Development of a transcriptional regulator-based bioreporter - towards a generic selection method for novel enzymes

Teunke van Rossum

Thesis committee

Promotor

Prof. Dr John van der Oost

Personal Chair at the Laboratory of Microbiology

Wageningen University & Research

Co-promotor

Dr Servé W.M. Kengen

Assistant Professor at the Laboratory of Microbiology

Wageningen University & Research

Other members

Prof. Dr Willem J.H. van Berkel, Wageningen University & Research

Prof. Dr Jan Roelof van der Meer, University of Lausanne, Switzerland

Prof. Dr Maarten Merkx, Eindhoven University of Technology

Dr Ruud A. Weusthuis, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Development of a transcriptional regulator-based bioreporter - towards a generic selection method for novel enzymes

Teunke van Rossum

Thesis

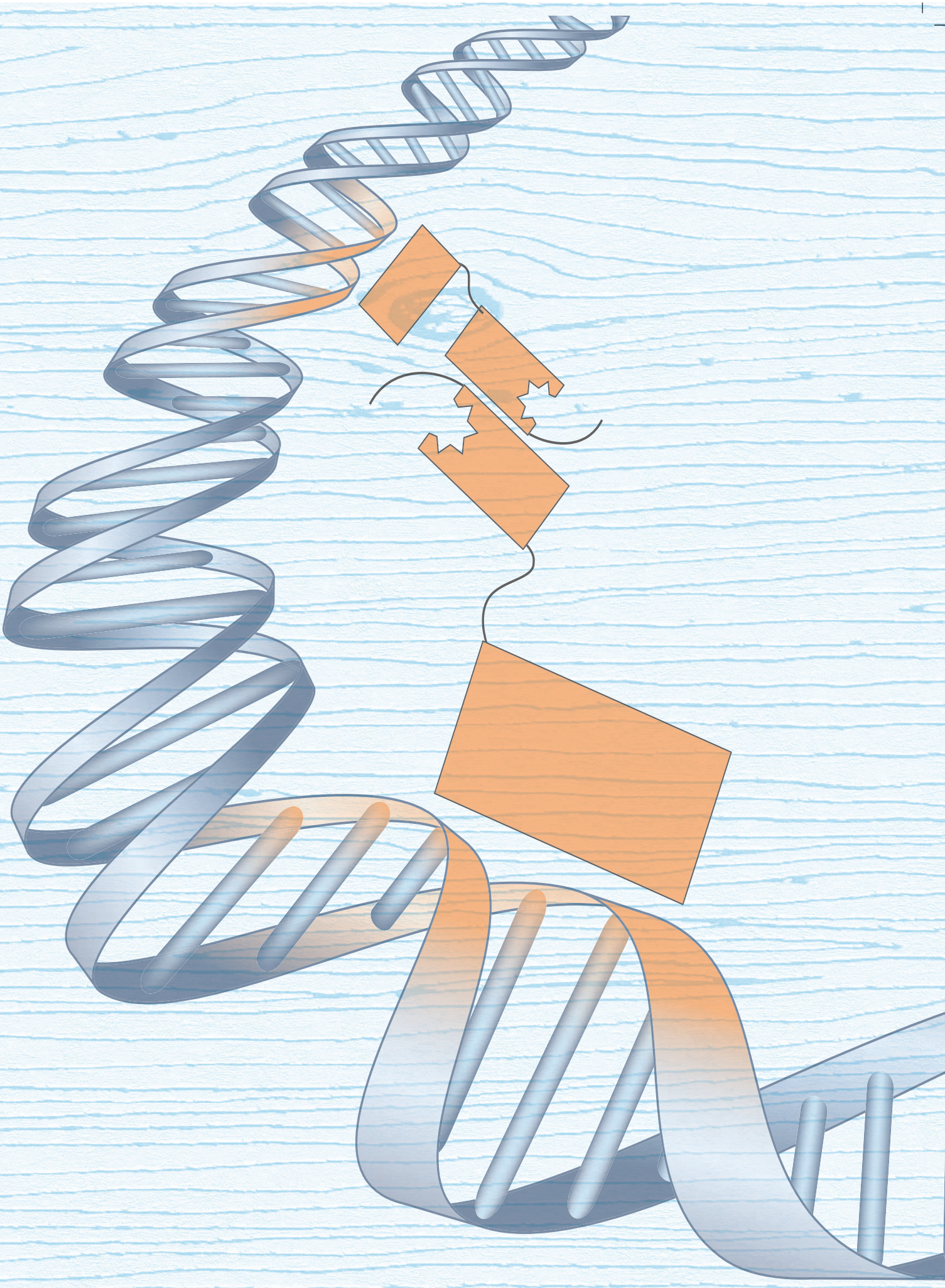
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 14 September 2018
at 11 a.m. in the Aula.

Teunke van Rossum
Development of a transcriptional regulator-based bioreporter - towards a generic selection method for novel enzymes, 242 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summary in English

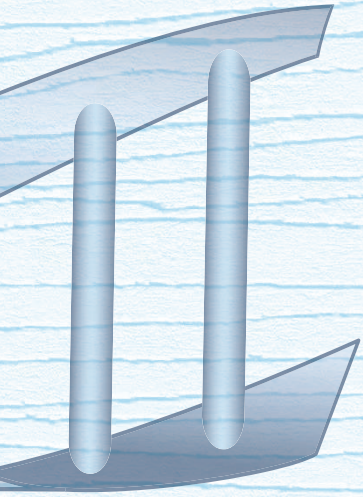
ISBN: 978-94-6343-326-6
DOI: https://doi.org/10.18174/456660

Table of contents



Chapter 1

General introduction and Thesis outline



General introduction

A brief history of enzymes

Enzymes are active proteins that are required for living organisms to catalyse their biochemical reactions. They speed up a reaction by lowering the activation energy, thereby allowing the equilibrium to be reached more quickly[1]. More than 8000 years ago, people unknowingly already made use of enzymatic conversions via fermentation by whole-cell microorganisms to make early forms of bread and beer (Fig. 1). The first application of a cell-free enzyme was probably the use of chymosin, a protease that is part of rennet from animal stomachs, for cheese making more than 7000 years ago[2,3]. However, the elucidation of the underlying mechanisms of these processes was comparatively recent. It started in 1830, when Gerardus Johannes Mulder first mentioned the name proteins[4]. Only a few years later, two major discoveries were made. Anselme Payen discovered the first enzyme, the starch degrading diastase, and Charles Cagniard de Latour discovered yeast, after being prompted into biology when the French Academy of Science promised a prize of one kilogram of gold for a solution of the mystery of fermentation. The second half of the 19th century was characterized by the step by step unravelling of what enzymes are and how they function, as well as by the first industrial enzyme production. In 1869, Friedrich Miescher used highly impure pepsin, extracted from pig stomach, in his discovery of DNA[4,5]. Five years later, Christian Hansen marketed the first industrially produced enzymes, crude rennet[6]. In 1876, Wilhelm Friedrich Kühne discovered trypsin, a substance in pancreatic juice that degraded other biological substances. One year later, he was the first person to call biological catalysts 'enzymes', meaning 'in yeast', to distinguish enzymes from the micro-organisms that produce them[1,4]. A crucial step in unravelling the mechanism of enzyme function was made by Emil Fischer in 1894. He postulated the key-lock principle, in which substrate and enzyme fit perfectly to one another, a theory 60 years later extended by the induced fit theory[7]. Another major step was made three years later by Eduard Buchner, who demonstrated that fermentation was possible with an extract of yeast in the absence of intact yeast cells.

In the 20th century, enzyme research accelerated and increasingly more enzymes were discovered. In 1909, Wilhelm Johannsen introduced the term 'gene' for the carrier of heredity and Sir Archibald Garrod described enzyme deficiencies as cause of certain human diseases[4]. In 1913, Leonor Michaelis and Maud Menten showed that the enzyme-catalysed

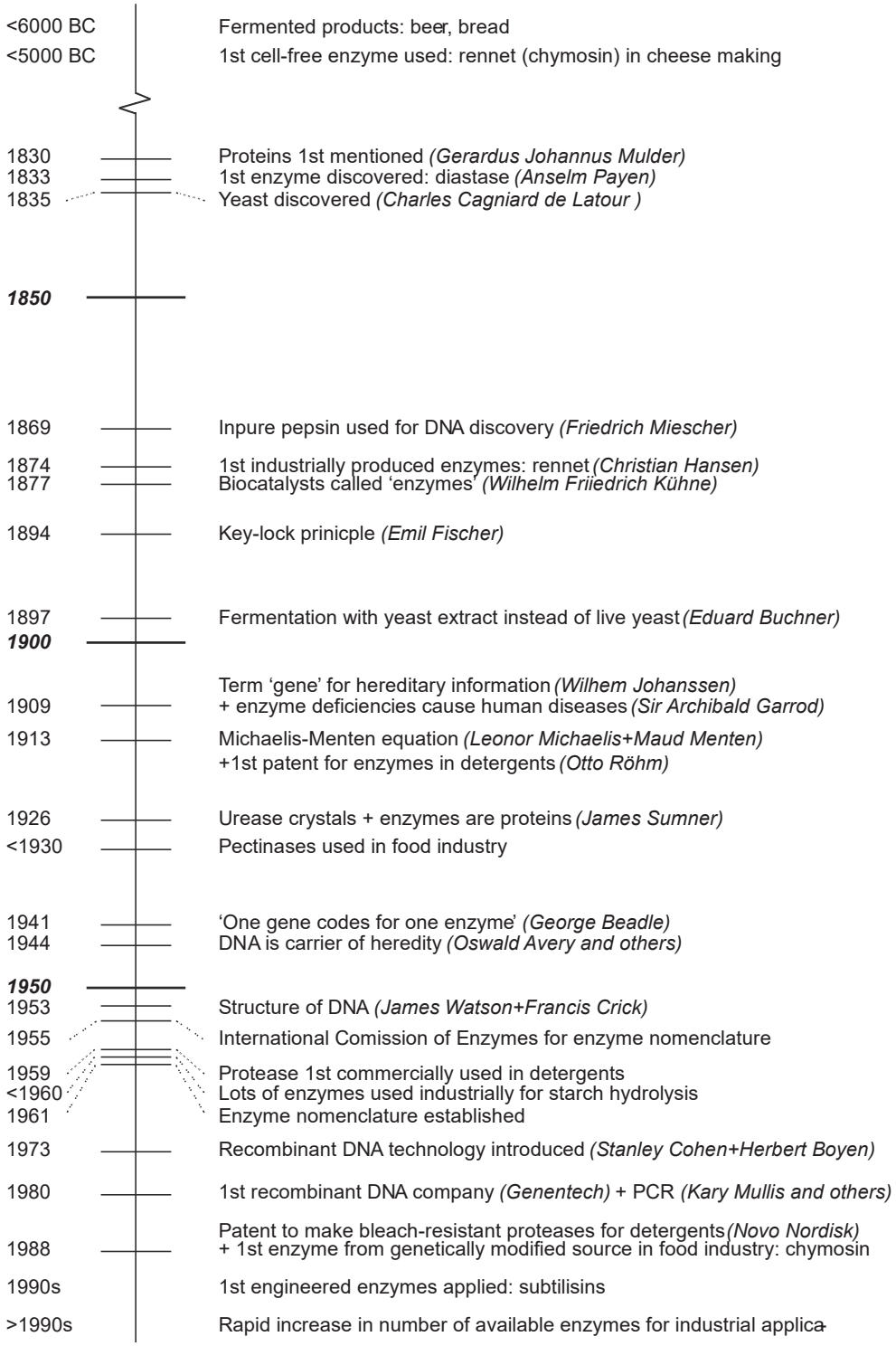


Fig 1: Timeline of a summarized history of enzymes

reaction rate was proportional to the concentration of the enzyme-substrate complex, leading to the well-known Michealis-Menten equation[8]. The 20th century also dates the onset of the industrial application of enzymes with, in 1913, the first patent on pancreatic enzymes in detergents by Otto Röhm[9], and by 1930, the use of pectinases in the food

industry for clarifying fruit juices[2]. In the following decades, four important discoveries were made not only for the future of enzyme research, but more importantly for biology in general. James Sumner demonstrated the protein basis of enzymes after he crystallized urease in 1926. In 1941, George Beadle postulated 'one gene codes for one enzyme'. In 1944, Oswald Avery and co-workers proved that DNA is the carrier of heredity[10] and in 1953, James Watson and Francis Crick unravelled the structure of DNA[11]. These discoveries form the basis of all molecular biology today. Meanwhile, more and more enzymes were discovered, leading to much confusion in enzyme names. As a result, the General Assembly of the International Union of Biochemistry (IUB) decided in 1955 to set up the first International Commission on Enzymes in order to formulate a nomenclature for enzymes. This nomenclature was established in 1961 and since then updated regularly (http://www.sbcs.qmul.ac.uk/iubmb/enzyme/). How is it structured? Enzymes are divided into six main classes based on the reaction type (Table 1) and each class with further subdivisions is indicated with an Enzyme Commission or EC number. The main classes are Oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5) and ligases (EC 6)[12]. With this nomenclature in place, much confusion like equal names for enzymes with different functions was prevented.

Table 1: Main enzyme classes[12,13]

EC number[a]	Enzyme name	Type of reaction	Reaction schema
EC 1	Oxidoreductases	Transfer of electrons (or hydride ions or H atoms)	$A^{e-} + B \rightarrow A + B^{e-}$
EC 2	Transferases	Transfer of a functional group	$A\text{-}B + C \rightarrow A + B\text{-}C$
EC 3	Hydrolases	Transfer of a functional group to water	$A\text{-}B + H_2O \rightarrow A\text{-}OH + B\text{-}H$
EC 4	Lyases	Addition of groups to double bonds or formation of double bonds by removal of groups	$A\text{-}B\text{-}C\text{-}D \rightarrow A\text{-}D + B{=}C$
EC 5	Isomerases	Transfer of groups with molecules, giving isomers	$A\text{-}B \rightarrow B\text{-}A$
EC 6	Ligases	Formation of a bond by condensation coupled to ATP cleavage	$C + D + ATP \rightarrow C\text{-}D + ADP + P_i$

[a]EC number is the Enzyme Commission number.

In the second half of the 20th century, increasingly more enzymes were industrially applied. In 1959, the first commercial protease in detergents was used[4] and by 1960 a lot of enzymes were used in the food industry for starch hydrolysis[2]. A very important development for the industrial application of enzymes was the introduction of recombinant

DNA technology by Stanley Cohen and Herbert Boyer in 1973. This followed the discoveries of DNA modifying enzymes, starting with DNA polymerase I by Arthur Kornberg in 1958. The first recombinant DNA company, Genentech, was founded in 1980. In the same year, Kary Mullis and others developed PCR technology. In 1988, a patent was awarded to Novo Nordisk for a process to make bleach-resistant proteases for the use in detergents; in the same year, the first enzyme from a genetically-modified source, chymosin, was approved for use in the food industry[4]. In the 1990s, the first engineered enzymes, subtilisins, were applied[14] and with the use of 'omics' and protein engineering technologies the number of enzymes available for industrial applications increased enormously from the 1990s onwards.

Current interest in enzymes

The current interest in enzymes is mainly based on their high specificity and selectivity as well as the more sustainable nature of processes that use enzymes instead of chemical catalysts. In addition to their high substrate specificity, enzymes regularly have a high enantio- and regioselectivity, which allows for a relative pure product formation and less waste compared to chemical synthesis[1,15,16]. In some applications though, e.g. pulp and textile production, a broader substrate specificity is desired[17]. Next to a reduction in waste, enzymatic processes are usually more sustainable, because they use less water, less harsh chemicals, no blocking and deblocking steps, they are very efficient and they can be operated under relatively mild conditions like pH and temperature. The enzymes themselves are biodegradable and can be produced from renewable sources by (micro)organisms[1,15,16]. However, it should be kept in mind that an enzymatic process is not per se more sustainable than a chemical process. It is not the enzymatic reaction alone that should be considered, but also the whole pipeline, including the upstream- and downstream processes and the enzyme production[18]. With the finite nature of fossil fuels and their negative effects on the climate and the environment, a shift should be made from a fossil fuel based economy to a biobased economy. To be able to make this shift, the implementation of enzymes in industry is very important. This enzyme-based industrial sector is called 'white biotechnology' and not only includes the direct application of enzymes themselves, but also the indirect application in the form of whole-cell biocatalysis. Currently, only 5% of chemical products is produced biologically and of the 3000 existing enzyme types, only 150-170 are being applied[1]. Next to existing enzymes, new enzymes can be found to generate any product (natural or non-natural), because, theoretically, proteins can catalyse any conversion that is

thermodynamically feasible. Thus there is much room for improvement towards a more sustainable and circular economy. This year, a global increase to more than 7.4 million metric tons of biobased materials and chemicals is expected (Lux Research analysts).

The global enzyme market in 2016 was well over $5 billion (report GMI743 by Global Market Insights, 2017) and growing with a compound annual growth rate (CAGR) of 8.2% from 2013 to 2018 (report BIO030H by BCC Research, 2014). Of this market, more than 50% was dominated by the companies Novozymes, DSM and Danisco (report GMI743 by Global Market Insights, 2017). The application areas of enzymes are very diverse and include food and beverages, textile, detergents, pulp and paper, animal feed, leather, biofuel, and fine and commodity chemicals, but also specialty applications like pharma, analytical devices and DNA technology[1,15,17,19]. Of all these areas, food and beverages held with 36% the biggest share of the global enzyme market in 2016 (report GMI743 by Global Market Insights, 2017). The most applied enzyme types are carbohydrases and proteases, followed by phytases, lipases, polymerases/nucleases and some others (report GMI743 by Global Market Insights, 2017). The majority of these enzymes belongs to the enzyme class of hydrolases and is used for the breakdown of molecules, e.g. stains in case of enzymes in detergents.

Obtaining novel enzymes

As indicated above, only a small part of the currently produced chemicals is made biologically. Enzymes can play a major role in increasing this number, but how to find novel enzymes for industrial applications? Three strategies can be followed: natural evolution, laboratory evolution and computational design (Fig. 2). In natural evolution, one makes use of the biodiversity that nature has to offer by enriching (micro)organisms with the desired bioconversion activity or by metagenomics (sequencing and/or cloning of isolated DNA into libraries followed by screening). The latter method is currently more used to be able to exploit the diversity of the non-culturable microorganisms, since less than 1% of all microorganisms is culturable[20,21] and only about 1% of all enzymes is known[15]. Screening can be done based on a known sequence with an oligonucleotide or based on the function of the target enzyme[16,22]. In addition, due to the exponential growth of available sequence data, new enzyme structures and the rapid development of bioinformatics tools, identification of new enzymes by mining of this enormous amount of data is increasing every year[16]. A very interesting group of microorganisms to exploit are extremophiles. These organisms live in extreme environments and therefore have enzymes with the ability to function under these

conditions, which is often a desirable characteristic for industrial applications[22-24]. For example, thermostable proteins are in generally more stable, but can also reduce cooling costs and allow the process to be performed at elevated temperatures where non-natural substrates are more soluble. In general, more stable enzymes maintain proper functionality during industrial biotransformation due to extended half-lives[23,24].

Instead of searching for enzymes in nature, however, one could also engineer enzymes by laboratory evolution or by computational design. In laboratory or directed evolution, a known enzyme is used as starting point and enzyme variants are generated in a randomly (e.g. epPCR) or semi-randomly (e.g. site-saturation mutagenesis) fashion, followed by screening of the variants[17,25,26]. The choice of the screening method is essential to obtain the desired variants, since 'you get what you screen for' [27]. Lately, a shift occurs towards directed evolution strategies that do not just give improved variants, but also give insight into the molecular mechanism of these variants[26]. In computational design, the variation is generated *in silico*, followed by *in silico* screening and eventually experimental verification of some selected variants[16,22,28]. Enzymes can be designed *de novo* or using an existing enzyme as starting point. Overall, there are a lot of different methods for enzyme engineering and not one single approach is suitable for each enzyme[15]. Since both laboratory evolution and computational design have their own advantages and disadvantages, a combination of the two is currently most fruitful[16].

