Identification and characterization of novel promoters in *Ralstonia eutropha* H16

Diplomarbeit

durchgeführt von

STEFFEN GRUBER

Institut für Molekulare Biotechnologie

der Technischen Universität Graz

Leiter: Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Mag.rer.nat. Dr.rer.nat.Petra Köfinger

Graz, März 2012

Acknowledgement

First of all I want to thank Prof. Helmut Schwab for his support and for giving me the opportunity to successfully complete my work. My personal gratitude goes to my supervisor Dr. Petra Köfinger for her friendship and perpetual assistance in scientific and personal matters. It was a pleasure to be a part of team “Knallgas”. Furthermore I would like to thank Daniel Schwendenwein for his friendship and for introducing me to this project. I also want to thank Zalina Magomedova, Karina Treppe and Elisabeth Dornisch for their friendship and help during this work. Finally I want to thank my parents and my grandparents for their financial and more importantly their personal support at all times.

Abstract

Ralstonia eutropha H16 is a gram-negative, strictly respiratory facultative chemolithoautotrophic bacterium which uses H_2 and CO_2 as sole energy and carbon sources in absence of organic substrates. *R. eutropha* H16 has attracted significant interest for its ability to metabolize heavy metals, to degrade a variety of chloroaromatic compounds or chemically related pollutants. It was also found to be a good candidate for polymer production. The production of the biodegradable polymer polyhydroxyalkanoates (PHA) for example was applied in *R. eutropha* H16 on an industrial scale. *R. eutropha* H16 does furthermore serve as a model organism for mechanisms involved in control of autotrophic carbon dioxide fixation, hydrogen oxidation and denitrification. On the basis of this the use of *R. eutropha* H16 as a production organism under environmentally friendly chemolithoautotrophic growth conditions is favored. However, an inducible expression system is yet required for gene expression in *R. eutropha* H16. The aim of this work was to identify inducible expression systems originating from *R. eutropha* H16 facilitating heterologous gene expression. The main focus was hereby set on the identification of promoters that exhibited activity either under lithoautotrophic or heterotrophic growth conditions in *R. eutropha* H16. Two approaches were defined to generate a set of promoters based on the genome of *R. eutropha* H16, random and direct approach respectively. Numerous putative promoter regions derived from the genome of *R. eutropha* H16 were hereby identified. Based on comparatively high promoter activity ten promoter regions derived from *R. eutropha* H16 were identified of which the most active promoters were characterized in more detail.

Zusammenfassung

Ralstonia eutropha H16 ist ein gram negatives, streng fakultativ chemolithotrophes Bakterium, dass H_2 und CO_2 als einzige Energie- beziehungsweise Kohlenstoffquelle in der Abwesenheit von organischen Wachstumssubstraten nutzten kann. *R. eutropha* H16 besitzt zudem die Fähigkeit eine Reihe chlorinierter Aromaten, sowie strukturell nah verwandter, umweltschädlicher Substanzen abzubauen. Weitere Eigenschaften von *R. eutropha* H16 sind das Metabolisieren von Schwermetallen und die Biosynthese des biologisch abbaubaren Kunststoffpolymers Polyhydroxyalkanoat (PHA). Die Produktion von PHAs im industriellen Maßstab mit *R. eutropha* H16 als Wirtsorganismus ist bereits etabliert. *R. eutropha* H16 dient als Modellorganismus für die Beschreibung essentieller Mechanismen bei der autotrophen CO_2 Fixierung, der H_2 Oxidation und der Denitrifikation. Die Fähigkeit von *R. eutropha* H16 unter chemolithoautotrphen Bedingungen zu wachsen macht den Organismus für die Biotechnologie als Produktionsstamm interessant. Um eine gute Expression für homologe und heterologe Proteine zu erreichen ist ein induzierbares Expressionssystem von Vorteil. Da für *R. eutropha* H16 kein effektives und induzierbares Expressionssystem beschrieben ist war die Identifizierung eines induzierbaren Expressionssystems auf der Basis des *R. eutropha* H16 Genoms das Ziel dieser Arbeit. Um passende Promoterregionen zu finden wurde einerseits eine Promoter-Bank angelegt und andererseits wurden gezielt Promoterregionen von Genen, die unter litoautotrophen Bedingungen hochreguliert sind, verwendet. Im Zuge dieser Arbeit konnten eine Reihe von Promoterregionen auf dem Genom von *R. eutropha* H16 identifiziert und charakterisiert werden.

Table of contents

List of figures

Table directory

1. Introduction

General introduction

The chemolithoautotrophic, gram negative bacterium *Ralstonia eutropha* H16 was isolated almost 50 years ago from its natural soil habitat [1]. Ever since *R. eutropha* H16 attracts significant interest for instance due to its facultative chemolithoautotrophic growth abilities or its polyhydroxyalcanoate (PHA) production capacities .*R. eutropha* H16 is categorized as part of the burkholderiaceae family in the order of burkholderiales which contains a variety of diverse microorganisms isolated from various ecological habitats. Like many other members of the burkholderiaceae family *R. eutropha* H16 possesses a multi replicon genome [2, 3]. The multi replicon genome of *R. eutropha* H16 consists of two chromosomes and a megaplasmid. It was shown that the megaplasmid maintains self-transmissible properties which enable the propagation of metabolic attributes like chemolithoautotrophy amongst related Ralstonia strains [4, 5]. Structurally closely related megaplasmids isolated from different lithoautotrophic Ralstonia strains also suggest a shared conserved structure of the megaplasmid [2]. In case of the 452156 bp long megaplasmid pHG1 originating from *R. eutropha* H16 variations from the conserved structure of megaplasmids in Ralstonia strains include for instance the presence of genes responsible for aromatic compound decomposition or genes facilitating growth on different nitrogen based compounds [6]. The Hox operon coding for the hydrogen oxidation apparatus is also found on the megaplasmid pHG1. Other essential metabolic functions like genes promoting general housekeeping such as replication, translation or transcription are instead encoded on chromosome 1 (4052032 bp) of *R. eutropha* H16 [7, 8]. Chromosome 2 on the other hand is 2912490 bp long and encodes genes for a various number of alternative metabolic pathways including genes for pathways enabling the use of alternative carbon sources [2, 8].

Lithoautotrophic and heterotrophic growth of *R. eutropha* H16

Supported by these findings recent genome sequencing projects indicate that *R. eutropha* H16 is a fairly versatile microorganism [6, 9]. Among better known properties are for instance its denitrifying capacities [10, 11] or the ability to synthesize considerable amounts of polyhydroxyalcanoates (PHAs) as intra cellular carbon and energy storage source, a quality for which *R. eutropha* H16 received notable attention [12, 13]. In addition genes coding for enzymes being responsible for the mineralization of aromatic compounds were identified on the genome of *R. eutropha* H16 implying the use of aromatic compounds as growth substrates [6, 14]. Next to all these various abilities *R. eutropha* H16 is also known to be a facultative chemolithoautotroph. The organism is able to solely grow on molecular hydrogen as an energy source and CO_2 as carbon source. The ability to take up and utilize H_2 is facilitated by two NiFe hydrogenases [7, 15], a dimeric membrane bound hydrogenase (MBH) located at the cytoplasmic membrane and a soluble hydrogenase (SH) present in the cytoplasm of *R. eutropha* H16. A complex hydrogen sensing system is employed to activate the expression of MHB and SH operons located on megaplasmid pHG1 [15]. The hydrogen sensing system relies on a regulatory hydrogenase to activate a histidine kinase in the presence of H_2. The histidine kinase again interacts with HoxA, a transcription factor of the NtrC family, to trigger the expression of MHB and SH operons [6, 7]. SH is then deployed to reduce NAD^+ to NADH in the cytoplasm of *R. eutropha* H16 whereas MHB is mainly involved in electron transport coupled phosphorylation. The assimilation of CO_2 under chemolithoautotrophic growth conditions is accomplished by the enzymes of the Calvin–Benson-Bassham cycle includingribulose-1,5-bisphosphate carboxylase oxygenase (RubisCO) as the CO_2 fixing enzyme. The cbb operon codes for all enzymes of the Calvin–Benson-Bassham cycle and is located on chromosome 2 [16]. A second, partially incomplete copy of the cbb operon sharing high homology with the chromosomal cbb operon can also be found on the megaplasmid pHG1 [6, 16]. On the contrary the genetic information for growth under heterotrophic conditions is found to be distributed over the entire multi replicon genome of *R. eutropha* H16. Based on the variety of genetic information a range of growth substrates is accepted under

heterotrophic conditions. This includes fructose, gluconate, some aromatic compounds and several organic acids [14, 17].

The use of expression systems in *R. eutropha* H16

The ability of *R. eutropha* H16 to adopt different growth strategies and hereby respond to environmental influences is based on an effective regulatory system, which enables the cell to adapt its metabolism accordingly. Regulation of gene transcription in prokaryotes is generally mediated by the promoter region (see figure 1). The promoter region is located upstream of the initial transcription start and directly controls gene transcription under the influence of the RNA polymerase (RNAP) and various transcription factors [18, 19]. Main motifs that are involved in the regulation of gene transcription related to the promoter region are the -10 hexamers and -35 hexamers, which are located 10 bp or 35 bp upstream of the transcription start. Both hexamers are recognized by the sigma factor, a subunit of the RNAP. The -10 box is recognized by domain 2.4 of the sigma factor whereas the -35 box is recognized by domain 4.2 (see figure 2) [20]. The closer the nucleotide sequence of these hexamers is to the consensus sequence the stronger are the interactions with RNAP, leading to enhanced transcription [21]. Another important part of a promoter is the -10 extended element, which is 3 – 4 basepairs in size and located directly upstream of the -10 hexamer interacting with domain 3 of the sigma factor (see figure 2) [22].The UP-element is a region located approximately 20 bp upstream of the -35 box that is also recognized by the C-terminal domains of the RNAP. It has been shown to increase promoter activity for some promoter regions [23, 24]. All these regions together confine the binding region of the RNAP to the promoter. The binding properties of RNAP are influenced by the sum of interactions induced by the UP-element, the -10 box, the -10 extended box and the -35 hexamer and vary from promoter to promoter. Other elements influencing transcription are transcription factors. In general transcription factors are proteins that possess the ability to interact with specific DNA sequences to influence the transcription of genes [25]. Transcription factors can act as activators promoting or enhancing transcription for instance by improving the recruitment of RNAP. On the contrary transcription can also be repressed by

transcription factors for example by steric hindrance or looping of DNA. In either case recruitment of RNAP to the promoter region is prevented [26, 27]. Another element that plays a role in protein expression is the Shine-Dalgarno sequence. The Shine-Dalgarno sequence is a ribosomal binding site on the mRNA, that is encoded approximately 8 bp upstream of the startcodon ATG and intensifies translation levels by recruiting ribosomes [28]. The closer the sequence of the Shine-Dalgarno sequence is to its determined consensus sequence (AGGAGG)the more enhanced is the translation process [29]. Consequently the presence and conformation of a Shine-Dalgarno sequence have a significant impact on protein expression [30]. However, the predominant way to conduct major changes to transcription patterns of a cell is accomplished by sigma factors. Sigma factors act as dissociable promoter recognition parts of the multisubunit RNA polymerase holoenzyme [31]. Depending on the type of sigma factor that is attached to the RNAP holoenzyme, the binding properties of RNAP and also its transcription patterns of genes are alternated. As mentioned in the abstract before the sigma factor recognizes the -10 hexamer, -10 extended box and the -35 hexamer of a promoter sequence (see figure1) [18].

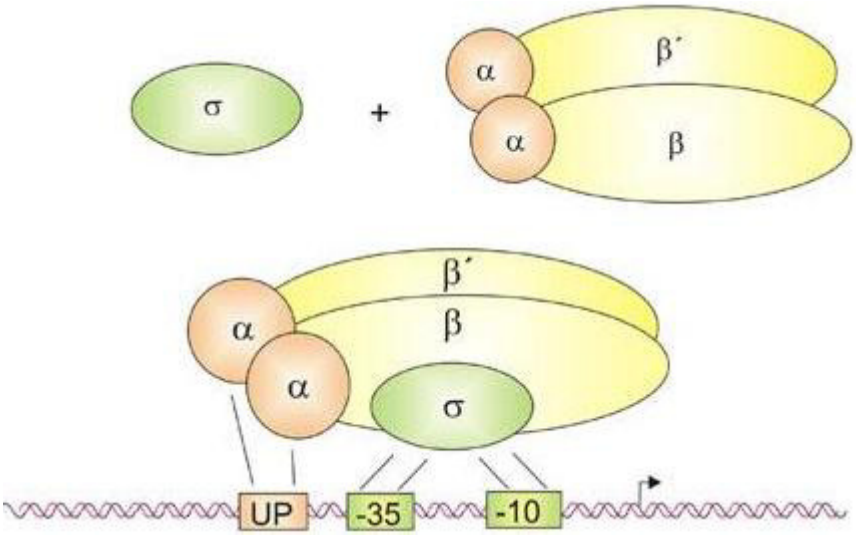

Figure 1: Illustration of a transcription unit including a σ^{70} based promoter. Subunits $\alpha_2\beta\beta'\omega$ of RNAP and a sigma factor are shown, as well as the UP-element, -35 hexamer and -10 hexamer. The arrow is indicating the transcription start point [18].

However, sigma factors are only active when attached to RNAP. The core structure of the cell's RNAP is always alike and consists of $\alpha_2\beta\beta'\omega$ subunits. The ω and α subunits play a role in general RNAP assembly, whereas the second α subunit enables binding of RNAP to regulatory factors like for instance enhancers, β' subunits enable nonspecific binding to DNA and the β subunit possesses the polymerase activity [32, 33].The sigma factor is employed in order to complete the RNAP holoenzyme. Most bacterial strains contain a pool of multiple sigma factors that compete for a limited amount of free RNAP. The composition of the sigma factor pool can be influenced based on regulated synthesis, degradation, activation or deactivation of sigma factors [34]. The composition of the sigma factor pool hereby influences the composition of the RNAP pool which in turn affects the transcription pattern of the cell [35]. The sigma factor σ^{70} for instance is responsible for the regulation of most housekeeping genes expressed during exponential growth in a majority of bacterial strains. RNAP holoenzymes containing σ^{70} are for instance responsible for the transcription of the lac operon or the phosphate sensing operon in *E. coli* [36]. The binding pattern of σ^{70} holoenzymes to the corresponding promoter region is mainly based on the amino acid sequence of σ^{70} which facilitates interaction of σ^{70} regions 2.4, 3.0 and 4.2 with the -10, -10 extended and -35 region of the promoter DNA upstream to the transcription start (see figure 2) [19, 37]. The polymerase activity of σ^{70} holoenzymes is initiated upon binding to the DNA promoter regions and in some cases also depends on the presence or absence of regulatory transcription factors.

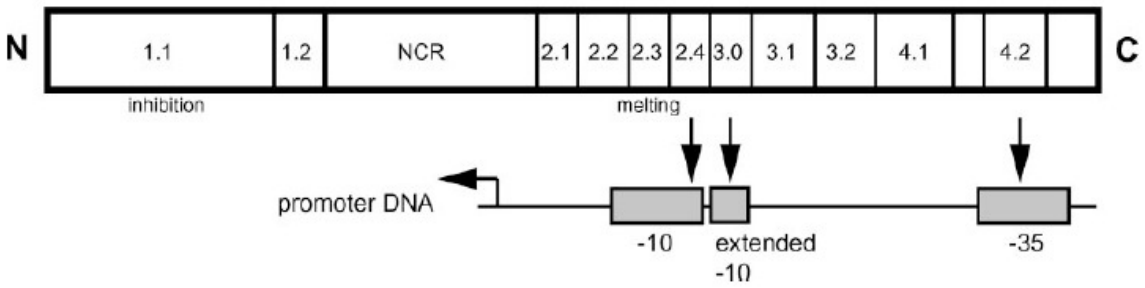

Figure 2: Illustration of the interaction of sigma 70 with -35, -10 and extended -10 DNA binding domains. Binding domains 1 to 4 and their interaction with the DNA strands are shown. Interactions of sigma 70 domains with DNA are indicated by arrows. NCR relates to a non crystalline region of the sigma 70 factor [19].

The main promoter DNA sequence motifs involved in the σ^{70} holoenzymes binding process are the -35 region and the -10 region, also referred to as Pribnow box [37]. Based on sequence analysis of σ^{70} promoter regions across different bacteria strains, consensus sequences for the -35 and -10 region could be defined. Consensus sequences of other important sigma factors like σ^{32} and σ^{54} were also defined (see table 1) [38, 39].

Table 1: Listing of promoter consensus sequences related to the corresponding sigma factors σ^{70} σ^{32} and σ^{54}.

σ Factor	**Promoter consensus sequence**	
	- 35 region	- 10 region
σ^{70}	TTGACA	TATAAT
σ^{32}	TCTCNCCCTTGAA	CCCCATNTA
	- 24 region	- 12 region
σ^{54}	CTGGNA	TTGCA

As a response to environmental stress caused by heat shock σ^{32} initiates transcription of genes encoding for chaperones or enzymes involved in DNA repair [40], whereas σ^{54} employs transcription of genes required for nitrogen utilization [41]. Sigma factor σ^{38} on the other hand triggers gene expression as stress responses of the cell to reply to oxidative stress, UV-radiation, pH shifts or nutrient deprivation [42]. In case of *E. coli* a set of six sigma factors and furthermore 350 transcription factors were identified to regulate the cell's transcription pattern [43]. Efficient transcription regulation is also known in *R. eutropha* H16 for instance concerning the expression of the cbb operon which is tightly regulated by the cbbR repressor molecules blocking P_{cbb}, a σ^{70} similar promoter, during growth on organic substrates [16]. Other regulatory systems have for instance been described related to PHA synthesis relying on regulation of P_{phaC} or P_{phaP}[44] or the activation of NiFe hydrogenase expression under lithoautotrophic growth conditions of *R. eutropha* H16 [45]. In order to promote heterologous gene expression in *R. eutropha* H16 regulatory expression systems as described in the abstract above are required. However, all of the expression systems native to *R.eutropha* H16 underlie regulatory mechanisms that are not favorable for heterologous gene expression under particular growth conditions of *R. eutropha*

H16 or share complex inducing mechanisms; which excludes them from the use as a simple expression system in *R. eutropha* H16. Alternatively expression systems on the basis of lac and tac promoters are already successfully applied for expression of heterologous genes in *R. eutropha* H16 [14, 44]. However, due to lacking transport capacities for inducers like IPTG or lactose neither P_{lac} nor P_{tac} can be used as inducible expression systems in *R. eutropha* H16 [44].

Par region

pKR-tac-PAR vectors encoded the partitioning system *par* derived from the broad host range Birmingham IncP-alpha plasmid, which was introduced to enhance the stability of the low copy number vector pKR-tac in *E. coli* cells [46, 47]. The par region stabilizes plasmids in a variety of gram-negative bacteria and consists of *parA1, parA2*, *parB*, *parC*, *parD* and *parE* encoding a complex site-specific recombination system that prohibits the formation of plasmid multimers (see figure 3). The par region has been shown to promote correct plasmid distribution during cell division and preventing the formation of plasmid multimers. Furthermore the par region was already used in pBR322 and pACYC177 low copy number plasmid where the stabilization effect of the par region was demonstrated. It was shown that pBR322 and pACYC177 did not exhibit plasmid loss under nonselective growth conditions after 200 generations in *E. coli* [46-48].

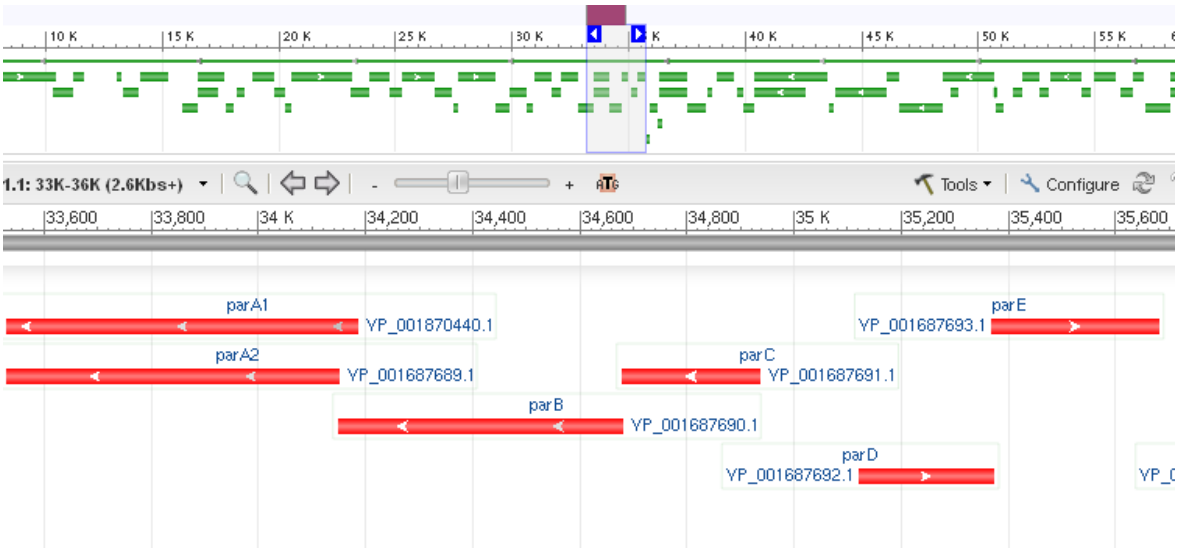

Figure 3: Illustration of the par sequence encoded on the Birmingham IncP-alpha plasmid showing *parA1, parA2, parB, parC, parD* and *parE* genes. Basepairs 33166 to 35781 are shown of the Birmingham IncP-alpha plasmid.

The aim of this work

In order to promote inducible heterologous gene expression in *R. eutropha* H16 the aim of this work was to identify and characterize parts of gene regulatory systems originating from *R. eutropha* H16. The main focus of this work was set on the search for promoter regions that were active only under lithoautotrophic or heterotrophic growth conditions in *R. eutropha* H16. Finally a plasmid based inducible gene expression system under the control of a promoter derived from *R. eutropha* H16 itself was to be set up. Two approaches were defined to generate a set of promoters based on the genome of *R. eutropha* H16 applying promoter activity screening vectors pRS415 and pKR-tac-PAR derivatives. The first approach named the random approach was used to set up a promoter library based on the analysis of partially digested genomic DNA of *R. eutropha* H16 for promoter activity. Therefore genomic DNA was partially digested by *Bsp*143I, cloned into the pRS415 (see figure 13) promoterless screening vector and transferred into *E. coli* cells. The promoter activity of *R. eutropha* H16 genomic DNA fragments was determined under heterotrophic growth conditions in *E. coli* cells on the basis of a β-galactosidase activity assay. In this case the degree of promoter activity is defined by Miller Units. The direct approach on the contrary was defined to characterize putative promoter regions of genes significantly upregulated under lithoautotrophic growth conditions in *R. eutropha* H16. The selection of potential promoter regions was based on the findings of Friedrich et al. [7, 17] who examined differences in protein levels in the soluble part of the proteome of *R. eutropha* H16 by comparing lithoautotrophic to heterotrophic growth. The findings of Friedrich et al. suggest significant upregulation under lithoautotrophic growth conditions of *R. eutropha* H16 for following loci: GyrA, SerS1, Lon, H16_1526 and PHG027. In order to examine expression regulation of these loci a 1000 bp long DNA fragment was amplified 5' upstream to the corresponding genes at each loci. These fragments were cloned into pKR-tac-PAR-EGFP (Figure 8) cloning vector with *Not*I and *Bam*HI for promoter activity screening under lithoautotrophic and heterotrophic growth conditions in *R. eutropha* H16. The promoter activity was determined on the basis of fluorescence units relating to EGFP expression levels.

