTAGGTATGGATACCAGCCTATTGCGAGCTCTACAGACCCGCTAAATCTTGTCAGTGA
TCAAGACTCAACCTCGGCCGCTGTAAAACTTCTAGATTTTAGCGCAAAAGAGGGAAAAAAAAAAAGGAGAGC
AAATTAATGCGGCAAATGCACCGATGACCAGATTTGATGCACCAAATGTACACCGCGACGCAATTAAATCTGC
AACTATTCCTAGCAAAAACAATAACTTATTCAATTGCAGGGCTGTAAGGATTGTCCTCGAGGCGCTACTAAAA
CTGTGACTGCCTGCATGCTAGTCACCAGCCCTACCCACGAGCCGAGTAATAACCGTGGGCAACATGAAGTAAA
TAACCGAAACTCAACAGAAACGGCTGTCAAAGCTTATAACCAGCGGGTTTAGACCTCCTTCGCAGGAGGGTCC
AATAGGGAAGGGTTGATTCAAGATCAAAAGTAGTGTATTTGTCATCAATTTTGAAGATAAGCACATTAGAGAG
ACGCATATGTAGCAGTAATTCGATGAAACAAAGCTATTGAGGCGGTGCTCAAAGCCCGCTTCAATAGAACG

int7, 1338 bp

AGTCGAACATGGTCACTATTGCACTAGGCCTGAGTATTGGTAATAATAGCACTGTAATATCAATCATCACGTG
AAATTATACAAGTAAGGACGGACAAAAAGGGAACTTTAGTAAGGACAACCGTTTAATGGTAGGTACACTAAA
GATGTTAGATGTTGGTATGCGAGAAATTAAGGATTGATAACCAGATCGTTAAGCCAAATATGTATTGCGACTT
GCTTTTTCAAGACAGGGATGAATAGGAACAGAGGCAAAGACCCATAGCTTGTGAATATTGACTGAATGTGAG
CAGTTTATGCACAGGAAAAGAAAGGAAAATTTGTCAAAGAGATAGCTTTACTGCGTTGCTATCCTGTATTTCA
GTGATCGGCAAACCTGCATGTTCCTGTCAACATAAATCTTAGCTCCGGCATCACATCGAAAGAACACACATTG
GTATCACCAACTGTTGGAACAACCTAGGAGAGCTGGTACTTCAGTTTTGATCGATTCCATCACTGAATGTAGCC
ATGTAACCATGTTACCATCGTCGATACAGTTACACCATCATGAAGCAAACCTTTTTCAATCCAAGTAGAAAACG
TCAACTTTACAAAGATCTACTTTGTCTAGCTGCTAGATCCACGACGGCTGATGTGCAACTTTATAATGGCGTTG
GCACTCTGAATGGTTAGTCATAATACAGGAAGATCAATGTGTAGCAAATTCATGGAACTATCAGATAGATGCA
ATAGAAGAGTAGCCGAGCTCAATTAGCCCCAACCACGATGTGAAGTCGTTGAGAAGAAGAAGGCACAACAA
GATGCTGGGATGAGCTCGTCAAATATACAATCAAGAATTTGTGAAATGCAAATCATGATCTCGTTGTAGTTGT
TAGGTTGCAACTGTTGCTAAAATCGAGCTCATCCTCCCAAGTTTACCTGAACACCGTTCTTCAAGATAAGGGCT
TTGACAGTAGTTGTAAACTAAATTGGCTCAATTAAAAGCATTCTTACTCCGTTGTCGTAAGTTACTGAAGTAGT
GGGGACGTCTTGCTTCTTGTATTTTCAGTTCCCAGTTGCCAAGTATGAAAAGTTTGTAATGTCATCCGATTTTCA
TTCAGGCACTGAAACTAAATTCTTACGTCTTATCACATTGAATTAGAAAGACTGCACTTTACACATTGTATTCCC
TCTGTGTGCTTCAAAGACACAAAAGGGGCAAAGACTGGCATAAGTACAACCTGTCGAACCAACATTTGCCGG
CCAAAAGGATAACATTCCCAAAGACCTTTCCTTTCCCTTCGCCTAGCAAACCTATTTGCAAACCTATTTGACAAC
GAATCGTTCTTTGTGCAG