Although many methods exist to find or generate enzyme variants, either by metagenomics or by enzyme engineering, the limitation is the ability to screen these large libraries for enzymes with the desired functionality[16,25]. Often screening methods are time-consuming, complicated and/or require expensive equipment. The methods range from simple and low-throughput, like screening on agar plates based on physical properties such as pH, to complex and high-throughput, like fluorescence activated cell sorting (FACS). Microtiter plate screening is medium-throughput, but it allows many different analytical methods to study enzyme activity. Growth complementation is high-throughput, because it is selection rather than screening, meaning that only positive cells stay in the pool and therefore less clones need to be screened further on. However, a high false positive rate is a risk and it is dependent on the reaction whether this approach is possible at all. *In vivo* methods have the advantage that functional proteins can be produced, e.g. with cofactor and proper folding, as long as there are no issues with functional heterologous expression, but they are limited by the transformation efficiency of the host organism and the ability of substrate, product or enzyme to cross the cell boundary. In contrast, *in vitro* methods do not

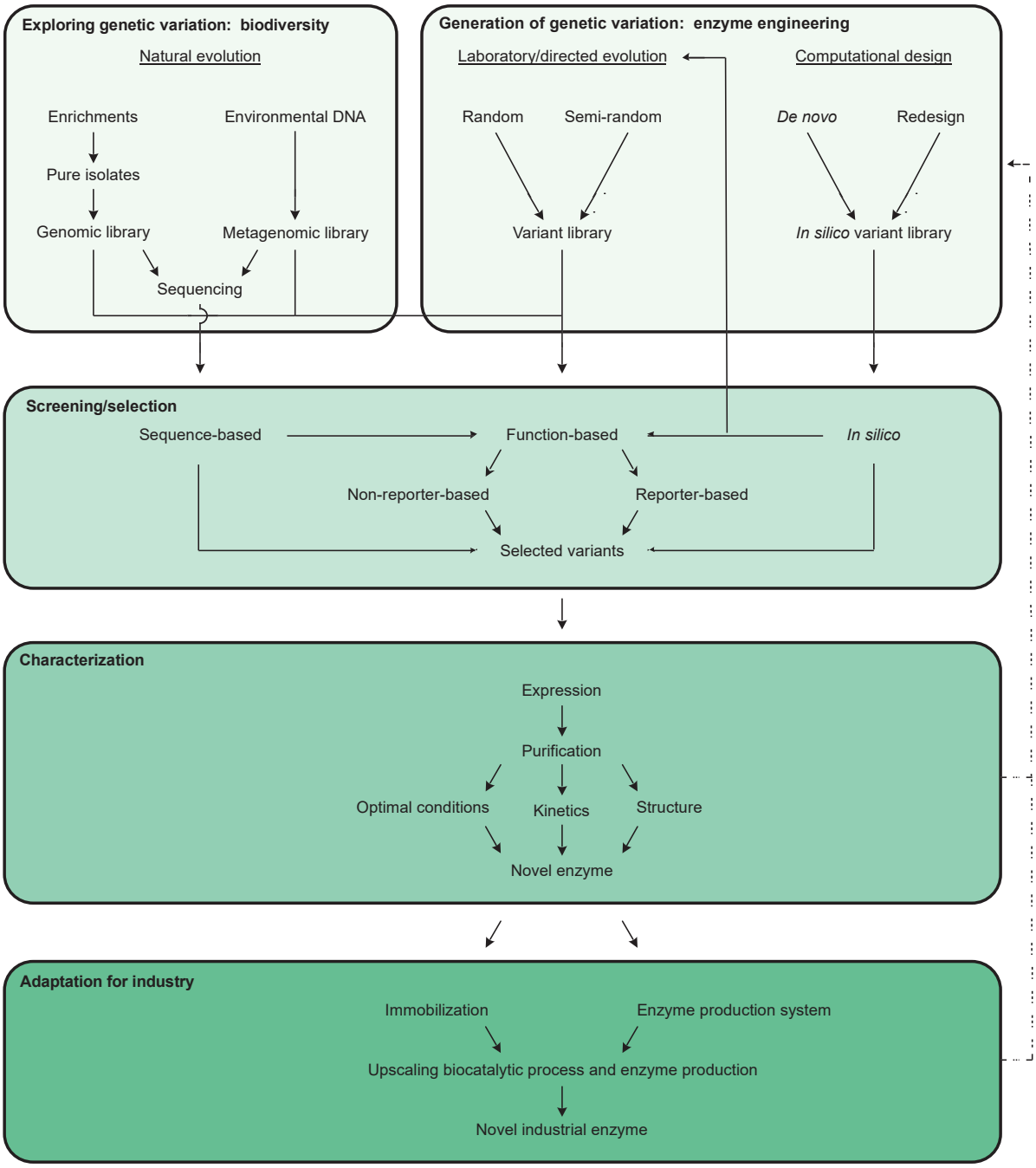


Fig. 2: Overview of approaches for obtaining novel enzymes

have these disadvantages and are more high-throughput, but they are only suitable for a select number of enzymes and can be more complex. To deal with the large numbers in screening, current trends are reporter-based screening, microfluidics, robotics, and smaller and more focused libraries in engineering[16,17,29,30]. The most important criterion for success in screening is the ability to link phenotype to genotype. Without this link the positive hit cannot be traced back.

Unfortunately, once the target enzyme is obtained, it is not immediately ready for

application in industry. Often further optimization of the enzyme properties, like activity and stability, is still required for functional performance under industrial conditions. For this optimization, directed evolution is a common approach, but the stability can also be increased by immobilization of the enzyme. Immobilization also allows for easier recovery of the enzyme and thus usage through multiple rounds of biocatalysis[15,22]. Since the enzyme costs are mostly the highest costs of the biocatalytic process, it is very important to develop a system to produce sufficient quantities of the enzyme against a reasonable production cost[23,24]. This development involves the choice of the production host and the expression method, possibly followed by optimization steps, but also the choice for the fermentation method, the downstream processing approach and the final product formulation[1]. Of course, also the biocatalytic process should be designed from upstream- to downstream processing. A last and crucial step is the upscaling of both the enzyme production and the biocatalysis processes. As upscaling means that process parameters change, this often results in different outcomes than with laboratory scale set-ups and thus requires some further optimization[31].

Reporter-based screening and selection for novel enzymes

An interesting approach to deal with the large numbers in screening for novel enzymes is reporter-based *in vivo* screening or selection. In reporter-based strategies, it is not the product of an enzymatic reaction or the enzymatic conversion itself that results in a measurable property, but rather a genetically encoded reporter that gives a discriminating phenotype. Since the enzyme activity is monitored indirectly, this strategy is independent of the reaction and thus can be applied for different enzyme types. The whole-cell system that functions as reporter is called whole-cell bioreporter or simply bioreporter and its detection output can be chosen based on the reporter choice, e.g. bioluminescence or fluorescence[32]. It thereby allows for high-throughput screening methods like FACS for a wide range of enzymes, not just for enzymes that can convert a substrate to a fluorescent product. For reporter-based screening, various strategies are possible, like riboswitch-based or posttranslational-modification-based strategies, but the most common strategy uses a transcriptional regulator as sensor for enzyme activity. The product produced by the enzyme is very specifically bound by the transcriptional regulator, which undergoes a conformational change that alters its DNA binding capacity, switching on expression of the reporter gene[33,34]. For each product, the specificity of the sensor should be modified. Also the

development of other reporter-based strategies takes time and effort. Nevertheless, the wide applicability, the possibility to screen based on enantioselectivity, the possibility for signal enhancement, the lacking requirement for artificial substrates and the possibility for high-throughput screening make this a very powerful screening method. A more elaborate review of reporter-based screening and selection is given in **Chapter 2**.

The most common reporter-based screening method uses GFP as reporter and screening by FACS. This method is high-throughput and has proven very successful in finding novel enzymes. However, it does require expensive equipment and experienced people to handle this equipment. The aim of this thesis is to simplify this technology by providing a generic and high-throughput reporter-based selection system. The advantage of selection over screening is that only positive cells, containing the active enzyme, stay in the pool of variants. This allows for a rapid reduction of the initially large library size. Although reporter-based selection systems are being used, they are often not applicable for a wide range of enzymes. In this thesis, the modular set-up of the system should make it generic. The system is based on the most common reporter-based strategy, namely the transcriptional regulatory-based strategy. An *in vivo* transcriptional regulator-based selection system or bioreporter is developed that couples enzymatic activity to growth of the bacterium *Escherichia coli*. Since a high false positive rate is often a problem encountered for growth-based selection, this system was designed with dual reporters, a selection and a screening reporter. The transcriptional regulator binds the product of the enzymatic reaction and switches on transcription of both reporters. After reducing the pool size by growth-based selection, positive cells can be screened based on bioluminescence to exclude false positives and to quantify the response.

Thesis outline

The general aim of this thesis is to develop a generic and high-throughput *in vivo* reporter-based selection system or bioreporter as a simpler and alternative method for the currently available enzyme screening methods. To this end, a dual reporter system is developed, in which the transcriptional regulator AraC (the regulator of L-arabinose metabolism in *E. coli*) regulates expression of both a selection and a screening reporter. AraC is chosen as transcriptional regulator, because it has been well studied, a protein structure with and without ligand is available, and it has been a topic of several engineering studies. Once

developed (**Chapter 3**), a proof of principle of this system should be provided. This involves several aspects: (1) the ability to detect an enzymatic activity, (2) the applicability at library scale, and (3) the changeability of the specificity to make the system applicable for a wide range of enzymatic products and thus enzymes. The first two aspects are described in **Chapter 3**, whereas the latter aspect is dealt with via two approaches in **Chapters 4** and **5**. One approach is the replacement of AraC by another transcriptional regulator, namely LacI (**Chapter 4**). The second approach focuses on engineering the ligand specificity of AraC (**Chapter 5**). In addition, unexpected inhibitory and stimulatory effects of L-arabinose on growth of *E. coli* are discussed in **Chapter 6**.

Firstly, in **Chapter 2**, an elaborate review of *in vivo* screening and selection strategies for finding novel enzymes is provided. The focus is on *in vivo* reporter-based systems in which the activity of a reporter is controlled by the activity of an enzyme of interest. The different mechanisms of these systems, including those based on transcriptional regulators, are described. A comparison is made of the various *in vivo* screening and selection strategies as well as the various reporter-based mechanisms. Some general advantages and disadvantages of reporter-based screening and selection are discussed.

In **Chapter 3**, the development of a growth- and bioluminescence-based bioreporter for the *in vivo* detection of novel biocatalysts is described. The sensor part of this bioreporter is based on the transcriptional regulator AraC that controls expression of both a selection reporter (LeuB or KmR; enabling growth) for rapid reduction of the initially large library size and a screening reporter (LuxCDABE; causing bioluminescence) for further quantification of the positive variants. The characteristics of four systems that differ in the used selection reporter and in the origin of replication are compared. Most importantly, a proof of principle is provided using the best performing dual reporter system and a to be screened enzyme, L-arabinose isomerase, either from mesophilic or thermophilic origin.

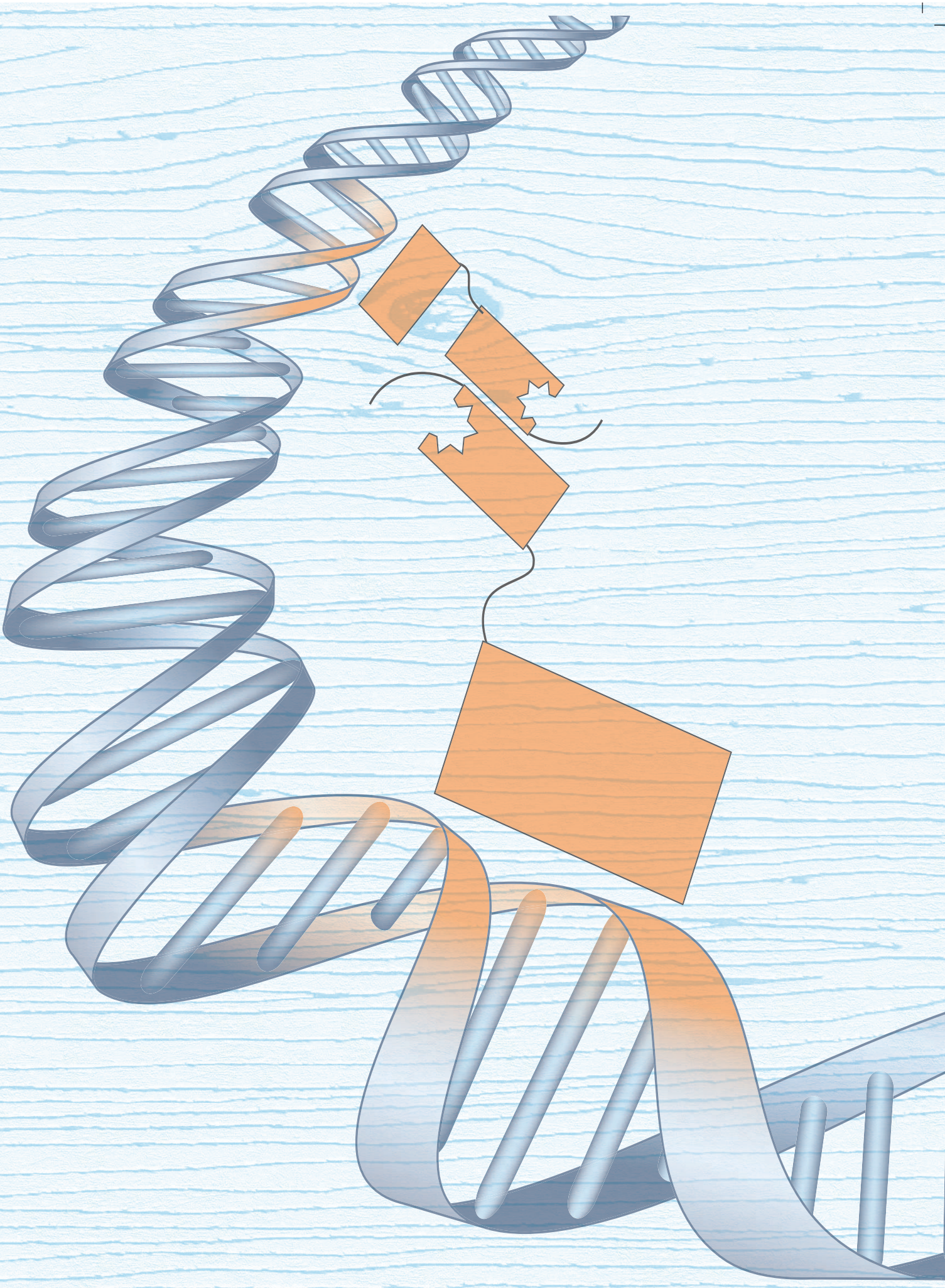
However, to show that the developed bioreporter is generic and thus applicable for a wide range of enzymes, its specificity should be adaptable towards the product of the enzyme. In **Chapters 4** and **5**, two different approaches to change the system's specificity are outlined. In **Chapter 4**, the replacement of the transcriptional regulator AraC by a different transcriptional regulator, LacI (the regulator of lactose metabolism in *E. coli*), is described. The characteristics of four different systems are compared, all having LacI as transcriptional regulator, but varying in the selection reporter and in the origin of replication. A specificity test with the best performing system is included, using previously described weak inducers and anti-inducers. In addition, the newly developed LacI-based system is compared with the

original AraC-based system.

A second approach to alter the system's specificity is described in **Chapter 5**. This approach focuses on engineering the ligand specificity of AraC to D-xylose by targeting residues in the ligand binding pocket with combinatorial site-saturation mutagenesis. Although others already successfully modified the specificity of AraC, the aim here is to offer a simpler and alternative method to the commonly used GFP- and FACS-based screening of transcriptional-regulator variants by using growth-based selection instead. To this end, the dual reporter system itself is applied for selection and screening of variants. A description of the complete process is provided, starting from library design and formation and ending with kanamycin resistance-based selection and bioluminescence-based screening of these libraries in the presence of D-xylose. Finally, the response of the resulting variants to various monosaccharides is discussed.

During the experimental work with the AraC-based dual reporter system (**Chapters 3** and **5**), inhibitory and stimulatory effects of L-arabinose on growth of the system strain were observed. **Chapter 6** provides an overview of these observations as well as follow-up experiments to understand the underlying regulatory mechanisms of these effects. The growth effects elicited by L-arabinose are described for the system strain, its parent strain *E. coli* BW25113 and various single knockout strains in both LB medium and M9 minimal medium. Based on the different genotypes and phenotypes of the various strains, hypothetical regulatory mechanisms are discussed that may explain the L-arabinose effects on growth of *E. coli*.

Chapter 7 provides a summary of the work presented in this thesis and a general discussion on the developed screening and selection system or bioreporter. The bioreporter is compared to other screening methods based on various comparison criteria such as the handling and the success and false positive rates. In addition, suggestions to further improve the bioreporter are provided, including the construction method and the reporter choice. To conclude, perspectives are discussed of the developed method specifically and transcriptional regulator-based bioreporters in general.

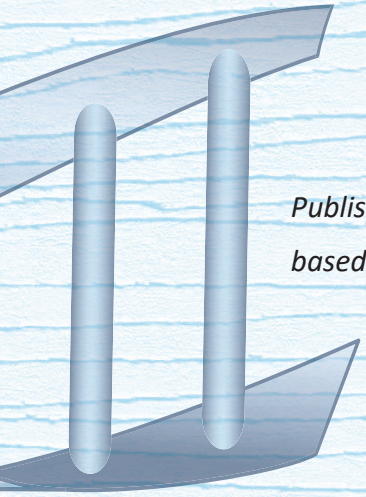


Chapter 2

Reporter-based screening and selection of enzymes

Teunke van Rossum[a], Servé W.M. Kengen[a] and John van der Oost[a]

[a]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands



Published as: van Rossum, T., Kengen, S.W.M., and van der Oost, J. (2013) Reporter-based screening and selection of enzymes. FEBS J *280: 2979-2996.*

Abstract

The biotech industry is continuously seeking for new or improved biocatalysts. The success of these efforts is often hampered by the lack of an efficient screening assay. Thus, to be able to extend the number of enzymes available for industrial applications, high-throughput screening and selection methods are required. In the last few years an impressive range of screening and selection strategies has been developed. In this review, we will mainly focus on *in vivo* reporter systems in which the activity of a reporter is controlled by the activity of an enzyme of interest. Different mechanisms can be distinguished: (a) binding of the product of the enzymatic reaction to a transcriptional regulator and thereby turning on transcription of the reporter; (b) direct modification of a transcriptional regulator by the enzyme resulting in expression of the reporter; (c) binding of the product to a regulatory riboswitch or ribozyme, resulting in translation of the reporter; and (d) direct modification of the reporter by the enzyme, altering the reporter's activity. The choice for either a selection or a screening strategy depends on the type of reporter, e.g. providing antibiotic resistance (selection) or transmitting a fluorescent signal (screening). Although developing the specificity of each of these reporter-based selection or screening systems towards a certain enzymatic reaction is not yet straightforward, their adjustable modular design appears to be a promise for general applicability in the near future.

Introduction

Enzymes are unique because of their catalytic power as well as their extraordinary specificity and (enantio and regio-) selectivity[35,36]. In addition, they can be employed under relatively mild temperatures and pH values with water as solvent, conditions that are energy efficient and environmentally friendly[36,37]. In theory, proteins can catalyse any thermodynamically feasible conversion. Hence, any product (natural or unnatural) can be generated with biocatalysts that possess the appropriate catalytic features. These characteristics make them very interesting for industrial processes. Examples of applications range from the production of pulp and paper, textiles and leather, to fine chemicals, food additives and pharmaceutical intermediates[35,36]. The fact that such enzymatic solutions are preferable over the traditional chemical ones has resulted in a shift towards biocatalysts in recent years[36]. The industrial enzyme market, comprising about 100 enzymes (over half originating from fungi, over one-third from bacteria and the remainder from archaea, animals and plants), increased between 1998 and 2009 from $1.6 billion to $5.1 billion[36]. Still, biocatalysts are often not efficient enough, too costly or just not available[21,36]. Industrial processes are often operated under rather extreme conditions, such as high temperature or pressure, non-neutral pH and non-aqueous solutions. Although enzymes are faster and more environmentally friendly than traditional chemical catalysts[38], these harsh conditions are demanding, especially on the enzyme stability; this implies that the performance on an industrial scale is often insufficient. So, novel or improved biocatalysts are urgently required[38,39].