2. Materials and Methods

2.1 Overview of strains and plasmids used in this work

Table 2: Overview of bacterial strains in this work

Strain	Genotype	Source
E. coli XL1	Tc-r; recB, recI, sbcC201, uvrC, umuC::Tn5(Km-r), mcpA, mcrB, mrr, lac, Δ(hsdRMS), endA1, gyrA96, thi, relA1, supE44(F´), proAB, lacI9ZΔM15, Tn10(Tc-r)	Invitrogen
E. coli TOP10	F´(proAB, lacIq, lacZΔM15, Tn10(tet-r)), mcrA, Δ(mrr-hsdRMS-mcrBC), Φ80ΔlacZΔM15, ΔlacX74, deoR, recA1, araD139(ara, leu), 7697, galU, galK, λ-, rpsL(streptomycin-r), endA1, nupG	Invitrogen
R. eutropha H16	Wiltype	DSM428

Table 3: Overview of plasmids used in this work

Plasmid	Description	Source
pRS415	Amp^r, *lacZ,* promoterless screening vector for *E. coli*	6437
pKR-tac	Kan^r, P_{tac}, cloning vector for *E. coli* and *R. eutropha H16*	This work
pKR-tac-PAR	Kan^r, P_{tac}, *par29.1*,cloning vector for *E. coli* and *R. eutropha* H16	This work
pKR-PAR-EGFP	Kan^r, *egfp, par29.1,* promoterless screening vector for *R. eutropha* H16	This work
pKR-tac-PAR-EGFP	Kan^r, P_{tac}, *egfp*, *par29.1*, cloning vector for *E. coli* and *R. eutropha* H16	This work
pKR-tac-PAR-EGFP *ΔlacI*	Kan^r, P_{tac}, *egfp*, *par29.1*, *ΔlacI*, cloning and expression vector for *E. coli* and *R. eutropha* H16	This work
pKR-lac-PAR-EGFP *ΔlacI*	Kan^r, P_{lac}, *egfp*, *par29.1*, *ΔlacI*, cloning and expression vector for *E. coli* and *R. eutropha* H16	This work

2.2 Overview of strains forwarded to IMBT strain collection

Table 4: Overview of strains forwarded to IMBT strain collection

Strain	Plasmid	Strain collection number IMBT
E. coli TOP 10	pRS415	6437
E. coli TOP 10	pKR-tac-PAR	6438
E. coli TOP 10	pKR-tac-PAR-EGFP	6439
E. coli TOP 10	pKR-tac-PAR-EGFP *ΔlacI*	6440
E. coli TOP 10	pKR-lac-PAR-EGFP *ΔlacI*	6441
E. coli TOP 10	pKR-PAR-EGFP	6442
E. coli TOP 10	pKR-clone 3 fragment 1-PAR-EGFP	6443
E. coli TOP 10	pKR-clone 3 fragment 2-PAR-EGFP	6444
E. coli TOP 10	pKR-clone 5 fragment 1-PAR-EGFP	6445
E. coli TOP 10	pKR-clone 5 fragment 2-PAR-EGFP	6446
E. coli TOP 10	pKR-clone 5 fragment 3-PAR-EGFP	6447
E. coli TOP 10	pKR-clone 5 fragment 4-PAR-EGFP	6448
E. coli TOP 10	pKR-clone 10 fragment 1-PAR-EGFP	6449
E. coli TOP 10	pKR-clone 10 fragment 2-PAR-EGFP	6450
E. coli TOP 10	pKR-clone 10 fragment 3-PAR-EGFP	6451
E. coli TOP 10	pKR-clone 10 fragment 4-PAR-EGFP	6452
E. coli TOP 10	pKR-gyrA-PAR-EGFP	6453
E. coli TOP 10	pKR-sers1-PAR-EGFP	6454
E. coli TOP 10	pKR-lon-PAR-EGFP	6455
E. coli TOP 10	pKR-phg027-PAR-EGFP	6456
E. coli TOP 10	pKR-h16_A1526-PAR-EGFP	6457

2.3 Overview of primers used in this work

Table 5: Overview of primers used in this work

Name	Sequence	TM of primer
PHG027-NotI fwd	5' - tt**gcggccgc**atgtaacacttgccgatgtg - 3'	61.5°C
PHG027-BamHI rev	5' - cg**ggatcc**tctccaaagattgcttgagg - 3'	62.3°C
H16_A1526-NotI fwd	5' - tt**gcggccgc**tcagcgcgaaccagtactc - 3'	64.7°C
H16_A1526-BamHI rev	5' - cg**ggatcc**tgctggttcctttggtgtc - 3'	64.3°C
Lon-NotI fwd	5' - tt**gcggccgc**cggcatcgtctacatcgat - 3'	64.2°C
Lon-BamHI rev	5' - cg**ggatcc**ttttccccaagtcagtcattt - 3'	62.6°C
gyrA-NotI fwd	5' - tt**gcggccgc**gcacgacttcggatcgac - 3'	65.1°C
gyrA-BamHI rev	5' - cg**ggatcc**ttgcggcgtggtttatg - 3'	64.9°C
serS1-NotI fwd	5' - tt**gcggccgc**agcctggacgatgccg - 3'	67.0°C
serS1-BamHI rev	5' - cg**ggatcc**tgttgctttggtgaagtgca - 3'	63.1°C
CIV14f1-NotIfwd	5' - tt**gcggccgc**catggaatgataaatcccggt - 3'	64.7°C
CIV14f2-NotIfwd	5' - tt**gcggccgc**cacgcccggactcaaa - 3'	65.4°C
CIV14f3-NotIfwd	5' - tt**gcggccgc**agttcgccgcactcctcta - 3'	65.4°C
CIII39f1-NotIfwd	5' - tt**gcggccgc**tttccgattctgagatccaca - 3'	64.9°C
CIII39f2-NotIfwd	5' - tt**gcggccgc**ttgtttgcatttgagccga - 3'	65.7°C
CIII39f3-NotIfwd	5' - tt**gcggccgc**attcttacggcacccaatcc - 3'	65.1°C
CIII39_BamHIrev	5' - cg**ggatcc**gacattccctgag - 3'	
CIII39_NotIfwd	5' - tt**gcggccgc**gatccttttatgttccccgc - 3'	65.6°C
CIV14_NotIfwd	5' - tt**gcggccgc**tcgagatcggccttcg - 3'	64.5°C
CIV14_BamHIrev	5' - cg**ggatcc**atcgaattgcccagagagg - 3'	64.5°C
CIV1a_NotIfwd	5' - tt**gcggccgc**aattcgccgagatcattg - 3'	61.7°C
CIV1b_NotIfwd	5' - tt**gcggccgc**tcaacagcgacgaatacagc - 3'	64.4°C
CIV1_BamHIrev	5' - cg**ggatcc**tcttctgcatgattggcttc - 3'	63.4°C
EGFP_NdeI_fwd	5' - cc**catatg**gtgagcaagggcg - 3'	64.3°C
EGFP_HindIII_rev	5' - cc**aagctt**ttacttgtacagctcgtccatgc - 3'	65.4°C
LacZ-rev	5' - cgacggccagtgaatccgtaa - 3'	
pRS415-fwd	5' - gccataaactgccaggaattgg - 3'	

Name	Sequence	TM of primer
Pkrtac_fwd	5' - caatgcttctggcgtgc - 3'	
PAR1_rev	5' - acattagcacatgtgtgggcg - 3'	
PAR2_fwd	5' - acgcccacacatgtgctaatg - 3'	
par29 fwd	5' - cgg**actagt**gcatgccagctt - 3'	56.6°C
par29 rev	5' - cgg**actagt**gatccgacgacc - 3'	55.3°C
par29 fwd länger	5' - cgg**actagt**gatcctctacgc - 3'	48.2°C
PAR lang fwd neu	5' - cgg**actagt**gatcctctacgc - 3'	48.2°C
PAR kurz rev neu	5' - cgg**actagt**gatccgacgacc - 3'	55.4°C
PAR Seq 1281 fwd	5' - tgttcgagccgcagcatttc - 3'	
PAR Seq 1380 rev	5' - agccgcgcaacgtcaacgcc - 3'	
Tac Seq 5712	5' - gtgagcggataacaatttcacaca - 3'	
KanR Seq 969 rev	5' - tatcagaccgcttctgcgttct - 3'	

Bold characters represent restriction sites of endonucleases. Underlined sequences refer to the binding part of the primer on the template. Melting temperature of primers was determined by Finnzymes TM calculator (Finnzymes, Vantaa, Finland).

2.3 General protocols

Medium and cultivation protocols

Cultivation of *E. coli* strains and *R. eutropha* H16

E. coli TOP10 and *E. coli* XL1 cells were cultivated at 37°C on regular LB medium or minimal medium A (MinA) with 100 µg/ml ampicillin or 40 µg/ml of kanamycin. *R. eutropha* H16 was cultivated at 28°C on LB medium or mineral salt medium (MSM) with a concentration of 200 µg/ml of kanamycin. The ingredients for one liter of MinA are 10.5 g of K_2HPO_4, 4.5 g of KH_2PO_4, 1.0 g of $(NH_4)_2SO_4$, 0.5 g of sodium-citrate $2H_2O$ and 3.0 g of casamino acids. These components were diluted in 1 liter of bidest. H_2O and autoclaved. After autoclaving 1 ml of 1 M $MgSO_4$, 20 ml of 20 % glucose solution and 2ml of thiamin (10 mg/ml) were added. MSM was made by mixing solution A and B with bidest.H_2O at a ratio of 1:1:28 before autoclaving. The pH was adjusted at 7 with diluted phosphorus acid. After autoclaving 0.01 volume solution C is added. Solution A (30x) containing 270g of Na_2HPO_4 x 12 H_2O and 45g of KH_2PO_4 diluted in 1 liter of bidest. H_2O. One liter of solution B (30x) contains 6 g of $MgSO_4$ x $7H_2O$, 60 g of NH_4Cl, 30ml of SL7, 125 mg of Fe(III)NH_4-citrate. Components of solution B were also diluted in one liter of bidest H_2O. Solution C (100x) contained250 mg $CaCl_2$ x H_2O diluted in 250 ml of bidest. H_2O.

Protocol for competent *R. eutropha* H16 cells

R. eutropha H16 is used to inoculate 30 ml of liquid LB kanamycin (200 µg/ml) medium. The ONC is incubated over night at 28°C and 500 rpm using a triangular shaped magnetic stirrer. 4 ml of ONC were used to inoculate 100 ml LB kanamycin (200 µg/ml) of main culture. The main culture was incubated at 28°C, 500rpm until OD_{600} readings reach 0.6 to 0.8. When OD_{600} values were reached the main culture was put on ice for 15 minutes. After the incubation on ice the main culture was centrifuged for 15 minutes at 4°C and 4000 rpm. The supernatant was discarded and cells were washed with 5ml of ice cold 0.3 M

sucrose solution. The centrifugation step was repeated at 4°C and 4000 rpm for 10 minutes. The supernatant was discarded and cells were washed with 2.5ml of ice cold 0.3 M sucrose solution. Cells were centrifuged once more for 5 minutes at 4°C and 4000 rpm. In a final step the supernatant was discarded and *R. eutropha* H16 cells were diluted with 0.3 M sucrose solution to a final OD_{600} of 30.

Transformation of *E. coli* strains and *R. eutropha*H16

An aliquot of 40µl of *E. coli* competent cells were mixed with an amount of approximately 100 ng of DNA and were incubated on ice for 10 minutes. *E. coli* TOP10 and *E. coli* XL1 *cells* were transformed with Micro Pulser™ (Bio-Rad, Hercules, California, U.S.) program EC 2 (0.2 mm cuvettes, 2.5 kV). *E. coli* cells were regenerated in 1 ml of LB medium at 37°C and 750rpm in a thermomixer for one hour. An aliquot of 100µl *R. eutropha* H16 cells at an OD_{600} of 30 was set on ice with approximately 100 ng of DNA for 30 minutes. *R. eutropha* H16 cells were transformed using the EC2 (0.2 mm cuvettes, 2.5 kV) program with Micro Pulser™ (Bio-Rad, Hercules, California, U.S.). The cells were regenerated in 1 ml of LB medium at 28°C and 950 rpm on a thermomixer for 1.5 hours. Best results were obtained when 100µl of a 1/10 dilution were plated out on solid medium in a petri dish.

Extraction of genomic DNA

The genomic DNA of *R. eutropha* H16 was extracted with Easy-DNA™ Kit (Invitrogen, Carlsbad, California, USA).

Preparation:
1.5 ml of overnight culture was centrifuged at 5000 rpm for 5 minutes to pellet the cells. The cell pellet was resuspended in 200 µl of 1 X PBS

Isolation of DNA:

1. 350 µl of solution A were added to the cell suspension and vortexed in 1 second intervals until evenly dispersed. 2µl of proteinase K were added after vortexing.
2. The cell suspension was incubated at 65°C for 30 minutes
3. 150 µl of solution B were added. The sample was vortexed vigorously until the precipitate moved freely in the tube and the sample was uniformly viscous
4. 500µl of chloroform were added and the sample was vortexed until viscosity decreased and the mixture was homogenous.
5. The samples were centrifuged at maximum speed for 20 minutes at 4°C

DNA Precipitation:

1. 1 ml of 100% ethanol (-20°C) was added to the sample and vortexed briefly
2. The samples were incubated on ice for 30 minutes.
3. Sample was centrifuged for 15 minutes at maximum speed and 4°C. Ethanol was removed from the pellet with a pipette.
4. 500µl of 80% ethanol (-20°C) were added and the sample was mixed by inverting the tube 3 – 5 times.
5. Sample was centrifuged at maximum speed for 5 minutes and 4°C. Ethanol is removed from the pellet with a pipette.
6. Sample was centrifuged for 3 minutes at 4°C. Residual ethanol was removed with a pipette. Additionally the sample was air dried for 5 minutes.
7. The pellet was resuspended in 100µl TE buffer. 2 mg/ml RNase were additionally added to a final concentration of 40 µg/ml.
8. The sample was incubated at 37°C for 30 minutes. DNA was stored at 4°C.

General PCR set-up

PCR reactions were performed using Finnzymes Phusion® High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland) and following the recommended protocol of the manufacturer.PCR mixture contained 5 µl of 5 x Phusion HF Buffer, 0.5 µl of 10mM dNTPs, 0.25 µl of Phusion®High-Fidelity DNA Polymerase, approximately 15 ng of template DNA, 1.25 µl of forward primer(10 pmol/µl) (see 2.2) and 1.25µl of reverse primer(10 pmol/µl) (see 2.2).Bidest. H_2O was added to a final volume of 25 µl. The PCR cycles were chosen as followed:

98°C	30 sec	
98°C	10 sec	25 cycles
X°C	20 sec	
72°C	30 sec	
72°C	10 min	

The annealing temperatures for primer (see 2.3) were determined by Tm calculator (Finnzymes, Vantaa, Finland).

Overlap extension PCR protocol

In order to eliminate an *Nde*I restriction site on *par29* an overlap extension PCR was performed. Therefore a 50 µl PCR approach was set up using Finnzymes Phusion® High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland).In this case 40 ng of each DNA template were added to the reaction mix. The first 6 cycles were run without outer primers. Afterwards outer primers were added to the reaction mix and another 18 PCR cycles were carried out. PCR mixture contained 10 µl of 5 x Phusion HF Buffer, 1 µl of 10mM dNTPs, 0.25 µl of Phusion® High-Fidelity DNA Polymerase, 40 ng of template 1, 40 ng of template 2, 2.5 µl of (10 pmol/µl) forward outer primer - PAR lang fwd neu (see 2.2) and 2.5µl of (10 pmol/µl) reverse outer primer - PAR kurz rev neu (see 2.2). Bidest. H_2O was added to a final volume of 50 µl.

Sequence of PCR cycles:

98°C	30 sec	
98°C	10 sec	25 cycles
58°C	20 sec	
72°C	30 sec	
72°C	10 min	

2.4 Construction of cloning and screening vectors suitable for *E. coli* and *R. eutropha* H16

General cloning strategies

General cloning strategies were accomplished by using Fermentas restriction enzymes (Fermentas, St. Leon-Rot, Germany), Promega T4 DNA Ligase (Promega, Mannheim, Germany) and Phusion® High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland). Wizard® SV Gel and PCR Clean-Up System Kit (Promega, Mannheim, Germany) was used for gel and PCR product purification. Direct PCR product cloning was carried out with FermentasCloneJET™ PCR Cloning Kit (Fermentas, St. Leon-Rot, Germany). Sequencing of DNA molecules was performed by Agowa (Agowa, Berlin, Germany). Dephosphorylation of DNA molecules was accomplished by using Shrimp Alkaline Phosphatase (SAP) (Fermentas, St. Leon-Rot, Germany). In case of DNA strand end repair and blunting the Fast DNA End Repair Kit was used (Fermentas, St. Leon-Rot, Germany). Plasmid preparations were accomplished by using GeneJET Plasmid Miniprep Kit (Fermentas, St. Leon-Rot, Germany).

Protocols for cloning strategies

Protocol for restriction mixtures with (Fermentas, St. Leon-Rot, Germany):

Preparation of a restriction reaction mixture:

1,5 µl of suitable reaction buffer (Fermentas, St. Leon-Rot, Germany)

0,5 µl of restriction enzyme (Fermentas, St. Leon-Rot, Germany)

x µl of 300 ng of DNA

x µl of bidest. H_2O to a final reaction mixture volume of 15µl.

The restriction reaction mixture was incubated at 37°C overnight. Thermal inactivation was performed according to suggested method by Fermentas. Preparative restrictions were conducted in larger volumes.

Protocol for ligation mixtures with Promega T4 DNA Ligase (Promega, Mannheim, Germany):

Preparation of the reaction mixture for a ligation reaction with insert to vector ratio of 1:3 or 1:4:

1,5 µl of T4 DNA Ligase Buffer

0,5 µl of T4 DNA Ligase

x µl of 90 ng of vector backbone DNA

x µl of appropriate amount of insert DNA

x µl of bidest. H_2O to a final reaction mixture volume of 15µl.

The ligation reaction mixture was incubated at 16°C overnight. Thermal inactivation was performed at 65°C for 20 minutes.

Protocol for gel and pcr product clean-up with Wizard® SV Gel and PCR Clean-Up System Kit (Promega, Mannheim, Germany):

A. Dissolving of the gel slice

    1. 10 µl of membrane binding solution were added per 10 mg of gel slice. The sample was vortexed and incubated at 65°C until the gel slice was completely dissolved.

B. Processing PCR reactions

    1. An equal volume of membrane binding solution was added to the PCR reaction

C. Binding of DNA

1. SV Minicolumn was inserted into the collection tube
2. The dissolved gel mixture or the prepared PCR product were transferred to the minicolumn assembly and incubated at room temperature for 1 minute.
3. The minicolumn was centrifuged for 1 minute at 16,000 x *g*. The flow-through was discarded and the minicolumn was reinserted into the collection tube.

D. Washing

1. 700 µl of membrane was solution added. The sample was centrifuged for 1 minute at 16,000 x *g*. The flow-through was discarded and the minicolumn was reinserted into the collection tube.
2. Step D.1 was repeated with 500 µl of membrane wash solution. The sample was centrifuged for 5 minute at 16,000 x *g*. The flow-through was discarded and the minicolumn was reinserted into the collection tube.
3. The column assembly was recentrifuged for 1 minute at 16,000 x *g*.

E. Elution

1. The minicolumn was transferred to a clean 1.5 ml microcentrifue tube.
2. 30 µl of bidest. H_2O were added to the minicolumn and incubated at room temperature for 5 minutes. The sample was recentrifuged for 1 minute at 16,000 x *g*.

Protocol for cloning of PCR products with Fermentas CloneJET™ PCR Cloning Kit (Fermentas, St. Leon-Rot, Germany):

A. The ligation reaction was set up on ice:

   10 µl of 2X Reaction buffer

   1 µl (0.15 pmol) ends of purified PCR product

   Add 1 µl (0.05 pmol ends) of pJET1.2/blunt cloning vector (50 ng/µl)

   Add bidest. H_2O to 19 µl.

   Add 1 µl of T4 DNA Ligase

The reaction mixture was vortexed briefly and centrifuged for 5 seconds. Afterwards the sample was incubated at 22°C for 5 minutes. The ligation mixture was used directly for transformation into *E. coli* cells.

Protocol for DNA dephosphorylation with Shrimp Alkaline Phosphatase (SAP) (Fermentas, St. Leon-Rot, Germany):

A. Preparation of the reaction mixture:

   1 µg (~1 pmol termini) of Linear DNA (~3 kb plasmid)

   2 µl of 10X SAP reaction buffer

   1 µl (1 u) of Shrimp Alkaline Phosphatase

   And bidest. H_2O to a final volume of 20 µl

B. The sample was mixed thoroughly, spun briefly and incubated at 37°C for 15 minutes. 1 µl (1 u) of Shrimp Alkaline Phosphatase was added and the sample was was mixed thoroughly, spun briefly and incubated at 37°C for another 15 minutes.

C. The reaction was stopped by incubation of the sample at 65°C for 15 minutes.

Protocol for DNA end repair and blunting with Fast DNA End Repair Kit (Fermentas, St. Leon-Rot, Germany):

A. Preparation of the reaction mixture

   x µl of 0.5-5 µg of DNA fragments

   5 µl of 10X End Repair Reaction Mix

   2.5 µl of End Repair Enzyme Mix

   Add bidest. H_2O to 50 µl final volume.

The reaction mix was incubated at 20°C for 5 minutes. The product was purified using Wizard® SV Gel and PCR Clean-Up System Kit (Promega, Mannheim, Germany).

Protocol for plasmid preparation with GeneJET™ Plasmid Miniprep Kit (Fermentas, St. Leon-Rot, Germany):

1. Resuspend cells, lyse and neutralize

   Added 250 µl of resuspension solution and vortexed

   Added 250 µl of lysis solution and inverted tube 4 – 6 times

   Added 350 µl of neutralization solution and inverted tube 4 – 6 times

2. Bind DNA

   Loaded supernatant to GeneJET™ spin column and centrifuged for 1 minute.

3. Wash column

   Added 500 µl of wash solution and centrifuged for 1 minute, the flow-through was discarded. This step was performed twice. Afterwards the empty column was centrifuged for 1 minute.

4. Elute purified DNA

   The column was transferred to a new tube. 50 µl of bides.H_2O was added to the column and incubated for 2 minutes. The spin column was centrifuged for 2 minutes and the flow-through collected.

Construction of *E. coli* and *R. eutropha* H16 cloning and screening vectors

All cloning and screening vectors were constructed on the basis of the pKR-tac vector (Figure 6). The pKR-tac vector encodes P_{tac}, a kanamycin resistance gene and a replication (REP) sequence originating from the pBBR1MCS plasmid of Bordetella bronchiseptica (x66730.1) which enables plasmid replication in *E. coli* as well as *R. eutropha* H16 cells [49].

Construction of pKR-tac-PAR vector

pKR-tac-PAR (Figure 7) was constructed from pKR-tac by inserting the *par29* region derived from the plasmid pCm470 Dsbc APLE C8P par29. *Par29* consists of several genes promoting plasmid stability and maintenance within *E. coli* cells [46, 48]. Before introducing *par29* into pKR-tac an *Nde*I restriction site located in the promoter sequence of *parD* on the native par29 region was eliminated by overlap extension PCR. Elimination of *Nde*I restriction site on *par29* was necessary to enable insert cloning via *Nde*I and *Hind*III restriction sites on pKR-tac-PAR vectors. The elimination of the *Nde*I restriction site on par29 was accomplished by introducing a nucleotide exchange to the recognition sequence of *Nde*I. The nucleotide exchange was done by overlap extension PCR using modified primers (PAR2_fwd and PAR1_rev, see 2.2) located at the *Nde*I restriction site.

PAR1_rev along with par29 fwd länger primer amplified 2040 bp fragment of the par29 region whereas PAR2_fwd and par29 rev primers amplified a 620 bp fragment of the par29 region. An overlap of 25 bp was generated including the nucleotide exchange. The two DNA templates were assembled by overlap extension PCR as described under 2.3. The *Nde*I restriction site was modified by exchanging an adenine for a guanine (Figure 4). The modified par29 region is referred to as par29.1. The par29.1 region was cloned into pKR-tac using *Spe*I restriction sites to construct the new vector pKR-tac-PAR (Figure 7).

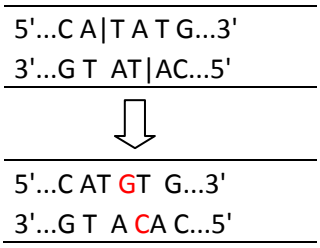

Figure 4: Illustration of the nucleotide exchange introduced by PAR2_fwd and PAR1_rev primers to eliminate an *Nde*I restriction site and therefore converting par29 into par29.1. Red characters are referring to the nucleotide exchange introduced. *Nde*I restriction pattern shown in upper sequence.

Construction of pKR-tac-PAR-EGFP and pKR-PAR-EGFP vector

The pKR-tac-PAR vector was turned into pKR-tac-PAR-EGFP (Figure 8) a promoter screening vector by introducing *egfp* with *Nde*I and *Hind*III. *Egfp* is coding for the enhanced green fluorescent protein (EGFP). EGFP is like GFPmut2 a modified variant of the wild type green fluorescent protein (GFP) isolated from the jellyfish *Aequorea* victoria. The amino acid sequences of EGFP and GFPmut2 were modified at several locations to improve the emission signals of the fluorescent proteins in order to facilitate their detectability [50, 51]. Accordingly the EGFP amino acid sequence differs in four positions from the one of wild type GFP. In the EGFP amino acid sequence valine is added as second amino acid extending the total sequence length from 238 in wild type GFP to 239 amino acids in EGFP. Position 64 of the wild type GFP is changed from phenylalanine to leucine in EGFP, position 65 is changed from serine to threonine and position 231 encodes histidine in wild type GFP and leucine in EGFP. On the contrary

GFPmut2 differs in three amino acids from wild type GFP. Positions 65 and 72 of the 238 long amino acid sequences are encoded for serine in wild type GFP whereas both positions are alanine amino acids in GFPmut2. Additionally position 68 is changed from valine in wild type GFP to leucine in GFPmut2.The variations of the amino acid sequences of EGFP, GFPmut2 compared to GFP are visualized in figure 5.

I. Wild type GFP isolated from *Aequorea* victoria

II. GFPmut2 mutant GFP variant [51]

III. EGFP mutant GFP variant

I. M SKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL

II. M SKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL

III. MVSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL

I. VTTFSYGVQC FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV

II. VTTFAYGLQC FARYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV

III. VTTLTYGVQC FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV

I. NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD

II. NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD

III. NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD

I. HYQQNTPIGD GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK238

II. HYQQNTPIGD GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK 238

III. HYQQNTPIGD GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT LGMDELYK 239

Figure 5: Illustration of the variation in the amino acid sequences comparing wild type GFP to GFPmut2 and EGFP.

Ultimately EGFP was selected as a reporter gene due to simple and reproducible clone analysis based on OD_{600} readings and the measurement of EGFP fluorescent units. Selected promoter fragments were subsequently cloned into pKR-tac-PAR-EGFP in front of *egfp* via *Not*I and *Bam*HI restriction sites for promoter activity screenings. The promoterless screening vector pKR-PAR-EGFP (Figure 9) was generated by cutting pKR-tac-PAR-EGFP with *Not*I and *Eco*RI. The remaining vector backbone was blunted and religated. pKR-PAR-EGFP was used for screening of *R. eutropha* H16 partially digested genomic DNA under heterotrophic and lithoautotrophic growth conditions. pKR-tac-PAR-EGFP vectors were modified once more by removing *lacI* in order to obtain EGFP expression in *R. eutropha* H16. The expression system based on LacI as a repressor was not inducible in *R. eutropha* H16. It was shown that the inducer molecule IPTG is not able to enter *R. eutropha* H16 cells to active EGFP expression [44]. In order to obtain expression from P_{lac} and P_{tac} in *R. eutropha* H16 *lacI* was removed from the pKR-tac-PAR-EGFP vector. pKR-tac-PAR-EGFP Δ*lacI* (Figure 11) was constructed by removing *lacI* with *Not*I and *Pst*I. The vector backbone was blunted and religated. The resulting vector, named pKR-tac-PAR-EGFP Δ*lacI* was further modified. pKR-lac-PAR-EGFP Δ*lacI* (Figure 12) was constructed by exchanging P_{tac} with P_{lac} on pKR-tac-PAR-EGFP Δ*lacI*. These vectors were used to establish P_{tac} and P_{lac} activity readings as a reference for promoter activity assessments in this work. The sequences of pKR-tac-PAR, pKR-tac-PAR-EGFP, pKR-tac-PAR-EGFP Δ*lacI*, pKR-lac-PAR-EGFP Δ*lacI* and pKR-PAR-EGFP vectors were verified by sequence analysis (Agowa, Berlin, Germany).

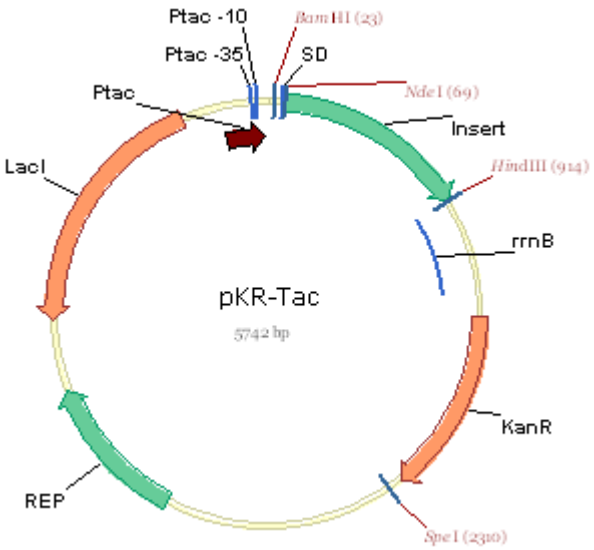

Figure 6: Illustration of pKR-tac vector. Used as the basic vector for further vector construction in this work. Encodes the REP sequence of the pBBR1MCS plasmid form Bordetella bronchiseptica (x66730.1) which enables plasmid replication in *E. coli* and *R. eutropha* H16 cells [49]. SD is indicating the location of the Shine-Dalgarno sequence.