int8, 1290 bp

CAATCATCCACTCTGCCTCTCTGTAAGTGAACAAGAAAAAGAAAATAATTCATGCAAGTACTCTGGAAGTAGT
ATCCGAAAAAAGCTGTTATCTGTTGCAAAAAACCCAAGATGCAGACTCTGGTCCAATGCGGGGGGTGTTAGCT
GATATCCATGCCGGATGCAGGCAGTGACAATAAACGTCGTCTCGGCTACACTATATGCATGCAAGAAAAAGT
AAAAAAAATAAATCAAGGAACATCATCATGCGTCGAGAAAACTTGAAGATCAAGTCATTGAATCGATAATTAC
CAAAAAGAGGAGGTCATGTTACAAACAGGAATGACTCAAAACGATGCGTTGCCGATGTTGTCTACTTCGAAG
CAAAAAGAAGATGAGAAACTAAATTATGTTACAAATCTAATTGAGTAGTTTGTCCGCAGAATTTTCCCACTAAT
TGCAAACCTCTCTTGATAACCGTTGCTGCTCACAAGTTTTTCGGAAACCCGAAGCAAACAGCTACATGAGTCGT
TTGCCAACCAAAAATGAAGAAGGTCAAGATCTCACTTTTTTTCCAGGCGATATGCGGGTAAATGCAGAAGCAT
GTAAGATTAATTACCGATGTACTTTGCTGCAACTAATTGTTTCATCGAACATTCCTATTGCAAGTCACTATTATT
TCTGTAGGTCGTTCACATTTTGTGGCTATAATCGTATAGAGATAGCTATTTTGGTGATAGTAAATATGCACTTC
AACTTCCTGATGGCCTGTTATTTTGGTTTTATAATTGTGATTCCGAATCCAGTCAGGAGATTGATGAGATTATA
ACCTTTTTGGGGTTGGCAAAGTCTTGAGGAGGTGCAAAAACGGAAAGAACAAAAGGAATACATATTATGCTG
AAACTGAAGTGTACAGATGGATGCCTCGACTCCGAACAGGTTTAGCTGGTTTTGGCTGTTTTGATGCTCTTCTG
CCTTAACTTGGCATTCTTTGCTATTATCGTAGCTACACCTCAATCCCGTGTACGCATTCTTTCAAATAAAGGAAA
ATGCAGCTTGAGCAAAAGTTTAGTGGCAATTTTCTGATGGAGGTAGTTCGATTCGTTAGATTATTTTCGCTCAG
AGTTCTCGTCTACATTAATCTCTATCACTAGTAGTAGTCTTTGTTGAATTACAGTTCGTCTAGTATATCCTCAGA
ACATTTAGTTGTGTTTGGAAAGAAGAAAGGTGGGGTAATTTGATTATGGCTTTGTGTTTTCACTATGCCAAGCT
GTAATAAGCCCGTAGCCTTGGGCATGAAAGCACTATCCT