New enzymes can be obtained (a) by making use of natural evolution (enriching (micro)organisms with desired bioconversion activity, generating (meta)genomic libraries, and subsequent screening and selection), (b) by performing laboratory evolution (screening and selection of libraries of randomly generated enzyme variants) or (c) by conducting computational evolution (*in silico* variation, followed by *in silico* screening, and eventually experimental verification of a few selected variants)[40]. Nature is an excellent resource for novel biocatalysts as it has had billions of years to evolve enzymes for a range of reactions. When searching for enzymes with certain characteristics, one could explore those environments that most probably host microorganisms that require those enzymes. For example, for novel lignin-degrading enzymes one could isolate lignin-degrading microorganisms from rain forest soils[41], and for enzymes stable at extreme conditions such as high temperature or high salt concentrations one could look in extreme environments like hot springs or salt marshes[38]. Unfortunately, it is estimated that < 1% of all microorganisms

are culturable[20,21,42]. Metagenomic libraries are therefore particularly valuable, but the number of available libraries exceeds the possibilities of investigating them[43]. Moreover, sequence-based *in silico* library screening may run into problems due to functional mis-annotation and a bias for previously gained available information, actually preventing discovery of novel enzymes[44]. On the other hand, the probability of identifying a certain gene in experimental screens or selections depends on several practical features: a gene's abundance in the generated metagenomic library, the size of the target gene, the presence of a full-length sequence, the selected heterologous host, the expression system, and last but not least the assay method[20,43].

Although nature is a very good source for enzymes, these enzymes generally perform optimally in the context of a living cell (moderate activity, narrow specificity, moderate stability and short life span), but they are often less suited for desired performance in an industrial setting (high activity, broad specificity, high stability and long life span). Laboratory evolution (often referred to as directed evolution) is a powerful approach to alter enzyme characteristics, such as substrate specificity, enantioselectivity and stability. This iterative process involves the generation of random genetic diversity by introducing point mutations or by recombination, followed by high-throughput screening or selection for desirable variants[45]. Remarkable progress has been made in this field, and after the initial harvest of low-hanging fruit Goldsmith and Tawfik[46] stated that 'directed evolution is now ready to tackle high-hanging fruit'. The major advantage is that significant changes in enzyme characteristics are possible in the absence of the enzyme's structure or detailed knowledge on the catalytic mechanism. This may work when relatively small changes (e.g. single amino acid substitutions) already contribute to the improvement of an enzyme for a certain feature or when multiple changes are cumulative, e.g. thermostability[47-49]. However, when more complicated adjustments are desired (e.g. adaptations that require introducing > 10 amino acid substitutions at once), sampling of sequence space without any prior knowledge is an impossible task, because the size of the library one needs for such an experiment is just too big to synthesize, let alone to screen[28]. In *in vivo* screening or selection the library size is limited by the transformation efficiency, which in practice implies a value of 10^9 for *Escherichia coli*. Enlarged capacities can be obtained by adjusting the overall procedure, either by carrying out library creation and screening both *in vitro* or by performing development of library diversity and screening both *in vivo*[32]. In addition, increasing the manageable library size and making the handling of large libraries more straightforward may contribute to solving high-throughput problems in screening (likewise applicable for

metagenomic libraries). Interestingly, a recent trend directs towards smaller but smarter libraries, for which information on sequence, structure, function and evolution is integrated, and sometimes even combined with computational design[39,46].

Computational design by itself is also an interesting approach to extend the number of available enzymes for industrial applications, ranging from relatively simple enzyme improvements to the more challenging design of biocatalysts for completely new reactions[21,28,50]. Depending on the computer power, astronomical numbers of variants can be efficiently screened *in silico*. Despite impressive advances in computational enzyme design (both *de novo* and by re-designing existing systems), the actual improvements of enzyme performance obtained by the designed systems are rather poor, and still far from that of analogous systems resulting from natural evolution[28,50,51]. Unexpected behaviour or inactivity of designed enzymes in wet-lab experiments generally relates to insufficient insight in an enzyme's overall structure (including poorly structured elements), in an enzyme's active site and in the catalytic mechanism[21,28,52]. A recent development in this field is the movement towards *in silico* directed evolution, including *in silico* screening of variants[28,53]. This and other computational design methods are a very important step towards creating smart libraries for directed evolution. The combination of computational design and directed evolution potentially is a very powerful approach in enzyme engineering, certainly when combined with enzymological insights[21,28,37,46,50].

The search for novel and improved biocatalysts benefits from a wide range of recently developed approaches. Although many hurdles are still faced, the main obstacle remains the screening of large mutant libraries or metagenomic libraries for variants with the desired functionality. This can be a complicated and time-consuming effort, especially in the absence of a high-throughput screening or selection assay. In recent years various *in vitro* and *in vivo* screening and selection systems have been developed, which have been covered in some excellent reviews[37,43,45]. These reviews focused on *in vitro* and *in vivo* systems for screening of directed evolution or metagenomic libraries. *In vivo* methods involving reporter-based strategies were only briefly discussed. The current review will focus on the different *in vivo* screening and selection strategies as well, but with special emphasis on reporter-based strategies.

Together with the development of novel screening methods, there is also a growing awareness that smaller, more focused libraries are needed. However, as the focus of this overview is on screening/selection aspects, practical issues on library size and formation are only briefly mentioned. For a more elaborate discussion on this topic, the reader is referred

to other reviews[32,37,46].

General overview of *in vivo* screening and selection strategies

In vivo systems are defined as replicating cellular entities, in most cases bacterial cells that produce a library of protein variants of interest. As in all functional screening and selection systems, the phenotype and the genotype of the protein(s) of interest are linked in *in vivo* strategies. A major advantage of *in vivo* systems may be the functional production of the protein(s) of interest, for instance correctly folded and with the incorporation of a cofactor; in the case of heterologous expression, functional enzyme production may depend on the choice of the production host. In addition, by changing the screening or selection conditions of an *in vivo* system, one can tune the properties of the desired biocatalysts[54], e.g. expression in a thermophilic bacterium at elevated temperatures for obtaining variants with enhanced stability[47,48]. However, *in vivo* strategies are limited by (a) the transformation efficiency of the host, (b) functional expression of the protein of interest, (c) difficulties in substrate uptake, (d) less sensitive product detection because of complex intracellular background[42,55] and (e) the growth rate of the microbial host.

Before describing the details of the various strategies, it is important to indicate the difference between 'selection' and 'screening'. In selection approaches, negative clones are not present in the final pool, e.g. because they do not survive (Fig. 1A). The main advantage is that usually a much smaller number of variants has to be screened. Of course, one should realize that some false positives may arise as well (see below). In contrast, in screening approaches, all clones, negative and positive, are maintained, meaning that all library variants need to be screened, which makes this approach significantly less efficient[45]. However, screening may also have some advantages, like a better dynamic range, precision of activity measurements, tailored reaction conditions and the possibility to monitor multiple parameters[46]. The screening step is of course followed by selection, i.e. picking the positive clones. This can be done either manually with toothpicks or automatically with for example fluorescence activated cell sorting (FACS). In FACS, individual cells, emulsions of cells in aqueous droplets in oil or emulsions of aqueous droplets in oil (containing sophisticated *in vitro* expression systems with some colorimetric product detection) are separated in narrow channels, where illumination of cells or droplets occurs one by one by a focused laser beam. When the desired fluorescence is detected, a charge will be applied to the cell/droplet, resulting in deflection of the positive clone by an electrostatic field into a

collection tube[56].

Thus compared to screening, the selection method allows for analysing much larger libraries, namely ~ 10^9 versus ~ 10^5 [37,45,57]. It should be mentioned that more recent screening techniques such as cells in- droplet screens coupled to FACS are also used to screen relatively big libraries (~ 10^9 [32,37]). In the next paragraphs a general overview will be given of the different *in vivo* screening and selection strategies, discussing only non-reporter-based approaches. The reporter-based strategies will be discussed separately in a later section: Reporter-based *in vivo* screening and selection strategies.

Selection

The selection strategies employed vary with respect to the selective principle they are based on, but often the desired enzymatic activity is coupled to cell survival and growth. This makes selection efficient (libraries of ~ 10^9 [37]) and applicable to both metagenomic and enzyme variant libraries. However, selection on growth appears to result in more false positives compared with screening. The higher selective pressure may select for variant cells that circumvent the coupling of the enzyme activity to growth and are able to survive by a different mechanism.

Three approaches have been described in which enzyme activity is coupled to growth: development of enzymes that complement auxotrophy, development of enzymes that neutralize lethal conditions, and use of a specific enantioselective counter selection. In genetic complementation, microbial strains are used which are auxotrophic for the product of the enzyme of interest[32,45]. Hence, this approach is limited to enzymes that catalyse the synthesis of an essential product and for which an auxotrophic host is available or can be constructed by deleting or mutating the corresponding gene. Otten *et al.*[58] applied auxotrophy complementation to evolve the glutaryl acylase of *Pseudomonas* SY-77 into an adipyl acylase with an improved activity towards adipyl-7-aminodesacetoxycephalosporanic acid (adipyl-7-ADCA). The β-lactam component of the substrate is replaced by leucine to enable selection. Leucine auxotroph *E. coli* cells expressing an error-prone PCR library of the glutaryl acylase are grown in the presence of adipyl-leucine as sole leucine source. In this way, only enzymatic hydrolysis of adipyl-leucine allows for growth on minimal medium. Activity of the selected variants towards the desired substrate adipyl-7-ADCA is confirmed in a biochemical assay.

A second approach is the neutralization of increasing concentrations of toxic compounds (e.g. antibiotic resistance markers) or other lethal conditions (e.g. cold

shock)[43,45]. This is of course restricted to the subset of enzymes which have such neutralizing activity. The third manner to couple enzyme activity to growth is used to select for enantioselective enzymes. Formation of the desired enantiomer enhances growth, whereas the wrong enantiomer is toxic[37,45]. Fernandez-Alvaro *et al.*[59] applied this principle to select for enantioselective esterases. Two substrates are added to the medium, namely (*R*)-3- for

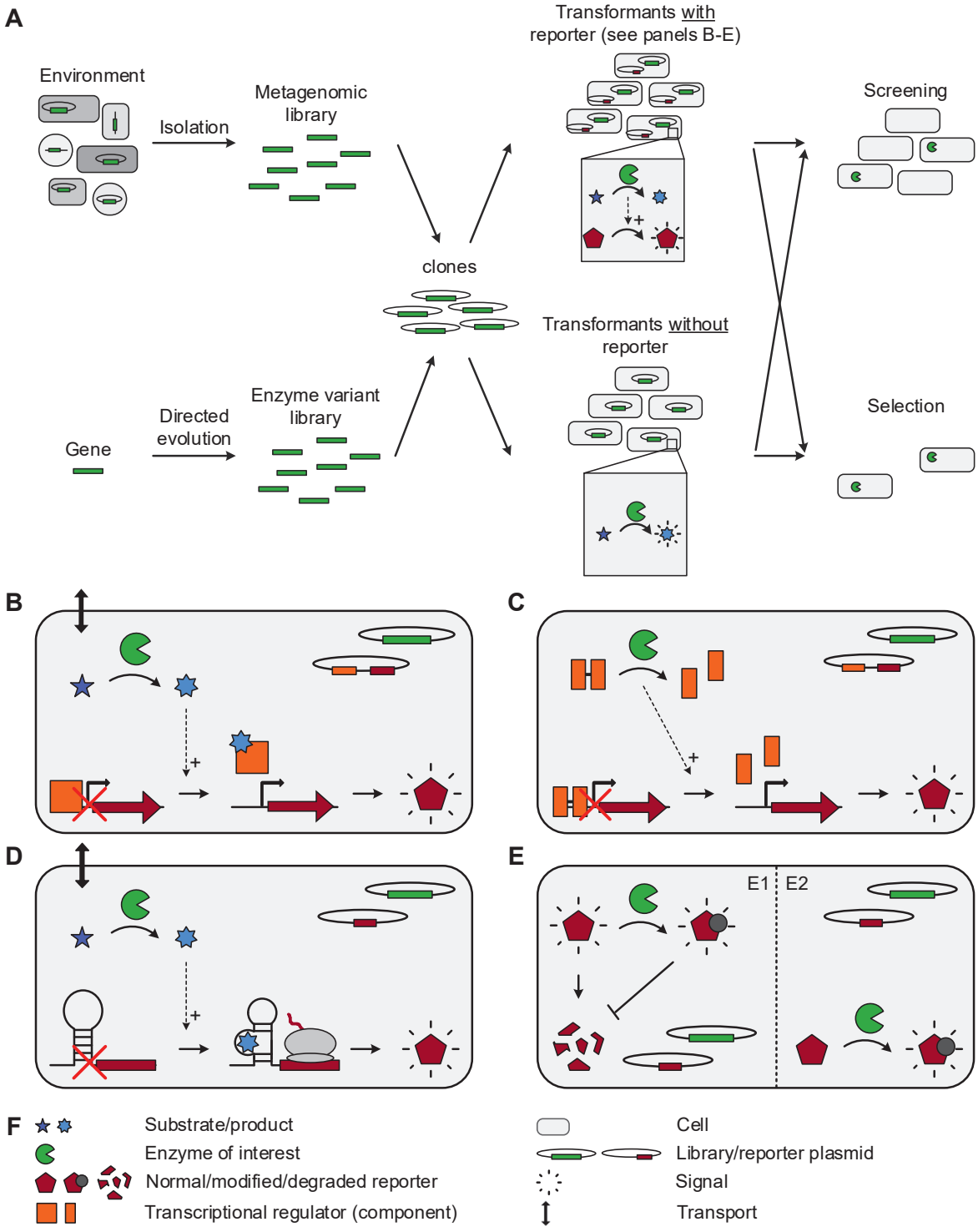


Fig. 1. Reporter-based *in vivo* screening and selection strategies. (A) Metagenomic libraries or enzyme variant libraries are created from DNA isolated from the environment or by mutating a gene through directed evolution respectively. The libraries are cloned into a vector and are subsequently transformed to cells either with or without a plasmid, encoding a reporter. In reporter-based strategies it is not the product or the enzymatic reaction itself that directly results in a measurable characteristic, but rather a genetically encoded reporter that gives a conditional phenotype. The activity of the reporter is dependent on the activity of the enzyme of interest via interference at several regulatory levels as shown schematically in (B)–(E). These strategies are defined as 'selection' or 'screening' depending on the type of reporter. The 'signal' the reporter gives can thus vary from growth to fluorescence when the reporter is an antibiotic resistant protein or GFP respectively. Non-reporter-based strategies are defined as 'selection' or 'screening' depending on the enzyme activity and/or product. The latter strategies are displayed in more detail by Leemhuis *et al.*[37]. (B) Natural transcriptional regulator based: binding of the product alters the conformation of the regulator, resulting in dissociation from the DNA and activation of transcription. (C) Synthetic transcriptional regulator based: the enzyme directly acts on the regulator, resulting in dissociation of its components and activation of transcription. (D) Riboswitch based: binding of the product to the riboswitch changes its secondary structure and activates translation. (E) Post-translational modification based: the enzyme directly acts on the reporter and modifies it, preventing degradation (E1) or resulting in detectability of the reporter (E2). (F) Legend.

enantioselective esterases. Two substrates are added to the medium, namely (*R*)-3-phenyl butaric acid that is covalently coupled to glycerol and (*S*)-3-phenyl butaric acid coupled to 2,3-dibromopropanol. Depending on the enantioselectivity of the esterase, either the growth-supporting glycerol or the toxic 2,3-dibromopropanol is released. Only *E. coli* cells with an esterase selective for the (*R*)-enantiomer survive and are selected based on live/dead staining and FACS.

Agar plate screening

Agar plate screening is the simplest format of screening: incubating colonies with a chromogenic substrate[37], which leads to colouring of the colonies themselves and/or the surrounding agar upon product formation (e.g. release of *o*-nitrophenol), or a coloured substrate that results in clearing halos around the colony upon substrate hydrolysis (e.g. disappearing Congo Red stained carbohydrates). Also indicators that react with the product can be added to the agar, e.g. Schiff's reagent that reacts with aldehydes[43]. In general, it is easy to operate and straightforward to identify active variants. However, low dynamic ranges of these screening approaches generally do not allow for accurately distinguishing differences in catalytic rates. Besides, only libraries of up to ~ 10^5 can be screened[32,37] and potential intracellular accumulation decreases the screening's sensitivity. The latter can be solved by coupling the initial *in vivo* screening to a second *in vitro* round after cell lysis[42]. For example, Böttcher *et al.*[60] aimed at obtaining enantioselective esterase variants. To do so,

overlaid replicated agar plates were covered with soft agar that contained the substrate 1-naphthyl-acetate and the staining salt Fast Red TR. Esterase-positive clones were coloured brownish/red by coupling of Fast Red TR with the esterase product 1-naphthol (product 1); the corresponding cells were lysed and tested in an enzyme cascade assay. Acetyl-CoA was formed from CoA and acetic acid (product 2) by acetyl-CoA synthetase, and CoA was regenerated by citrate synthase, converting oxaloacetate to citrate. The required oxaloacetate was derived from L-malate by L-malate dehydrogenase, meanwhile reducing NAD^+ to NADH. NADH generation was measured in a microtiter plate spectrophotometer. It was possible to do an enantioselectivity screen by comparing the hydrolysis of two pure enantiomers in separate wells[60].

Enantioselectivity can also be established already during the initial screening on plates via coupling to some downstream product-converting reaction(s). In an enzyme variant library (alanine racemase) expressed in *E. coli*, Willies *et al.*[61] detected active variants that catalysed the racemization of L- to D-alanine. Cells were grown on Hybond-N membranes laid on top of agar plates. D-alanine-producing clones were detected by using a coupling assay, based on a D-amino acid oxidase generating H_2O_2 that is monitored in a horseradish-peroxidase-catalysed colour reaction. Two subsequent screens were performed to remove false positives. Plasmid-encoded D-amino acid oxidase localized colour change to the cell, whereas an oxidase overlay led to colour diffusion over the plate[61]. In general, agar plate screening has successfully been employed to screen for multiple enzyme classes, such as cellulases, lipases/esterase, proteases, laccases and oxidoreductases[42]. Moreover, both enzyme variant and metagenomic libraries can be screened.

Microtiter plate screening

The most commonly applied screening strategy is based on microtiter plates. Single transformants are grown in standard 96-well microtiter plates[37]. Deepwell plates may result in elevated yields (2-mL wells, typically with 0.2 mL culture volume), whereas more wells per plate (384, 1536) enable a higher throughput. An interesting alternative is the micro-Petri dish designed by Ingham *et al.*[62]. This microbial culture chip of 36 x 8 mm has up to $1 \cdot 10^6$ wells with 10-µm-high laminate side walls on top of porous aluminium oxide strips; for supply of nutrients the chips can be placed on agar-like matrices. This design allows for rapid changes in cellular environment by simply transferring the chip from one medium to another. As a proof of principle *E. coli* cells were screened for expression of *lacZ*.