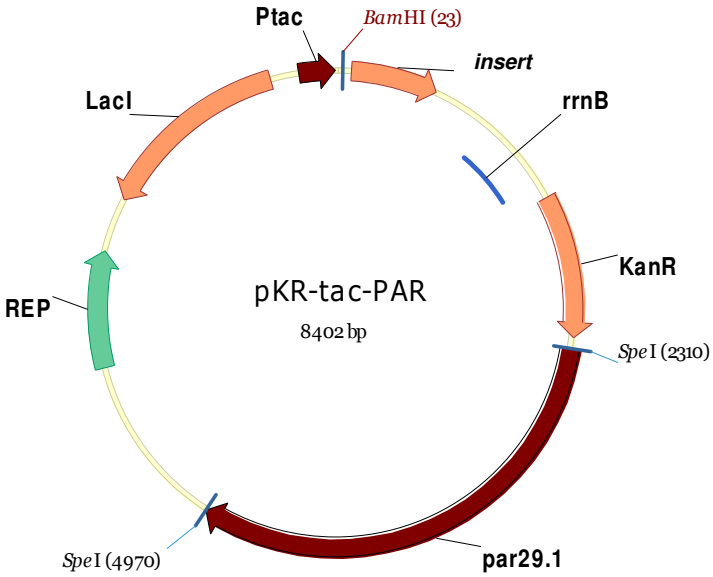

Figure 7: Illustration of pKR-tac-PAR vector. pKR-tac-PAR derived from pKR-tac by inserting *par29.1*. pKR-tac-PAR is used as the basic cloning vector and encodes next to the REP sequence also for par29.1 region, which promotes plasmid stability in *E. coli* and *R. eutropha* H16 cells.

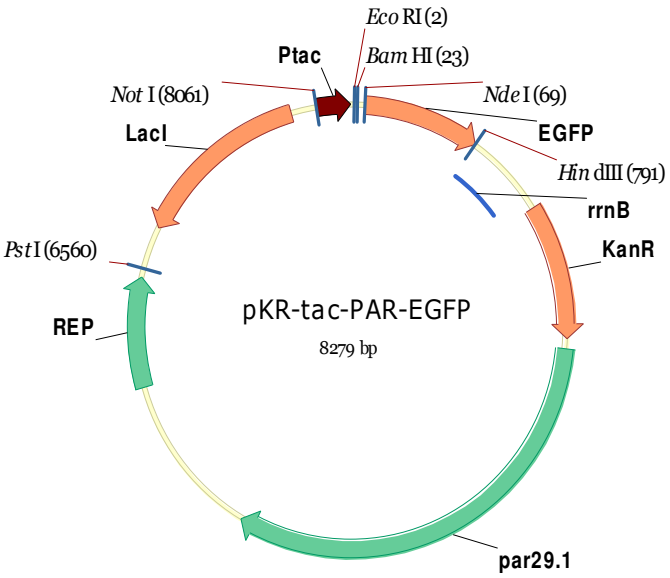

Figure 8: Illustration of pKR-tac-PAR-EGFP vector. Constructed from pKR-tac-PAR by introducing *egfp* via *Nde*I and *Bam*HI restriction sites. pKR-tac-PAR-EGFP was used as a cloning vector for promoter inserts. Promoter inserts were cloned in the vector using *Not*I and *Bam*HI restriction sites.

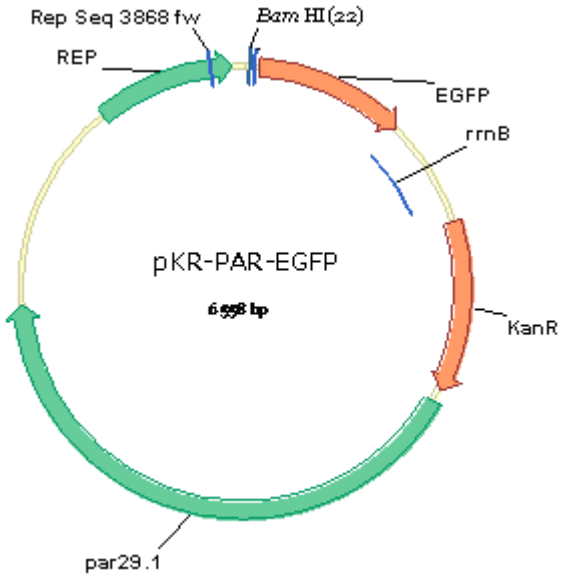

Figure 9: Illustration of pKR-PAR-EGFP vector. Constructed by removing *lacI* and P_{tac} from pKR-tac-PAR-EGFP. Used as a promoter screening vector. Partially digested genomic DNA of *R. eutropha* H16 was inserted via *Bam*HI restriction site.

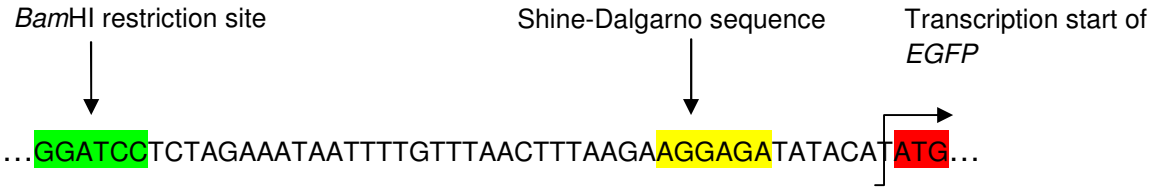

Figure 10: Illustration of the location of the Shine-Dalgarno sequence encoded pKR-PAR-EGFP prior to EGFP. Basepairs 22 to 72 of pKR-PAR-EGFP are illustrated. The Shine-Dalgarno sequence is located 7 bp upstream of the EGFP transcription start.

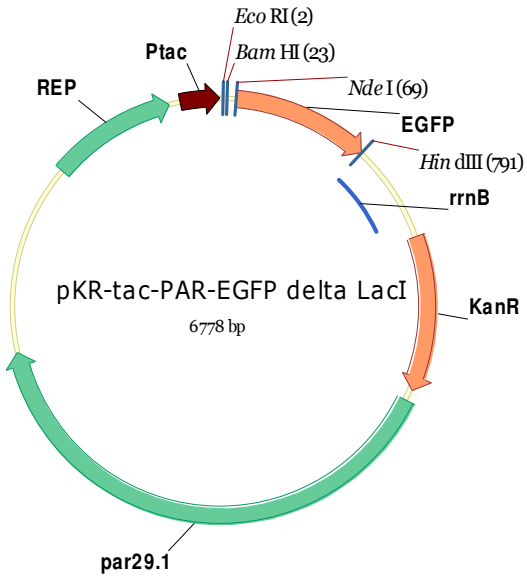

Figure 11: Illustration of pKR-tac-PAR-EGFP *ΔlacI* vector. Constructed by removing *lacI* from pKR-tac-PAR-EGFP. Used as a reference for P_{tac} activity readings.

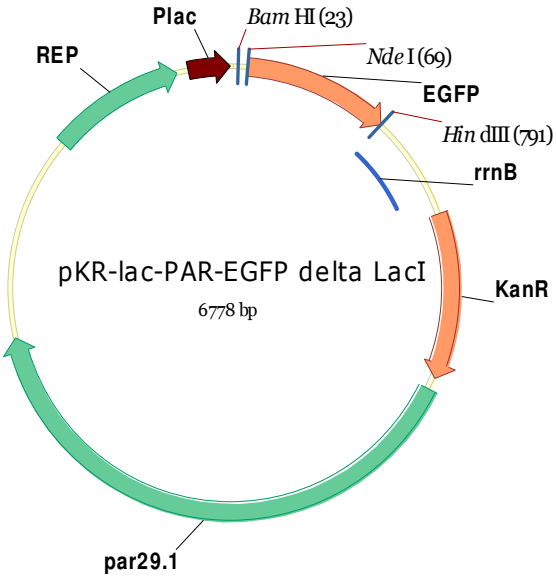

Figure 12: Illustration of pKR-lac-PAR-EGFP *ΔlacI* vector. Constructed by removing *lacI* from pKR-tac-PAR-EGFP and exchanging Ptac with Plac. Used as a reference for P_{lac} activity readings.

2.5 Promoter activity screenings with pRS415 vectors in *E. coli*

Screening for activity of different *R. eutropha* H16 promoters in *E. coli* XL1 under heterotrophic growth conditions

In a first step promoter activity of *R. eutropha* H16 derived genomic DNA fragments was determined in *E. coli* XL1using a β-galactosidase assay. In order to be able to screen for promoter activity the genomic DNA of *R. eutropha* H16 had to be extracted and partially digested. The partial digest of genomic *R. eutropha* H16 DNA was performed using *Bsp*143I restriction enzymes. *Bsp*143I cuts λ-DNA statistically every 250 basepairs, the amount of restriction enzyme was determined to generate DNA fragments with the size in the range of 1000 to 3000 basepairs. *Bsp*143I furthermore shares the restriction pattern with *Bam*HI, which enables cloning into pRS415 vector via the *Bam*HI restriction site in order to screen for promoter activity.

Partial digest of *R. eutropha* H16 genomic DNA

2 µg genomic DNA of *R. eutropha* H16 was incubated with 2 units of *Bsp*143I for 2.5, 5 and 7.5 minutes at 37°C. The reaction was stopped by thermally inactivating *Bsp*143I at 65°C for 20 minutes. 5µl of the reaction mix were loaded on a 1% agarose gel in order to determine the most abundant size of DNA fragments. A vector insert ratio of 1:3 and 100 ng of pRS415 were used to define the amount of insert DNA originating from the partial digest. For that calculations were made to determine ligation conditions for average insert sizes of 1500 bp, 2000 bp and 3000 bp. 42 ng of insert DNA were added to the ligation mixture at an average insert size of 1500 bp, 56 ng for 2000 bp and 84 ng for 3000 bp. The concentration of the partial digest was determined by running 2 µl of partial digest reaction mix on a 1% agarose gel for 5 minutes. The DNA concentration was estimated by comparing to the GeneRuler^{TM} DNA ladder Mix (Fermentas, St. Leon-Rot, Germany). The pRS415 vector was cut with *Bam*HI and dephosphorylated with SAP (Fermentas,St. Leon-Rot, Germany) prior to the ligation step. The ligation mixture of the partial digest and the screening vector pRS415 was set up and incubated at 16°C over night. The ligation mixture was then thermally inactivated at 65°C for 20 minutes and afterwards desalted using dialyses membranes (Millipore, Billerica, USA). 2µl of the desalted ligation mixture were transformed into *E. coli*XL1 cells. After regeneration the cells were plated out on LB ampicillin 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-Gal) plates. The screening plates contained 100 µg/ml ampicillin and 40 µg/ml of X-Gal. Incubation was done at 37°C over night. Clones exhibiting a blue color and hereby β-galactosidase activity were selected for further characterization.

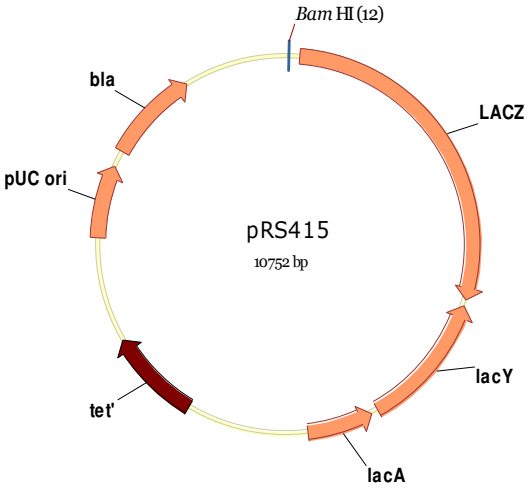

Figure 13: Illustration of pRS415 promoterless plasmid encoding *lacZ*, ampicillin resistance and pUC ori for *E. coli* replication. *Bam*HI restriction site is located in front of *lacZ*.

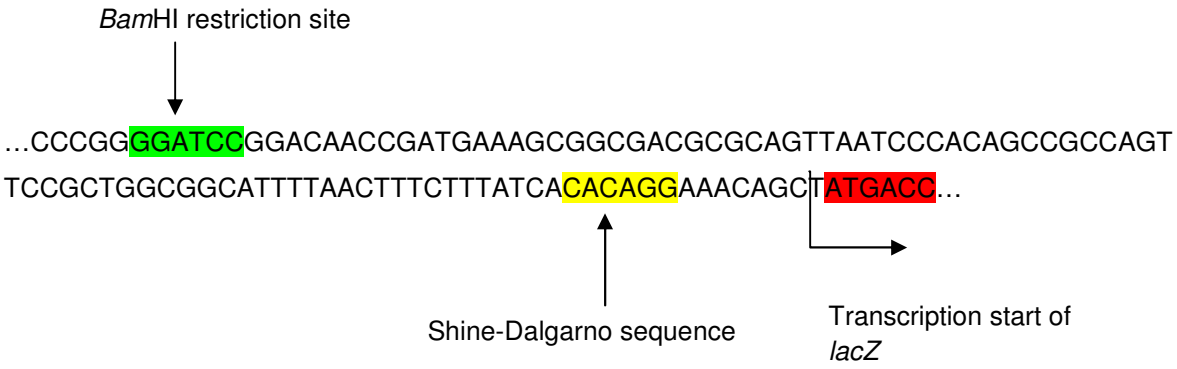

Figure 14: Illustration of the location of the Shine-Dalgarno sequence encoded on pRS415 prior to *lacZ*. Basepairs 52 to 162 of pRS415 are illustrated. The Shine-Dalgarno sequence is located 8 bp upstream of the *lacZ* transcription start.

β-galactosidase assay

Selected clones were used to inoculate 0.6 ml of MinA (100 µg/ml ampicillin) in a 96 deep well plate with toothpicks. The samples in the 96 deep well plate were incubated on a shaking plate with 900 rpm on 37°C over night. After overnight incubation100 µl of the cell cultures grown in the 96 deep well plate were transferred to a microtiterplate for OD_{600} determination. Another 150 µl of overnight culture were transferred into an additional 96 deep well plate. 850 µl of Z-Buffer were added to each sample. One liter of Z-Buffer solution based on bidest. H_2O

contains 60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl and 1mM $MgSO_4$. The pH of the solution was adjusted to 7 with 1 M NaOH. Furthermore 50mM of ß-mercaptoethanol was added freshly before each use. In the next working step 50 µl of chloroform and 25 µl of 0.1% SDS were added to the samples mixed with 850 µl of Z-buffer. The samples were mixed thoroughly by pipetting using a multichannel pipette. After the separation of phases 125 µl of the upper phase were transferred into a fresh microtiter plate and the samples were incubate at room temperature for 5 minutes.

In order to screen for β-galactosidase activity 25 µl of 4 mg/ml 2-nitrophenyl-β-D-galactopyranoside (ONPG) dissolved in 0,1 M sodium phosphate buffer pH=7 were added to each sample on the micortiter plate. The samples were incubated for two minutes at room temperature before adding 60 µl of 1M Na_2CO_3. OD_{600}, OD_{550} and OD_{420} were measured at room temperature using a platereader. The activity of the promoter fragments is determined by Miller Units. The Miller Unit formula applies the reaction time t in minutes, the volume of the culture used in activity assay (in ml) and OD readings taken at 420 nm, 550 nm and 600 nm.

Equation (1) was used to determine the Miller Units:

$$Miller\ Units = 1000 \times \frac{(OD420 - 1.75 \times OD550)}{t \times V \times OD600} \qquad (1)$$

Selected *E. coli* pRS415 clones exhibiting significant β-galactosidase activity were sent to sequencing (Agowa, Berlin, Germany) using LacZ-rev and pRS415-fwd primers. After blasting the sequencing results against the genome of *R. eutropha* H16; selected regions on the *R. eutropha* H16 genome containing potential promoters were amplified by PCR. Additionally to that primer walking was performed for selected amplified regions in order to confine the promoter region and characterize potential regulatory elements. Following primers were used for the amplification and primer walking of selected regions:

Table 6: Illustration of PCR amplified fragments by primer walking based on *E. coli* pRS415 clones 3, 5 and 10 as part of the random approach.

***E. coli* pRS415 clone**	**Fragment**	**Forward primer**	**Reverse primer**	**Length**
clone 3	1	CIV1a_NotIfwd	CIV1_BamHIrev	548 bp
clone 3	2	CIV1b_NotIfwd	CIV1_BamHIrev	138 bp
clone 5	1	CIV14f1-NotIfwd	CIV14_BamHIrev	1000 bp
clone 5	2	CIV14f2-NotIfwd	CIV14_BamHIrev	549 bp
clone 5	3	CIV14f3-NotIfwd	CIV14_BamHIrev	301 bp
clone 5	4	CIV14_NotIfwd	CIV14_BamHIrev	167 bp
clone 10	1	CIII39f1-NotIfwd	CIII39_BamHIrev	1000 bp
clone 10	2	CIII39f2-NotIfwd	CIII39_BamHIrev	613 bp
clone 10	3	CIII39f3-NotIfwd	CIII39_BamHIrev	354 bp
clone 10	4	CIII39_NotIfwd	CIII39_BamHIrev	168 bp

Three regions were amplified in total, for each region additional PCR products were amplified by primer walking. PCR products were designed to be cloned into pKR-tac-PAR-EGFP with *Not*I and *Bam*HI for further screening in *R. eutropha* H16 under heterotrophic and lithoautotrophic growth conditions.

2.7 Characterization and identification of selected hypothetical promoter regions derived from *R. eutropha* H16

Selection of hypothetical promoter regions from *R. eutropha* H16

In earlier work the abundance of soluble proteins of *R. eutropha* H16 under different growth conditions was examined by B. Friedrich [7, 17]. The soluble fraction of the proteome of *R. eutropha* H16 was therefore examined under lithoautotrophic growth conditions with H_2 and CO_2 as growth substrates and under heterotrophic growth conditions with glycerol and succinate as growth substrates. In a following step the soluble protein fraction of *R. eutropha* H16 grown under lithoautotrophic conditions was compared to the soluble protein fractions of cells grown under different heterotrophic conditions, in order to identify

differences in protein expression patterns. Based on these findings six 5' upstream regions of genes that were significantly upregulated under lithoautotrophic growth conditions were selected for further work (see table 7) [7]. The gene loci and region for amplification were identified via NCBI. In order to ensure the presence of a proximate gene related promoter an approximately 1000 basepair DNA region was amplified 5' upstream of the selected loci. The hypothetical promoter containing DNA fragments were amplified by PCR and sequenced (Agowa, Berlin, Germany). The fragments were subsequently introduced into the pKR-tac-PAR-EGFP vector via *Not*I and *Bam*HI restriction sites. The six different ligation preparations of pKR-tac-PAR-EGFP and the selected promoter regions were transformed in *E. coli* TOP10. In a second step plasmid preparations were made of the promoter variant constructs of pKR-tac-PAR-EGFP in *E. coli* TOP10 and were transformed into *R. eutropha* H16. Table 7 lists the locus tags of genes which 5' upstream regions were amplified for promoter screening as described in the abstract above. It also indicates the replicon and difference in gene expression levels under lithoautotrophic andheterotrophic growth conditions. Gene expression levels for heterotrophic growth conditions are shown for glycerol and succinate as carbon sources.

Table 7: Selection of hypothetical promoter regions on *R. eutropha* H16 based on upregulated expression levels under lithoautotrophic growth conditions. Growth conditions: A, H_2-CO_2 ; G, Glycerol; S, Succinate [7].

Name:	Locus Tag:	Replicon:	LOG2(A/S)	LOG2(A/G)
Sers1	H16_A0764	Chr1	3,8	3,3
GyrA	H16_A0789	Chr1	2,8	2,0
Lon	H16_A1485	Chr1	4,5	1,7
A1526	H16_A1526	Chr1	5,1	5,4
PHG027	PHG027	PHG1	3,4	1,6

Preparation method for EGFP fluorescence unit measurement

Overnight cultures of *R. eutropha* H16 clones containing pKR-tac-PAR-EGFP Δ *lacI* vectors were harvested from liquid culture medium by centrifugation and resuspended in the same volume of 50 mM Tris-HCl buffer. The cell suspension was incubated on ice for 10 minutes. Fluorescence units of the samples were determined afterwards. This method was already successfully applied for *R. eutropha* H16 containing plasmids encoding for GFPmut2 in order to obtain GFPmut2 fluorescence [44].

Overnight cultures of *E. coli* clones were used directly for EGFP fluorescence unit measurement.

EGFP fluorescence unit measurement

EGFP fluorescence units were determined with Biotek Synergy Multi-Mode Microplate Reader (Biotek, Winooski, VT, USA). 200 µl of *E. coli* or *R. eutropha* H16 overnight cultures were loaded onto a UV-microtiter plate and analyzed with the Biotek Synergy Multi-Mode Microplate Reader at an excitation wavelength of 488 nm. The fluorescence of EGFP was detected at and emission wavelength of 509 nm.

3. Results

3.1 Vector construction

Vector construction was based on pKR-tac as basic cloning vector. pKR-tac-PAR was constructed by introduction *par29.1* to pKR-tac. Control restriction of pKR-tac-PAR is shown in figure 15. Lane 1 shows uncut pKR-tac-PAR vector DNA. Lane 2 shows the *Spe*I restriction with *par29.1* at 2600 bp and the remaining vector backbone of pKR-tac-PAR at 5800 bp. Lane 3 shows a *Nde*I linearization of pKR-tac-PAR with a length of 8400 bp. The restriction pattern of pKR-tac-PAR-EGFP is shown in figure 16. Lane 1 refers to uncut pKR-tac-PAR-EGFP. Lane 2 shows *Nde*I linearized pKR-tac-PAR-EGFP vector DNA with a size of 8300 bp. Lane 3 shows a *Nde*I and *Hind*III cut with pKR-tac-PAR-EGFP vector DNA backbone at a size of 7580 bp and *egfp* at 720 bp.

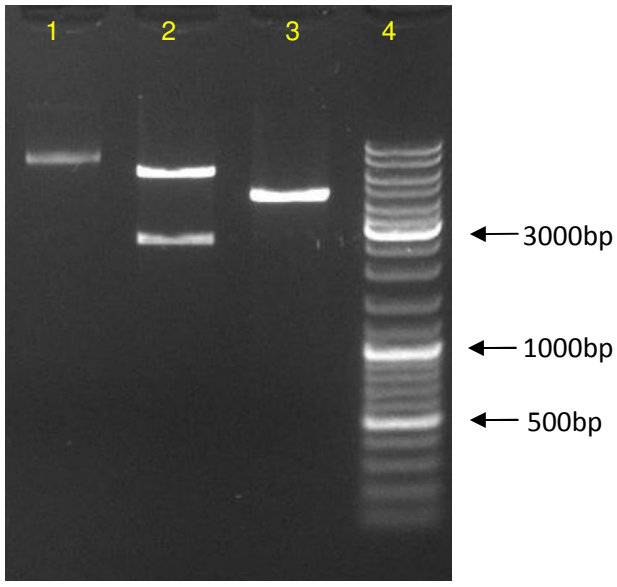

Figure 15: 1% Agarose gel of different pKR-tac-PAR restrictions. 1. pKR-tac-PAR uncut, 2. pKR-tac-PAR cut with *Spe*I, 3. pKR-tac-PAR cut with *Nde*I, 4. 500 ng of Gene Ruler DNA Ladder Mix

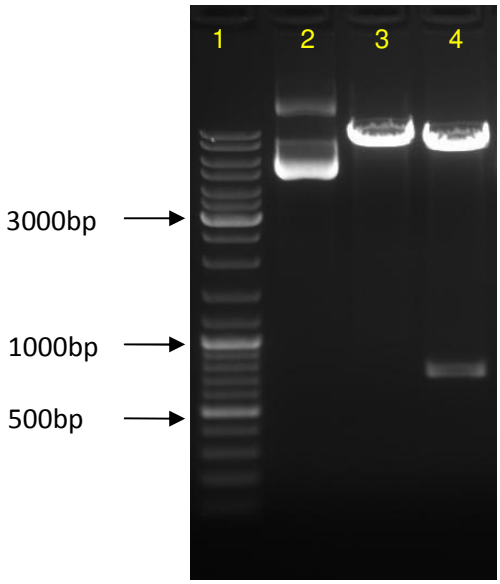

Figure 16: 1% Agarosegel of different pKR-tac-PAR-EGFP restrictions. 1. 500 ng of Gene Ruler DNA Ladder Mix, 2. pKR-tac-PAR-EGFP uncut, 3. pKR-tac-PAR-EGFP cut with *Spe*I, 4. pKR-tac-PAR-EGFP cut with *Nde*I and *Hind*III

EGFP expression in *E. coli* and *R. eutropha* H16

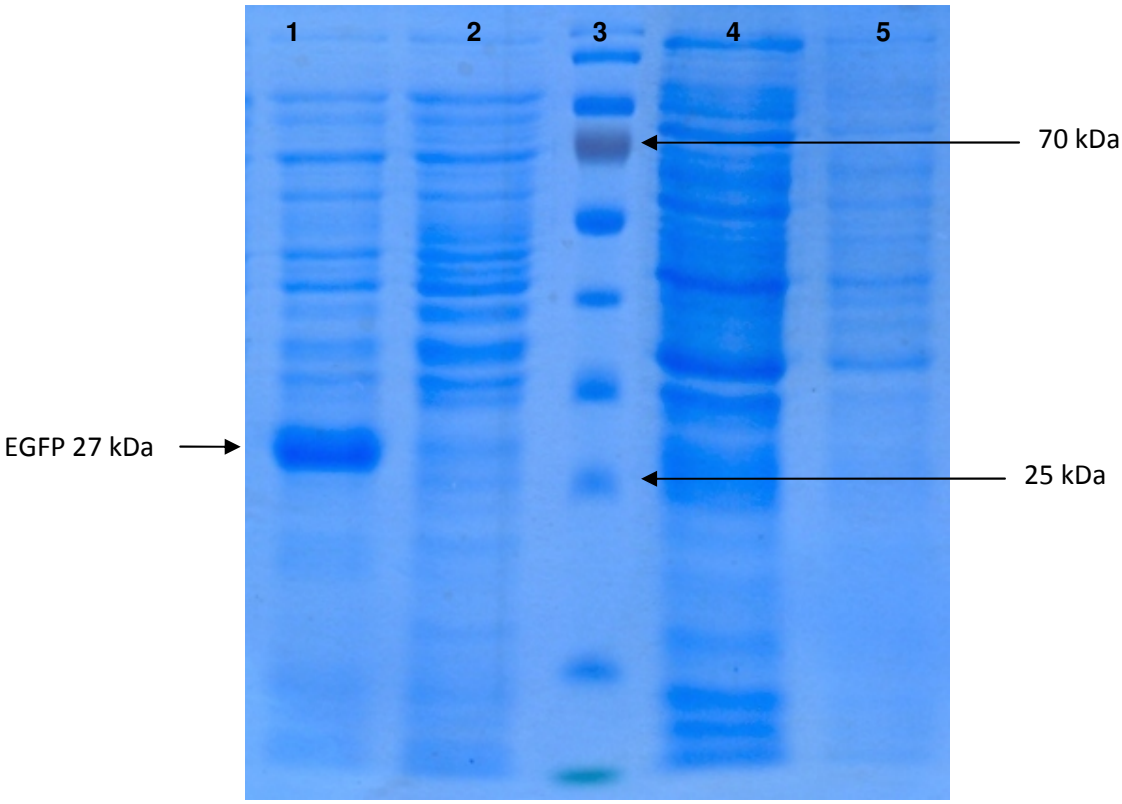

Figure 17: 12 % SDS-PAGE. For all samples crude extracts were used. 1. *E. coli* pKR-tac-PAR-EGFP *ΔlacI*, 2. *E. coli* pKR-tac-PAR-EGFP, 3. PageRuler^{TM} Prestained Protein Ladder, 4. *R. eutropha* H16 pKR-tac-PAR-EGFP *ΔlacI*, 5. *R. eutropha* H16 pKR-tac-PAR-EGFP. The red band of the PageRuler^{TM} Prestained Protein Ladder represents 70 kDa, 25 kDa is represented by the third blue band from the bottom. The EGFP protein is visible at 27 kDa in lane 1.