int9, 1023 bp

AGGACCTAATGGCTAGAAGTTGCTTAAGAATGAGTTCAAACAGCATGCTGAGCAGCGTACCTACTCGAGAGA
GAACCCTTTCCAACAAACAAATCTTCCCCTTGAGACTTATTAGTCACTTTGTGAGGATTCGTGTCATGAATGCTC
CCCCACCCTCCACCTCCTGGAGTCATCAGCACAATTCGATCTCCTGGAATCATATCAACAGTATTTTTTCCTCCT
AAGTTGTAATTTTTGTATCCCATAATGTTCCCTTGAAGATCCAAACTTTGTCTCTTTAAAATGTTCAGACCTCTTG
CACCGCAGCCACCTCCTTTCAAGCCAAATGGGGCATTGACTCGTCGTTCCGACAAAATAGAAGCCTTCACTGG
TATACGGAATTCAATCTCTCGGACCACTCCACAACCTCCCACATGAAGGCCTTGCCCTCCGGAACCTGGCCTGA
CACTAAATTCATGAACAATAGCAGGGTAGCGCTTCTCAAAGATCTCAACATCAGTCATCCTAGTGTTAGTCATA
TTGGTATGAACTGCATCAGTACCGTTCCAACCAGGCCCAGCTCCGTGCCCACCGGCAATAGTTTCGTAATATGA
ATTAGTCTTCCTTGCTCCTTGAGGAAGATTGTAGCTCGACCCAAAGGTGAAATTGTTACAGTCACCTTGAGAG
GCGGCCATAACTCTTAATGTTTTCAATATCACATCTGTGACCCTCTGTGAAGTGATGACGTTCCCTGCTACTACA
GCACATCCATCATCAGGGCAAAGTATACTGCTTGGCGGAATGTTGACCCGAACTGGCTTAAGACATCCTTGAT
TCAACGGAATGTTGTCATCCACCAAGCAACGAAGACAATATATAATTGCGGAATAGGTAATAGACTCCGGTGC
GTTCAAGTTTCCATAAACTTGATGCGAAGTTCCCGAAAAATCAAAAATGGTTTCTTCTGTCTTTAAGTTATAAG
AGATTCGTAAATGAATCTGAGAACCATCGTCTAAAAATTCAGAGCCCTCTGCATAAAGGATG

Table A3: Used primers for amplification of the targeting sites

Primer No.	Primer name	Sequence 5' → 3'
-	integrsite1 fwd	ACTGCTGCAGGAAGTCGGTGTAAGATGTTCGTTG
-	integrsite1 rev	ACTGCTGCAGCTTGCCTCAGAGATTGGGTAAGAG
P39	integrsite2 fwd	ACTGCTGCAGGAAAACCGGTGTCCAGAGAGATTC
P40	integrsite2 rev	ACTGCTGCAGAGTGACGACGAATCTATTGCTGAC
P41	integrsite3 fwd	ACTGCTGCAGGACTTGATTGTCGAGATTGGTGAG
P42	integrsite3 rev	ACTGCTGCAGTCTGCTAGCCAGATGGTGATTATG
-	integrsite6 fwd	ACTGCTGCAGTTCTCTGAGTCTGTTGCTTGGATC
-	integrsite6 rev	ACTGCTGCAGCGTTCTATTGAAGCGGGCTTTGAG
-	integrsite7 fwd	ACTGCTGCAGAGTCGAACATGGTCACTATTGCAC
-	integrsite7 rev	ACTGCTGCAGCTGCACAAAGAACGATTCGTTGTC
P43	integrsite8 fwd	ACTGCTGCAGCAATCATCCACTCTGCCTCTCTG
P44	integrsite8 rev	ACTGCTGCAGAGGATAGTGCTTTCATGCCCAAG
-	integrsite9 fwd	ACTGCTGCAGAGGACCTAATGGCTAGAAGTTGC
-	integrsite9 rev	AgTcCTGCAGCATCCTTTATGCAGAGGGCTCTG

References

Henrissat B, Driguez H, Viet C and Schülein M (1985) Synergism of cellulases from Trichoderma reesei in the degradation of cellulose. *Nat* Biotechnol 3: 722–726

Kickenweiz T, Glieder A, Wu JC (2018) Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase. Appl Microbiol Biotechnol 102: 1297–1306. doi: https://doi.org/10.1007/s00253-017-8656-z

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506-577 doi: 10.1128/MMBR.66.3.506-577.2002

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

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

Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, Purkarthofer T, Glieder A (2014) Synergistic modular promoter and gene optimization to push cellulase secretion by *Pichia pastoris* beyond existing benchmarks. J Biotechnol 191: 187–195. doi: https://doi.org/10.1016/j.jbiotec.2014.08.035

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* 7: e39720. doi: https://doi.org/10.1371/journal.pone.0039720

Sturmberger L (2013) Master's Thesis: Characterization of synthetic regulatory elements for protein expression in *Pichia pastoris*. Graz University of Technology, Faculty of Technical Chemistry, Chemical and Process Engineering, Biotechnology

Vogl T (2015) Doctoral Thesis: Synthetic biology to improve protein expression in *Pichia pastoris*. Graz University of Technology, Institute of Molecular Biotechnology

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5: 172-186. doi: 10.1021/acssynbio.5b00199