In general, a wide range of analytical methods can be employed for microtiter plate

screening, such as colorimetry, liquid/gas chromatography, NMR or mass spectrometry. In addition, the dynamic range of this strategy is excellent[32,37] and both enzyme variant and metagenomic libraries can be screened. Although often cells are lysed in order to enhance sensitivity or to make substrate available to the enzyme, it is not a strict requirement, as shown by Wagschal and Lee[63]. They developed a screen in which a 4-methylumbelliferyl (mu) labelled substrate is added to a ~ 50-100 µL *E. coli* culture, expressing variants of a biomass-degrading enzyme (wild-type or error-prone PCR library). The change in fluorescence is measured with a top-reading microplate spectrophotometer (no interference by turbidity). A particularly attractive feature in this case is the fact that a single microplate is used for cultivating picked colonies, enzyme expression, the fluorescence assay, and finally storage of the library in a freezer after dispensing a cryoprotectant such as glycerol. Besides, if cell lysis is required, a simple cell permeabilization protocol has been developed in which only a single reagent, polymyxin-B-sulfate, has to be added prior to a direct assay, with no further manipulation in the same microplate well[63]. So, for this permeabilization protocol it is not necessary to lyse the cells, centrifuge them and transfer the cell-free extracts to a second plate. The main limitations of this type of assay for the activities tested relate to the reaction rate with mu-tagged substrates and background substrate hydrolysis. Moreover, general applicability of this method to other hydrolytic activities is limited by the commercial availability of appropriate mu-labelled substrates[63]. Other drawbacks of microtiter plate screening are that only libraries of up to ~ 10^4 can be screened[32,37] and that the method is limited to reactions of which the product has some measurable characteristic. Artificial substrates can increase the applicability, but as they differ from the substrate in the desired application other enzymes than the one demanded can be found. After all, 'you get what you screen for'[27].

Cell-in-droplets

A more recently developed screening strategy involves single cells in water-in-oil emulsions or cells in water-in-oil-in-water double emulsions. In a hydrophobic environment, water droplet compartments contain cells with enzyme variants, as well as substrates and products. This greatly enhances the screening capacity compared with microtiter plate screening (factor 100 000), since the reactor volume is much smaller, typically in pico- and femto-litre scale[37,64]. This strategy requires a flow-cytometry-based screening that allows for high-throughput analysis of the library. Further developments couple the droplet compartmentalization strategy with microfluidics. This nanotechnology approach enables

exact control over the micro-reactor droplets' lifetime and allows for the addition of substrates and quenchers at a desired time point. In addition, small improvements in enzyme activity may become detectable[37,64].

A proof of principle study using this approach concerns the detection of cellulase activity in *Saccharomyces cerevisiae* cells by Ostafe *et al.*[65]. Positive cells (5%), expressing the endo-cellulase Cel5A from *Trichoderma reesei*, were mixed with negative cells (95%). After emulsification of the mixed population with assay components in water-in-oil-in-water double emulsions, cellulase activity was detected by coupling the release of reducing sugars to the formation of a fluorescent product in a coupled enzyme assay involving a hexose oxidase and a vanadium bromoperoxidase. One round of sorting with FACS enriched the positive cell population 12-fold, as was demonstrated by a subsequent screening on carboxymethyl cellulose containing agar plates and a Congo Red assay.

Another good example of the cell-in-droplet approach is the work of Kintses *et al.*[64] on the improvement of a promiscuous activity of arylsulfatase from *Pseudomonas aerigunosa* (PAS) towards the non-native substrate phosphonate. Single *E. coli* cells expressing a PAS variant library are compartmentalized with substrate and cell lysis agents in water-in-oil droplets. These are incubated to generate a fluorescent product and hits are sorted after detection by laser-induced fluorescence. Plasmid DNA is isolated and transformed to *E. coli*. After seven rounds a six-fold improvement in both activity and expression was achieved.

A drawback of the cell-in-droplets strategy is that the substrates should readily enter the cell and that the (fluorescent) hydrophilic products should remain inside the droplets. Although the quality of the screens is comparable to microtiter plate screening, measurements at single cell level are influenced by cell-to-cell variation in expression levels. The signal variance is therefore slightly increased compared with microtiter plate screening[64]. In addition, this technique does not allow the identification of variants with improved properties, generated by for example directed evolution[59], unless coupled to microfluidics[64].

Cells as micro-reactors

Instead of using microtiter plate wells or cells in droplets as small reactors, one can also use a single microbial cell as micro-reactor, reducing the reactor volume even to the femto-litre scale. Also with this strategy libraries of ~ 10^9 can be screened, either enzyme variant or metagenomic libraries. Similar to the cell-in-droplets approach, the method is limited by the

fact that the substrate should be able to enter the cell, but in contrast to cell-in-droplets the product should now also remain within the cell[37]. Besides, the product itself should be detectable, e.g. by fluorescence. Coupling to for example FACS makes this method high-throughput and since cells can be used straightaway without compartmentalization it is less laborious.

A good example of this strategy, showing the entrapment and the additional required washing away of substrate, is the work of Yang *et al.*[66]. They developed a screen to improve a β-1,3-galactosyltransferase (CgtB) which transfers galactose from UDP-galactose to oligosaccharides, thereby demonstrating that cellular entrapment is not restricted to charged products. Two acceptor substrates, bearing the same sugar but chemically distinct fluorophores, are imported into the *E. coli* cell by a sugar transporter (permease). After an incubation period during which the acceptor may have been modified by a CgtB variant, unreacted acceptor is washed out of the cells with LB medium and phosphate-buffered saline. Cells containing catalytically-active glycosyltransferases retain the fluorescent product inside the cell as it is no longer a substrate for the permease. The use of two distinct fluorophores minimizes the chance of selecting for improved fluorophore binding (you get what you screen for[27]). Tailored alterations in substrate specificity may be possible by using also two different sugar moieties, enabling positive and negative screening. A library of $2 \cdot 10^7$ was created by random mutagenesis and catalytically active enzymes were identified and isolated by three rounds of FACS. Subsequently, a second library ($5 \cdot 10^6$) was created by gene shuffling, combining the parent gene and the best hits from the first step. After screening an improved enzyme was obtained with a higher substrate tolerance and a 300-fold increased catalytic activity compared with the parent enzyme[66].

Cell surface display

A completely different screening strategy is cell surface display, in which the enzyme variant is displayed on the outside of the cell by fusing it to an anchor motif, making it freely accessible for the substrate. Also here, relatively large libraries of ~ 10^9 can be used[37]. However, the required fusion of the gene of interest to the anchoring motif limits the applicability to enzyme variant libraries. Both the enzyme variants and the products are displayed on the surface of the cells[37]. The choice for the anchoring motif to display the enzyme of interest as well as the choice for the host organism should be carefully made. The enzyme anchor needs to have an efficient signal sequence, a relatively stable structure, it should be compatible with the enzyme of interest, and it should be resistant to protease

attacks. Also the host organism should be compatible with the enzyme of interest. In addition, the host should be easy to cultivate and possess as few as possible cell-wall-associated and extracellular proteases. Gram-positive bacteria, e.g. *Bacillus* or *Staphyloccocus*, are preferred over Gram-negative bacteria because they lack an outer membrane and they have a more rigid cell wall. However, *E. coli* is often employed because of its high transformation efficiency[45]. Enantioselectivity can be obtained by using (*S*)- and (*R*)-enantiomeric substrates, either simultaneously via coupling to different coloured fluorescent labels or separately[37]. The disadvantage of cell surface display in combination with FACS is that display may lead to loss of the enzyme's activity and that fluorescent substrates/products are required that remain bound to the cell surface[37]. Another drawback is the potential unspecific labelling of negative cells in the vicinity of a positive cell. A recent investigation by Prodanovic *et al.*[67] combines this selection strategy with *in vitro* compartmentalization in order to solve this problem. *S. cerevisiae* cells, expressing a library of glucose oxidase variants, were encapsulated in water-in-oil emulsions. Glucose oxidase entrapped within the cell wall matrix converts glucose to gluconolactone, releasing H_2O_2. The H_2O_2 is used by a displayed horseradish peroxidase to activate extracellular tyramide fluorescein, which can subsequently form a covalent link with tyrosine residues on the cell surface. After removal of the oil phase, FACS is applied and positive cells are further tested by microtiter plate screening[67]. This is a nice example of exploiting the best of several strategies.

Reporter-based *in vivo* screening and selection strategies

In reporter-based strategies, it is not the product of an enzymatic reaction or the enzymatic conversion itself that results in a measurable property, but rather a genetically encoded reporter that gives a discriminating phenotype (Fig. 1A). As the enzymatic activity is thus monitored indirectly, these strategies can in theory be applied for each enzymatic reaction and are therefore regarded as reaction independent. Both 'selection' and 'screening' is possible, depending on the type of reporter gene chosen by the researcher[32]. Some reporter types are colorimetric (e.g. LacZ), fluorescent (e.g. green fluorescent protein, GFP), bioluminescent (e.g. LuxCDABE) or they result in conditional survival (e.g. CAT), cell motility (e.g. CheZ), acidification (e.g. AraBAD), ice nucleation (e.g. InaZ) and cell display (e.g. LamB).
The activity of the reporter is dependent on the activity of the enzyme of interest via interference at the transcription, translation, post-translational modification, degradation or

solubility level. Based on general signal transduction systems one could think of the following strategies (examples will be described in detail below): (a) binding of the product to a transcriptional regulator and thereby turning on transcription of a reporter; (b) direct modification of a transcriptional regulator by the enzyme and thereby turning on transcription of a reporter; (c) binding of the product to a riboswitch or ribozyme, resulting in translation of a reporter; and (d) direct modification of the reporter by the enzyme, altering the reporter's activity. Alternatively, transcription repression, reporter inactivation and translation inhibition upon enzyme activity is also possible; however, a stimulated reporter is preferred because an appearing signal is more readily detectable than a disappearing signal.

Such reporter-based strategies have been exploited for a range of applications, such as (a) making signal-responsive genetic parts with transcriptional regulators and riboswitches in synthetic biology[68], (b) small molecule detection in metabolic engineering[32], (c) pathway optimization in metabolic engineering[69], (d) drug discovery[70], (e) tracing explosives in soil[71] or (f) pollutant detection either alone[72] or coupled to activation of bioremediation pathways[73]. Until recently, only a limited number of studies that aim for identification of certain enzyme variants have used a reporter-based approach. Some proof of principle studies have been performed, e.g. changing the specificity of the transcriptional regulator HbpR from 2-hydroxybiphenyl to 2-chlorobiphenyl[74]. The potential of different reporter-based strategies employed in other applications can be very useful for the rapidly developing field of reporter-based strategies in enzyme discovery and optimization. Here an overview will be given of the strategies that are already employed in this field.

Although general and reporter-based strategies are treated separately here, the two are in fact integrated, depending on the reporter that is used. For example, a method in which a transcriptional regulator turns on *gfp* followed by FACS screening is described as 'cells as micro-reactor', but if the same regulator turns on the gene encoding an antibiotic resistance marker the method is referred to as 'selection'.

Transcriptional regulator-based strategy

The most employed *in vivo* screening or selection strategy involving a reporter is the transcriptional regulator-based approach. The activity of the enzyme is transduced through product-dependent activation or de-repression of the transcription of a reporter gene by the transcriptional regulator. For detailed information about the response profiles, describing the relation between small molecule and reporter, the reader is referred to Dietrich *et al.*[32].

The transcriptional regulators used can be divided into two subgroups: (a) a natural transcriptional regulator (its ligand specificity can potentially be adapted by laboratory evolution; Fig. 1B) or (b) a synthetic transcriptional regulator, composed of multiple components that are either designed or derived from natural systems, the association or dissociation of which is triggered by the enzyme activity (Fig. 1C).

The natural regulators generally depend on an allosteric event: binding of the enzyme's substrate or product in one domain alters the conformation and switches the DNA binding capacity of the regulator. Although this strategy is generally product-based, there are examples of substrate-based systems. In substrate-induced gene expression screening (SIGEX), for example, metagenomic fragments are cloned in front of *gfp* and the resulting library is screened with FACS, after addition of a substrate[75,76]. The rationale behind SIGEX is the possibility that catabolic operons are substrate or intermediate induced and that regulatory elements are often situated in close proximity to catabolic genes. However, as this rationale is not a rule and regulatory systems may even evolve separately from the metabolic genes they control[77], SIGEX does not always prove useful[78]. Besides, SIGEX is only suitable for metagenomic screening, not for enzyme optimization.

The strength, but also the limitation, of this natural transcriptional regulator-based strategy is the specificity towards the product (or substrate) of the enzymatic reaction. This specificity makes it possible to obtain only the enzyme which produces the product of interest without false positives, caused for example by binding of the substrate instead of the product to the regulator. When choosing a proper transcriptional regulator for a specific screen or selection, it is recommended to look first at already described natural regulators. For example, Uchiyama and Miyazaki[79] screened metagenomic libraries in *E. coli* for benzamidases with the natural benzoate-responsive transcription activator BenR controlling expression of the reporter *gfp*. With this product-induced gene expression (PIGEX), they obtained 11 hits, of which three were genes with low sequence similarity towards known amidases. For all 11 hits benzamidase activity was confirmed. In contrast to most transcriptional regulator-based studies, this group used two sets of cells, one with the metagenomic library and one with the sensor (*benR* and *gfp*), because they believed that, with single-cell intracellular screens, cross-talk between negative and positive clones might occur. The two sets of cells were grown separately and later combined in wells on ten 96-well plates, each well containing 100 fosmid library clones. So, several sequential screening steps were necessary. GFP fluorescence was only observed when the sensor cells were in log phase[79].

However, not for every enzyme substrate/product of interest is a fitting regulator known, and the ones which are known may be promiscuous. Probably this problem exists for most regulators and further engineering of the regulator is often required to obtain the desired functionality of specific binding of the product of interest and meanwhile preserve its DNA binding properties. A first step towards the required drastic adaptation of a regulator's ligand binding site might be accomplished by computational design. Unfortunately, efforts in this direction are not yet as sophisticated as one would like. For transcriptional regulators, no examples are known yet. Although ligand specificity has been changed for periplasmic binding proteins[80,81], problems in verifying these designs experimentally show that many hurdles still need to be taken[82]. The difficulties lie amongst others in loss of protein stability upon mutation and limitations in the description of molecular interactions between protein, ligand and water, such as long-range electrostatics and dynamics[81-83].

Alternatively, one could also strike the golden mean by using a less drastic, stepwise approach such as directed evolution. Here the specificity of a known regulator is changed by mutagenesis methods like saturation mutagenesis or gene shuffling. A few examples will be given here.

van Sint Fiet *et al.*[84] employed a previously described mutant of the transcriptional activator NahR to detect *E. coli* cells with XylC activity. This enzyme forms benzoate or 2-hydroxybenzoate from the corresponding aldehydes. Binding of these products to a mutant NahR turned on the expression of *tetA*. The colony size was related to the product concentration, and the optimal ratio of true and false positives could be established by adapting the tetracycline concentration. This selection system was also turned into a screening system by replacing *tetA* with *lacZ*. Changing the specificity of NahR from benzoate to salicylate by PCR-based saturation made this detection possible[85].

Mohn *et al.*[86] evolved the toluene-responsive transcription activator XylR by gene shuffling plus mutagenic PCR to be optimally responsive to 1,2,4-trichlorobenzene, the major product of γ-hexachlorocyclohexane dehydrochlorination. With *lacZ* as reporter and *E. coli* as host, the activity of the dehydrochlorinase LinA was demonstrated. Selection was possible by introducing the lactose transporter LacY: only cells which were able to produce LacZ could grow on lactose.

With saturation mutagenesis, Tang and Cirino[87] changed the arabinose-responsive AraC to respond to mevalonate. Reporting mevalonate synthesis is a handy tool in metabolic engineering, e.g. to improve the mevalonate-dependent isoprenoid pathway enzymes. Also

here *E. coli* was the host and *lacZ* or *gfp* the reporter.

Although cases have been described in which the ligand specificity of transcriptional regulators are successfully engineered, obtaining the proper regulator for the enzyme of interest is certainly not simple. Another approach to acquire such a regulator is to design one from multiple components. Strategies based on these synthetic transcriptional regulators rely on the fact that the binding of the multiple components of the regulator to one another is dependent on the enzyme activity. A few examples are described here.

In the QUEST system (QUerying for EnzymeS using the Three-hybrid System), catalysis is detected by coupling substrate turnover to a transcriptional event. The DNA binding domain of the transcriptional regulator AraC is fused to a domain that can bind either a substrate or a chemical inducer of dimerization (CID). If the substrate (the scytalone analogue 2,3-dihydro-2,5-dihydroxy-4H-benzopyran-4-one) is present, it competes with the CID (although chemically different from substrate/product) for binding to the domain, resulting in monomerization of the activating regulator, dissociation from the DNA and downregulation of the *araBAD* operon. Conversion of the substrate to the product by the enzyme shifts the equilibrium towards CID binding and activated expression of *araBAD*, enabling the bacteria to grow on arabinose and thus acidifying the medium. Fungal scytalone dehydratase was detected using pH indicators on plates. A second screen was done to eliminate false positives. In theory this system could be tailor-made for other enzymes as well, but general applicability is limited as protein-CID pairs for the substrate of interest might not be available[88].

Baker *et al.*[89] developed a yeast–three hybrid system to detect enzyme catalysis. The transcriptional regulator is composed of two fusion proteins: a LexA DNA binding domain fused to a dihydrofolate reductase (LexA-DHFR), and a B42 activation domain fused to a glucocorticoid receptor (B42-GR). These two fusions are linked via the substrate, which consists of three parts, namely Mtx bound at one side to LexA-DHFR, Dex bound at the other side to B42-GR, and the substrate of interest positioned between Mtx and Dex. Breaking or formation of the bond by the enzyme monomerizes or dimerizes the transcriptional activator, resulting in repression or de-repression of transcription respectively. So, this system is limited to bond breaking or bond forming reactions. It has been applied for selecting glycosynthase activity from a Glu197 saturation library of the endoglucanase Cel7B, using a gene involved in leucine biogenesis (*LEU2*) as reporter to complement a leucine auxotroph. This resulted in a five-fold increase in glycosynthase activity[90]. Furthermore, a variant with six-fold increased catalytic efficiency (k_{cat}/K_M) was selected from a cellulase

library created by DNA family shuffling of genes encoding Cel7B variants. *URA3* was the reporter, converting the substrate 5-fluoroorotic acid to the toxic product 5-fluorouracil (5-FU). So, upon cleavage of the Mtx-Cel-Dex substrate, dimerization of the transcriptional activator was disrupted and toxic 5-FU was no longer produced[91].

Also Verhoeven *et al.*[92] developed a system in which the enzyme disrupts dimerization of a transcriptional regulator. In this case two DNA binding domains of the repressor cI of bacteriophage 434 were linked with a flexible region. An active protease could cleave the linker of this single chain repressor, releasing repression of the reporter gene by losing DNA binding. Three reporters were used, *HIS3*, *kan* and *lacZ*, of which the first two enabled growth of *E. coli* and the last was for quantitation. With this system a variant of the tobacco etch virus protease with changed substrate specificity was selected from an error-prone PCR library.

Riboswitch/ribozyme-based strategy

Riboswitch- or ribozyme-based approaches are not yet widely applied for finding novel and improved biocatalysts. Therefore additional information from other fields is included to get a more complete story. Several forms of regulation of the reporter by the aptamer are possible, on either transcriptional or translational level. They can be divided into riboswitches and ribozymes. In riboswitches an aptamer is often located in the 5'-UTR of the transcribed gene (often encoding an enzyme in natural cases or a reporter in synthetic constructs), in such a way that binding of a specific ligand to the aptamer triggers a change in its tertiary structure, switching (on or off) the translation of the reporter (Fig. 1D). Natural examples are ligand dependent accessibility of the ribosomal binding site, ligand-dependent change from intrinsic transcription terminator to anti-terminator, or ligand-dependent accessibility of mRNA[55]. In the last case a conformational change in trans-acting RNA results in its inability to block the mRNA via its antisense aptamer sequence. In contrast, ribozymes have a catalytic activity, such as hydrolysis of the nucleic acid phosphodiester bonds in the backbone, usually of the aptamer-gene fusion transcript. When the self-cleaving hammerhead ribozyme, for example, is coupled to an aptamer and together inserted in the mRNA, ligand binding to the aptamer results in a structure switch, preventing self-cleavage of the ribozyme. Thus, the mRNA stays intact, allowing translation to occur[55].