3.2 Random Approach

3.2.1 Results of β-galactosidase activity assay screening with pRS415 in *E. coli* cells

A series of more than twelve β-galactosidase activity assay screening rounds covering over 800 clones were carried out. The Miller Units of *E. coli* pRS415 clones containing partially digested *R. eutropha* H16 genomic DNA were determined as described in section 2.5. Collected data and Miller Units of significant screening rounds are shown in tables below. Miller Units and OD readings of clones selected for further work were color coded.

Results of screening round

OD_{550} readings of screening round:

	1	2	3	4	5	6	7	8	9	10	11	12
A	0,011	0,012	0,010	-0,007	0,008	0,008	0,010	0,014	0,005	0,015	0,014	0,031
B	0,009	0,012	0,010	0,000	0,010	0,008	0,013	0,007	0,005	0,017	0,016	0,034
C	0,009	0,010	0,003	0,008	0,007	0,020	0,010	0,017	0,016	0,015	0,013	0,026
D	0,005	0,004	-0,002	0,003	0,004	0,000	0,007	0,005	0,011	0,001	0,002	0,019
E	0,015	0,016	0,010	0,011	0,019	0,014	0,016	0,015	0,015	0,015	0,015	0,000
F	0,013	0,014	0,006	0,008	0,017	0,012	0,011	0,013	0,008	0,012	0,011	-0,004
G	0,024	0,017	0,010	0,004	0,046	0,019	0,007	0,020	0,064	-0,001	0,015	0,011
H	0,158	0,012	0,008	-0,006	0,014	0,009	0,001	0,015	0,006	-0,005	0,004	0,000

OD_{420} readings of screening round:

	1	2	3	4	5	6	7	8	9	10	11	12
A	0,229	0,584	0,591	0,073	0,185	0,438	0,623	0,105	0,100	0,534	0,303	0,194
B	0,208	0,572	0,602	0,062	0,193	0,395	0,540	0,097	0,101	0,499	0,290	0,202
C	0,141	1,067	0,130	0,186	0,150	0,398	0,178	0,272	0,177	0,121	0,143	0,399
D	0,198	1,043	0,117	0,153	0,157	0,372	0,182	0,147	0,168	0,102	0,112	0,377
E	0,531	0,479	0,080	0,107	0,303	0,296	0,488	0,094	0,085	0,080	0,432	0,000
F	0,555	0,419	0,079	0,104	0,271	0,283	0,397	0,108	0,072	0,073	0,449	-0,002
G	0,500	0,186	0,270	0,274	0,249	0,397	0,086	0,297	1,076	0,013	0,059	0,011
H	0,749	0,171	0,245	0,152	0,133	0,347	0,044	0,142	0,330	-0,004	0,019	-0,001

OD_{600} readings of screening round:

	1	2	3	4	5	6	7	8	9	10	11	12
A	0,250	0,274	0,311	-0,035	0,288	0,248	0,347	0,244	0,161	0,437	0,344	0,622
B	0,331	0,322	0,929	0,024	0,362	0,293	0,370	0,347	0,216	0,513	0,423	0,726
C	0,289	0,257	0,267	0,333	0,249	0,263	0,270	0,429	0,460	0,423	0,344	0,563
D	0,369	0,314	0,288	0,526	0,315	0,285	0,220	0,528	0,460	0,497	0,424	0,704
E	0,305	0,535	0,336	0,420	0,498	0,465	0,433	0,522	0,346	0,460	0,437	0,000
F	0,319	0,424	0,365	0,448	0,511	0,492	0,421	0,557	0,377	0,460	0,450	0,048
G	0,389	0,308	0,376	0,149	0,735	0,412	0,207	0,537	1,256	0,465	0,446	-0,023
H	0,437	0,297	0,437	0,126	0,696	0,471	0,218	0,471	1,262	0,460	0,471	-0,002

Miller Units of screening round:

	1	2	3	4	5	6	7	8	9	10	11	12
A	1117	2741	2460	-3230	794	2276	2324	441	753	1551	1082	298
B	774	2281	839	3467	648	1736	1865	327	566	1218	826	262
C	575	5450	624	692	740	1842	796	754	432	298	469	838
D	685	4395	556	373	633	1738	1029	349	433	268	340	652
E	2209	1122	250	275	722	780	1416	173	228	157	1239	0
F	2227	1238	252	269	631	712	1195	205	207	153	1272	108
G	1570	678	895	2391	305	1178	475	648	1024	39	98	468
H	1443	672	703	1706	206	939	262	328	337	13	35	383

Results of screening round

OD_{550} readings of screening round:

	1	2	3
A	0,040	0,048	0,031
B	0,033	0,035	0,042
C	0,034	0,029	0,034
D	0,023	0,032	0,023
E	0,020	0,013	0,025
F	0,030	0,034	0,026
G	0,033	0,037	1,283
H	0,003	0,042	0,009

OD_{420} readings of screening round:

	1	2	3
A	0,311	0,418	0,228
B	0,272	0,265	0,272
C	0,193	0,422	0,134
D	0,176	0,325	0,283
E	0,114	0,269	0,339
F	0,221	0,274	0,035
G	0,225	0,419	1,558
H	0,037	0,581	0,006

OD_{600} readings of screening round:

	5	6	7
A	0,569	0,533	0,406
B	0,393	0,432	0,505
C	0,447	0,152	0,321
D	0,315	0,343	0,360
E	0,282	0,120	0,465
F	0,385	0,407	0,016
G	0,281	0,444	-0,002
H	-0,002	0,539	-0,002

Miller Units of screening round:

	1	2	3
A	1409	2090	1426
B	1827	1573	1311
C	996	8154	783
D	1431	2611	2252
E	933	6858	2123
F	1456	1762	-2046
G	1986	2663	1347255
H	-49881	3135	16270

Results of pooled screening round

OD_{550} readings of pooled screening round:

	1	2	3	4	5	6	7
A	0,021	0,010	0,013	0,021	0,022	0,019	0,019
B	0,038	0,016	0,013	0,022	0,026	0,023	0,024
C	0,018	0,018	0,002	0,024	0,036	0,038	0,031
D	0,022	0,027	0,018	0,032	0,045	0,045	0,040
E	0,010	0,021	0,023	0,026	0,027	0,031	0,024
F	0,008	0,021	0,019	0,025	0,026	0,026	0,024
G	0,009	0,032	0,013	0,024	0,027	0,026	0,031
H	0,003	0,031	0,015	0,022	0,028	0,022	0,027

OD_{420} readings of pooled screening round:

	1	2	3	4	5	6	7
A	0,677	0,427	0,120	0,365	0,491	1,223	1,227
B	0,740	0,497	0,093	0,385	0,480	1,105	1,139
C	0,067	0,999	0,276	0,167	2,181	1,348	0,513
D	0,067	1,065	0,359	0,187	2,020	1,270	0,516
E	0,013	0,976	1,401	0,679	0,398	1,472	0,255
F	0,010	0,851	1,034	0,620	0,456	1,739	0,288
G	0,011	1,667	2,358	0,478	0,544	0,939	0,974
H	0,006	1,722	2,414	0,468	0,674	0,959	0,770

OD_{600} readings of pooled screening round one:

	1	2	3	4	5	6	7
A	0,501	0,234	0,24	0,493	0,535	0,407	0,401
B	0,602	0,29	0,312	0,537	0,593	0,363	0,472
C	0,234	0,27	0,116	0,492	0,53	0,357	0,6
D	0,302	0,244	0,187	0,441	0,412	0,341	0,71
E	0,234	0,346	0,4	0,511	0,536	0,463	0,541
F	0,206	0,34	0,361	0,57	0,499	0,366	0,603
G	0,143	0,45	0,079	0,437	0,541	0,5	0,563
H	0,143	0,53	0,062	0,466	0,552	0,433	0,542

Miller Units of pooled screening round one:

	1	2	3	4	5	6	7
A	2835	3899	903	1480	1884	6496	6612
B	2485	3599	501	1435	1628	6520	5164
C	332	7967	5216	563	8884	7981	1701
D	212	9271	3898	660	10469	7763	1394
E	-37	6033	7565	2755	1455	6804	873
F	-46	5323	6164	2248	1832	10286	910
G	-65	7954	65684	2223	2042	3973	3629
H	-2	6989	85570	2053	2512	4723	2968

Data of clones selected for additional characterization

This section shows the pooled data of the final clone selection which was used for further work. The final clone selection contains *E. coli* pRS415 clones of different screening rounds. All color coded clones were sequenced for additional characterization.

OD_{550} readings of final clone selection:

	1	2	3
A	0,009	0,019	0,026
B	0,015	0,026	0,032
C	0,028	0,015	0,034
D	0,032	0,020	0,040
E	0,012	0,024	0,033
F	0,009	0,020	0,034
G	0,008	0,021	0,000
H	0,008	0,020	0,002

OD_{420} readings of final clone selection:

	1	2	3
A	0,300	0,894	1,722
B	0,313	0,759	1,626
C	0,744	0,400	1,819
D	0,773	0,445	1,894
E	1,624	0,398	0,917
F	1,577	0,355	0,878
G	0,135	0,596	0,000
H	0,139	0,569	0,001

OD_{600} readings of final clone selection:

	1	2	3
A	0,2365	0,3955	0,3955
B	0,3475	0,4485	0,4725
C	0,5125	0,2265	0,5385
D	0,3615	0,2925	0,4035
E	0,0635	0,4285	0,5765
F	0,0545	0,4385	0,5655
G	0,1565	0,4025	0,0105
H	0,1415	0,3885	-0,0105

Miller Units of final clone selection:

	1	2	3
A	3998	7253	14126
B	2741	5302	11081
C	4519	5500	10884
D	6610	4664	15069
E	84194	2772	4970
F	95471	2441	4827
G	2576	4622	0
H	2948	4579	0

Color code for clones in final screening round:

	1	2	3
A	Clone1	Clone5	Clone9
B	Clone1	Clone5	Clone9
C	Clone2	Clone6	Clone10
D	Clone2	Clone6	Clone10
E	Clone3	Clone7	Clone11
F	Clone3	Clone7	Clone11
G	Clone4	Clone8	Blank
H	Clone4	Clone8	Blank

E. coli pRS415 clones selected for further characterization

E. coli pRS415 clones 1, 3, 5, 6, 10 and 11 were sequenced (Agowa, Berlin, Germany) using pRS415-fwd and LacZ-rev sequencing primers (see 2.3). The sequencing results are shown in section 3.2. Selected clones and corresponding Miller Units are shown in table 8.

Table 8: Selection of *E. coli* pRS415 clones with corresponding Miller Units which were used for continuous work in the random approach.

Clone	Miller Units
Clone1	5629
Clone 3	99275
Clone 5	8246
Clone 6	8220
Clone 10	11905
Clone 11	6345

3.2.2 Characterization of selected *E. coli* pRS415 clones based on β-galactosidase activity screening

The following section gives an overview of putative promoter regions identified on the genome of *R. eutropha* H16. The results base on blast searches conducted against the genome of *R. eutropha* H16 with pRS415 insert sequences of *E. coli* pRS415 clones 1, 3, 5, 6, 10 and 11. The orientation of identified DNA fragments on the pRS415vector is indicated by a black arrow pointing from 5' to 3'. *E. coli* pRS415 clones 1, 3 and 5 exhibiting the highest Miller Unit values were used for further characterization.

Clone 1

An average Miller Unit value of 5629 was determined for clone 1. The sequencing results of clone 1 revealed two blast hits on the genome of *R. eutropha* H16. Clone 1 was not used for further work due to unclear location of the putative promoter and low Miller Unit values (see table 8).

Blast hit 1: *R. eutropha* H16, chromosome 1, positive strand, basepairs 2706568 to 2707685

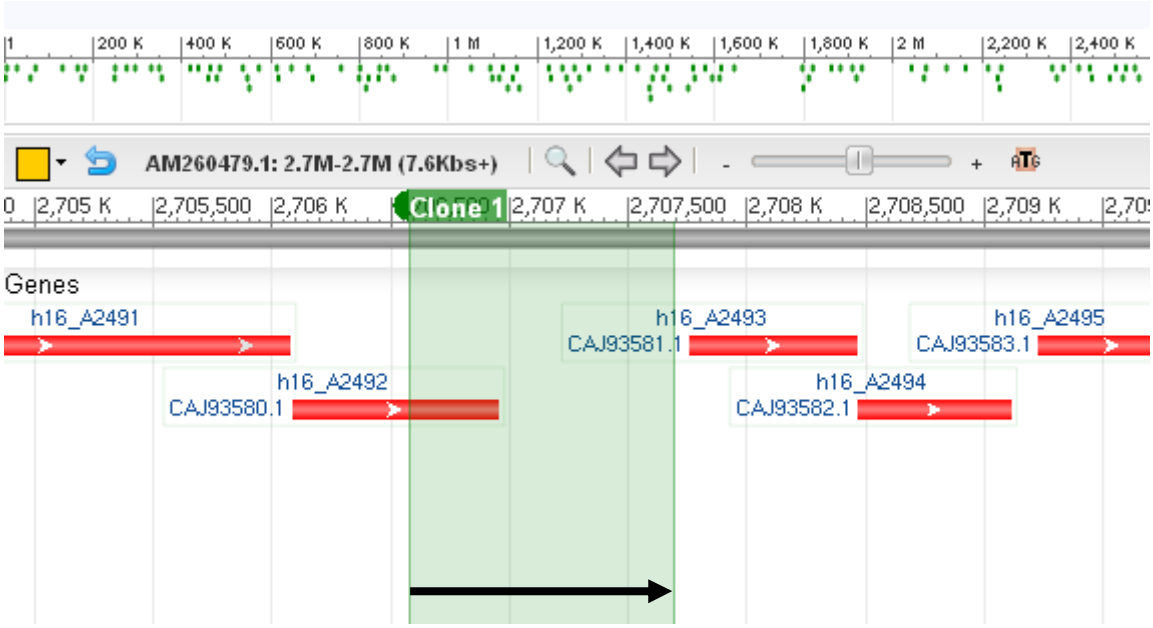

Figure 18: Illustration of clone 1, blast hit 1.*R. eutropha* H16, chromosome 1, positive strand, basepairs 2706568 to 2707685

Blast hit 2: *R. eutropha* H16, chromosome 1, positive strand, basepairs 2688569 to 2688615

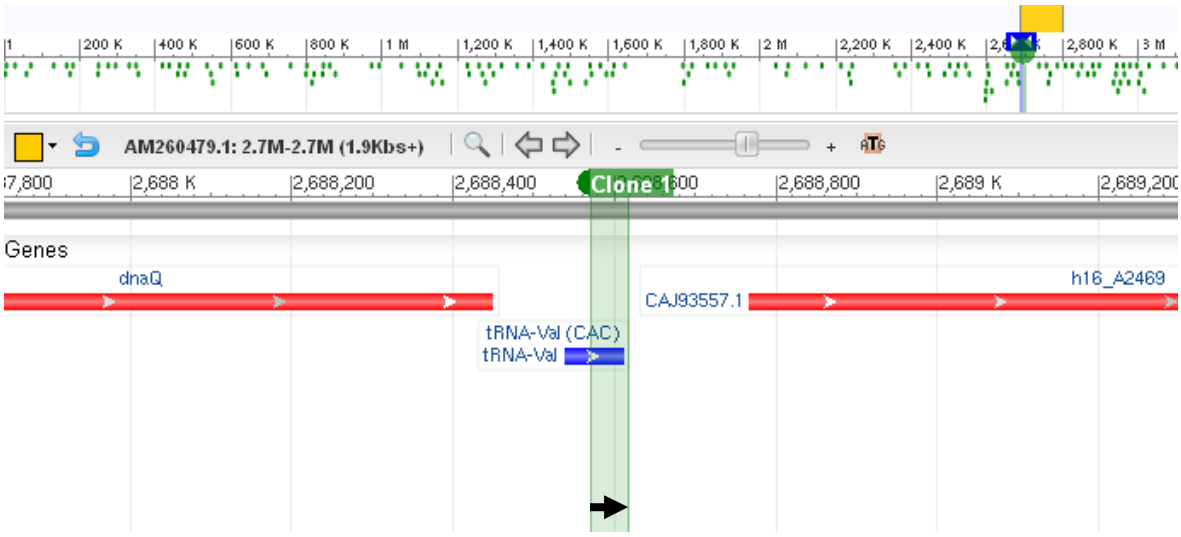

Figure 19: Illustration of clone1,blast hit 2. *R. eutropha* H16, chromosome 1, positive strand, basepairs 2688569 to 2688615

Clone 3

An average Miller Unit value of 99275 was determined for clone 3. The sequencing results of clone 3 revealed two blast hits on the genome of *R. eutropha* H16. Both DNA regions identified for clone 3 were located upstream of a predicted transcriptional regulator, one blast hit was found on chromosome 1 the other on chromosome 2 (h16_A1712; h16_B1772). Both regions were amplified by PCR and cloned into pKR-tac-PAR-EGFP for further characterization in *R. eutropha* H16 under lithoautotrophic and heterotrophic growth conditions.

Blast hit 1: *R. eutropha* H16, chromosome 1,positive strand, basepairs 1869061 to 1869196

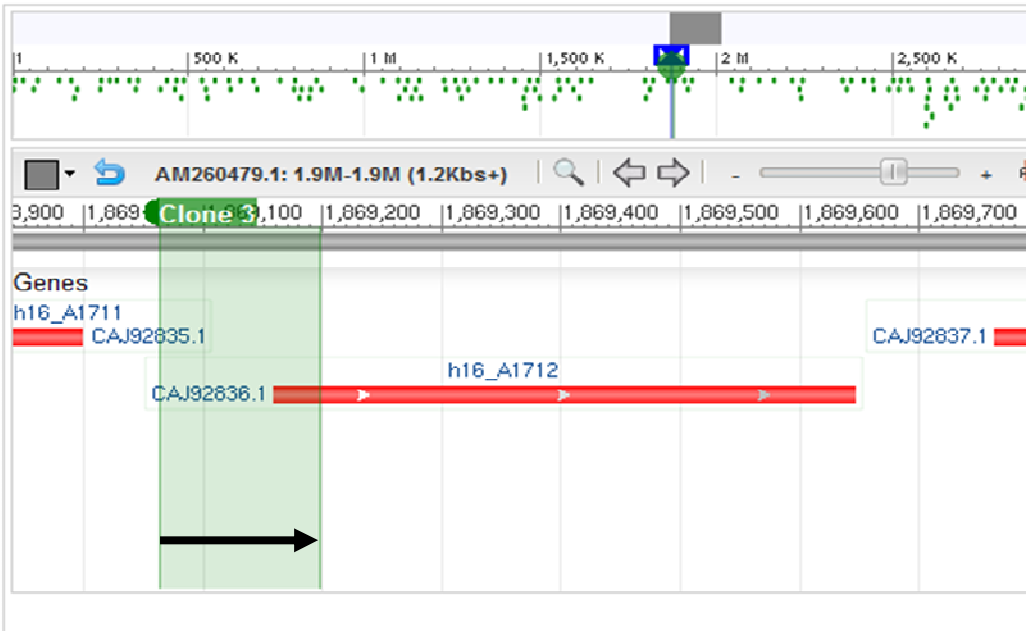

Figure 20: Illustration of clone 3 blast hit 1: *R. eutropha* H16, chromosome 1, positive strand, basepairs 1869061 to 1869196. h16_A1712 is described as a predicted transcriptional regulator.

Blast hit 2: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2013148 to 2012600

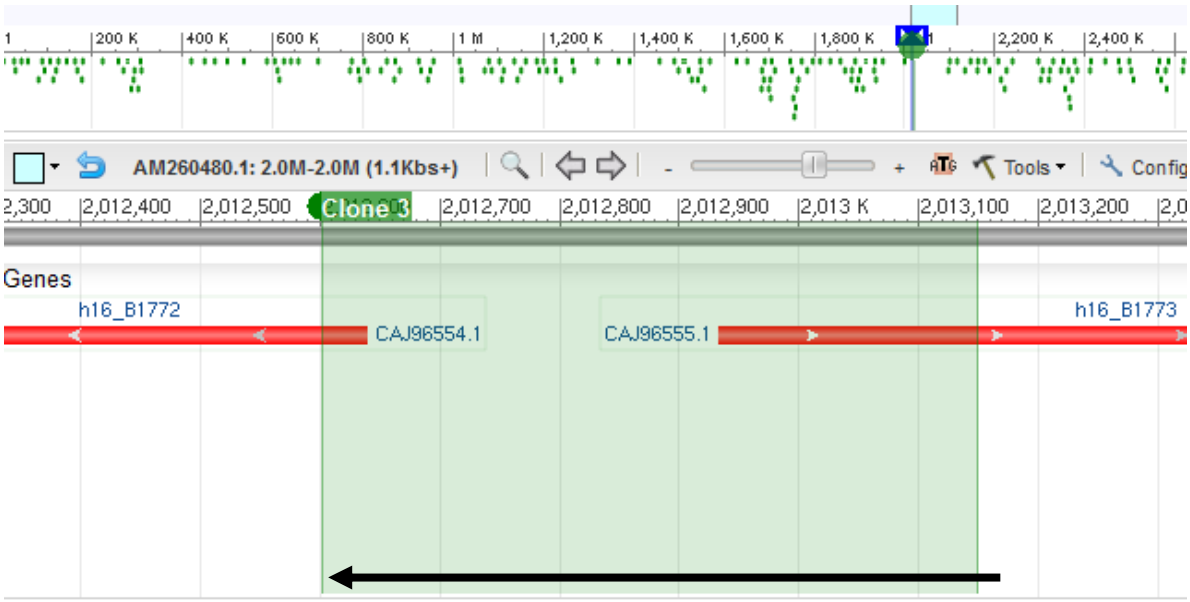

Figure 21: Illustration of clone 3, blast hit 2. *R. eutropha* H16, chromosome 2, negative strand, basepairs 2013148 to 2012600.h16_B1772 is described as a predicted transcriptional regulator.

Clone 5

An average Miller Unit value of 8246 was determined for clone 5. The sequencing results of clone 5 revealed one blast hit on the chromosome 2 of *R. eutropha* H16. The DNA region identified for clone 5 was located upstream of a predicted alpha-ketoglutarate-dependent taurine dioxygenase (h16_B1034). The DNA fragment was amplified by PCR and cloned into pKR-tac-PAR-EGFP for further characterization in *R. eutropha* H16 under lithoautotrophic and heterotrophic growth conditions.

Blast hit 1: *R. eutropha* H16*,* chromosome 2, negative strand, basepairs 1169503 to 1168709

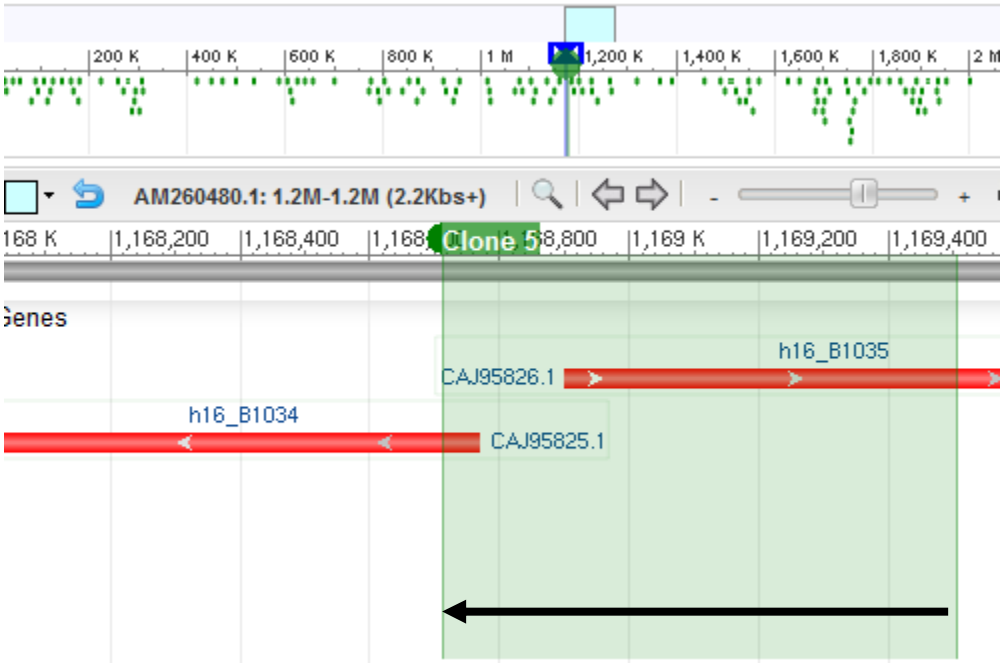

Figure 22: Illustration of clone 5, blast hit 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169503 to 1168709.

Clone 6

An average Miller Unit value of 8246 was determined for clone 6. The sequencing results of clone 6 revealed three blast hits on the genome of *R. eutropha* H16. Clone 6 was not used for further work due to unclear location of the putative promoter and low Miller Unit values (see table 8).

Blast hit 1: *R. eutropha* H16, chromosome 1, negative strand, basepairs 754241 to 754121

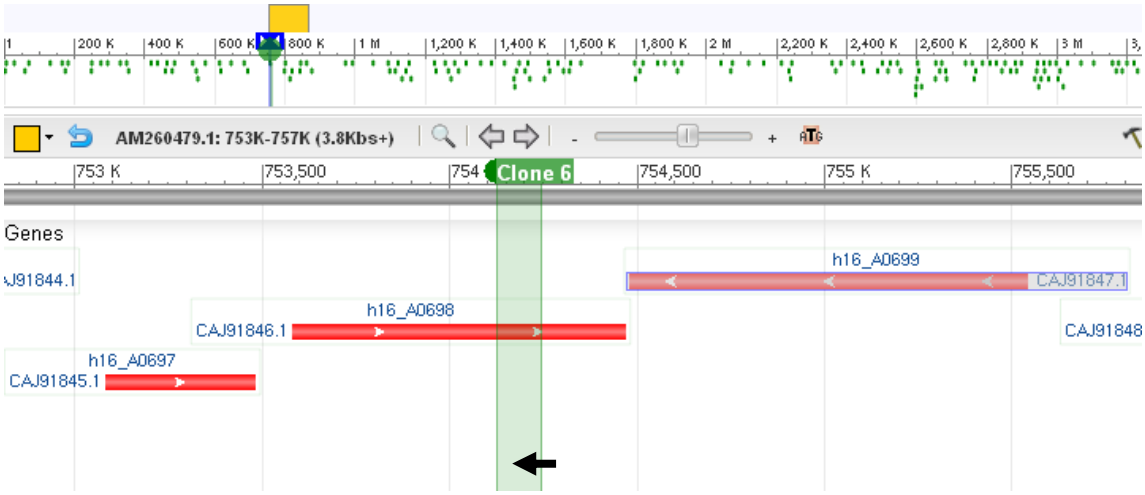

Figure 23: Illustration of clone 6, blast hit 1: *R. eutropha* H16, chromosome 1, negative strand, basepairs 754241 to 754121

Blast hit 2: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1871208 to 1871075

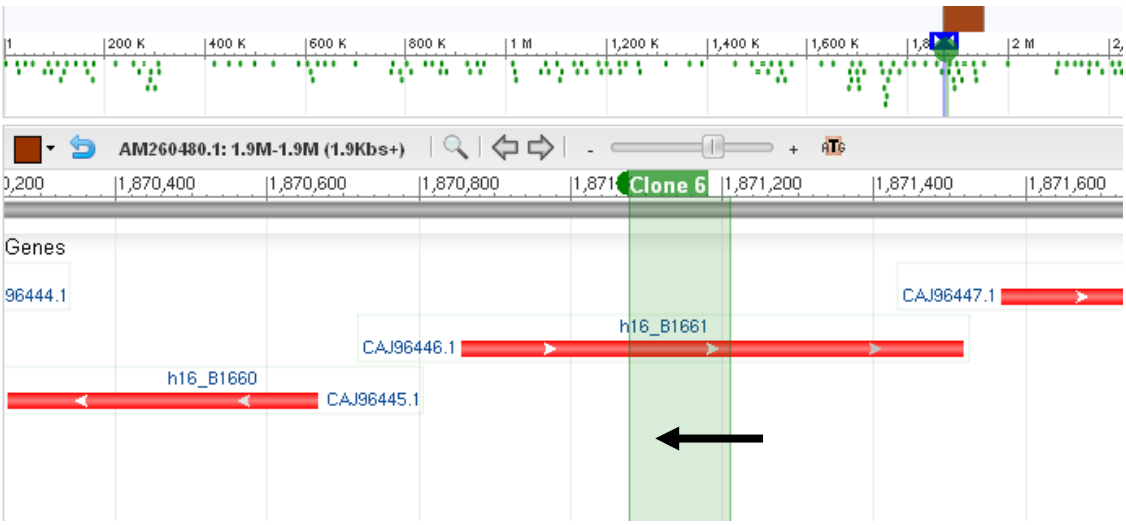

Figure 24: Illustration of clone 6, blast hit 2: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1871208 to 1871075

Blast hit 3: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2691170 to 2690927

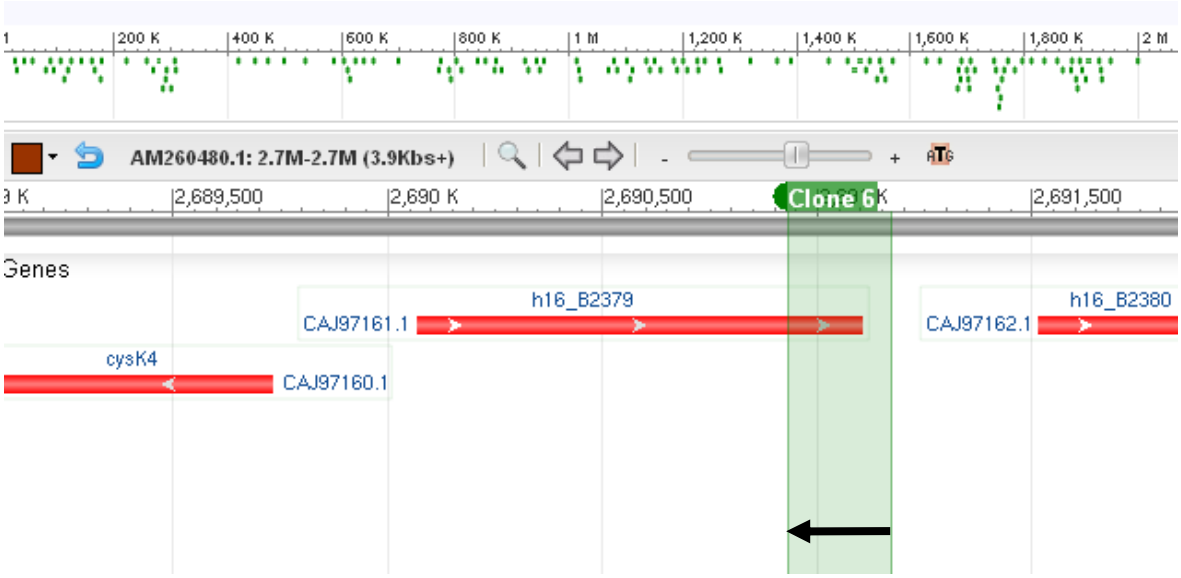

Figure 25: Illustration of clone 6, blast hit 3: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2691170 to 2690927

Clone10

An average Miller Unit value of 11905 was determined for clone 10. The sequencing results of clone 10 revealed one blast hit on the chromosome 2 of *R. eutropha* H16. The DNA region identified for clone 10 was located upstream of a putative uncharacterized protein (h16_B2204). The DNA fragment was amplified by PCR and cloned into pKR-tac-PAR-EGFP for further characterization in *R. eutropha* H16 under lithoautotrophic and heterotrophic growth conditions.

Blast hit 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502395

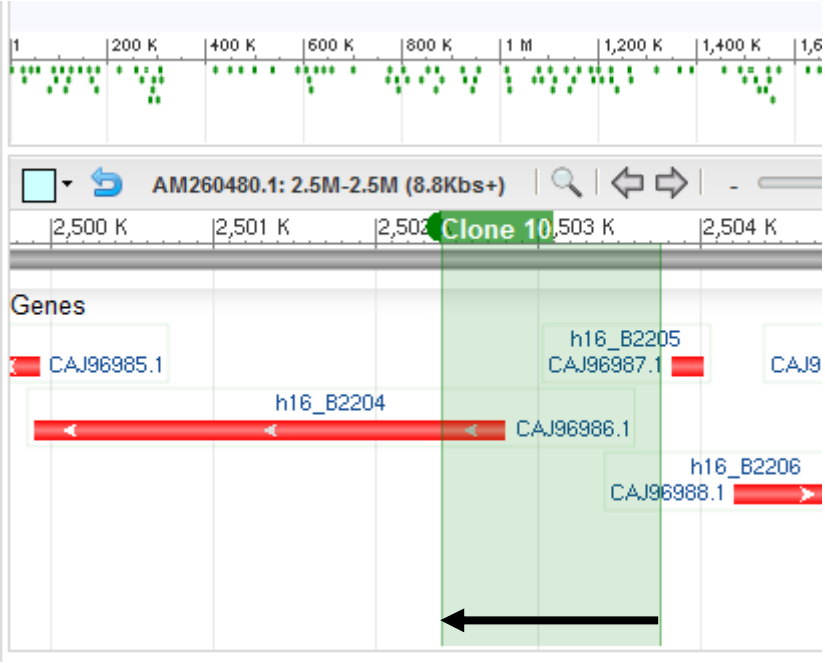

Figure 26: Illustration of clone 10, blast hit 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502395

Clone 11

An average Miller Unit value of 6345 was determined for clone 11.The sequencing results of clone 11 revealed seven blast hits on the genome of *R. eutropha* H16. Work with clone 11 was not continued due to unclear location of the putative promoter and low Miller Unit values (see table 8).

Blast hit 1: *R. eutropha* H16, chromosome 2, positive strand, basepairs 245767 to 246394

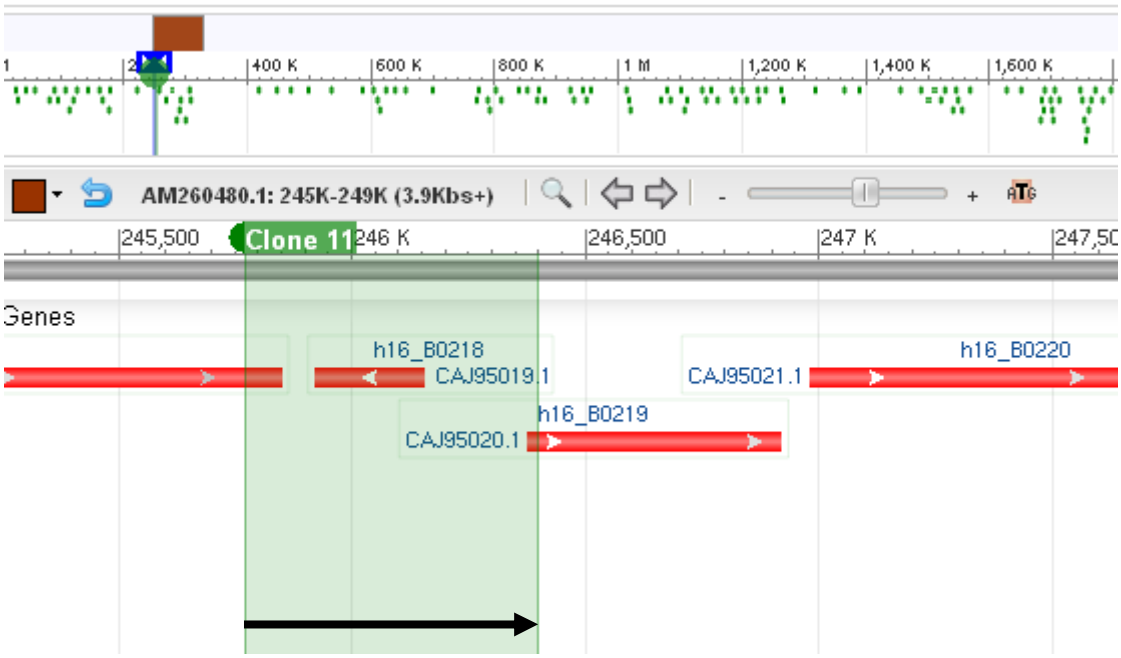

Figure 27: Illustration of clone 11, blast hit 1: *R. eutropha* H16, chromosome 2, positive strand, basepairs 245767 to 246394

Blast hit 2: *R. eutropha* H16, chromosome 1, positive strand, basepairs 3042471 to 3042678

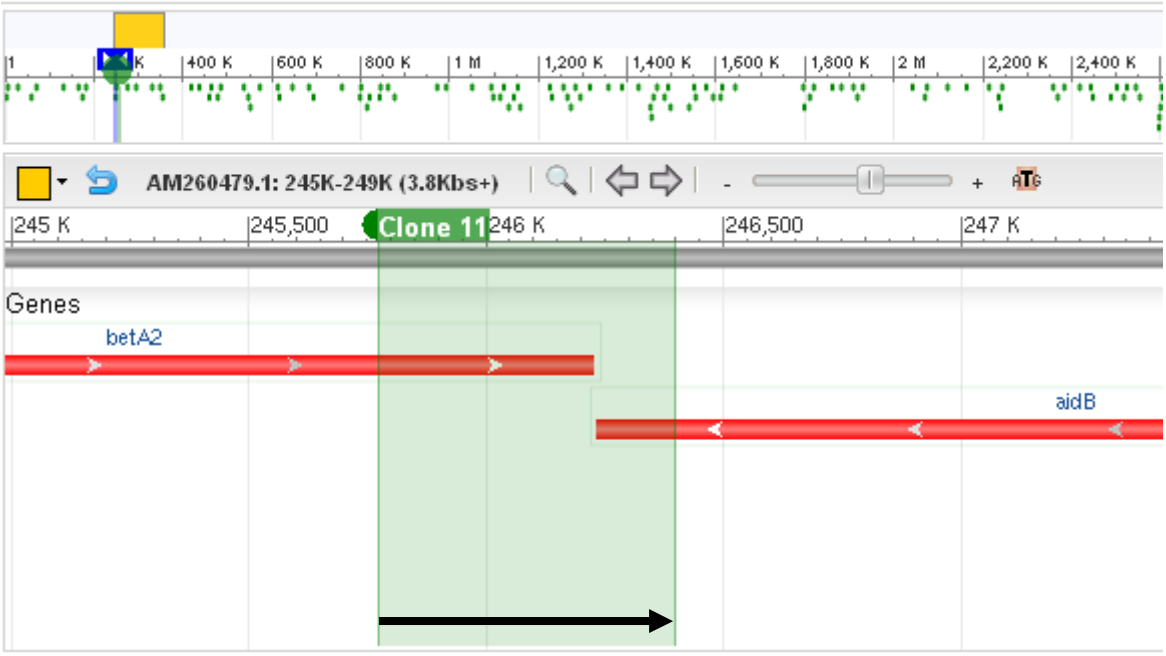

Figure 28: Illustration of clone 11, blast hit 2: *R. eutropha* H16, chromosome 1, positive strand, basepairs 3042471 to 3042678

Blast hits 3 to 7: All blast hits share the recurring motive on rrlB genes. Identified loci on the genome of *R. eutropha* H16:

3. chromosome 1, positive strand, basepairs 3575948 to 3577090

4. chromosome 1, positive strand, basepairs 3781285 to 3782427

5. chromosome 1, negative strand, basepairs 1810891 to 1809749

6. chromosome 2, negative strand, basepairs 179328 to 178186

7. chromosome 2, negative strand, basepairs 872058 to 870916

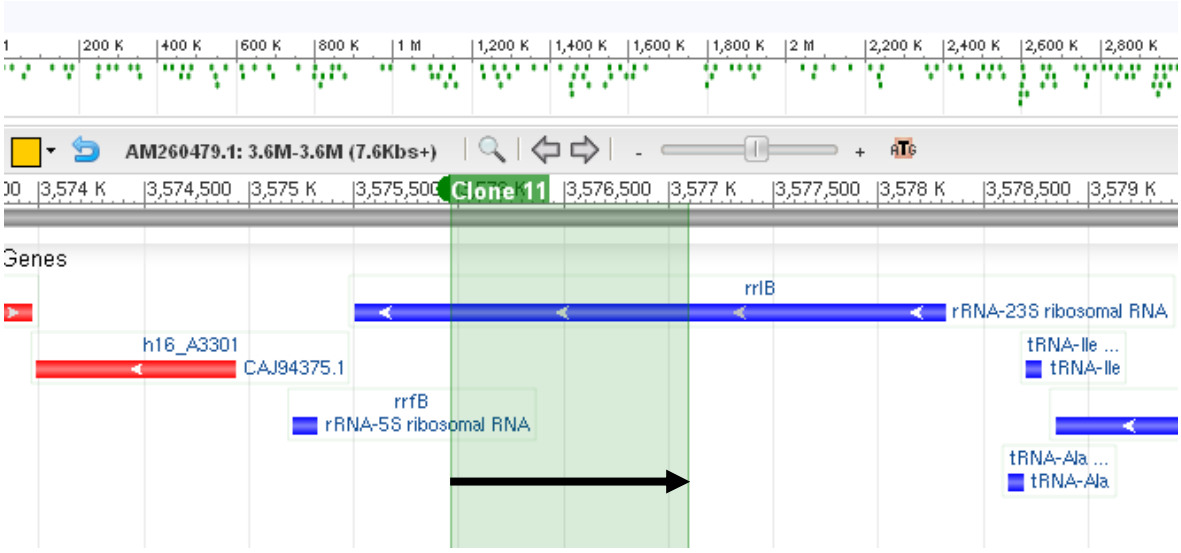

Figure 29: Illustration of clone 11, blast hit 3: *R. eutropha* H16 chromosome 1, positive strand, basepairs 3575948 to 3577090.

3.2.3 Amplification and characterization of putative promoters based on β-galactosidase activity screenings

E. coli pRS415 clones 3, 5 and 10 were selected for further work after sequence analysis due to defined locations of the putative promoters and high Miller Unit values (see table 8). The sequenced DNA regions of *E. coli* pRS415 clones 3, 5 and 10 were amplified by PCR, shown in the following section. Furthermore primer walking was performed for the selected sequences. The amplified DNA molecules were cloned into pKR-tac-PAR-EGFP via *Not*I and *Bam*HI restriction sites in order to determine promoter activity levels under lithoautotrophic and heterotrophic growth conditions in *R. eutropha* H16. The arrows listed in the figures indicate the length and orientation (from 5' to 3') of all amplified DNA fragments.

Clone 3

Putative promoter regions amplified from *R. eutropha* H16 genome based on sequence analysis of *E. coli* [pRS415 clone 3]. Two DNA fragments were amplified for *E. coli* [pRS415 clone 3] in order to contain the promoter region:

Fragment 1: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs2013138 to 2012630. Length: 548 bp.

Fragment 2: *Ralstonia eutropha* H16, chromosome 1, positive strand, basepairs 1869061 to 1869196. Length: 138 bp.

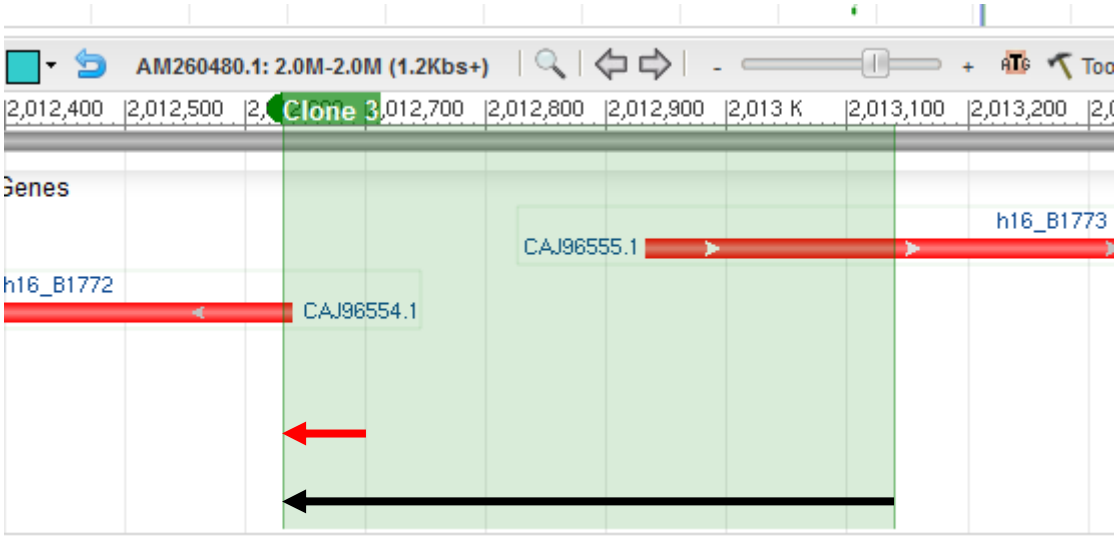

Figure 30: Illustration of PCR amplified regions based on sequencing results of *E. coli* [pRS415clone 3]. Fragment 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2013138 to 2012630. Fragment 2: *R. eutropha* H16, chromosome 1, positive strand, basepairs 1869061 to 1869196. Length: 138 bp.

Clone 5

Putative promoter regions amplified from *R. eutropha* H16 genome based on sequence analysis of *E. coli* [pRS415 clone 5]. Four DNA fragments were amplified for *E. coli* [pRS415 clone 5] in order to contain the promoter region:

Fragment 1: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 1169503 to 1168709. Length: 1000bp.

Fragment 2: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 1169258 to 1168709. Length: 549 bp.

Fragment 3: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 1169010 to 1168709. Length: 301 bp.

Fragment 4: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 1168876 to 1168709. Length: 167 bp.

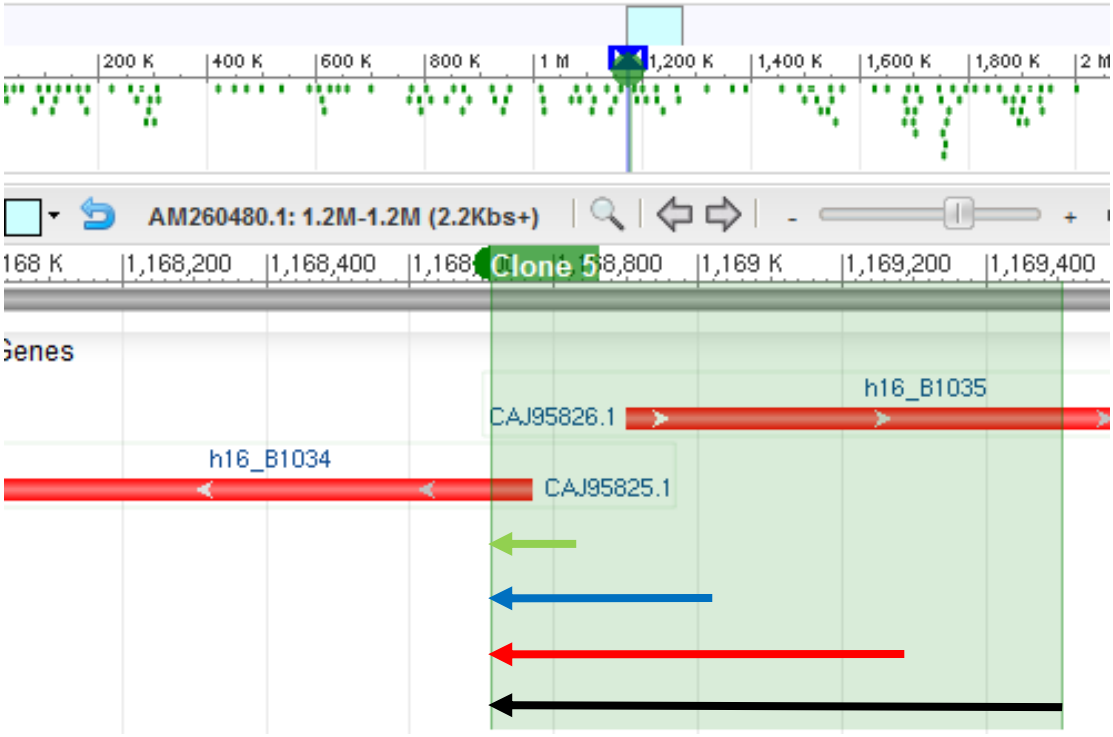

Figure 31: Illustration of PCR amplified regions based on sequencing results of *E. coli* [pRS415 clone 5]. Fragment 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169503 to 1168709 Fragment 2: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169258 to 1168709. Fragment 3: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169010 to 1168709. Fragment 4: *R. eutropha* H16, chromosome 2, negative strand, basepairs 1168876 to 1168709.

Clone 10

Putative promoter regions amplified from *R. eutropha* H16 genome based on sequence analysis of *E. coli* [pRS415 clone 10]. Four DNA fragments were amplified for *E. coli* [pRS415 clone 10] in order to contain the promoter region:

Fragment 1: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502787

Fragment 2: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 2503400 to 2502787

Fragment 3: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 2503141 to 2502787

Fragment 4: *Ralstonia eutropha* H16, chromosome 2, negative strand, basepairs 2502955 to 2502787

Figure 32: Illustration of PCR amplified regions based on sequencing results of *E. coli* [pRS415 clone 10].Fragment 1: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502787. Fragment 2: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503400 to 2502787. Fragment 3: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503141 to 2502787. Fragment 4: *R. eutropha* H16, chromosome 2, negative strand, basepairs 2502955 to 2502787

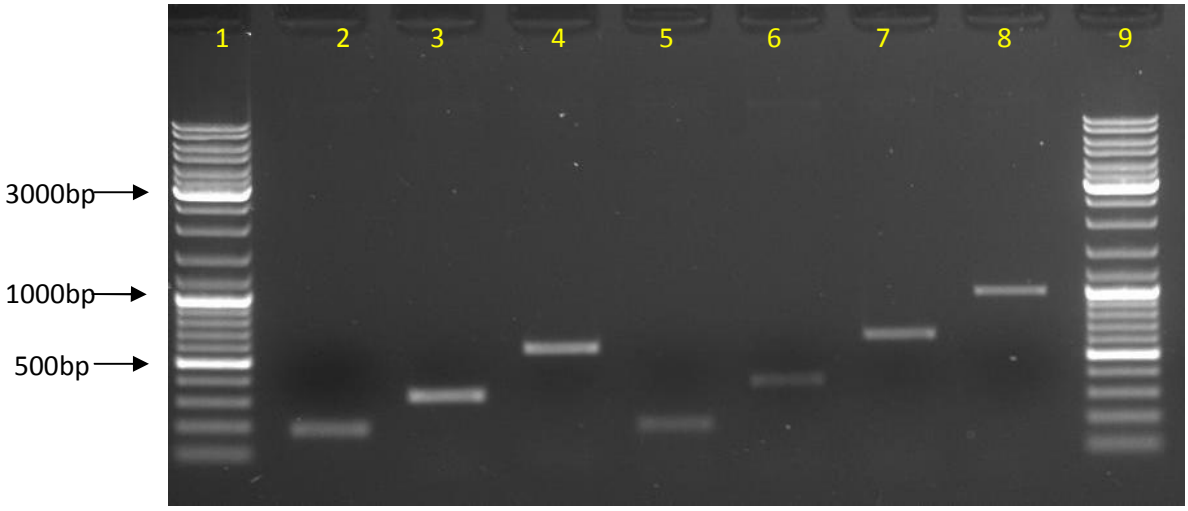

Figure 33: 1% agarose gel of PCR products gained from primer walking based on *E. coli* pRS415 clone 5 and 10. 1 + 9:500 ng of Gene Ruler DNA Ladder Mix, 2: *E. coli* pRS415 clone 5 fragment 1; 3: *E. coli* pRS415 clone 5 fragment 2; 4: *E. coli* pRS415 clone 5 fragment 3; 5: *E. coli* pRS415 clone 10 fragment 1; 6: *E. coli* pRS415 clone 10 fragment 2; 7: *E. coli* pRS415 clone 10 fragment 3;8: *E. coli* pRS415 clone 10 fragment 4.