Also for choosing a proper aptamer for a certain screen/selection it is convenient to look at already described aptamers of natural origin. Although new aptamers/riboswitches are frequently identified, the repertoire is still quite limited. The most common method to

obtain new aptamers is by systematic evolution of ligands by exponential enrichment (SELEX). In this *in vitro* approach a random RNA library goes through several rounds of selection, with or without the ligand[55,93,94]. However, aptamer binding *in vitro* does not ensure activity *in vivo*. To overcome this problem a combination of initial *in vitro* selection, to reduce the library size, and subsequent *in vivo* screening or selection, to find the proper *in vivo* activity, is possible[55].

Although not yet applied in finding or improving new biocatalysts, Desai and Gallivan[95] strongly suggest that *in vivo* screening or selection based on riboswitches is a good strategy for this purpose. They describe a theophylline-responsive riboswitch in the 5'- UTR of either *lacZ* or *cat*. The expression level is dependent on the distance between the ribosomal binding site and the aptamer. They demonstrate that synthetic riboswitches can be used to perform either genetic screening or genetic selection experiments to detect the presence of a specific, non-endogenous small molecule in *E. coli*, which in theory could be the product of an enzymatic reaction. In addition, they used the system to select for riboswitches with different ligand specificity. Moreover, it is demonstrated that a cell harbouring a synthetic riboswitch with particular ligand specificity can be selectively amplified from a million-fold larger pool of cells containing mutant riboswitches that respond to a closely related ligand. This indicates the promise of successfully using this approach for selecting riboswitches with desired ligand specificities.

A recent study demonstrates the actual applicability of this strategy for biocatalyst improvement. Michener and Smolke[54] describe a theophylline-responsive ribozyme incorporated in the 3'-UTR of *gfp*, which couples product concentration to reporter expression in *S. cerevisiae*. The ribozyme consists of an input or product binding domain and an output or self-cleaving domain. The connection between the two domains is such that they cannot be folded properly simultaneously. When the output domain folds correctly it cleaves itself, resulting in removal of the poly-A tail and rapid degradation of the mRNA. Ligand binding favours the conformation with a properly folded input domain and an incorrect folded output domain. So, less cleavage and more gene expression occur. Coupled to flow cytometry (measuring the average fluorescence of the culture) or FACS (sorting the cells based on the fluorescence per cell), libraries of ~ 10^3 or ~ 10^7 respectively can be screened. Even small changes in fluorescence or enzyme activity are measurable with flow cytometry, whereas FACS measurements are less precise. Iterative application of this FACS screen to libraries of caffeine demethylase in yeast led to a series of beneficial mutations that ultimately increased enzyme activity *in vivo* 33-fold and product selectivity 22-fold.

A big advantage of this strategy is that multiple signals originating from different enzymes can be implemented by using for instance different GFP variants. This can be valuable when not one enzyme but a whole metabolic pathway needs improvement[54].

Post-translational modified reporter-based strategy

As screening or selection strategy, not only interference by the enzymatic reaction on the level of reporter transcription or translation is used, but also posttranslational modification and degradation approaches have been developed (Fig. 1E). In contrast to transcriptional regulator- and riboswitch/ribozyme-based strategies, this strategy is less generally applicable. It is not the binding of the product but the actual activity of the enzyme which lies at the basis of the screening or selection. This limits each strategy to certain groups of enzymes. One enzyme group, the proteases, is very well represented. Kostallas *et al.*[96] designed a GFP with a protease substrate peptide and an ssrA-tag, co-expressed with the protease of interest. Proteases which can process the substrate peptide and remove the ssrA-tag rescue GFP from degradation by the protease ClpXP, increasing the fluorescence of the *E. coli* cell. In this case the system was used to find new substrates for the tobacco etch virus protease with FACS, but it could also be employed for directed evolution of proteases by changing the substrate peptide for the activity of interest[96].

Another approach was taken by O'Loughlin *et al.*[97]. Negative and positive selections are combined to alter the substrate specificity of HIV protease. In the first step, protease variants created by error-prone PCR are induced during the mid-logarithmic phase of growing *E. coli* cells. Based on the idea that HIV protease with broad specificity is cytotoxic by cleaving essential *E. coli* proteins, selection already reduced the pool from 60 000 to 15 000 cells. In the second step, HIV variants that could cleave β-galactosidase or an altered β-galactosidase with the protease substrate peptide embedded decreased the blue colouring of *E. coli* cells grown in the presence of the β-galactosidase artificial substrate X-Gal. Combination of these two steps more rapidly alters the protease specificity.

Another class of enzymes, 4'-phosphopantetheinyl transferase (PPTase), is of interest for finding new secondary metabolite biosynthetic clusters via association, because they are needed for the activation of nonribosomal peptide synthetases (NRPS) and polyketide synthase (PKS) enzymes, both members of such a cluster. PPTases catalyse the post-translational attachment of a 4'-phosphopantetheine group to a conserved serine residue in the carrier protein domains of NRPS and PKS enzymes, a modification that is essential for activity. Owen *et al.*[98] employed the NRPS enzyme BpsA as reporter, as it generates a

coloured product, indigoidine, through cyclization of two L-glutamines upon activation by a PPTase. By selecting the indigoidine-positive clones, the number of clones is reduced such that lower throughput secondary screening methods, like *in vitro* activity measurements, become feasible. Although *E. coli* is used here, also applying this approach for screening other hosts such as *Pseudomonas* is possible. In addition, the carrier protein domain of BpsA can be modified in order to screen for a wider range of PPTases.

Fusion-based strategy

The fusion-based approach is not a screening or selection strategy by itself, but it can certainly improve the ratio of active versus inactive clones in a pool, making fewer high-throughput methods possible in further steps. Fusing GFP to the enzyme variant and measuring the fluorescence with FACS or a fluorometer is a way to remove the insoluble and therefore inactive clones from the pool, because GFP only gives a fluorescent signal when soluble. Gupta and Tawfik[99] improved the activity of serum paraoxonase PON1 towards several substrates via small and effective neutral drift libraries. In these libraries the protein function is maintained but mutations are accumulated leading to highly polymorphic, stable and evolvable variants, which can be used as the starting point for directed evolution. To reduce the pool size they fused PON1 to a certain variant of GFP, sfGFP-F12, to have a balance in stability of GFP and the fused enzyme. The fraction of positive clones was enhanced from ~ 15% to > 85% by sorting, indicating the loss of a significant number of inactive variants. The fraction of false positives, variants showing high fluorescence but no activity, was ~ 10%. The activity was measured with a chromogenic substrate after growth and lysis of the *E. coli* cells in 96-well plates.

Heterologous expression frequently results in low yields of functional protein due to incorrect folding of the polypeptide chain; a majority of the proteins may end up in insoluble aggregates, inclusion bodies. When that is the case, fusion-based strategies can also be used to improve the folding and increase the solubility of an enzyme. Japrung *et al.*[100] first selected active variants of pfDHFR–GFP that could rescue growth of DHFR-deficient mutants of *E. coli*. These variants were subsequently sorted with FACS based on their fluorescence. Only variants which exhibited the highest 10% fluorescence, and thus the highest solubility, were selected.

A drawback of this approach is that false positives may be obtained if the fusion proteins end up in inclusion bodies. Both enzyme and GFP can still be (partially) functional in inclusion bodies[101]. However, it is not guaranteed that the activity of the two goes hand in

hand. GFP may still be fully functional without the enzyme of interest being soluble/active. In addition, it is important to test whether the fusion itself does not interfere with the activity of the enzyme. Testing both N-terminal and C-terminal fusions might help. Also the improved activity of the enzyme variants found with the screen should be verified without the fusion.

Comparison of the *in vivo* screening and selection strategies

First the general *in vivo* screening and selection strategies will be compared with respect to library type, library size, enantioselectivity and experimental requirements, followed by a comparison of reporter-based strategies. An overview of the characteristics of both strategy types is summarized in Tables 1 and 2 respectively.

A comparison of the general *in vivo* screening and selection strategies shows that most of them are suitable for both metagenomic (m) and enzyme variant (ev) libraries. Due to the required fusion between enzyme and anchoring motif, cell surface display is the only strategy that is not compatible with screening a metagenomic library.

All approaches, except for agar and microtiter plate screening, are high-throughput (frequently due to the use of FACS), although the library size is still limited by the transformation efficiency (~ 10^9). Concerning the expression host for *in vivo* screening, *E. coli* is used in the majority (~ 90%) of all directed evolution studies. However, other bacterial hosts such as *Bacillus subtilis* and *Thermus* species are also used, as well as eukaryotic hosts including *S. cerevisiae* (second host in directed evolution studies, 9%) and even insect and mammalian cells[102]. No clear relationship exists between a specific host and a certain strategy. Nevertheless, bacterial hosts are often preferred over eukaryotic hosts because they have higher transformation efficiencies, have a faster growth rate and are easier to manipulate[103,104]. On the other hand, as discussed earlier, for cell surface display Gram-positive species like *Bacillus* are favoured because of their single cell membrane and rigid cell wall[45]. Another important aspect in choosing a proper host is the origin of the genes in the library. Since expression mechanisms (e.g. codon bias) vary between organisms, a host which is closely related to the organism from which the genes originate is preferred. However, with metagenomic libraries, covering genes from multiple organisms, this criterion is no longer feasible. Luckily, more hosts and improved hosts are still being developed[43,102].

All general strategies can be used to screen or select for enantioselective enzymes as long as it concerns reactions in which the enzyme is enantioselective with respect to the

Table 1. Characteristics of general *in vivo* screening and selection strategies[a]

Strategy[b]	Library type[c]	Library size	Requirements[d]	Advantages	Disadvantages
Selection	m/ev	~10^9	None	Only desired variants Simple to operate	Low activity variants might be missed Difficult to quantify Higher false positive risk
Agar plate screening	m/ev	~10^5	Reagents[e] Detection device[e]	Simple to operate	Low dynamic range Difficult to quantify
Microtiter plate screening	m/ev	~10^4	Detection device	Multiple analytical detection methods possible Increased sensitivity by cell lysis Large dynamic range	
Cell-in-droplets (screening)	m/ev	~10^9	Detection device Cell sorter	Pico/femto-litre scale (not much reagents needed)	Product should stay in the droplet Technically challenging
Cells as micro-reactors (screening)	m/ev	~10^9	Detection device Cell sorter	Femto-litre scale (not much reagents needed) No compartmentalization needed	Product should stay in the cell Technically challenging
Cell surface display (screening)	ev	~10^9	Detection device Cell sorter	Substrate can stay extracellular	Product needs to attach to cell surface Technically challenging Careful selection of host and anchoring motif needed Fusion can alter enzyme structure/activity

[a]Adapted from[37]. [b]It is the product of the enzymatic reaction or the enzymatic conversion itself that need to be detected, making these strategies reaction dependent (in contrast to most reporter-based strategies). [c]m, metagenomic library; ev, enzyme variant library. [d]Requirements encompass everything except basic microbial techniques/reagents and substrate. [e]In some cases.

substrate. In that case two substrates with different chirality can be added. In some elegant cases this has been done in a single experiment where one of the two enantiomer products is toxic[59]. Another option is the use of blocking of those enzyme variants that have the unwanted enantioselectivity with suicide inhibitors of the corresponding chirality. A more laborious option is the performance of two subsequent screens, each with one substrate. Unfortunately, these approaches are useless when the enzyme converts a non-chiral substrate to either the L- or the D-product. Here only a coupled enzyme assay, such as coupling to the D-amino acid oxidase as described above[61], is able to determine the enantioselectivity of the enzyme.

The general strategies which are the simplest to operate and have minimum requirements are selection and agar plate screening. Selection is of course a special strategy, as in theory only desired variants are kept in the pool. However, in some cases a range of conversion rates is demanded, meaning that one tries to find multiple enzymes performing the same conversion but having different conversion rates. Selection may not be the right choice in this case, because enzymes with a lower activity might not be kept in the pool and quantification is not always possible. Reporters with different efficiency might offer an alternative.

A major hurdle of the general *in vivo* strategies may be that the substrate should enter the cell, and in some cases (such as cells as micro-reactors) the product should also stay inside the cell. Therefore, not every enzymatic reaction can be screened or selected. Cell surface display circumvents this limitation partially by presenting the enzyme on the outside of the cell. Of course, the product now has to be contained at the outside of the cell as well, for otherwise the link between genotype and phenotype is lost. Thus the type of *in vivo* strategies that can be used are often reaction or at least product dependent, meaning that the enzymatic activity itself or the product needs to be detectable, e.g. as a fluorescent product or as a pH change. The use of artificial substrates which are converted into a measurable product can reduce this problem. However, another solution is the use of certain reporter-based strategies, as discussed above.

The type of reporter determines the general strategy a reporter-based method belongs to and thus its characteristics such as library size and requirements (Table 1). The unique characteristics of the reporter-based strategies are described in a separate table (Table 2). In cases where expression of the reporter is dependent on the enzyme activity, all available reporter types can be employed. However, for fusion-based and posttranslational modification-based strategies, the reporters need to report on solubility and be modifiable

Table 2. Characteristics of reporter-based *in vivo* screening and selection strategies

Strategy based on[a]	Library type[b]	Reporter types	Advantages	Disadvantages
Natural TR[c]	m/ev	All	High specificity Adaptation TR possible (directed evolution) Reaction independent No artificial substrates required Extra possibilities for enantioselectivity Detection low activities Signal enhancement possible via transcription/translation	Proper TR needs to be found/designed: ·number natural TRs available limited ·computational design TR to be developed
Synthetic TR	m/ev	All	Reaction independent[e] No artificial substrates required[e] Relative simple design TR due to multiple components Signal enhancement possible via transcription/translation Detection low activities	Reaction dependent[e]
Riboswitch/ ribozyme	m/ev	All	High specificity Initial *in vitro* selection of aptamer possible Reaction independent No artificial substrates required Extra possibilities for enantioselectivity Integration multiple signals possible Detection low activities Signal enhancement possible via transcription/translation	Proper aptamer needs to be found/designed: ·number natural aptamers available limited No guaranteed translation *in vitro* to *in vivo* activity
PTM[d]	m/ev	Modifiable by enzyme of interest	No artificial substrates required	Reaction dependent
Fusion	ev	Reporting solubility	Increase ratio active/inactive variants in pool possible Reaction independent No artificial substrates required	Fusion can alter enzyme structure/activity False positives when in inclusion bodies

[a]The actual screening or selection method is dependent on the reporter type and will fall within the general strategies described in Table 1. [b]m, metagenomic library; ev, enzyme variant library. [c]TR, transcriptional regulator. [d]PTM, post-translational modification. [e]In some cases.

by the enzyme of interest respectively. Of course, the choice for a specific reporter is also by the enzyme of interest respectively. Of course, the choice for a specific reporter is also determined by the available equipment, the necessary reagents and the complexity of the protocol that needs to be followed in the measurements, and the desired characteristics of the reporter–measurement combination with respect to, for example, sensitivity and dynamic range. Comparing the different reporter-based strategies shows that, except for the fusion-based approach, all are suitable for selection and screening of both metagenomic and enzyme variant libraries.

Involvement of reporters in screening or selection strategies has some advantages. (a) Extra selection or screening criteria can be implemented, e.g. solubility in the case of a fusion-based approach. (b) Artificial substrates are often not necessary because the product is not measured directly but via a reporter. This is not always the case for synthetic transcriptional regulator-based strategies, as in some approaches the substrate needs to be attached to parts of the regulator. (c) One has the ability to screen or select for enantioselective enzyme variants which convert a non-chiral substrate to either the L- or the D-product, provided the binding of the transcriptional regulator or the riboswitch/ribozyme to the product is enantioselective. (d) They are more widely applicable, because they are reaction independent and the screening or selection strategy is dependent on the reporter choice, as explained earlier. (e) The signal may be enhanced by coupling enzyme activity to reporter expression. Altering features such as promoter strength, ribosomal binding site and codon bias may allow for detection of weak signals. This can be very valuable for further laboratory evolution, as improving an existing activity is much more feasible than changing towards a new activity.

Despite these interesting features of reporter-based strategies, there are two major downsides. (a) Developing such systems requires time and effort. In particular, the development of transcriptional regulators and riboswitches/ribozymes to specifically bind to the product of interest is challenging and time consuming. The number of known natural regulators[43] and aptamers is limited. However, the possibility of doing an initial riboswitch/ribozyme selection *in vitro* may greatly simplify the development of riboswitches/ribozymes compared with transcriptional regulators. For modifying the transcriptional regulators the most successful method at the moment is directed evolution. (b) Coupling to selection on growth may result in more false positives compared with screening. The higher selective pressure may select for cells that circumvent the coupling of the enzyme activity to growth and are able to survive by a different mechanism. For

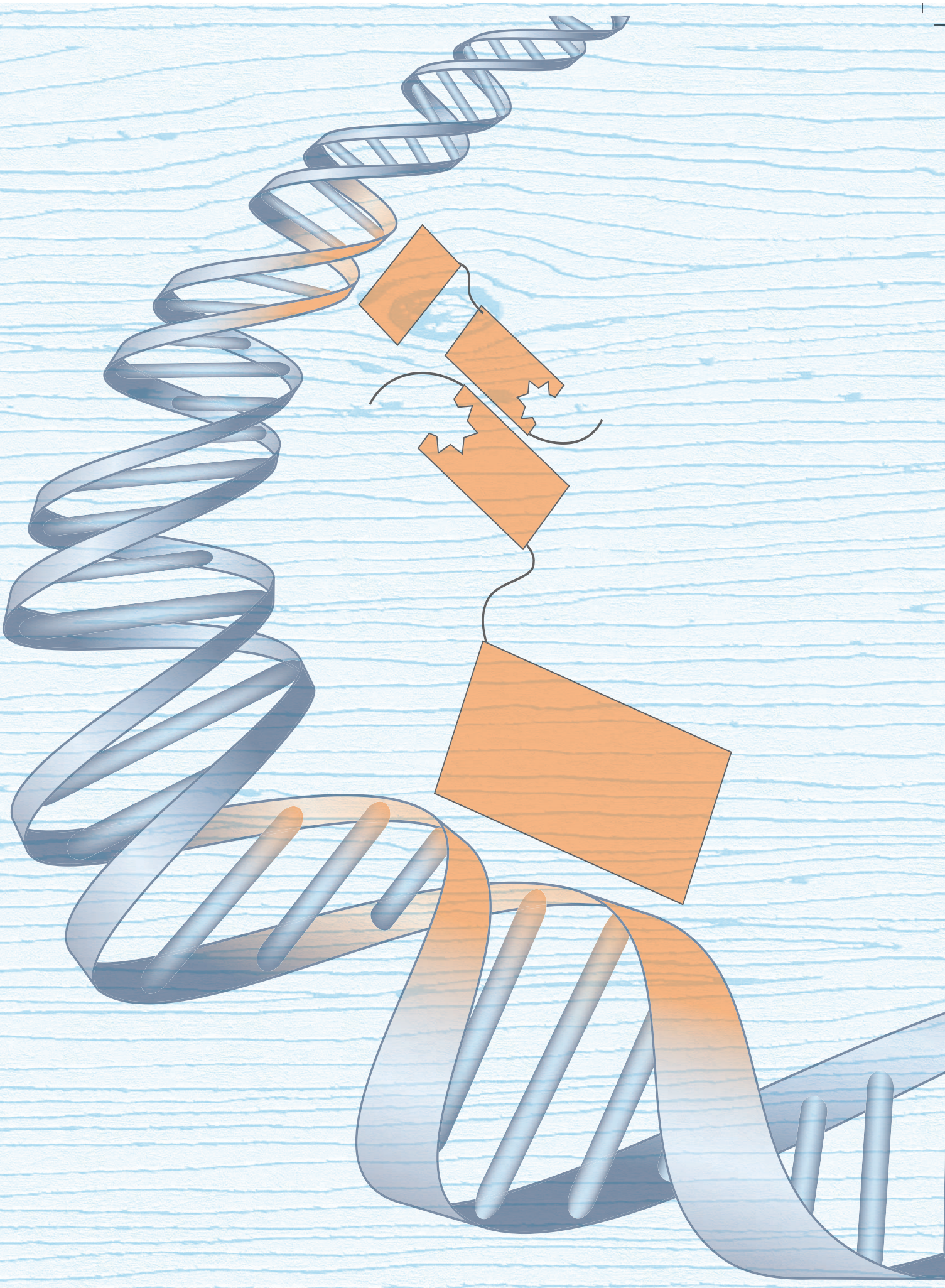
example, false positives may result from spontaneous escape mutants, such as a variant transcriptional regulator that turns on expression independent of product binding.