All selected regions obtained from *E. coli* pRS415 clones 3, 5 and 10 (see figure 30 - 32) were amplified by PCR and cloned into pKR-tac-PAR-EGFP via *Not*I and *Bam*HI restriction sites. In a first step pKR-tac-PAR-EGFP clones were introduced into *E. coli* where they exhibited EGFP activity, indicating the presence of a promoter region in all amplified PCR products. However, EGFP activity could not be detected in *R. eutropha* H16.

3.3 Direct approach

3.3.1 Characterization of preselected putative promoter regions on *R. eutropha* H16 genome

The DNA regions selected for promoter screening in *R. eutropha* H16 were based on increased transcription levels of genes on the genome of *R. eutropha* H16 under lithoautotrophic growth conditions (see table 7) [7]. An approximately 1000

bp 5' upstream region of PHG027, SerS1, GyrA and H16_A1526 was amplified by PCR and cloned into pKR-tac-PAR EGFP with *Not*I and *Bam*HI for additional screening of promoter activity under lithoautotrophic and heterotrophic growth conditions in *R. eutropha* H16. The black arrows listed in the figures indicate the length and orientation (from 5' to 3') of the amplified DNA fragments.

PHG027

Based on homology models PHG027 codes for fructose-1,6-bisphosphate /seduheptolose-1,7-bisphosphatase and is most likely active as a part of the calvin cycle in *R. eutropha* H16. PHG027 was upregulated more than 1.6 times under lithoautotrophic growth conditions compared to heterotrophic growth conditions in *R. eutropha* H16 (see table 7). The following region was amplified to screen for promoter activity under lithoautotrophic conditions in *R. eutropha* H16.

R. eutropha H16, megaplasmid, positive strand, basepairs24972 to 25947. Length: 975 bp.

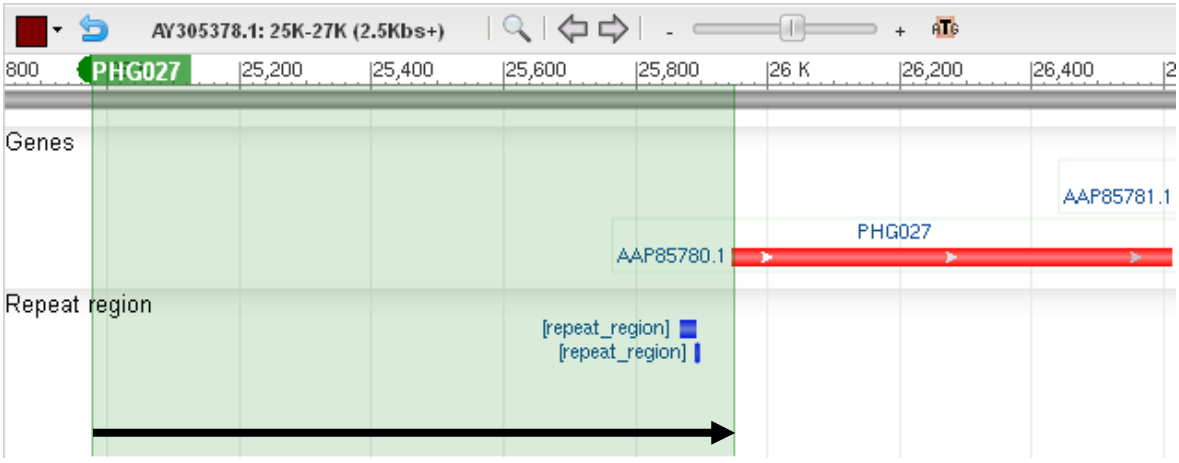

Figure 34: Illustration of region of amplification upstream of *phg027* locus containing putative promoter*: R. eutropha* H16, megaplasmid, positive strand, basepairs 24972 to 25947.

H16_A1526

Based on homology models H16_A1526 encodes for a Enoyl-CoA-hydratase/Delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase. H16_A1526 was upregulated more than 5.1 times under lithoautotrophic growth conditions compared to heterotrophic growth conditions in *R. eutropha* H16 (see table 7). The following region was amplified to screen for promoter activity under lithoautotrophic conditions in *R. eutropha* H16.

R. eutropha H16, chromosome 1, positive strand, basepairs 1658248 to 1659248. Length: 1000 bp.

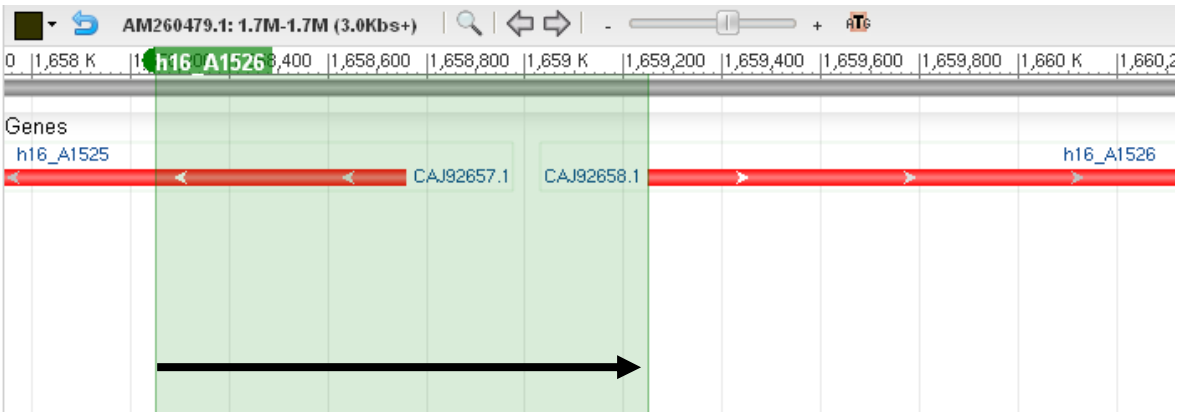

Figure 35: Illustration of region of amplification upstream of *h16_A1526* locus containing putative promoter*: R. eutropha* H16, chromosome 1, positive strand, basepairs 1658248 to 1659248.

lon

Lon codes for ATP-dependent Lon protease, a serine peptidase. *Lon* was upregulated more than 1.7 times under lithoautotrophic growth conditions compared to heterotrophic growth conditions in *R. eutropha* H16 (see table 7). The following region was amplified to screen for promoter activity under lithoautotrophic conditions in *R. eutropha* H16.

R. eutropha H16, chromosome 1, positive strand, basepairs 1610548 to 1611548. Length: 1000 bp.

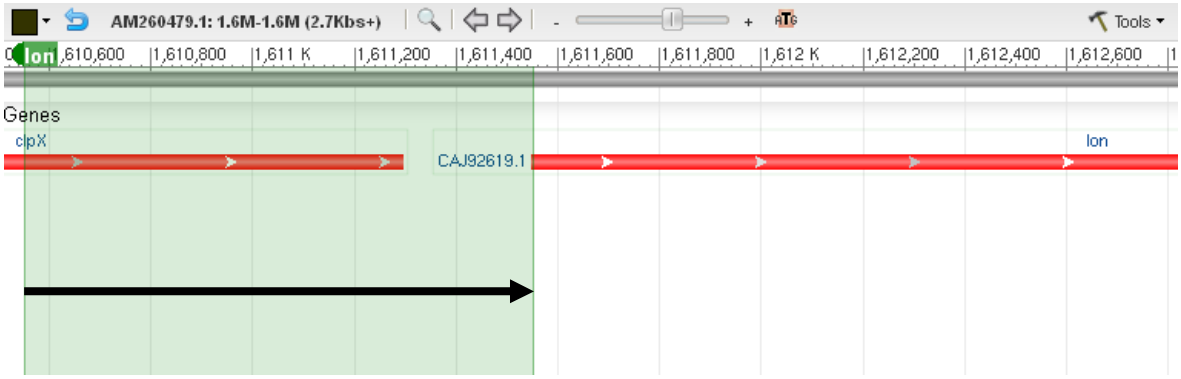

Figure 36: Illustration of region of amplification upstream of *lon* locus containing putative promoter: *R. eutropha* H16 chromosome 1 positive strand, basepairs 1610548 to 1611548.

gyrA

GyrA coding for the DNA gyrase A subunit of DNA gyrase, an enzyme that negatively supercoils closed circular double-stranded DNA. *GyrA* was upregulated more than 2.0 times under lithoautotrophic growth conditions compared to heterotrophic growth conditions in *R. eutropha* H16 (see table 7). The following region was amplified to screen for promoter activity under lithoautotrophic conditions in *R. eutropha* H16.

R. eutropha H16, chromosome 1, positive strand, basepairs 859088 to 860088. Length: 1000 bp.

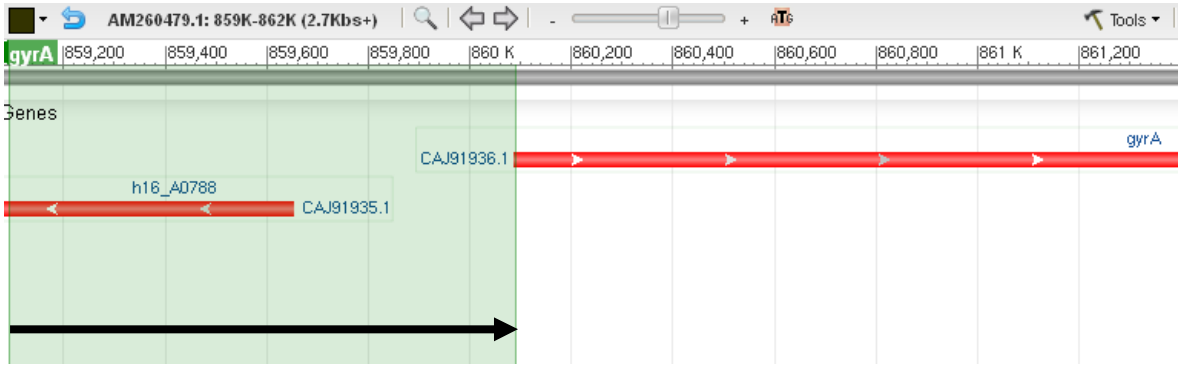

Figure 37: Illustration of region of amplification upstream of *gyrA* locus containing putative promoter: *R. eutropha* H16 chromosome 1 positive strand, basepairs 859088 to 860088.

serS1

SerS1 encodes a Seryl-tRNA synthetase 1.*SerS1* was upregulated more than 3.3 times under lithoautotrophic growth conditions compared to heterotrophic growth conditions in *R. eutropha* H16 (see table 7). The following region was amplified to screen for promoter activity under lithoautotrophic conditions in *R. eutropha* H16.

R. eutropha H16, chromosome 1, positive strand, basepairs 829233 to 830233. Length: 1000 bp.

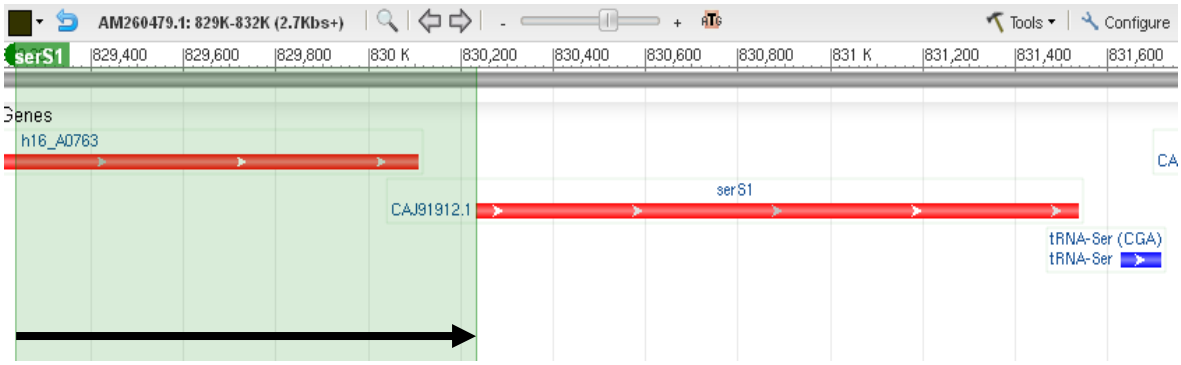

Figure 38: Illustration of region of amplification upstream of *serS1* locus containing putative promoter: *R. eutropha* H16 chromosome 1 positive strand, basepairs 829233 to 830233.

All selected regions (see figures 34 – 38) were amplified by PCR and cloned into pKR-tac-PAR-EGFP via *Not*I and *Bam*HI restriction sites. In a first step pKR-tac-PAR-EGFP clones were introduced into *E. coli* where they exhibited EGFP activity, indicating the presence of a promoter region in all amplified PCR products. However, EGFP activity could not be detected in *R. eutropha* H16.

3.4 DNA sequences of identified promoters

The following section illustrates the sequences of the identified promoter regions derived from the random approach and the direct approach. The sequences were analyzed by applying the following promoter prediction software tools:

Berkeley Drosophila Genome Project: Neural Network Promoter Prediction (Berkeley, CA,USA) [52]

Softberry: BPROM (Softberry, Mount Kisco, NY, USA) [53]

University of Groningen: Prokaryote Promoter prediction (University of Groningen, Groningen, Netherlands) [54]

The hereby identified -10 hexamers were labeled yellow, the -10 extended box was labeled grey, -35 hexamers were labeled red, and the start ATG of the subsequent gene is indicated in green. Anticipated Shine-Dalgarno sequences were labeled in turquoise. Promoter regions could be predicted for all selected regions originating from the random and direct approach.

3.4.1 DNA sequences of identified promoters based on the random approach

Clone 3 fragment 1

A σ^{70} based promoter could be identified for *E. coli* [pRS415 clone 3] fragment 1 by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. BPROM revealed a score of 43 out of 77 for the -10 box and a score of 33 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.91 out of 1. A Shine-Dalgarno sequence could be anticipated 6 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The ATG marked in green resembles the first amino acid coding for the subsequent gene *h16_B1772* located on *R. eutropha* H16*,* chromosome 2, negative strand, basepairs 2013148 to 2012600. *H16_B1772* codes for a predicted transcriptional regulator.

From basepair 2013148 to 2012600 on chromosome 2:

AATTCGCCGAGATCATTGCCGGCCGTAAAGCTGTCCCCGTCCCCCTGCAGCAGCAA
TACCCGGATAGCCCGGTCCTGCTGTGCCTGCTCGATGGTGTCGGCAAGTGCGCCGT
ACATCTCGTTCGTCAAGGCGTTCATCTTGTCCGGCCGCGCAAGCGTCATGGTCAGCA
CGCCGCCGGTGAGGTCAATGCGGATTTCGGGGGTCATTGGTCGGTCTCCAGGTGAG
AGTTGCACGCTGCCGTGCAGGCATCGCGATTAAGTTCGCATGCGAAGTATATAGATA
GTTCGTGTGCGAAGTAAACGGTTGAGGCCGGGCAGTGTGCAGAATCGGCGTCTGGG
CGGCGGTGCCCTCCACATAGACGGTGTGGGCAGGGGCGTGCTTCCGGTGCCTCGC
GGGCGCGGAATACAGCACAAGCGAAAAACGAACTTGCACGATGATAGTGACTACCC
TAGACTGACTTTCAAATCGATAGCAAGCAGGTCTCTCGAGGAAGCCAATCATGCAGA
AGA

Figure 39: Illustration of the promoter sequences identified for *E. coli* [pRS415 clone 3] fragment 1 The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha H16,* chromosome 2, negative strand, basepairs 2013148 to 2012600.

Clone 3 fragment 2

A σ^{70} based promoter could be identified for *E. coli* [pRS415 clone 3] fragment 2 by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. BPROM revealed a score of 43 out of 77 for the -10 box and a score of 33 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.91 out of 1. A Shine-Dalgarno sequence could be anticipated 6 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The ATG marked in green resembles the first amino acid coding for the subsequent gene *h16_A1712* located on *R. eutropha* H16*,* chromosome 1, positive strand, basepairs 1869061 to 1869196. *H16_A1712* codes for a predicted transcriptional regulator.

From basepair 1869061 to 1869196 on chromosome 1:

GGGCGTGCTTCCGGTGCCTCGCGGGCGCGGAATACAGCACAAGCGAAAAACGAAC
TTGCACGATGATAGTGACTACCCTAGACTGACTTTCAAATCGATAGCAAGCAGGTCT
CTCGAGGAAGCCAATCATGCAGAAGA

Figure 40: Illustration of the promoter sequences identified for *E. coli* [pRS415 clone 3] fragment 2 The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16*,* chromosome 1, positive strand, basepairs 1869061 to 1869196.

Clone 5

Two σ^{70} based promoters could be identified for *E. coli* [pRS415 clone 5] by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The predicted σ^{70} promoter distant to the start ATG revealed BPROM scores of 41 out of 77 for the -10 box and a score of 26 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.89 out of 1. The predicted σ^{70} promoter closer to the start ATG revealed BPROM scores of 23 out of 77 for the -10 box and a score of 27 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.84 out of 1. A

Shine-Dalgarno sequence could be anticipated 8 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *h16_B1034* located on *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169503 to 1168709. *H16_B1034* encodes a predicted alpha-ketoglutarate-dependent taurine dioxygenase.

From basepair 1169503 to 1168709 on chromosome 2:

TCGAGATCGGCCTTCGCGCAGATGATGCAGTTGAAGTCATAGTCGAACTCGACGCAA
GGGCCAAAAAACCGTTGGTGCGTACCAAGGTCGCGCGGCGCCGGGTGCTCGAAGC
ATACGCGCCGCGGCTCCCAGTCGGGCTTGAGAAGCTGGCGTATCAGCCGCACCATT
ACCCCCAGCGCCAGTTCAACCCTTTGGCGCGTGGGCTGATGTGCGCTGCCAGCGAT
GACCGCCTCGCGAATAATGACCAGTTCGCCGCACTCCTCTACCGCCAGCGACAGCG
CACCGTTGAGCAACGCCTGATAGCGCATCAGCATGCGAAGCGAGTCGCGCAGCGTC
GCCTGGTCGCGGATCAGCAATCCTACGGCCCCAAGATTGGACAACAGGCGCGACTC
GGCCATGCACAGTCCGAAACTCTCGTTGCCTGACATGGTTGCCGACGCCTGCAGCA
GTCGTCCGACACGCTCAACCGGAATCATGAGGTCCGGCTCGCGCAGCACGCCCGG
ACTCAAACCCGCGTCAAGGAGCATCCGCACCGGGTCCAGTCCGGCGGCTCGGGCG
ACCTCGCTGTAGTTGGTGAGGGCCGCAGCGCGGACAAGTGACGACATTAGCGAGCA
GAGTGGAACATTGACATGGAATGATAAATCCCGGTCATGTAGTGTCAAGCGGGAGAA
ACCCGCTGGATCCAGACTCGCAGGAATCGATCGGCCTGTTCGAAAAATCGGAGACA
ACCTTCATGATCCAACGCCTCAAAGCGGCACCCGCAGTGGTCCGCGCAACCTCTCT
GGGCAATTCGAT

Figure 41: Illustration of the promoter sequences identified for *E. coli* [pRS415 clone 5]. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 2, negative strand, basepairs 1169503 to 1168709.

Clone 10

Two σ^{70} based promoters could be identified for *E. coli* [pRS415 clone 10] by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The predicted σ^{70} promoter distant to the start ATG revealed BPROM scores of 11 out of 77 for the -10 box and a score of 34 out of 66 for the -35 box, the Neural Network Promoter Prediction tool did not predict a promoter at this region. The predicted σ^{70} promoter close to the start ATG revealed BPROM scores of 39 out of 77 for the -10 box and a score of 66 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.84 out of 1. A Shine-Dalgarno sequence could be anticipated 6 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *h16_B2204* located on *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502395. *H16_B2204* codes for a putative uncharacterized protein.

From basepair 2503750 to 2502395 on chromosome 2:

GATCCTTTTTATGTTCCCCGCCAGCGCATTCCATATGACTTCGATATGTCGCTGTCGT
GACAACCGAAAACGTCACGGAACGACAACACCAAAAGACACACCGCATGGGCGCTG
CACAAGTGGGAATCATTCTGCCGAATTTTCCCGGCGTGTCAATCAGGCTAAGCGCCT
AATCATGATGAGAGGTTCCGATAGGAGATATTAGCGTTTTCTACTTCAGCTTCAGTTC
AGTGAATTCCCTGATTTTTTTATCGCGGGTTTGGTTGTGCGTGCACCGCCAGCGAGC
GCGATGCCACACCGCCGTGCATAATGGCGGTCGTCCGGTGCATGGGCGTGCCATCT
TTTTTACAAATTCTTACGGCACCCAATCCCTACCGCGTTCGGGGGAAGCCGGCGCG
GCCCGGATATATCATTATGCTGACAATAAAGGCCAGGCCGGTGGCCGCGCTCGCGT
CGTGCACAGACATACATGAGCGCGCCCACTGTATGCGGGCCCATACATAAAAGCTC
GGAACGGGGAATCCATCATGGCGTTTCAAACCTAGCTGTCCGGAATTTCGGTTGCG
GACAGCCCCACCACCCGTCGCGCACGGGACCTGGACTGTTTTCGTTGTTTGCATTTG
AGCCGAGCCGAGCCGAGCCGCGCCCGCCGTACCGCAAGCGCGGCTGACGTGCGC
CCGCAGCACGGTGGCGCGCGTTAACGCGCGAGCCGGCTCAATGTCTCAGTGTTGG

GACAATTGTTGACAATGTGATCCATGCGTCTGTCCTAGCATTCGTTTCCGGATACCAT
CAGATTTTTCCGATTCTGAGATCCACAGTCGGACACAAGGCGGCAGCGACCGCCTC
AGGGAATGTCGTGCGTGAGGCACTTGCCGGCGTGCGCAGACAAGAGGAAGTAAGC
CTGTGAAGCAGCTGGAGGAAGCACAAGGGAACCGGGGCGGCCGCATGCCGTTCCG
ATCGGGGT

Figure 42: Illustration of the promoter sequences identified for *E. coli* [pRS415 clone 10]. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 2, negative strand, basepairs 2503750 to 2502395.

3.4.2 DNA sequences of identified promoters based on the direct approach

PHG027

Several σ^{70} based promoters could be identified for the 5' upstream region of PHG027 by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The predicted σ^{70} promoter distant to the start ATG revealed BPROM scores of 44 out of 77 for the -10 box and a score of 25 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the predicted promoter region of 0.82 out of 1. Another promoter region was predicted with a score of 0.84 out of 1 by the Neural Network Promoter Prediction tool immediately after the first predicted promoter. The predicted σ^{70} promoter close to the start ATG revealed BPROM scores of 21 out of 77 for the -10 box and a score of 55 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not indicate a promoter at this region. A Shine-Dalgarno sequence could be anticipated 8 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *phg027* located on *R. eutropha* H16, megaplasmid pHG1, basepairs 24972 to 25947. *Phg027* codes for a fructose-1,6-bisphosphate/seduheptolose-1,7-bisphosphatase.

From basepair 24972 to 25947 on megaplasmid pHG1:

ATGTAACACTTGCCGATGTGGCGAGACTGTACAACCCAGTGATACAGGGGTGGTGG
AATTACTATGGCGCGTTCTACAAGACTGCGATGCTTCACATCTTCCAGCACATCGACT
GCGCTCTTGAGCGTTGGGCCCGACGGAAGTACAAGGGTCTGCACAGGCGCAAGCG
ACGAAGTGCCGAATGGCTGGACAGGATGCGCCGTAACAACCCGATGCTGTTCGCTC
ACTGGCGTGTGGTTACGCAACCGGCTGGATAACGGGAGCCCTATGACGCGAGAGTG
TCACGTACGGTTCTGCGAGAGGCTCGGGGGGCGGTTCCCCGGGTCTACTTACCACT
TCGGCATGAAGCTCCACATTGGCGTGGACAGTCAGATTGGGTTGGCCCATTCGGCA
GTCGTAACGCCAGCCAATGTTCATGACAAGCATGCAGTGCACCAGCTATTGCATGGC
AGGAGCAGCGAGTCTATGGCGATAGCGCCTACGCCAGTCAGCAGGCGCTGATCGC
GTCCAAGGCGCCCCGCGCCAAGGACTTCACCAATCAGCGTACGCGCAAGAGCGGG
CAGATTGACGAGGTGCAACGTGGGAAGAACCGCAACAAATCGAAGATCCGTGCTCG
GGTCGAGCATGTGTTTGCCGTAGTCAAACGCCTGTGGGGCTTCACTAAAGTACGCTA
CCGTGGCTTGAAGAAGAATGCCGGCCGGGCCTTCACGGCCCTGGCGCTGGCCAAT
CTGTACCTCAGCCGAGGGCATTTGATGGGAGCAGTCCGTCCGTAAAGGGCGAATGT
GGGGCAAGTGCTCCGCCAACACGCCCCATTTGGGATGCGGCAGGAAAAATTCGCGT
CATCAGAGCCGAGTTTTTCGACCGCACACGTTCTTATCAGTCTGAACACACCAATTCA
CGCGGCTTGTTCAGCATAGCCTTAGTCGCCAGCCGTACTCCGTAGAAGACCAACTTG
CCTCAAGCAATCTTTGGAGATG

Figure 43: Illustration of the promoter sequences identified for the 5' upstream region of PHG027. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, megaplasmid pHG1, basepairs 24972 to 25947.

H16_A1526

One σ^{70} based promoter could be identified for the 5' upstream region of H16_A1526 by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The predicted σ^{70} promoter revealed BPROM scores of 58 out of 77 for the -10 box and a score of 38 out of 66 for the -35 box. The Neural Network Promoter Prediction tool revealed an overall score for the

predicted promoter region of 0.94 out of 1. A Shine-Dalgarno sequence could be anticipated 6 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *h16_A1526* located on *R. eutropha* H16, chromosome 1, basepairs 1658248 to 1659248. *H16_A1526* codes for a Enoyl-CoA-hydratase/Delta(3)-cis-delta(2)-trans - enoyl-CoA isomerase.