Conclusions and future prospects

None of the *in vivo* selection and screening strategies described here is perfect, and none is able to target each possible enzyme. Introducing reporters certainly increases the general applicability of a strategy, but the fact that each strategy only targets a subset of enzymes still holds. For each study one has to choose the proper approach on the basis of preferred features. It is expected that this field of reporter-based strategies will increase in the coming years. Existing strategies will be further improved and new strategies will be developed. Detection of the product by binding to a cytoplasmic sugar binding protein with incorporated GFP and subsequent alteration of its fluorescence could be one of them[105]. Besides, research on reporters themselves is ongoing[106]. Since transcriptional regulator- and riboswitch/ribozyme-based strategies are reaction independent, these will gain the most interest among reporter-based strategies.

In vivo methods have advantages such as enhanced success rates of functionally produced (correctly folded, cofactor incorporated, catalytically active) protein(s) of interest and the possibility of varying the properties of the desired biocatalysts by changing the screening or selection conditions. However, one should keep in mind that these methods are limited by problems associated with heterologous expression, transformation efficiency, transport of substrate and/or product over the cell boundaries, reduced sensitivity and host growth rate. It is therefore important not only to further develop and improve the selection and screening strategies themselves, but also to improve for example heterologous expression by changes in library construction (e.g. codon optimization in enzyme variant libraries) and increasing the number of expression hosts. In addition, progress at the level of computational design and *in silico* and *in vitro* selection and screening strategies will be very valuable. Both smaller but smarter libraries as well as making the handling of large libraries more straightforward or increasing the manageable library size are interesting approaches. Also combining different strategies can help in the search for novel or improved biocatalysts: e.g. cells as micro-reactors screened with FACS, followed by microtiter plate screening.

Although there are still many problems to be solved, spectacular developments are initiated. Certainly the combination of computational design and laboratory evolution will have an enormous influence and will in time lead to more universal strategies.

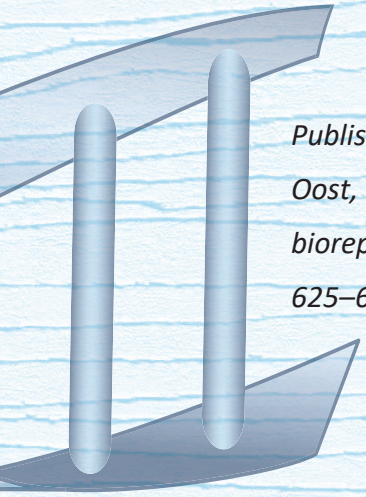


Chapter 3

A growth- and bioluminescence-based bioreporter for the *in vivo* detection of novel biocatalysts

Teunke van Rossum[a], Aleksandra Muras[a], Marco J.J. Baur[a], Sjoerd C.A. Creutzburg[a], John van der Oost[a], Servé W.M. Kengen[a]

[a]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands



Published as: van Rossum, T., Muras, A., Baur, M. J. J., Creutzburg, S. C. A., van der Oost, J., and Kengen, S. W. M. (2017) A growth- and bioluminescence-based bioreporter for the in vivo *detection of novel biocatalysts.* Microb Biotechnol. *10, 625–641.*

Abstract

The use of bioreporters in high-throughput screening for small-molecules is generally laborious and/or expensive. The technology can be simplified by coupling the generation of a desired compound to cell survival, causing only positive cells to stay in the pool of generated variants. Here, a dual selection/screening system was developed for the *in vivo* detection of novel biocatalysts. The sensor part of the system is based on the transcriptional regulator AraC, which controls expression of both a selection reporter (LeuB or KmR; enabling growth) for rapid reduction of the initially large library size and a screening reporter (LuxCDABE; causing bioluminescence) for further quantification of the positive variants. Out of four developed systems, the best system was the medium copy system with KmR as selection reporter. As a proof of principle, the system was tested for the selection of cells expressing an L-arabinose isomerase derived from mesophilic *Escherichia coli* or thermophilic *Geobacillus thermodenitrificans*. A more than a millionfold enrichment of cells with L-arabinose isomerase activity was demonstrated by selection and exclusion of false positives by screening. This dual selection/screening system is an important step towards an improved detection method for small molecules, and thereby for finding novel biocatalysts.

Introduction

Research aiming at the development of whole-cell bioreporters for a wide range of applications has increased substantially over the last few decades. Applications include detection of pollutants[107-109], the search for novel biocatalysts[110-112], and the improvement of strains for the industrial production of small molecules[113-115]. A whole-cell bioreporter (hereafter referred to as bioreporter) is a living microorganism containing a sensor molecule that upon binding of a small molecule of interest switches on a reporter, resulting in a detectable phenotype[116-118]. The high specificity of the sensor towards this small molecule together with the option to choose the reporter and thereby the way of measuring, makes this method attractive. The potential to use bioreporters for high-throughput screening, explains the increased interest in these systems[113,119-121]. For instance, various mutagenesis techniques lead to large numbers of altered production strains, but without a high-throughput screening method, only a limited number of variants can be analyzed[114,121,122]. In the search for novel biocatalysts, screening large metagenomic or biocatalyst mutant libraries, can be complicated and time-consuming without a high-throughput screening method, although in this field smart and focused libraries are emerging as well[39,46]. Also other advantages have led to an increase in the use of bioreporters. These include high specificity, high enantioselectivity, lower costs, reduced handling, measuring bioavailability instead of actual concentration, no requirement of artificial substrates and the possibilities of on-line monitoring and signal enhancement[113,123-125].

The sensor part of the bioreporter can either function on transcriptional, translational or post-translational level. Examples of sensors on the first two levels are transcriptional regulators and riboswitches/ribozymes respectively. On post-translational level various set-ups are possible, for example, a FRET (Förster Resonance Energy Transfer) sensor, or a sensor directly coupled to enzyme activity[69]. The specificity of the sensor towards the target molecule is essential in the functioning of the bioreporter. Obtaining the proper specificity can be time-consuming. One can exploit nature, but for many small molecules no sensor is known yet[116,123] and if there is one known, it cannot always be expressed heterologously[111]. Another option is to engineer the specificity of a sensor, which may, however, demand a lot of time[69,115,125]. Moreover, problems may arise, like the loss of protein stability[82], or difficulties translating *in vitro* to *in vivo* if the initial screening is done *in vitro*[69]. Despite these hurdles, but due to their interesting properties, bioreporters are a growing practice and a lot of bioreporter related research is going on[55,69,121,126,127].

The reporter part of the bioreporter gives the cell a distinguishable phenotype, such as fluorescence, bioluminescence, colour, conditional survival, acidification of the environment, or cell motility. Which type of reporter is used, mainly depends on the available equipment and the desired characteristics such as dynamic range and sensitivity. Reporters that are most often used are green fluorescent protein (GFP), bacterial luciferase (LuxAB or LuxCDABE) and β-galactosidase (LacZ). All three reporters are screening reporters, meaning that all cells, both negative and positive, stay in the pool[45]. Also with all three methods, the concentration of the molecule of interest can be quantified. However, high-throughput screening with these reporters is often still laborious or expensive because of the requirement of microtiter plate assays or of fluorescence-activated cell sorting (FACS) respectively. A simple, high-throughput alternative is the use of a selection reporter instead of a screening reporter, which, by providing cell survival, causes only positive variants to stay in the pool. Although, selection based on growth is rather straightforward and cheap, these are not yet broadly applied[84,110].

The aim of this study was to develop a selection-based reporter system for the detection of small molecules or more particularly for products of novel biocatalysts, and characterize its behaviour with respect to leakiness, maximal signal, dynamic range and sensitivity. More specifically, the developed system makes use of double reporters, consisting of both a selection reporter and a screening reporter, which allow for a rapid reduction of the initially large library size based on growth as well as subsequent quantification of the positive hits. Detection is based on the binding of the product of an enzyme reaction to a transcriptional regulator, resulting in a conformational change that alters its DNA binding capacity. This allows expression of the two divergently transcribed reporter genes. The selection reporter enables growth of the *Escherichia coli* cell, meaning that only cells in which the enzyme product is present, and thus express the active enzyme, will survive. The survivors can subsequently be screened using the screening reporter.

Here, different versions of the developed selection and screening system, varying in plasmid copy number and selection reporter, were compared in induction assays. The best performing system was the medium copy system with KmR as selection reporter. This system was used to detect the L-arabinose isomerases derived from mesophilic *Escherichia coli* and thermophilic *Geobacillus thermodenitrificans* with L-ribulose as substrate. Moreover, making use of the selection reporter, cells with one of the two L-arabinose isomerases were enriched over cells without L-arabinose isomerase. The screening reporter enabled the distinction of true from false positives.

Results and Discussion

Components of the system

To develop a sensitive double reporter system, with a broad dynamic range, high sensitivity and no leakage, four different versions were constructed and their performance was compared. To simplify the comparison, a plasmid-based system was chosen, but for future work chromosomal integration might be preferred, to enhance stability and to reduce the use of antibiotics. Each system consisted of a host strain (*E. coli* BW25113 derivatives) and a regulator-reporter plasmid, encoding the transcriptional regulator and both reporters (Fig. 1). The two reporters were divergently transcribed to prevent read-through transcription from one to the other. In the different system versions the selection reporter and the plasmid copy number were varied.

As transcriptional regulator, we selected AraC, because it has been extensively studied and a protein structure is available with and without ligand. Especially the last criterion is important in further studies in which we want to design variants in which the binding specificity of the regulator towards a small molecule of interest is adjusted. Also, this regulator has previously been engineered to alter its ligand specificity[87,88,128,129]. In short, AraC is a dimer of which each monomer binds to one of two distant operator half sites upstream the *araBAD* operon, repressing its expression. Upon binding of L-arabinose to AraC, DNA-binding domains are reoriented to bind two more closely located half sites, allowing the *araBAD* operon to be transcribed and L-arabinose to be metabolized. AraC also regulates its own gene, a gene of unknown function (*araJ*), genes involved in L-arabinose transport (*araFGH* and *araE)* and several genes that are not directly implicated in arabinose metabolism[130,131]. The arabinose regulon is also activated by the global regulator CRP (cAMP receptor protein) in response to low glucose levels[130,132]. In this study, the natural inducer L-arabinose was used for AraC and the pBAD promoter had a randomized CRP recognition site to make sure that reporter transcription was only regulated by AraC.

For selection, two different strategies for cell survival were compared, namely antibiotic resistance (kanamycin, KmR) and auxotrophy complementation (leucine, LeuB). Kanamycin resistance is realized by the aminoglycoside 3'-phosphotransferase that impairs kanamycin binding to the 30S ribosomal subunit by adding a phosphate group to this aminoglycoside [133]. LeuB is a 3-isopropylmalate dehydrogenase and is essential for L-leucine biosynthesis[134]. Only when this protein is present, cells can survive in absence of L-leucine. As the plasmid copy number may affect the behaviour of the reporter system, we

constructed a low and a medium copy system, by introducing the replication origins p15A and ColE1 respectively. For screening, bioluminescence was chosen, because it is very sensitive, has a broad dynamic range and is quickly detectable after induction. Moreover, no substrate is required when the whole operon *luxCDABE* is present (except $FMNH_2$ and O_2). The screening reporter genes used were in all systems *luxCDABE* from *Photorhabdus luminescens*, encoding the luciferase LuxAB and the multienzyme complex LuxCDE (LuxC, reductase; LuxD, transferase; LuxE, synthetase) that converts myristol-acyl-carrier protein to myristyl aldehyde, the substrate for the luciferase[127,135,136].

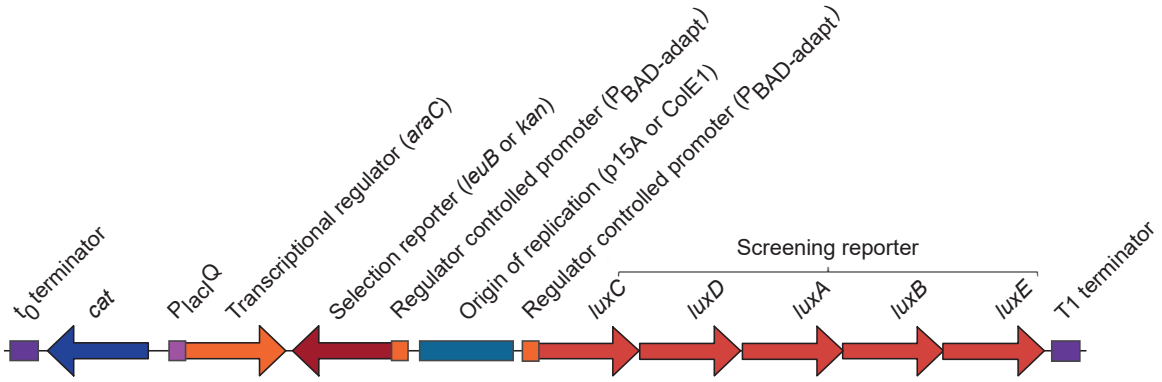


Fig. 1. Linear representation of the regulator-reporter plasmid. Different versions of the plasmid vary in the selection reporter (*leuB* or *kan*) and the copy number of the regulator-reporter plasmid (ColE1 or p15A origins of replications for medium or low copy number respectively). The t0 terminator blocks read-through transcription coming from the selection reporter or the chloramphenicol resistance marker (*cat*), whereas the T1 terminator blocks read-through transcription from the screening reporter *luxCDABE*. P_{lacI}^{Q} is a moderate constitutive promoter. $P_{BAD\text{-}adapt}$ is regulated by AraC.

Construction of the system

The construction of the system involved a series of cloning steps (Fig. S1) to make the regulator-reporter and control plasmids (Fig. 1), and the formation of several knockout strains. Each system module in the plasmids is flanked by unique restriction sites, allowing individual replacements. For each of the four regulator-reporter plasmids two control plasmids were constructed, one for selection and one for screening. The ideal control would be an active site mutant of the reporter, because it is most similar to the actual system in terms of plasmid size, copy number, transcriptional and translational burden and therefore growth rate. However, as these reporter mutants were unavailable, an alternative approach was chosen here. A frameshift was made, either in the selection reporter gene (*kan*/*leuB*) or in one of the screening reporter genes (*luxA*). Compared to, for example, removal of the coding sequence (CDS), these controls are very similar to the parent plasmids regarding plasmid size and transcriptional and translational burden. The obtained sequences for the

frameshift in the *kan* and the *leuB* genes differed from the expected fill in and removal of 5' and 3'overhangs respectively. Details and explanations are given in Table S1.

E. coli BW25113[137] was used as host strain for the regulator-reporter plasmids and the control plasmids. This strain has a deletion in the *araBAD* operon[138]. It is therefore unable to metabolize L-arabinose[139]. Here, the genes *araC*, *leuB*, and *recA* were deleted to exclude interference of endogenous AraC, to enable leucine auxotrophy complementation with LeuB and to prevent recombination events involving the plasmids respectively. Genes were replaced by a kanamycin resistance marker, which was later removed. Initially, the marker was removed by recombination of the flanking FLP recognition target (FRT) sites by FLP recombinase[137]. However, in subsequent gene deletions the scar FRT site is still recognizable by FLP and hence not suitable. Therefore, the marker was flanked with *lox71*/*lox66* sites instead, of which the scar after recombination by Cre recombinase is no longer recognizable by Cre[140]. The two obtained knockout strains Δ*araC* Δ*recA* and Δ*araC* Δ*leuB* Δ*recA* are indicated by AR and ALR in the rest of the text respectively.

After transformation of the knockout strains with the regulator-reporter or control plasmids, the relative copy numbers were determined. The relative plasmid copy number of the low and medium copy systems was 4-5 (Table S2). This ratio is slightly higher than copy number ratios reported for the pZ expression vectors, the parent plasmids of pFU98 from which the regulator-reporter plasmids and control plasmids were derived. pZ vectors with p15A or ColE1 replication origins, had copy numbers of 20-30 and 50-70 respectively[141]. However, since this study's plasmids are larger and have some different genes encoded, their demand on the cellular machinery and the building blocks might deviate, thereby altering the plasmid copy number. In addition, the pZ copy numbers were determined by comparing the activity of the plasmid-encoded with the chromosome-encoded luciferase (single copy). The ratio between frameshift control and parent plasmid was 1.0, confirming the expected similarity between the controls and their parent plasmids.

Characterization of the selection reporter LeuB

All systems were characterized to determine their performance in selection and screening. In this context, a good performance means a low leakiness, a high maximal signal, a broad dynamic range and a high sensitivity. In the selection step of this system, a high sensitivity and low leakiness are the most important criteria to detect even low concentrations of the small molecule of interest without many false positives. Every cell that survives is interesting and the reporter signal will subsequently be quantified in the screening step, in which all

four performance criteria are of importance, especially a high sensitivity and a broad dynamic range to obtain a relative ranking. In induction assays, the systems were induced by various concentrations of L-arabinose. LeuB-based assays were performed in minimal M9 medium, whereas KmR- and LuxCDABE-based assays were performed in rich LB medium. The reporter activity or output was quantified by measuring the optical density (OD600) and/or the bioluminescence. This paragraph describes the results of the selection assay based on leucine auxotrophy complementation by LeuB.

In the leucine auxotrophy complementation assay, the low and medium copy versions were analysed (Fig. 2). Three strains were tested for each system: (1) the system itself (auxotroph ALR + regulator-reporter plasmid), (2) a negative control (auxotroph ALR + regulator-reporter plasmid with a frameshift in *leuB*), and (3) a positive control (non-auxotroph AR + regulator-reporter plasmid with a frameshift in *leuB*). The strains were not induced in the precultures because pre-induction did not influence survival in the assay (Fig. S2). Bacteria were grown for 32 h (Fig. 2) and 48 h (Fig. S3) in minimal M9 medium. After 32 h the positive controls were in stationary phase (except at low L-arabinose concentrations), whereas most system strains were not (except for the low copy system at high L-arabinose concentrations). The higher the L-arabinose concentration, the faster system strains reached stationary phase. In addition, the low copy system grew faster than the medium copy system. The medium copy system did only barely grow after 48 h and in an unstable manner (large standard deviations and no definite relation between inducer concentration and growth). It could be that in minimal medium without leucine, the burden of the medium copy system was too high for the auxotrophic cells. Since growth of the positive controls was not much influenced by the copy number, it was the combination of the higher copy number and the dependence on the plasmid encoded LeuB that caused the troubled complementation in the medium copy system. Growth was somehow positively affected by higher L-arabinose concentrations (see positive controls), but growth on L-arabinose seemed unlikely as *E. coli* BW25113 does not have the *araBAD* operon. The increase in growth of the low copy system with higher L-arabinose concentrations was larger than for the positive control, as the increase was due to both the induction of *leuB* and the positive growth effect of L-arabinose. Under non-selective conditions, the frameshift-based controls indeed grew very similar to the system itself. Moreover, under selective conditions, their reporter activity, measured as growth was negligible. The frameshift approach is therefore a good method to make controls and may also be used in other studies

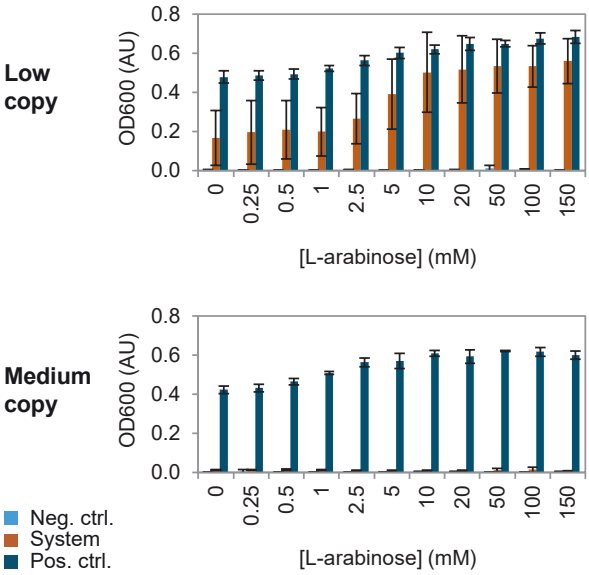


Fig. 2. Selection based on leucine auxotrophy complementation. The plasmid-encoded reporter gene *leuB* was induced in low and medium copy systems by various concentrations of the inducer L-arabinose. Bacteria were grown in M9 medium for 32 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*araC* Δ*leuB* Δ*recA* (ALR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph ALR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid with a frameshift in *leuB*.

Characterization of the selection reporter KmR

In the kanamycin resistance assay, the low and medium copy versions were analysed (Fig. 3). Two strains were tested for each system: (1) the system itself (AR + regulator-reporter plasmid), and (2) a negative control (AR + regulator-reporter plasmid with a frameshift in *kan*). The strains were induced in the precultures (only non-induced strains in the assays came from non-induced precultures), because pre-induction did affect survival in the assay (Fig. S4). The explanation of the pre-induction effect was that L-arabinose induces also expression of *araE*, encoding the low affinity L-arabinose transport system. This inducer-dependent transport control results in an all-or-nothing induction, in which intermediate L-arabinose concentrations give rise to subpopulations of cells that are fully induced or non-induced. The ratio of these subpopulations shifts over time towards full induction of all cells[142]. This stage is most likely reached in the precultures, explaining the positive effect of pre-induction on growth in the assay.

Bacteria were grown for 17 h in LB medium (stationary phase) in the presence of 0, 5, 15 and 30 µg mL^{-1} kanamycin. These concentrations were chosen based on death curves at a fixed inducer concentration (Fig. S5). The negative controls and non-induced system strains could not survive above 2.5 µg mL^{-1} kanamycin, a concentration comparable to literature (1-

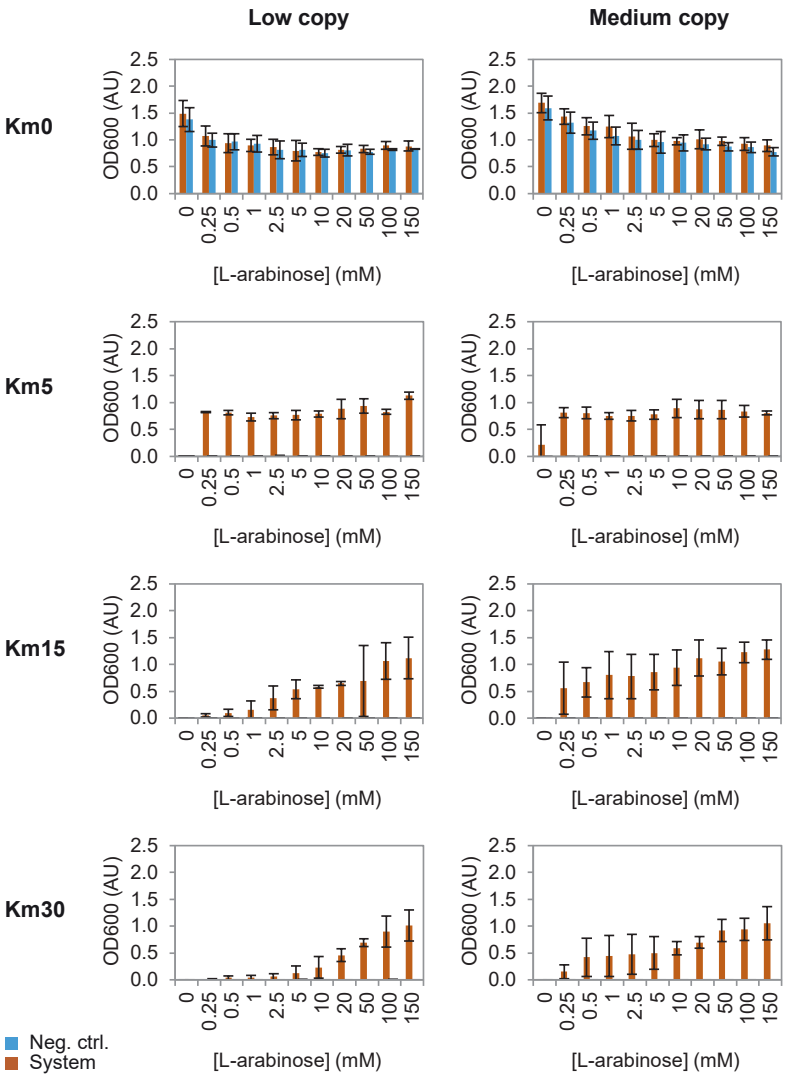


Fig. 3. Selection based on kanamycin resistance. The plasmid-encoded reporter gene *kan* was induced in the low and medium copy system with the inducer L-arabinose. Bacteria were grown in LB medium for 17 h in presence of 0, 5, 15 or 30 µg mL^{-1} kanamycin. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid. Neg. ctrl.: AR with the regulator-reporter plasmid with a frameshift in *kan*.

3 µg mL^{-1} kanamycin[143]). Induction by L-arabinose enabled the system strains to survive above 2.5 µg mL^{-1} and higher inducer concentrations allowed survival at higher kanamycin concentrations. However, at maximum induction none of the strains could cope with 50 µg mL^{-1} kanamycin, the concentration commonly used to maintain plasmids with the same kanamycin marker. Since in this study the plasmids were large and contained eight genes, the expression per gene was probably relatively low and not enough resistance was built up to deal with 50 µg mL^{-1} kanamycin. Consistent with this, the lower copy system needed higher inducer concentrations than the medium copy system to deal with the same kanamycin concentration. This phenomenon of more gene copies, more protein and thus

more resistance is called the gene dosage effect[144]. The relative low range of kanamycin concentrations should not be a problem, as long as future selections are performed within or just around this range. In contrast to the LeuB-based assay, increasing the L-arabinose concentration affected growth negatively (see 0 µg mL^{-1} kanamycin). The opposite effect in the two assay types might be caused by the difference in growth medium, rich versus minimal medium. Unfortunately, a more detailed explanation cannot be given. Under non-selective conditions, the frameshift-based controls once more grew very similar to the system itself and also here under selective conditions their reporter activity, measured as growth was negligible.

Characterization of the screening reporter LuxCDABE

In the bioluminescence assay, all four systems were analysed (Fig.4). Two strains were tested for each system: (1) the system itself (AR + regulator-reporter plasmid), and (2) a negative control (AR + regulator-reporter plasmid with a frameshift in *luxA*). Bacteria were grown in LB medium for 5.5 h. At this time point, cultures were in late log phase at a point for which signal production and wash out due to cell division were about equal. Higher inducer concentrations resulted in more bioluminescence with maximal induction at 50 mM. These concentrations were comparable with literature values, namely, 0.1-30 mM[128,145,146]. The maximal induction for medium copy systems was higher than for low copy systems; probably a gene dosage effect. The KmR and LeuB versions did not differ in signal. The frameshift-based controls again grew very similar to the system itself and also here their reporter activity, measured as bioluminescence was negligible. Comparing these systems with previous and future systems based on bioluminescence values will be difficult, because the energy state of the cell influences the bioluminescence. Slight differences in the protocol can already change the output. However, for the comparison of the systems within one study this is not an issue.

Comparison of the systems

To further compare the four systems, leakiness, maximal signal, dynamic range and sensitivity were determined (Table 1). Based on these characteristics, a comparison was made for (1) low versus medium copy and (2) LeuB versus KmR.

Low versus medium copy. In the LeuB-based assay, the growth rate of the medium copy system was unstable compared to the low copy system (Fig. 2 and Fig. S3), making determination of the four characteristics impossible. The medium copy system did not

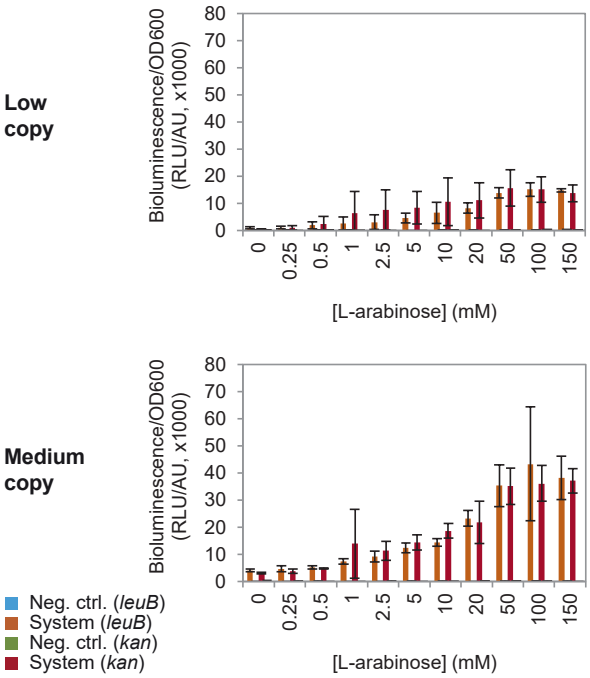


Fig. 4. Screening based on bioluminescence. The plasmid-encoded reporter operon *luxCDABE* was induced in four different systems by various concentrations of the inducer L-arabinose. The four systems were the low and medium copy systems with either LeuB or KmR as selection reporter. Bacteria were grown in LB medium under non-selective conditions for 5.5 h. The data are an average of three independent experiments (standard deviation indicated). System with LeuB: *E. coli* BW25113 Δ*araC* Δ*leuB* Δ*recA* (ALR) with the regulator-reporter plasmid with *leuB*. Neg. ctrl. with LeuB: ALR with the regulator-reporter plasmid with *leuB* and a frameshift in *luxA*. System with KmR: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid with *kan*. Neg. ctrl. with KmR: AR with the regulator-reporter plasmid with *kan* and a frameshift in *luxA*.

function very well, possibly because the auxotrophic cells were more burdened by the higher copy number in combination with the relative harsh condition of minimal medium without leucine. In the KmR-based assay in general, low and medium copy systems were both not leaky, they had a similar maximal signal, but the medium copy system was more sensitive than the low copy system and the relative dynamic range of the two systems depended on the kanamycin concentration. Most likely there was some expression in absence of inducer; only the amount of KmR was not enough to deal with the lowest tested kanamycin concentration of 5 μg mL^{-1}, appearing as if there was no leakiness. In contrast to the low copy system, the medium copy version had sufficient *kan* expression to survive 5 μg mL^{-1} kanamycin. This gene dosage effect is likewise observed in the leakiness in the LuxCDABE-based assay and also the probable cause of the difference in sensitivity in the KmR-based assay. The delicate balance of survival and death at 5 μg mL^{-1} kanamycin, promotes use of slightly higher kanamycin concentrations in future studies. In the LuxCDABE-based assay, the

low copy systems were less leaky, had a lower dynamic range and were less sensitive than the medium copy systems, due to an overall lower expression level (gene dosage effect).

LeuB versus KmR. LeuB-based selection was leakier than KmR-based selection, due to the threshold set by adding ≥ 5 µg mL^{-1} kanamycin. In addition, LeuB-based selection had a lower maximal signal, because growth in minimal medium compared to rich medium reduces the maximal OD600. The sensitivity and the dynamic range (latter only at higher kanamycin concentrations), were better with KmR than with LeuB. In the KmR-based assay, the sensitivity could be varied by changing the kanamycin concentration, and the assay time is much less than for the LeuB-based assay, due to a higher growth rate in rich medium. Both are interesting features for later applications. Remarkably, the ability to deal with the selection pressure was less than expected in both selection assays. For leucine auxotrophy complementation, the system strains grew much slower than the positive controls and for kanamycin resistance, system strains could not deal with the commonly used 50 µg mL^{-1} kanamycin. The explanation is twofold. On the one hand, the plasmids are large and multiple genes have to be expressed, lowering the expression per gene. On the other hand, the CRP binding site is absent, preventing regulation of reporter expression by CRP and thus by glucose. Normally, the presence of both cAMP (low glucose) and L-arabinose does result in a higher induction than with L-arabinose alone[147].

Overall. All systems were functional except for the medium copy system with LeuB as selection reporter. But which system functions best? Based on the different characteristics described above and the rationale that in selection a high sensitivity and a low leakiness are the most important criteria and in screening a high sensitivity and a broad dynamic range, the medium copy system with KmR as selection reporter was selected as best system. Since in the selection step a high sensitivity and a low leakiness are the most important criteria to detect even low concentrations of the small molecule of interest without much false positives, the total lack of leakiness at higher kanamycin concentrations is very valuable in future applications. Everything that survives is interesting and will subsequently be quantified in the screening step, in which a high sensitivity and a broad dynamic range are the most important criteria. The bit of leakiness in screening with the best system is therefore not detrimental. For screening, the fold change of the maximal signal over the leakiness was about ten. This fold change is similar to those in other transcriptional regulator-based systems[111,148], but it is higher than in riboswitch-based systems[95]. The sensitivity for both selection (<0.25 mM) and for screening (0.25-0.5 mM) is lower than the sensitivity of described screening-based bioreporters that were applied in for example

Table 1. Characteristics of the reporter systems[a]

Reporter	Copy number	Leakiness (AU)[b]		Maximal signal (AU)[c]		Dynamic range (mM)[d]		Sensitivity (mM)[e]	
LeuB	Low	0.16 ± 0.14	−	0.56 ± 0.12	− +	2.5-10	−	1.0-2.5	−
	Medium	NA	NA	NA	NA	NA	NA	NA	NA
KmR (Km5)	Low	0.00 ± 0.00	+ +	1.13 ± 0.07	+ +	0.25-0.25	− −	<0.25	− +
	Medium	0.22 ± 0.37	− −	0.80 ± 0.04	+	0.25-0.25	− −	<0.25	− +
KmR (Km15)	Low	0.00 ± 0.00	+ +	1.12 ± 0.38	+ +	0.25-100	+	<0.25	− +
	Medium	0.00 ± 0.00	+ +	1.27 ± 0.18	+ +	0.25-20	− +	<0.25	− +
KmR (Km30)	Low	0.00 ± 0.00	+ +	1.01 ± 0.29	+ +	10-150	−	5-10	−
	Medium	0.00 ± 0.00	+ +	1.05 ± 0.31	+ +	0.25-50	− +	<0.25	− +
LuxCDABE (*leuB*)	Low	1058 ± 414	+	14599 ± 599	+	5-50	−	2.5-5	−
	Medium	4087 ± 507	+	38168 ± 8032	+ +	1-50	−	0.5-1	−
LuxCDABE (*kan*)	Low	349 ± 203	+ +	13684 ± 3101	+	5-50	−	2.5-5	−
	Medium	2960 ± 385	+	37076 ± 4436	+ +	0.5-50	− +	0.25-0.5	−

[a]The systems vary in the selection reporter (LeuB or KmR) and the copy number of the regulator-reporter plasmid (medium or low). The KmR-based systems are characterized at three different kanamycin concentrations (5, 15 and 30 μg mL^{-1}). The LuxCDABE-based systems are characterized for both LeuB and KmR containing versions. The standard deviation is included. A qualitative ranking is made (− −, −, − +, +, + +) with − − indicating a poor system and + + a good system. For leakiness, this indication is relative to the maximal signal. Absolute numbers for leakiness and maximal signal cannot be directly compared between the selection reporters LeuB and KmR and the screening reporter LuxCDABE, because they represent growth and bioluminescence respectively. [b]Signal at 0 mM inducer. [c]Signal at saturating inducer concentration. [d]Range of concentrations giving a changeable signal. [e]Lowest detectable inducer concentration.

library screening or strain optimization (0.05-10 µM[115,123,149]), but is still of biological relevance (see section on isomerase detection below). The dynamic range of the medium copy system with KmR was satisfactory for both selection and screening (two orders of magnitude) and is comparable to those in other transcriptional regulator-based systems[107,115,149].

Proof of principle for application in enzyme screening

The next step was to obtain a proof of principle that the best performing system would be suitable for enzyme screening. As target, the enzyme L-arabinose isomerase or AraA was chosen, because this enzyme activity can be linked to the AraC-based system. Moreover, this type of enzymes is interesting for industrial production of rare sugars, like the sweetener D-tagatose, which is produced from D-galactose as a side reaction of L-arabinose isomerase[150]. L-arabinose isomerase catalyses the first reaction in L-arabinose breakdown, namely the conversion of L-arabinose to L-ribulose[151]. *E. coli* BW25113[137], the strain used to create the system, has a deletion in the *araBAD* operon[138] and thus no endogenous L-arabinose isomerase (AraA), L-ribulosekinase (AraB) and L-ribulose-5-phosphate 4-epimerase (AraD). Since the reaction equilibrium of the isomerase is in favour of L-arabinose (L-arabinose/L-ribulose = 5-15[152,153]) and the reaction is not pulled towards L-ribulose without AraB, it is likely that L-ribulose is converted to L-arabinose under the growth conditions in this study. Uptake of L-ribulose was expected, because *E. coli* MG1655, having an intact *araBAD* operon, could grow on L-ribulose. To show the applicability of the system for enzyme discovery of different origin, the L-arabinose isomerase from mesophilic *E. coli* and the predicted L-arabinose isomerase from thermophilic *G. thermodenitrificans* T12 were chosen. The latter was annotated as L-arabinose isomerase (60% and 93% amino acid identities with *E. coli* MG1655 AraA and *G. thermodenitrificans* CBG-A1 AraA respectively), but its function was not yet experimentally verified. For constitutive expression of *araA*, a second low copy plasmid was used next to the medium copy KmR-based reporter system.

To show that the system could indeed detect the activity of the two L-arabinose isomerases, KmR- and LuxCDABE-based assays were performed in which L-ribulose was added to the medium as substrate for AraA (Fig. 5). The negative control was the system strain with the second plasmid lacking the *araA* CDS. For the KmR-based assay, cells were grown in LB medium for 17 h with 0 or 15 µg mL^{-1} kanamycin. Only when one of the L-arabinose isomerases was expressed, cells survived the kanamycin, verifying the annotation of *G. thermodenitrificans* T12 *araA* and showing that the system is capable of detecting a

mesophilic and a thermophilic enzyme based on growth. However, a substantial amount of L-ribulose was needed to observe the enzyme activity, namely ~2 mM. This sensitivity differed an order of magnitude with the sensitivity for L-arabinose of cells without L-arabinose isomerase (~2 mM vs. ~0.25 mM; Table 1). It was unlikely that this decrease in sensitivity was a result of a difference in uptake between the two sugars, because the sensitivity in the LuxCDABE-based assay (see below) was in the same order of magnitude for extracellular added L-arabinose or L-ribulose converted to L-arabinose. A more probable explanation was the burden of expressing *araA* (Fig. 5; Km0, empty plasmid vs. *araA*), This burden had two components; the effect of *araA* on growth in absence and in presence of L-ribulose. In absence of L-ribulose, cells expressing *araA* were hindered in growth (Fig. 5, stationary phase; Fig. S6, exponential phase). Whether it was the activity of AraA or just its expression load, was not known, but the observation that *araA* was a burden to the cells was strengthened by the failure to make a plasmid with *E. coli araA* under the stronger P_{lacUV5} promoter. Cells expressing *E. coli araA* were more burdened than cells expressing *G. thermodenitrificans araA* (Fig. S6), possibly because they seemed to higher express *araA* (Fig. S7). Better expression of *E. coli araA* than *G. thermodenitrificans araA* was expected, because the latter was not expressed in its endogenous host. In presence of L-ribulose, cells were more burdened by *araA* than in absence of L-ribulose and with higher L-ribulose concentrations the burden increased (Fig. 5). As mentioned above, L-arabinose had a negative effect on growth and it is therefore most likely that the L-arabinose formed out of L-ribulose caused the concentration dependent growth defect. The system was slightly more sensitive for the *G. thermodenitrificans araA* than for the *E. coli araA* (<2 mM vs. 2-5 mM), probably due to the growth differences between the two strains. Cells with *G. thermodenitrificans araA* might have had a lower level of active *araA* due to a lower expression and a lower activity because of its thermophilic origin. Therefore these cells had a less negative effect on growth from L-arabinose compared to the cells with *E. coli araA*.