From basepair 1658248 to 1659248 on chromosome 1:

TCAGCGCGAACCAGTACTCGGCCTCGCCGAAGCCGCGCACCGAGAACGCATTGAG
CCCGAACATCAGCAGCAGGAACGCGGCGCTCCAGACCACGCCGGGCATTTCGGGG
AACCAGTACTGCATCACCAGCTGCGCCGCGGTCAGTTCCACGGCGATGGTCACCGC
CAGGCTGAACCAGTAGTTCCAGCCCAGCGCGAAGCCGAAGCCTTCGTCGACATAGA
GCGCGCTGTAGGTGACGAAGGAGCCGGCCACCGGCATATGCACCGCCAGCTCGCC
CAGGCTGGTCATCAGGCAATAGACCATCAGCCCGAGCAGCATATACATCAGCAGCG
CGCCGCCGGGGCCGGCCTGCGCGATCGAGGCGCCCGAGGCCACGAACAGCCCGG
TGCCGACCGCGCCGCCGATGGCAATCATGCTCAGGTGGCGGGCCTTGAGCCTGCG
CTTGAGGTGGTCGTGCTCGACCACCGGGTGCTGGTGCTCGTGGGCGTGGGAATGT
GAGGGGGAAGACATCAGGGTCGCTGGACCTTCCGCCGATCCGGCATCTTGTGGAAC
CCGGACGGCAAAAAACGAAAAGCCGCGAGTTTACAGGGTCGGCCCGCGCGATTTGT
CCGTGTCGGCCAAATAAAGTCCGATACTGGTTTAGAGGCGGCACTCAGTCGCACTC
GGCAAGTGCTTTATCCGCCATGGCGGGACATTTGGCACCGCGTCAGGCGCCCCGCC
ATGCATCGCGGTCAGGGGCGGCATTAGCGCCGCGCAGTCGCCGGAAAAGTCCGAA
AATTAAAATTGAACGACCGTGCTATTTTGTGTATGCTTGTTTCGCCCGGCGCCGGAA
GGCGTTGCGGGACGCGGGAAAGCGGCAGGTCGGGCGGAGGGTTTACCCGCTCTGA
CGGCAGCCTCCCGCAGCAGTCGCCGGCCGCTGCCGGCCGCGCCGTGCAGTAAAAA
TTGAATCGAGAATTCGAATCCCGAGACAACTTTTGACACCAAAGGAACCAGCATG

Figure 44: Illustration of the promoter sequences identified for the 5' upstream region of H16_A1526. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 1, basepairs 1658248 to 1659248.

Lon

Several σ^{70} based promoters could be identified for the 5' upstream region of *lon* by the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The predicted σ^{70} promoter distant to the start ATG revealed BPROM scores of 36 out of 77 for the -10 box and a score of 27 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not predict a promoter for this region. The second promoter located in front of the start ATG was predicted by BPROM and revealed scores of 65 out of 77 for the -10 box and a score of 3 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not predict a promoter for this region. The promoter predicted closest to the start ATG was predicted by the Neural Network Promoter Prediction tool with a score of 0.97 out of 1. A Shine-Dalgarno sequence could be anticipated 5 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *lon* located on *R. eutropha* H16, chromosome 1, basepairs 1610548 to 1611548. *Lon* codes for a ATP-dependent Lon protease.

From basepair 1610548 to 1611548 on chromosome 1:

CGGCATCGTCTACATCGATGAGATCGACAAGATCTCGCGCAAGTCCGACAACCCGT
CGATCACCCGCGACGTGTCGGGCGAGGGCGTGCAGCAGGCACTGCTCAAGCTGAT
CGAAGGCACCATGGCATCGGTGCCGCCGCAGGGTGGCCGCAAGCATCCGAACCAG
GACTTCCTGCAGGTGGACACGACAATATCCTGTTCATCTGCGGCGGCGCCTTCGAT
GGCCTGGAGAAGGTCATCATGCAGCGCTCGGCAAGACCGGCATCGGCTTTGCCGC
GCAGGTCAAGAGCAAGGAAGAGCGCGACGCCAGCGAAGTGCTGCCGCAGACCGAG
CCGGAAGACCTGATCAAGTTCGGCCTGATCCCCGAACTGATCGGCCGCCTGCCGGT
GGTCGCCACGCTGTCCAAGCTGGACGAGGCCGCGCTGATGCAGATCCTGGTCGAG
CCCAAGAACGCGCTGGTCAAGCAATACCAGAAGCTGCTGGCTATGGAAGGCGTCGA
GCTGGAAATCCGCCCGGGTGCGCTGAGCGCCATCGCCCGCAAGGCGATCCGCCGC
AAGACCGGCGCCCGTGGCCTGCGTTCGATCCTGGAGCAGTCGCTGATGGACGTCAT

GTACGACCTGCCCAACTACAAGGGCGTGCAGAAGGTGGTGATCGATGAAAATACGA
TCAACGGCGATGCACCTCCGCTGCTGATGTACGAAGAGCAGCAGCCCAAGGTGGCG
GGGTCCAACTGACGCGAGGGCTGCCAGACGTGGTAAGTGGTCGGGACCACACCGG
CAGCCGGGCGAAAAGGCCGTTCGCGTAGAAGCGAGCGGCTTTTTTTGTTTATTTCTA
CCGACGAGAGGGAGGAAACGCGGGTATGCTGTTTCAGAGGGGCCAGCAAGTGCAG
GGAGCGGCATCTGCCTGCCTTGATGAATCGATTGGCACGTATCTTGTAATCGATTCG
GGCGGCCCAATTTACGCCTTAAATGACTGACTTGGGGAAAATG

Figure 45: Illustration of the promoter sequences identified for the 5' upstream region of Lon. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 1, basepairs 1610548 to 1611548.

GyrA

Several σ^{70} based promoters could be identified for the 5' upstream region of GyrA by the promoter prediction software tool BPROM. The predicted σ^{70} promoter distant to the start ATG revealed BPROM scores of 41 out of 77 for the -10 box and a score of 20 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not predict a promoter for this region. The second promoter located close of the start ATG was predicted by BPROM and revealed scores of 41 out of 77 for the -10 box and a score of 20 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not predict a promoter for this region. A Shine-Dalgarno sequence could be anticipated 5 bp upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding for the subsequent gene *gyrA* located on *R. eutropha* H16, chromosome 1, basepairs 859088 to 860088. *GyrA* codes for a DNA gyrase A subunit of DNA gyrase.

From basepair 859088 to 860088 on chromosome 1:

GCACGACTTCGGATCGACCTTCAGCTGGCGCTTGCCCTTGCCTTCGGTGTAGACGC
GGTTGGCTTCAACGCCCTTGCTCACCAGGTAAGCCTTGACCGATTCAGCACGACGG
ATCGACAGGCGATCGTTGTACTTGTCCGAACCGAACGAGTCGGTGTGGCCAACGGC
GATGATAACTTCCAGGGTGATGCCTTGCAGTTTCGAGACCAGGTCGTCCAGCTTGGC
CTTGCCTTCCGGCTTCAGGACAGCCTTGTCGAAGTCGAACAGGGTGTCAGCAGCGA
AAGTGACCTTCTCGCTCGAGACGACCGGAGGAGCCACCGGAGCCGGTGCCGGTGC
CGGTGCGGCCGGGGCCAGGGCGCCGTCGCACAGCGCGTTGGCGGTGGCCGGGGT
CCAGGAGCTGTCGCGCCAGCAGAGCTCGTTCGTGCCATTCTTCCACACGTACTCGC
TCGTACCGTTGCCCCAGTTGTCGTTCACTGCCTTCTTTTCATAAGGGACGGACTGAG
CTTGAGCGGATGCAGCCATTACTGCGGTAGCTGCAACGAGCGCGAGCTTGGCAAAT
TTTTTCATATTTCTCCTCTCAGGATGAGATCACCGCAGGGTTACTGCGAGCAAATGGA
CGAAAACAAGTCATACATTCTTCGGAGTATAACATTCTCGATGTGGTGGATGCTCCAC
TTCGAACAATGTCTGGCTCTTCGCTGGGGAATTGTGCCACAAGGGTCGTTTCCTAAG
AAGTAAATATATTCGATTCGCCGCATGGGTTTTGGGCCTGTGGTGTTTGTGCAACATA
GGTCTTGTGTCGCCCGCAAAGCCTTGTGGCATCAGGGATTCCGGGTTTCAGCGAGG
GCCTCGGGAGGTGTGCGCAAGGGCTGTCCGGAGAGCGGCAGGCGGGGGGTGGCG
GCATGTTAGAATGACACGTTGCCCAGCCGTCAGGCGCCCACTGTGCCGCGCGCTAA
AGGGCCTGTCATTCGCTTAAACCTGCCGCATAAACCACGCCGCAATG

Figure 46: Illustration of the promoter sequences identified for the 5' upstream region of GyrA. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 1, basepairs 859088 to 860088.

Sers1

One σ^{70} based promoter could be identified for the 5' upstream region of Sers1 by the promoter prediction software tool BPROM. The predicted σ^{70} promoter revealed BPROM scores of 69 out of 77 for the -10 box and a score of 12 out of 66 for the -35 box. The Neural Network Promoter Prediction tool did not predict a promoter for this region. A Shine-Dalgarno sequence could not be anticipated upstream of the start ATG. The -10 region is indicated in yellow and the -35 region in red. The -10 extended box is shown in grey and the Shine-Dalgarno sequence in turquoise. The start ATG marked in green resembles the first amino acid coding

for the subsequent gene *sers1* located on *R. eutropha* H16, chromosome 1, positive strand, basepairs 829233 to 830233. *Sers1* codes for a Seryl-tRNA synthetase 1.

From basepair 829233 to 830233 on chromosome 1:

AGCCTGGACGATGCCGAGCTGACCCAGCTGGCGCTGCGCGCCAGCGAAGAACTCG
GCGCACAGGCGGGCGACCTGAGCTGGCAGGACGAAGCACTGCAGCTGATCGTGGC
CTCGGCCGATGGCGACGGGCGCAAGCTGCTGAACAATATCGAGATCGTGGCGCGT
GCCGCGCGCACCGCCGGCGCCGCGCAGATCGATACCGCGCTGCTGGGCAGCGCG
CTGTCGGAAAACCTGCGCCGCTTCGACAAGGGCGGCGACGCCTTCTACGACCAGAT
CAGCGCGCTGCACAAGTCCGTGCGCGGCTCGGACCCGGACGCGGCGCTGTACTGG
TTCTGCCGCATGATCGACGGCGGCGCCGACCCGCGCTACCTGGCGCGGCGCATCG
TGCGCATGGCATGGGAGGATATCGGCCTGGCCGATCCCCGTGCCGCCCGCATCAC
GCTGGACGCGGCCGAGACCTATGAGCGCCTTGGCTCGCCCGAGGGCGAGCTGGCG
CTGGGTCAGGCCCTGATCTACCTGGCCGTGGCGCCCAAGTCCAACGCCGGCTACAA
CGCCTACAACGCGGCGCGCGCCTTCGTCGCCAAGGACAAGTCGCGCGCGGTGCCG
GTGCACCTGCGCAACGCGCCGACCAAGCTGATGAAGGAGCTTGGCCACGGCCACG
CCTACCGCTACGCGCACGACGAGCCCGAGGCCTACGCCGCCGGCGAGCACTACTT
CCCGGACGACCTCAAGGCCCAGGACTGGTACCAGCCGGTGCCGCGCGGGCTGGAG
GGCAAGATCGCCGAGAAGCTGCGCCACCTGCGCGAGCTGGATGCCGCCTGGCATC
GCGAGCAGCGCGTGGCCAAGCCAAAGGACGGCGGCAAGGGCGGGGAGTGAGCGG
TGGCGTGAGCCGGCAACACGGCGCTGGCCTGGCATCGCCCGGTTCGCGCCGGACG
GGGCCGGTGCAGTAAAATGGCGGGTCCGCCGGCATGTCGCTGGCCCTGCACTTCA
CCAAAGCAACATG

Figure 47: Illustration of the promoter sequences identified for the 5' upstream region of Sers1. The identified -10 hexamers were labeled yellow, -35 hexamers were labeled red, Shine-Dalgarno sequences were labeled in turquoise, the start ATG of the gene is indicated in green and the -10 extended box was labeled grey. The sequence shown relates to *R. eutropha* H16, chromosome 1, positive strand, basepairs 829233 to 830233.

4. Discussion

The gram negative facultative lithoautotrophic bacteria *R. eutropha* H16 gained a lot of attention in recent years for its potential bioremediation capacities, its PHA production properties or for the ability to solely grow on CO_2 as carbon source and respectively H_2 as energy source [6, 14]. Based on these versatile abilities and the ability to grow to high cell densities, the use of *R. eutropha* H16 as a host for heterologous expression and furthermore as a production organism seems promising. High cell density fermentations with *R. eutropha* H16 could also be conducted under environmentally friendly lithoautotrophic growth conditions in order to obtain heterologous products produced by supplying CO_2 as the sole carbon source. Several attempts have been made so far in order to set up an efficient inducible expression system in *R. eutropha* H16. Lac and tac promoters are already used for expression of genes in *R. eutropha* H16 [14, 44]. However, due to lacking transport capacities for inducers like IPTG or lactose neither P_{lac} nor P_{tac} can be used in *R. eutropha* H16. Under constitutive expression conditions P_{lac} and P_{tac} were shown to work as strong promoters in *R. eutropha* H16 [14, 44]. Furthermore some promoter systems originating from *R. eutropha* H16 itself are already identified and characterized. P_{phaC} or P_{phaP} [44] and promoters based on nitric oxide [10, 55] or hydrogen [45] activation for instance are reported as inducible expression systems in *R. eutropha* H16. However, all of these expression systems underlie inducing mechanisms that are not favorable under particular growth conditions of *R. eutropha* H16 or share complex inducing mechanisms; which excludes them from the use as a simple expression system. In order to extend the quantity of selected promoter systems for *R. eutropha* H16 a promoter library based on the genome of *R. eutropha* H16 was established in this work. Accordingly the aim of this work was to identify and characterize promoter systems native to *R. eutropha* H16 based on their activity under different growth conditions. In this case the focus of this work was set to identify promoter regions that could be used to express homologoues and heterologous genes under lithoautotrophic growth conditions in *R. eutropha* H16. In order to enable fermentation processes under environmentally friendly conditions supplying CO_2

as the only carbon source. For this purpose a direct and a random approach were defined.

The aim of the random approach was to generate a promoter library based on the analysis of partially digested genomic DNA of *R. eutropha* H16. The initial screening rounds were carried out in *E. coli* cells applying pRS415 (Figure 13). Therefore genomic DNA of *R. eutropha* H16 was partially digested by *Bsp*143I, cloned into the promoterless vector pRS415 by *Bam*HI and transformed into *E. coli* cells. The promoter activity was subsequently determined on the basis of the β-galactosidase reporter gene encoded on pRS415. In order to determine promoter activity for pRS415 clones a β-galactosidase assay was performed to define the Miller Units. In a subsequent step selected putative promoter regions exhibiting high Miller Unit values were introduced into pKR-tac-PAR-EGFP with *Bam*HI and *Not*I to conduct promoter activity screening on the basis of EGFP in *R. eutropha* H16 (Table 8). The second approach, named the direct approach, was chosen to identify and characterize putative promoter regions on the genome of *R. eutropha* H16 based on increased expression levels of proteins monitored under lithoautotrophic growth conditions compared to heterotrophic growth conditions of *R. eutropha* H16 [17]. The analysis of the soluble part of the proteome of *R. eutropha* H16 under particular growth conditions was done by Friedrich et al. [17] and suggested significant upregulation of protein expression under lithoautotrophic growth conditions at following loci GyrA, SerS1, Lon, H16_1526 and PHG027 (Table 7). In order to examine expression regulation of these loci a 1000 bp long DNA fragment was amplified 5' upstream to the corresponding genes at each loci. These fragments were cloned into pKR-tac-PAR-EGFP (Figure 8) cloning vector with *Not*I and *Bam*HI for promoter activity screening under lithoautotrophic and heterotrophic growth conditions in *R. eutropha* H16. The promoter activity was determined on the basis of fluorescence units relating to EGFP expression levels. However, EGFP screening vectors on the basis of pKR-tac (Figure 6) had to be constructed first.

Construction of cloning and promoter screening vectors on the basis of pKR-tac

Cloning and promoter screening vectors on the basis of pKR-tac (Figure 6) had to be constructed prior to carrying out the promoter activity screening rounds. Vector construction of cloning and promoter screening vectors suitable for *E.coli* and mainly *R. eutropha* H16 were based on the pKR-tac vector. However, gel electrophoresis verified the formation of additional bands larger in size compared to the expected band size of unrestricted and restricted pKR-tac vector DNA derived from *E. coli* cells (data not shown). Incorrect partitioning of pKR-tac vectors in *E. coli* cells was suspected for this reason. Thereupon the *par29* region was introduced into pKR-tac. The *par29* region is reported to promote plasmid stability and correct plasmid partitioning during cell division in *E. coli* cells [46, 48]. The *par29* region needed to be modified by overlap extension PCR prior to introduction into pKR-tac in order to retain a unique *Nde*I restriction site on pKR-tac-PAR vectors, which enables standard insert cloning via *Nde*I and *Hind*III restriction sites on all pKR-tac derived vectors. The additional *Nde*I restriction site on *par29* was located in a non-coding region upstream of *parD*. The elimination of the *Nde*I restriction site on the *par29* region was accomplished by introducing a nucleotide exchange to this recognition sequence of *Nde*I (Figure 4). Overlap extension PCR applying modified inner primers was used to promote the nucleotide exchange at the *Nde*I restriction site on *par29*. The modified *par29* region was renamed to *par29.1*. After the integration of *par29.1* to pKR-tac the vector stability of pKR-tac-PAR (Figure 7) in *E. coli* cells was accomplished. Further deviation from the expected restriction patterns of pKR-tac-PAR vectors could not be observed in *E. coli* or *R. eutropha* H16 cells since. The pKR-tac-PAR vector served from here on as the basic cloning vector for this work. In a subsequent step *egfp* was chosen as a reporter gene on pKR-tac-PAR for promoter identification and characterization purposes in *E. coli* and *R. eutropha* H16. The choice of *egfp* as a reporter gene for promoter activity screening was based on the efficiency and simplicity of conducting EGFP promoter screening assays. Successful promoter screening assays were already accomplished in *R. eutropha* H16 with GFPmut2. Promoter activity screening assays on the basis of EGFP did solely require OD_{600} and EGFP fluorescent unit measurements of the

corresponding clones. On the contrary the β-galactosidase based promoter activity screening assay previously applied in the random approach required laborious and sensitive sample preparation steps prior to clone analysis. Consequently sample preparation was time consuming and more importantly the reproducibility of clone related activity readings was complicated. On this basis *egfp* was preferred over *lacZ* as a reporter gene for promoter activity on pKR-tac-PAR vectors. *Egfp* was introduced into pKR-tac-PAR using *Nde*I and *Hind*III restriction sites constructing the pKR-tac-PAR-EGFP (Figure 8) promoter screening vector.

Modifications of this vector are pKR-PAR-EGFP (Figure 9), pKR-tac-PAR-EGFP Δ*lacI* (Figure 11) and pKR-lac-PAR-EGFP Δ*lacI* (Figure 12). Firstly pKR-tac-PAR-EGFP Δ*lacI* and pKR-lac-PAR-EGFP Δ*lacI* were constructed to obtain constitutive promoter activity readings of P_{tac} and P_{lac} as reference promoter activity readings for identified promoters in *E. coli* and *R. eutropha* H16. Secondly the induction of P_{tac} and P_{lac} in *R. eutropha* H16 by IPTG could not be accomplished, as IPTG cannot be transported into *R. eutropha* H16 cells [44]. In order to achieve EGFP expression in *R. eutropha* H16 by P_{tac} and P_{lac} *lacI* had to be removed from the vector. Consequently *lacI* was removed from pKR-tac-PAR-EGFP by *Pst*I and *Not*I restriction to enable constitutive EGFP expression in *E. coli* and *R. eutropha* H16. The cut vector backbone of pKR-tac-PAR-EGFP was blunted and religated to obtain pKR-tac-PAR-EGFP Δ*lacI* in a subsequent step (Figure 11). pKR-lac-PAR-EGFP Δ*lacI* was constructed by exchanging the sequence of P_{tac} by P_{lac} (Figure 12). Additionally pKR-PAR-EGFP was constructed by removing the P_{tac} from pKR-tac-PAR-EGFP Δ*lacI* to create a promoterless screening vector as applied in the random approach (Figure 9). However, EGFP expression by pKR-tac-PAR-EGFP, pKR-tac-PAR-EGFP Δ*lacI* and pKR-lac-PAR-EGFP Δ*lacI* could only be reported for *E. coli* cells but not for *R. eutropha* H16. Accordingly differences in the amino acid sequence of EGFP compared to GFPmut2 are most likely responsible for inactivity of EGFP in *R. eutropha* H16, as described in the following abstract.

Stability of EGFP in *R. eutropha* H16

EGFP was initially chosen as a reporter to characterize promoter activity in *E. coli* as well as *R. eutropha* H16 due to efficiency and simplicity of clone analysis. However, EGFP fluorescence could only be reported by pKR-lac-PAR-EGFP Δ*lacI* and pKR-tac-PAR-EGFP Δ*lacI* in *E. coli* cells. *R. eutropha* H16 cells containing pKR-tac-PAR-EGFP Δ*lacI* and pKR-lac-PAR-EGFP Δ*lacI* vectors did not exhibit EGFP fluorescence. At first EGFP inactivity was suspected due to low pH values, as EGFP only exhibits fluorescence at a pH in the range of 7.5. In order to establish an environment at pH 7.5, overnight cultures of *R. eutropha* H16 clones containing pKR-tac-PAR-EGFP Δ*lacI* vectors were treated according to Toshiaki et. al. [44]. However, no fluorescence units of EGFP could be recorded in *R. eutropha* H16. A SDS-PAGE analysis comparing samples of *R. eutropha* H16 cells containing pKR-lac-PAR-EGFP Δ*lacI* and pKR-tac-PAR-EGFP Δ*lacI* with *E. coli* cells also containing pKR-lac-PAR-EGFP Δ*lacI* and pKR-tac-PAR-EGFP Δ*lacI* did only reveal EGFP expression in *E. coli* cells but not in *R. eutropha* H16 (Figure 17). In order to receive more information about the expression of EGFP in *R. eutropha* H16 a northern blot should be conducted to verify the presence of EGFP mRNA. The presence of EGFP mRNA could indicate degradation of EGFP on a protein based level. Reasons for this might for example be miss folding of EGFP due to different codon usage in *R. eutropha* H16 compared to *E. coli*. Also the vector stability and expression behavior of pKR-lac-PAR-EGFP Δ*lacI* or pKR-tac-PAR-EGFP Δ*lacI* under constitutive expression conditions in *R. eutropha* H16 could be monitored. Therefore expression of a homologous protein is suggested in order to guarantee correct expression and folding of the product in *R. eutropha* H16. In this case possible influences on the stability of the vector system could be exercised by *R. eutropha* H16 related to stress acting on the cell caused by constitutive protein expression. However, it was reported that plasmid based expression of the heterologous protein GFPmut2 alongside with P_{tac} and P_{lac} was successfully accomplished in *R. eutropha* H16 and furthermore in related *R. eutropha* species [44, 56-58]. This is a strong indication for the feasibility of successful plasmid based heterologous gene expression in *R. eutropha* H16 species. On the contrary no evidence for EGFP activity in *R. eutropha* H16 was reported in literature so far. GFPmut2 and EGFP both share the same basic amino

acid sequence with GFP, but differentiate in six amino acids (Figure 5). EGFP could be folded incorrectly or possibly be degraded in *R. eutropha* H16 upon the differences in the amino acid sequence compared to GFPmut2. Based on these findings the use of GFPmut2 instead of EGFP as a reporter for promoter activity is suggested for further work.

Identified promoter regions on *R. eutropha* H16 genome

In total eleven regions of the *R. eutropha* H16 genome with significant promoter activity were identified on the basis of the β-galactosidase assay in *E. coli*. The results base on the random approach in which genomic DNA of *R. eutropha* H16 was partially digested by *Bsp*143I, ligated into pRS415 and transformed into *E. coli*. The cells were grown under heterotrophic growth conditions and screened applying a β-galactosidase assay. The degree of promoter activity was defined by Miller Units. A selection of *E. coli* pRS415 clones, respectively *E. coli* pRS415 clones 1, 3, 5, 6, 10 and 11, exhibiting Miller Unit values in the range of 5000 to 95000 were sequenced for further analysis (Agowa, Berlin, Germany). The sequencing results of *E. coli* pRS415 clones 1, 3, 5, 6, 10 and 11 were blasted against the genome of *R. eutropha* H16 using NCBI blast search tools. Sequence analysis revealed that some of the selected clones appear to contain more than one DNA fragment of the partially digested genomic DNA of *R. eutropha* H16 (see results section 3.2). In most of these cases DNA fragments of a few hundred basepairs were attached to each other before being introduced into pRS415 vector during ligation, as in the case of *E. coli* pRS415 clones 1, 6 and 11. Sequencing results of *E. coli* pRS415 clones 1, 6 and 11 revealing several blast hits on the genome of *R. eutropha* H16 were evaluated visually to sort out potential promoter regions. DNA fragments were identified as putative promoter regions due to 5' upstream proximity to genes.

Both sequenced DNA fragments of *E. coli* pRS415 clone 1 are located on chromosome 1 of *R. eutropha* H16 and are likely to contain a promoter sequence. Blast hit 1 is possibly containing a promoter sequence due to the close location 5' upstream to *h16_A2493* coding for a hypothetical protein (Figure 18). Blast hit 2 could theoretically also contain a promoter sequence. It is located 5' upstream of

h16_A2469 (Figure 19). *H16_A2469* is representing a predicted phage integrase family protein. Miller Units of *E. coli* pRS415 clone 1 were 5629. *E. coli* pRS415 clone 1 was not chosen for further characterization due to comparatively low Miller Unit readings (Table 8). The sequencing results *E. coli* pRS415 clone 6 with Miller Units of 8220 revealed three blast hits on the genome of *R. eutropha* H16. Blast hit 1 and 3 are not likely to contain promoter regions due to their isolated and distant location to nearby genes (see figure 23 and 25). Blast hit 2 on the other hand is located roughly 350 bp 5' upstream of *h16_B1660* on chromosome 2 of *R. eutropha* H16 (Figure 24). An acetyl carrier function is predicted for *h16_B1660*. *E. coli* pRS415 clone 6 was not chosen for further characterization due to comparatively low Miller Unit readings (Table 8).

E. coli pRS415 clone 11 with Miller Units of 6245 had seven blast hits on the genome of *R. eutropha* H16. The first blast hit is located 5' upstream of *h16_B0219* on chromosome 2 of *R. eutropha* H16 (Figure 27). *H16_B0219* encodes for a GNAT family N-acetyltransferase. Blast hit 2 was not located near a potential promoter region (Figure 28) and blast hits 3 to 7 represent a recurring motive on *rrlB* genes on chromosome 1 and 2 of *R. eutropha* H16 (Figure 29). None of the blast hits 3 to 7 located on chromosome 1 or 2 are likely to contain a promoter sequence. *E. coli* pRS415 clone 11 was not chosen for further characterization due to comparatively low Miller Unit readings (Table 8). The sequencing results for *E. coli* pRS415 clones 3, 5 and 10 revealed clear blast hits of putative promoter regions located 5' upstream of particular genes on the genome of *R. eutropha* H16. *E. coli* [pRS415 clone 3] exhibiting a Miller Unit activity of 99275 had two blast hits of different lengths. Blast hit 1 was located 5' upstream of *h16_A1712* on chromosome 1 of *R. eutropha* H16 containing 5' sequence parts of *h16_A1712* (Figure 20). The sequence length was 138 bp. Blast hit 2 was 548 bp in length and located 5' upstream of *h16_B1772* on chromosome 2 of *R. eutropha* H16 also containing 5' sequence parts of *h16_B1772* (Figure 21). The genes *h16_A1712* and *h16_B1772* both encode for predicted transcriptional regulator. Furthermore both blast hits share an identical sequence of 138 bp including the first 40 bp of *h16_A1712* or *h16_B1772* and approximately 90 bp of their 5' upstream region. The sequence of the first blast hit with a length of 138 bp

and of the second blast hit 548 bp long were both amplified by PCR and cloned into pKR-tac-PAR-EGFP for further screening in *R. eutropa* H16.

Cloning was accomplished via *Not*I and *Bam*HI restriction sites, which exchanged P_{tac} on pKR-tac-PAR-EGFP for the PCR amplified fragments of blast hits 1 and 2. The vector constructs were then transformed into *R. eutropha* H16 for further screening. EGFP fluorescence units could not be monitored in *R. eutropha* H16. Sequence analysis of the *E. coli* [pRS415 clone 3] insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed a σ^{70} promoter region located 44 bp upstream of the transcription start of *h16_A1712* or *h16_B1772* (Figures 39 and 40). Based on the location of the predicted -10 and -35 hexamers and the anticipated Shine-Dalgarno sequence the presence of the predicted σ^{70} promoter 44 bp upstream of the transcription start of *h16_A1712* or *h16_B1772* is likely to exhibit the promoter activity measured in Miller Units for the insert of *E. coli* [pRS415 clone 3]. The Miller Units of *E. coli* [pRS415 clone 5] were determined to be 8246. The blast result for *E. coli* [pRS415 clone 5] revealed to be the 5' upstream region of *h16_B1034* (Figure 22). The sequence of the blast hit did also include approximately 60 bp of the 5' sequence of *h16_B1034*. *H16_B1034* is coding for a predicted alpha-ketoglutarate-dependent taurine dioxygenase. DNA fragments of 1000 bp, 549 bp, 301bp and 167 bp were amplified 5' upstream of *h16_B1034* by PCR (see figure 31). Primer walking was performed in order to narrow down the position of the promoter sequence and to examine possible regulatory elements. The four PCR products were cloned into pKR-tac-PAR-EGFP with *Not*I and *Bam*HI at the same time replacing P_{tac}. The vector constructs were then transformed into *R. eutropha* H16 for further screening. However, EGFP fluorescence units could not be monitored in *R. eutropha* H16. Sequence analysis of the *E. coli* [clone 5] insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed two σ^{70} promoter regions located 108 bp and 491 bp upstream of the transcription start of *h16_B1034* present on the insert (Figure 41). Based on the high scores of both promoter regions identified by the promoter prediction software tools it is possible that both predicted promoter regions 5' upstream of *h16_B1034* contributed to the Miller Unit readings obtained for *E. coli* [clone 5]. In order to relate the Miller Unit readings for *E. coli* [clone 5] to a single promoter

region, both predicted promoter regions should be amplified by PCR and evaluated separately for promoter activity.

The Miller Units of *E. coli* pRS415 clone 10 were determined to be 11905. The blast hit of *E. coli* pRS415 clone 10 was referring to the 5' upstream region of *h16_B2204*, a putative uncharacterized protein (Figure 26). The sequence of the blast hit did also include approximately 45 bp of the 5' sequence of *h16_B2204*. DNA fragments of 1000 bp, 613 bp, 354 bp and 168 bp were amplified 5' upstream of *h16_B2204* by PCR. Primer walking was performed in order to narrow down the position of the promoter sequence and to examine possible regulatory elements. The four PCR products were cloned into pKR-tac-PAR-EGFP with *Not*I and *Bam*HI for further screening in *R. eutropa* H16. The *Not*I and *Bam*HI double digest of pKR-tac-PAR-EGFP removes P_{tac} from the vector and at the time enables introduction of putative promoter fragments at the same location on pKR-tac-PAR-EGFP. The vector constructs were then transformed into *R. eutropha* H16 for further screening. EGFP fluorescence units could not be monitored in *R. eutropha* H16. Sequence analysis of the *E. coli* [clone 10] insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed two σ^{70} promoter regions located 172 bp and 708 bp upstream of the transcription start of *h16_B2204* present on the insert (Figure 42). The predicted promoter region 708 bp 5' upstream of *h16_B2204* is most likely a false hit by the BPROM promoter prediction tool. This is based on the low scores of the BPROM tool for the predicted promoter region and the fact that this region was not recognized by the Neural Network Promoter Prediction tool. Based on the high scores obtained by both promoter prediction tools, the σ^{70} promoter region identified 172 bp 5' upstream of *h16_B2204* can most likely be related to the Miller Units obtained for *E. coli* [clone 10].

Sequence analysis of the selected 5' upstream regions of PHG027, Sers1, GyrA, Lon and H16_A1526 based on the promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed hits on all selected sequences. Sequence analysis of the 5' upstream region of Sers1 revealed a σ^{70} promoter region located 47 bp upstream of the transcription start of *sers1* based on promoter prediction software tool BPROM (Figure 47). Even though the Neural Network Promoter Prediction tool did not predict a promoter region, the scores and

location of the predicted -10 and -35 hexamers obtained by BPROM indicate the presence of the predicted σ^{70} promoter 47 bp upstream of the transcription start of *sers1*. EGFP expression could be identified visually in *E. coli* clones containing pKR-sers1-PAR-EGFP vectors supporting the assumption of the presence of a promoter region encoded in the 5' upstream region of *sers1*. Sequence analysis of the 5' upstream region of H16_A1526 revealed a σ^{70} promoter region located 182 bp upstream of the transcription start of *h16_A1526* based on promoter prediction software tool BPROM and the Neural Network Promoter Prediction tool (Figure 44). The high scores of the promoter prediction tools, the location of the predicted -10 and -35 hexamers and the anticipated Shine-Dalgarno sequence indicate the presence of the predicted σ^{70} promoter 182 bp upstream of the transcription start of *h16_A1526*. Additionally EGFP expression could be identified visually in *E. coli* clones containing pKR-h16_A1526-PAR-EGFP vectors supporting the assumption of the presence of a promoter region encoded in the 5' upstream region of *h16_A1526*.

Sequence analysis of the 5' upstream region of PHG027 insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed three σ^{70} promoter regions located 909 bp, 811 bp and 312 bp upstream of the transcription start of *phg027* present on the insert (Figure 43). Based on the results of both promoter prediction tools the σ^{70} promoter region predicted 909 bp and 312 bp 5' upstream of *phg027* are believed to be active. Even though results for the predicted promoter region 312 bp 5' upstream of *phg027* could only be verified by BPROM, the location of the predicted promoter sequence and scores obtained by BPROM suggest the presence of this promoter region. The promoter region 811 bp 5' upstream of *phg027* was only predicted by the Neural Network Promoter Prediction tool and could not be verified by BPROM. In order to determine the influence of the predicted promoter regions on the promoter activity readings, each predicted promoter region should be amplified by PCR and tested for promoter activity. Alternatively primer walking could be conducted eliminating the hypothetical promoter regions step by step. Ultimately, promoter activity readings of PCR fragments obtained from primer walking could indicate the contribution or presence of the predicted promoter regions found in the 5'

upstream region of PHG027. EGFP expression could be identified visually in *E. coli* clones containing pKR-phg027-PAR-EGFP vectors supporting the assumption of the presence of a promoter region encoded in the 5' upstream region of *phg027*. Sequence analysis of the 5' upstream region of Lon insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed three σ^{70} promoter regions located 789 bp, 122 bp and 52 bp upstream of the transcription start of *lon* present on the insert (Figure 45). Based on the dissonant results of both promoter prediction tools the σ^{70} promoter region of the 5' upstream region of *lon,* the location of putative promoter regions cannot be predicted precisely. However, the predicted promoter region located 52 bp upstream of the transcription start of *lon* is suggested as the promoter region exhibiting promoter activity due to the alignment of the predicted σ^{70} promoter region and the Shine-Dalgarno sequence upstream of the transcription start of *lon.* In order to determine the influence of the predicted promoter regions on promoter activity readings, each predicted promoter region should be amplified by PCR and tested for promoter activity separately. Alternatively primer walking could be conducted eliminating the hypothetical promoter regions step by step. Ultimately, promoter activity readings of PCR fragments obtained from primer walking could indicate the contribution or presence of the predicted promoter regions found in the 5' upstream region of Lon. EGFP expression could be identified visually in *E. coli* clones containing pKR-lon-PAR-EGFP vectors supporting the assumption of the presence of a promoter region encoded in the 5' upstream region of *lon*.

Sequence analysis of the 5' upstream region of GyrA insert based on promoter prediction software tools BPROM and Neural Network Promoter Prediction revealed two σ^{70} promoter regions located 440 bp and 88 bp upstream of the transcription start of *gyrA* present on the insert (Figure 46). However, the predicted promoter region located 88 bp upstream of the transcription start of *gyrA* is suggested as the promoter region exhibiting promoter activity due to the alignment of the predicted σ^{70} promoter region and the Shine-Dalgarno sequence upstream of the transcription start of *gyrA*. In order to determine the influence of the predicted promoter regions on promoter activity readings, each predicted promoter region should be amplified by PCR and tested for promoter activity separately. Alternatively primer walking could be conducted eliminating the hypothetical

promoter regions step by step. Ultimately, promoter activity readings of PCR fragments obtained from primer walking could indicate the contribution or presence of the predicted promoter regions found in the 5' upstream region of GyrA. EGFP expression was identified visually in *E. coli* clones containing pKR-gyrA-PAR-EGFP vectors supporting the assumption of the presence of a promoter region encoded in the 5' upstream region of *gyrA*.

Sequence analysis of all selected putative promoter regions from the random and the direct approach revealed hits for σ^{70} promoter regions based on the promoter prediction software tools BPROM and Neural Network Promoter Prediction. The presence of σ^{70} promoter regions would also explain the high Miller Units values exhibited by the clones derived from the random approach and visible EGFP expression of all promoter regions encoded as inserts on pKR-tac-PAR-EGFP vectors in *E. coli* cells.

Conclusion and outlook

Several promoter regions could be identified on the genome of *R. eutropha* H16 applying the random and direct approach. The random approach for instance revealed numerous putative promoter regions derived from the genome of *R. eutropha* H16 of which a selection based on comparatively high Miller Unit values was sequenced and prepared for further analysis in *R. eutropha* H16. The final selection of clones consisted of *E. coli* pRS415 clones 1, 3, 5, 6, 10 and 11, for which sequence analysis revealed various loci of putative promoter regions on the genome of *R. eutropha* H16. Sequences of blast hits of *E. coli* pRS415 clones 3, 5 and 10 were amplified by PCR for further characterization based on blast hits indicating distinct promoter regions and high Miller Unit values. The amplified PCR products were transferred into pKR-tac-PAR-EGFP. The vector constructs of pKR-tac-PAR-EGFP containing the putative promoter regions of *E. coli* pRS415 clones 3, 5 and 10 and the corresponding PCR products derived from primer walking were successfully transformed into *E. coli* and *R. eutropha* H16 cells. EGFP expression could be identified visually in *E. coli* for pKR-tac-PAR-EGFP constructs

containing the putative promoter regions. However, EGFP expression could not be detected in *R. eutropha* H16. Samples of the direct approach were prepared accordingly for further analysis in *R. eutropha* H16. The 5' upstream regions of *sers1*, *gyrA*, *lon*, *phg027* and *h16_A1526* were amplified by PCR and cloned into pKR-tac-PAR-EGFP vectors. However, EGFP expression could only be identified visually in *E. coli* for pKR-tac-PAR-EGFP constructs containing the putative promoter regions, but not for *R. eutropha* H16. All loci were selected based on significant upregulated expression under lithoautotrophic growth conditions of *R. eutropha* H16. Samples of the random and direct approach encoded on pKR-tac-PAR-EGFP vectors were intended to be characterized in more detail in *R. eutropha* H16. However, promoter activity readings referring to EGFP could only be quantified in *E. coli*. As mentioned earlier in this work expression of EGFP seems to be affected in *R. eutropha* H16 thus no fluorescence units could be determined for promoter screening vectors encoding EGFP in *R. eutropha* H16. The amino acid sequence of EGFP only differs at six locations from GFPmut2 (Figure 5), however GFPmut2 is reported to be active in *R. eutropha* H16 [44, 58]. It is suggested that the difference in the amino acid sequences of GFPmut2 compared to EGFP influence EGFP expression in *R. eutropha* H16. The use of GFPmut2 is therefore recommended as a reporter gene for promoter activity screening in *R. eutropha* H16. Alternatively β-galactosidase could also be used as a reporter gene in *R. eutropha* H16; both proteins have been reported to maintain activity in *R. eutropha* H16 [2, 56]. When using β-galactosidase as a reporter gene laborious activity assays and the accompanying sensitive sample preparation steps need to be considered; which interfere with the generation of reproducible results. Accordingly the use of GFPmut2 as a reporter gene is recommended. Similar to EGFP GFPmut2 does not require any sample preparation steps sensitive to the activity of the reporter gene in order to determine the promoter activity of clones. Promoter activity screening assays applying screening vectors with GFPmut2 as a reporter gene could be carried out in a simple and time saving manner. The random and direct approach could both be adapted easily to the use of GFPmut2 as a reporter gene replacing EGFP. For this purpose EGFP needs to be exchanged by GFPmut2 on the screening vectors encoding EGFP and parameters of fluorescence unit measurements adapted to GFPmut2.

Further characterization of the already identified promoter regions derived from the direct and random approach could be successfully accomplished in *R. eutropha* H16 by applying screening vectors encoding GFPmut2. Additionally more screening rounds could be conducted in order to generate an extended promoter library of *R. eutropha* H16. In this case a random approach on the basis of a promoterless vector similar to pKR-PAR-EGFP encoding GFPmut2 could be used to screen for promoter activity in *R. eutropha* H16 under lithoautotrophic growth conditions.

References

1. Wilde, E., *Untersuchungen über Wachstum und Speicherstoffsynthese von Hydrogenomonas.* Archives of Microbiology, Volume 43, Number 2, 109-137, DOI: 10.1007/BF00406429, 1962.

2. Fricke WF, K.B., Bowien B., *The Genome Organization of Ralstonia eutropha Strain H16 and Related Species of the Burkholderiaceae.* J Mol Microbiol Biotechnol, 2009. **16**: p. 124-135

3. Vaneechoutte, M., *Wautersia gen. nov., a novel genus accommodating the phylogenetic lineage including Ralstonia eutropha and related species, and proposal of Ralstonia [Pseudomonas] syzygii (Roberts et al. 1990) comb. nov.* International Journal of Systematic and Evolutionary Microbiology, 2004. **54**(2): p. 317-327.

4. CHRISTINE HOGREFE, D.R., BÄRBEL FRIEDRICH, *Alcaligenes eutrophus Hydrogenase genes (Hox).* JOURNAL OF BACTERIOLOGY, Apr. 1984, p.43-48, 0021-9193/84/040043-06$02.00/0, Copyright X 1984, American Society for Microbiology, 1984.

5. B. Friedrich, E.S., *Molecular Biology of Hydrogen Utilization in Aerobic Chemolithotrophs.* Annu. Rev. Microbiol. 47, 351–383., 1993.

6. Schwartz, E., et al., *Complete Nucleotide Sequence of pHG1: A Ralstonia eutropha H16 Megaplasmid Encoding Key Enzymes of H2-based Lithoautotrophy and Anaerobiosis.* Journal of Molecular Biology, 2003. **332**(2): p. 369-383.

7. Schwartz, E., et al., *A proteomic view of the facultatively chemolithoautotrophic lifestyle of Ralstonia eutropha H16 supp.* Proteomics, 2009. **9**(22): p. 5132-42.

8. Kohlmann, Y., et al., *Analyses of soluble and membrane proteomes of Ralstonia eutropha H16 reveal major changes in the protein complement in adaptation to lithoautotrophy.* J Proteome Res, 2011. **10**(6): p. 2767-76.

9. Pohlmann, A., *Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16.* Nature Biotechnology 24, 1257 - 1262 2006.

10. R. Cramm, A.B.a.K.S., *NO-dependent transcriptional activation of gene expression in Ralstonia eutropha H16.* Biochemical Society Transactions 2006. **Volume 34, part 1**.

11. Anne Pohlmann, R.C., Karin Schmelz and and B. Friedrich, *A novel NO-responding regulator controls the reduction of nitric oxide in Ralstonia eutropha.* Molecular Microbiology, 38(3), 626±638, 2000.

12. Doi Y, K.Y., Koyama N, Nakamura S, Hiramitsu M, Yoshida Y, Kimura U . , *Synthesis and degradation of polyhydroxyalkanoates in Alcaligenes eutrophus.* FEMS Microbiol. Rev. 103: 103-108., 1992.

13. Peplinski, K., et al., *Genome-wide transcriptome analyses of the 'Knallgas' bacterium Ralstonia eutropha H16 with regard to polyhydroxyalkanoate metabolism.* Microbiology, 2010. **156**(Pt 7): p. 2136-52.

14. Sichwart, S., et al., *Extension of the substrate utilization range of Ralstonia eutropha strain H16 by metabolic engineering to include mannose and glucose.* Appl Environ Microbiol, 2011. **77**(4): p. 1325-34.

15. LAURA KLEIHUES, O.L., MICHAEL BERNHARD, THORSTEN BUHRKE, AND BÄRBEL FRIEDRICH, *The H2 Sensor of Ralstonia eutropha Is a Member of the Subclass of Regulatory [NiFe] Hydrogenases.* JOURNAL OF BACTERIOLOGY, 2000: p. 2716–2724.

16. THOMAS JEFFKE, N.-H.G., CLAUDIA KAISER, CLAUDIA GRZESZIK, and B.B. BERNHARD KUSIAN, *Mutational Analysis of the cbb Operon (CO2Assimilation) Promoter of Ralstonia eutropha.* JOURNAL OF BACTERIOLOGY,0021-9193/99/$04.0010 July 1999, p. 4374–4380, 1999.

17. Schwartz, E., et al., *A proteomic view of the facultatively chemolithoautotrophic lifestyle of Ralstonia eutropha H16.* Proteomics, 2009. **9**(22): p. 5132-42.

18. Browning, D.F. and S.J. Busby, *The regulation of bacterial transcription initiation.* Nat Rev Microbiol, 2004. **2**(1): p. 57-65.

19. Gruber, T.M. and C.A. Gross, *Multiple sigma subunits and the partitioning of bacterial transcription space.* Annu Rev Microbiol, 2003. **57**: p. 441-66.

20. Murakami, K.S., et al., *Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex.* Science, 2002. **296**(5571): p. 1285-90.

21. Campbell, E.A.e.a., *Structure of the bacterial RNA polymerase promoter specificity α subunit.* Mol. Cell. 9, 527–539 2002.

22. Andrew Sanderson, J.E.M., Stephen D Minchin, Stephen J.W Busb, *Substitutions in the EscherichiacoliRNApolymerase σ70factor that affectrecognition of extended −10elements at promoters.* FEBS Lett. 544, 199–205, 2003.

23. Doerks, T., et al., *Systematic identification of novel protein domain families associated with nuclear functions.* Genome Res, 2002. **12**(1): p. 47-56.

24. Richard L. Gourse, W.R.a.T.G., *UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition.* Molecular Microbiology (2000) 37(4), 687±695.

25. Müller-Hill, B., *Some repressors of bacterial transcription.* Curr. Opin. Microbiol. 1, 145–151 (1998).

26. Minsang Shin, S.K., Seok-Jin Hyun, Nobuyuki Fujita, *Repression of deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor.* The EMBO Journal Vol. 20 No. 19 pp. 5392-5399, 2001.

27. Latchman, D.S., *Transcription factors: An overview.* Int. J. Biochem. Cell. Biol. , 1997(29): p. 1305 - 1312.

28. Shine J, D.L., *Determinant of cistron specificity in bacterial ribosomes.* Nature 254 (5495): 34–8. doi:10.1038/254034a0. PMID 803646, 1975.

29. Malys N, M., *Translation initiation: variations in the mechanism can be anticipated.* Cellular and Molecular Life Sciences 68 (6): 991–1003. doi:10.1007/s00018-010-0588-z. PMID 21076851, 2011.

30. Ma, J., A. Campbell, and S. Karlin, *Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures.* Journal of Bacteriology, 2002. **184**(20): p. 5733-5745.

31. Allison, L.A., *The role of sigma factors in plastid transcription.* Biochimie 82 (2000) 537–548 © 2000 Société française de biochimie et biologie moléculaire / Éditions scientifiques et médicales Elsevier SAS. All rights reserved. S0300908400006118/FLA.

32. Murakami, K.S., S. Masuda, and S.A. Darst, *Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution.* Science, 2002. **296**(5571): p. 1280-4.

33. Vassylyev DG, S.S., LaptenkoO, Lee J, Vassylyeva MN, *Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 °A resolution.* Nature 417:712–19, 2002.

34. Mittenhuber, G., *An Inventory of Genes Encoding RNA Polymerase Sigma Factors in 31 Completely Sequenced Eubacterial Genomes.* J. Mol. Microbiol. Biotechnol. (2002) 4(1): 77–91.

35. Ishihama, A., *FUNCTIONAL MODULATION OF ESCHERICHIA COLI RNA POLYMERASE.* Annu. Rev. Microbiol. 2000. 54:499–518.

36. Mekler, V., O. Pavlova, and K. Severinov, *Interaction of Escherichia coli RNA polymerase sigma70 subunit with promoter elements in the context of free sigma70, RNA polymerase holoenzyme, and the beta'-sigma70 complex.* J Biol Chem, 2011. **286**(1): p. 270-9.

37. Tanja Gruber, D.M., Meghan Sharp, Brian Young, *Binding of the initiation factor sigma(70) to core RNA polymerase is a multistep process.* Molecular Cell, Vol. 8, 21–31, July, 2001.

38. MICHAEL LONETTO, M.G., CAROL A. GROSS, *The sigma 70 family sequence conservation and evolutionary relationships.* JOURNAL, OF BACTERIOLOGY, June 1992, p. 3843-3849 0021-9193/92/123843-07$02.00/0.

39. Elizabeth A. Campbell, O.M., Mark Chlenov, Jing L. Sun, C. Anders Olson, Oren Weinman, Michelle L. Trester-Zedlitz, Seth A. Darst, *Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit.* Molecular Cell, Vol. 9, 527–539, 2002.

40. Florence Arsèneb, T.T., Bernd Bukau, *The heat shock response of Escherichia coli.* International Journal of Food Microbiology 55 3–9, 2000.

41. MARTIN BUCK, M.A.-T.G., DAVID J. STUDHOLME, YULI GUO, AND JAY D. GRALLA, *The Bacterial Enhancer-Dependent s54 (sN) Transcription Factor.* JOURNAL OF BACTERIOLOGY, 0021-9193/00/$04.0010, p. 4129–4136, 2000.

42. Tamas Gaal, W.R., Shawn T. Estrem, Lam H. Nguyen, Richard R. Burgess, Richard L. Gourse, *Promoter recognition and discrimination by EsigmaS RNA polymerase.* Molecular Microbiology 42(4), 939–954, 2001.

43. Perez-Rueda E, C.-V., *The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12.* Nucleic Acids Res. 28:1838–47, 2000.

44. Toshiaki Fukui, K.O., Jun Mifune, Izumi Orita, Satoshi Nakamura, *Evaluation of promoters for gene expression in polyhydroxyalkanoate-producing Cupriavidus necator H16* Applied Genetics and Molecular Biotechnology, 2011. **Volume 89, Number 5, 1527-1536, DOI: 10.1007/s00253-011-3100-2**

45. OLIVER LENZ, A.S., ANDREA TRAN-BETCKE, BÄRBEL FRIEDRICH, *A hydrogen-sensing system in transcriptional regulation of hydrogenase gene expression in Alcaligenes species.* JOURNAL OF BACTERIOLOGY, 1655–1663, 0021-9193/97/$04.0010, 1997.

46. GERLITZ, M.H., O.; SCHWAB, H., *Partitioning of Broad-Host-Range Plasmid RP4 Is a Complex System Involving Site-Specific Recombination.* JOURNAL OF BACTERIOLOGY, 1990(Nov. 1990): p. 6194-6203.

47. Eberl L, K.C., Givskov M, Grohmann E, Gerlitz M, Schwab H, *Analysis of the multimer resolution system encoded by the parCBA operon of broad-host-range plasmid RP4.* Mol Microbiol. ;12(1):131-41, 1994.

48. Gerlitz, M., *DNA-Sequenz der par-Region des Plasmids RP4 und Charakterisierung von in den Stabilisierungsprozess involvierten Proteinen, insbesondere der Resolvase* 1990.

49. Kovach ME, P.R., Elzer PH, Roop RM 2nd, Peterson KM., *pBBR1MCS: a broad-host-range cloning vector.* Biotechniques, 1994 (16(5):800-2).

50. Ann Matthysse, S.S., Catherine Dandie, Nicholas McClure, *Construction of GFP vectors for use in Gram-negative bacteria other than Escherichia coli.* FEMS Microbiology Letters 145, 87-94, 1996.

51. Brendan Cormack, R.V., Stanley Falkow, *FACS-optimized mutants of the green fluorescent protein GFP.* Gene, 173, 1996: p. 33 - 38.

52. Reese, M.G. *Berkeley Drosophila Genome Project: Neural Network Promoter Prediction* 12.03.2012]; Available from: http://www.fruitfly.org/seq_tools/promoter.html.

53. Softberry, I. *Softberry: BPROM Prediction of bacterial promoters*. 12.03.2012]; Available from: http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb.

54. Department of Molecular Genetics, U.o.G. *Pepper: Prokaryote Promoter prediction*. 12.03.2012]; Available from: http://pepper.molgenrug.nl/index.php?option=com_content&view=article&id=1&Itemid=2.

55. Klink, A., et al., *Characterization of the signaling domain of the NO-responsive regulator NorR from Ralstonia eutropha H16 by site-directed mutagenesis.* J Bacteriol, 2007. **189**(7): p. 2743-9.

56. Barnard, G.C., et al., *Integrated recombinant protein expression and purification platform based on Ralstonia eutropha.* Appl Environ Microbiol, 2005. **71**(10): p. 5735-42.

57. Gregory M. York, B.H.J., JoAnne Stubbe, Anthony J. Sinskey, *Accumulation of the PhaP Phasin of Ralstonia eutropha Is Dependent on Production of Polyhydroxybutyrate in Cells* Journal of Bacteriology, 2001. **vol. 183 no. 14 4217-4226**

58. York, G.M., J. Stubbe, and A.J. Sinskey, *The Ralstonia eutropha PhaR Protein Couples Synthesis of the PhaP Phasin to the Presence of Polyhydroxybutyrate in Cells and Promotes Polyhydroxybutyrate Production.* J Bacteriol, 2002. **184**(1): p. 59-66.