For the LuxCDABE-based assay, cells were grown in LB medium for 5.5 h. Only when one of the L-arabinose isomerases was expressed, cells were bioluminescent, showing that the system is also capable of detecting a mesophilic and a thermophilic enzyme based on bioluminescence. The sensitivity of this assay was similar for both L-arabinose isomerases and about >50 fold higher than that of the KmR-based assay (0.01-0.1 mM vs.2-5 mM). This difference was most likely caused by the negative growth effect of both L-arabinose and AraA in the KmR- or growth-based assay. Since in this assay a threshold of expressed KmR had to be reached, a negative growth effect probably had a more detrimental effect than in

the LuxCDABE-based assay, having a more gradual response curve. Quantification of the different levels of enzyme activity was not as straightforward as envisioned due to the negative growth effect of L-arabinose and the difference in expression levels between the *E. coli* and the *G. thermodenitrificans* L-arabinose isomerase.

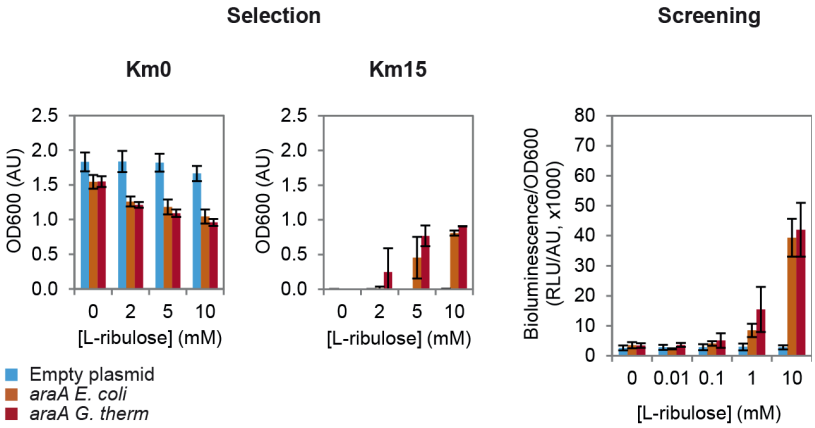


Fig. 5. L-arabinose isomerase detection by selection- and screening assays. Conversion of L-ribulose to L-arabinose by the L-arabinose isomerase AraA of *E. coli* or *G. thermodenitrificans* induced the system. Left (selection): detection based on kanamycin resistance. Bacteria were grown in LB medium for 17 h in presence of 0 and 15 µg mL^{-1} kanamycin. Right (screening): detection based on bioluminescence. Bacteria were grown in LB medium for 5.5 h. The data are an average of two or three independent experiments (standard deviation indicated) for selection or screening respectively. *araA E. coli* or *G. therm*: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid and the plasmid expressing *araA* of *E. coli* or *G. thermodenitrificans*. Empty plasmid: AR with the regulator-reporter plasmid and the empty plasmid.

Altogether, these assays showed that the system was capable of detecting a mesophilic and a thermophilic enzyme based on growth and on bioluminescence. However, to show that this system is suitable for application in enzyme screenings, it has to be able to enrich cells with the desired enzyme activity over cells that do not have this activity. For this purpose, selection and screening of an enzyme library was mimicked by mixing cells with the *E. coli araA, G. thermodenitrificans araA* or no *araA* (empty plasmid) in a $1:1:10^8$ ratio. Cells were selected based on kanamycin resistance for 6 h in liquid medium and 17 h on agar plates in the presence of 5 mM L-ribulose as substrate and 15 µg mL^{-1} kanamycin. Making use of the second reporter, the 68 selected colonies were analysed by a bioluminescence-based screening assay in presence of 0.5 mM L-ribulose to show the L-ribulose dependent bioluminescence as verification of *araA* presence. Six of these colonies gave L-ribulose dependent bioluminescence (Fig. S8) and were verified by PCR to contain *araA*. The other colonies were false positives; they did not give bioluminescence and were verified by PCR to contain the empty plasmid. Based on the control cultures with only one strain, 25 times

more *araA* containing cells were expected. The low number might have been caused by competition with false positives in the mixed culture. Of the six *araA* containing colonies, one colony had *araA* of *E. coli* and five colonies had *araA* of *G. thermodenitrificans* (Fig. 6). This advantage of the *G. thermodenitrificans araA* over the *E. coli araA* containing cells, was due to their faster growth. During the 6 h in liquid medium, the cells with *G. thermodenitrificans araA* grew about six times faster than the cells with *E. coli araA* in the control cultures containing only one strain. They were less burdened by AraA and growth inhibiting L-arabinose, as discussed above.

Starting from $2.0 \cdot 10^{-6}$% of the cells having *araA* and ending with 8.8% (Fig.6) meant an enrichment of $4.4 \cdot 10^{6}$ fold in only one round of selection and screening. Other systems were just tested with initial ratios up to $1:10^{6}$ and required at least two FACS rounds or one selection round to get to a more than 10^{5} fold enrichment[84,111,154]. Thus, the system described here is able to obtain a very good enrichment, and it is relatively easy, short and cheap, compared to for example FACS. In addition, it is able to distinguish the false positives from true positives with the subsequent screening assay, emphasizing the value of this dual reporter system. Dietrich *et al.* already published a dual reporter system with TetA for selection and GFP for screening, but unfortunately the combined use of the two reporters was not yet fully demonstrated[155]. Garmendia *et al.* successfully demonstrated another two stages approach, using *pyrF* as reporter gene in a Δ*pyrF* background. Positive selection was based on uracil auxotrophy complementation and negative selection based on fluoroorotic acid sensitivity [71,156]. The nature of the false positives was investigated by PCR and 74.2% (Fig. 6) of the false positives had a recombination in the regulator-reporter plasmid. A 17 base pairs region including the ribosomal binding site (RBS) in front of *kan* was recombined with the identical region in front of *araC*. This resulted in an exchange of the CDSs of these two genes, placing *kan* under the constitutive P_{lacI^Q} instead of under the AraC controlled $P_{BAD\text{-}adapt}$ and thereby enabling the cells to survive kanamycin in absence of *araA*. Interestingly, this recombination took place despite the deletion of *recA*. Although a fragment as short as 17 bp was not tested, *E. coli* is capable of RecA-independent recombination of short homologous regions[157]. Adaptation of the system to prevent this recombination was not considered useful, because in that case other escape mutants are likely to become dominant as is intrinsic to selection. The nature of the false positives made the screening by bioluminescence as second step better than a second selection step in which the false positives would survive again. The other 25.8% of false positives had an

unknown mutation giving constitutive resistance to 15 µg mL^{-1} kanamycin. One possibility is a mutation in $P_{BAD\text{-}adapt}$ to make expression of *kan* independent of AraC.

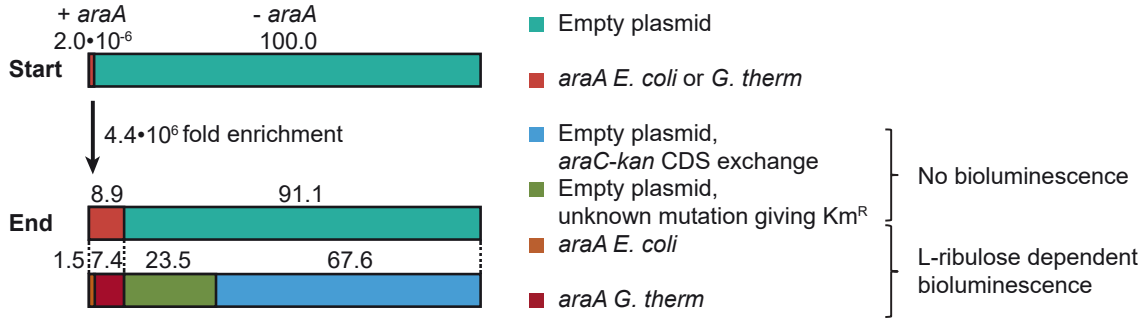


Fig. 6. Enrichment of cells with L-arabinose isomerase (AraA) activity. Cells with *E. coli araA, G. thermodenitrificans araA* or no *araA* were mixed in a 1:1:10^8 ratio ($2.0 \cdot 10^{-6}$% of cells with *araA*) and cells with L-arabinose isomerase activity were selected based on kanamycin resistance in presence of 5 mM L-ribulose as substrate and 15 µg mL^{-1} kanamycin. Selected colonies were analysed by a bioluminescence-based screening assay in presence of 0.5 mM L-ribulose to show the L-ribulose dependent bioluminescence as verification of *araA* presence. Further verification was done by PCR. *araA E. coli* or *G. therm*: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid and the plasmid expressing *araA* of *E. coli* or of *G. thermodenitrificans*. Empty plasmid, *araC-kan* CDS exchange: AR with the regulator-reporter plasmid with a recombination of the *araC* and *kan* CDSs that places *kan* under a constitutive promoter instead of the AraC controlled promoter, and the empty plasmid. Empty plasmid, unknown mutation giving Km^R: AR with the regulator-reporter plasmid and the empty plasmid and an unknown mutation giving kanamycin resistance. Values above bars are percentages.

Conclusions

In this study, a selection-based system for the detection of small molecules, or more particularly for products of novel biocatalysts, was developed and characterized. The system expresses two reporters under control of AraC, allowing for both selection (based on growth) and screening (based on bioluminescence). Growth-based selection allows for a rapid reduction of the initially large library size and subsequent positive hits can be quantified by bioluminescence. Different versions of the system with a low or medium plasmid copy number and leucine auxotrophy complementation (LeuB) or kanamycin resistance (KmR) as selection reporter were compared. The medium copy system with KmR as selection reporter was selected as best system, based on leakiness, maximal signal, dynamic range and sensitivity in both selection and screening. This system was used to detect L-arabinose isomerase derived from mesophilic *E. coli* and thermophilic *G. thermodenitrificans* with L-ribulose as substrate. Moreover, cells with one of the two L-arabinose isomerases were enriched over cells without L-arabinose isomerase with a factor $4.4 \cdot 10^6$, making use of the selection reporter. The screening reporter enabled the distinction of true from false positives.

Previous objections to bioreporters with growth-based selection were that growth assays can have a relatively low dynamic range or low sensitivity, and a high level of false positives due to escape mutants, unanticipated survival mechanisms or various influences on growth of the positive cells[111,125,155,158]. In the systems described in this study, however, the dynamic range and sensitivity in selection were similar or even slightly better than in screening. Both dynamic range and sensitivity are comparable to other reported systems, but the sensitivity of the here reported system might need some improvement, e.g. via adaptation of the relative expression levels of the system components. Overall, the best performing system has an appropriate working range as confirmed by its ability to detect an enzyme activity as proof of principle. Moreover, the system is able to enrich cells with the enzyme activity over cells that do not have the activity on a scale mimicking a library of 10^8, in a relatively easy, fast and cheap manner. The set-up as double reporter system reduces the number of false positives by having the selection and screening steps in series, which function therefore as double check. Although the enrichment is already much better than for other systems, further improvements like an additional selection reporter under control of AraC or using a selection reporter that allows for both negative and positive selection, could improve the selection potential and reduce the number of false positives even more. The modular make-up of the system makes the exchange of components like the selection reporter straightforward. Also the screening reporter could be exchanged, for example by GFP, in cases where the dependency of the reporter activity on the metabolism or growth phase is a problem. Genome integration of the reporters might be an option to enhance the stability of the system. Noteworthy, each of these alterations requires some fine tuning and characterization.

Although a proof of principle for the application in enzyme searches is shown here, the system developed in this study should be regarded as a prototype. Application of this system in detecting specific small molecules requires changing the specificity of the system by altering the transcriptional regulator. Two approaches can be used to adjust the specificity. Firstly, the system can be easily re-cloned to function with another transcriptional regulator, because the constructs have a modular design. In that case, the characteristics should be determined again, because they might differ due to distinct induction mechanisms or different transcriptional or translation rates of the regulators or dissimilar binding kinetics of the regulators to the DNA and to their inducers. Secondly, the transcriptional regulator can be engineered to change its inducer specificity as was done for AraC in other studies[87,88,128,129]. Although less drastic changes in characteristics are envisioned than for a

complete new regulator (promoter sequences, most of CDS, etc. stay the same), also in this case characteristics should be determined again. A most interesting feature of the system is that the system itself can be used to select and optimize a new regulator variant. A library of transcriptional-regulator variants can be made and with the system, the variant with the highest specificity towards the target small molecule can be selected. Additional rounds of library formation and selection can further optimize the specificity. Although the double reporter system with its subsequent selection and screening steps reduces the number of false positives when detecting small molecules, a good counter selection is still required to reduce the number of false positives that originate from regulators that allow transcription of the reporter in absence of the inducer. Also discrimination between variants that only differ slightly in specificity[159], might require a more tight selection as described above. A combination of negative and positive selection, preferably accommodated by one gene, might proof useful.

In conclusion, this study provided insight into various aspects of whole-cell bioreporters. The successful development is described of an alternative for the often expensive and/or laborious high-throughput novel biocatalyst detection, and more general for small molecule detection, by combining a selection and a screening reporter in a single system. Future research will focus on the next crucial step, namely using the system for the selection of regulator variants.

Materials and methods

Bacterial strains and media

E. coli DH10B $T1^R$ (Invitrogen, catalog number C6400-03) was used for plasmid propagation and was grown and transformed by standard methods[160]. *E. coli* BW25113 JW0063-1 of the KEIO-collection[161] was the parent strain for the constructed knockouts strains. The knockout strains hosted the regulator-reporter plasmids or their controls. Transformations were done by electroporation (ECM 630 electroporator (BTX), 2500 V, 200 Ω, 25 µF, 2 mm cuvettes, 20-50 µL of electrocompetent cells, recovery in LB medium). Cells were generally grown in LB medium with the appropriate antibiotics: 100 µg mL^{-1} ampicillin, 50 µg mL^{-1} kanamycin or 34 µg mL^{-1} chloramphenicol, unless stated otherwise. Leucine auxotrophy complementation assays and growth on L-ribulose were performed in M9 medium. Enrichments were done in LB medium with 4 g L^{-1} glycerol to reach a higher OD600.

Construction of regulator-reporter plasmids and control plasmids

The regulator-reporter plasmids pWUR766 and pWUR768 (~10 kb each) were obtained in seven subsequent cloning steps from pFU98[162]; kindly provided by Petra Dersch). pFU98 contains a chloramphenicol resistance marker (*cat* encoding chloramphenicol acetyltransferase), the pSC101* origin of replication protected from read-through transcription by two flanking terminators (t_0 and T1), a multiple cloning site and a very strong RBS (AGGAGG; -12 to -7 relative to translation start) in front of *luxCDABE*. The cloning steps were (1) replacement of the very low copy origin pSC101* by the medium copy ColE1 to ease further cloning steps, (2) insertion of the selection reporter gene *leuB* or *kan* (incl. RBS as above and PvuI site; for *leuB* silent mutation with same codon usage factor, TCG-->AGT, to remove AatII and PvuI site from CDS), (3), insertion of the moderately strong and constitutive P_{lacI}^{Q} promoter[163]; incl. CpoI site), (4) insertion of the transcriptional regulator gene *araC* (incl. RBS as above) behind P_{lacI}^{Q}, (5) insertion of the $P_{BAD\text{-}adapt}$ promoter and operator region in front of *luxCDABE*, (6) translocation of ColE1 in between the two reporters to prevent expression and/or recombination problems by the two almost identical promoter sequences next to one another (the terminators were left at the original location), and (7) insertion of $P_{BAD\text{-}adapt}$ in front of *leuB*/*kan*. $P_{BAD\text{-}adapt}$ (this study) had a randomized CRP binding site to make sure that the reporters are only regulated by AraC and it had an internal restriction site (NheI or PstI; Table S3). More details of the intermediary cloning steps and the primers are given in Fig. S1 and Table S4 respectively.

The origin ColE1 in pWUR766 and pWUR768 was replaced by p15A with Acc65I/AvrII to yield the low copy variants pWUR770 and pWUR772 respectively. From each of the four constructs, two control constructs were made containing a frameshift either in the selection reporter gene (*leuB* or *kan*) or in one gene of the screening reporter operon (*luxA*). The parent plasmids were digested inside the gene at a unique restriction site: Eam1105I in *leuB*, XagI in *kan* and Cfr42I in *luxA*. The ends were made blunt with Klenow Fragment, according to the protocol of Thermo Scientific.

For all cloning steps, plasmids were isolated with the Plasmid Miniprep kit of Thermo Scientific (#K0503). PCRs to create insert fragments were done with Pfu. Vector fragments were treated with Antarctic Phosphatase (NEB), according to the protocol of NEB. Insert or vector fragments were purified with the PCR purification kit of Thermo Scientific (#K0702), the DNA Clean&Concentrator-5 kit of Zymo Research (D4004), or the gel extraction kits of Thermo Scientific (#K0692) or Zymo Research (D4002). Ligation was performed for 1 h at room temperature with T4 ligase. Cloning events were verified by PCR with DreamTaq

and/or restriction analysis and by sequencing at GATC Biotech. All enzymes were obtained from Thermo Scientific, unless stated otherwise. The nucleotide sequences of the four regulator-reporter plasmids pWUR766, pWUR768, pWUR770 and pWUR772 were submitted to the Genbank database under accession numbers KX670545-8 respectively.

Construction of knockout strains

The kanamycin resistance gene *kan* from *E. coli* BW25113 JW0063-1 (Δ*araC*::*kan*) of the KEIO-collection[161] was eliminated by FLP recombinase encoded on pCP20[164] as described by Datsenko and Wanner[137].

The Δ*araC* Δ*leuB* double knockout was constructed according to Datsenko and Wanner[137], with the exception of the disruption cassette. A new disruption cassette was developed based on the recombination cassette from Westra *et al.*[165], replacing the FRT sites that flank *kan* with *lox71*(left)/*lox66*(right) sites[140] synthesized and cloned SfiI/SfiI in pMA-RQ by GeneArt AG, see Table S3 for description and sequence). With this plasmid, pMA-RQ_lox71_*kan*_lox66, as template, a linear cassette was created by PCR with Pfu (Thermo Scientific), introducing the homologous regions (same regions as in Baba *et al.*[161]. After direct DpnI treatment, the product was purified with the PCR purification kit of Thermo Scientific (#K0702). For elimination of *kan* by Cre recombinase, parts of the protocol from Datsenko and Wanner[137], were replaced by components of the protocol from Palmeros *et al.*[166]. Knockouts were transformed with pJW168[167], transformants were selected on LB medium with ampicillin and 0.5 mM IPTG at 30 °C and cured from plasmids at 37 °C. For the *leuB* deletion leucine auxotrophy was verified on minimal M9 medium with or without 20 mg L^{-1} L-leucine. The parent strain was taken as control.

The Δ*araC* Δ*recA* double knockout and the Δ*araC* Δ*leuB* Δ*recA* triple knockout (designated AR, and ALR respectively) were constructed as described above for Δ*araC* Δ*leuB* with one exception. The disruption cassette (same homologous regions as in Baba *et al.*[161]) was made with purified SfiI digested pMA-RQ_lox71_*kan*_lox66 as template in the PCR, making DpnI treatment unnecessary. Elimination of *kan* was the same as for Δ*araC* Δ*leuB*.

Recombination events were verified by PCR with REDTaq (Sigma) or DreamTaq (Thermo Scientific). Gene replacement by *kan* was verified with two primer sets, each set with one primer flanking the altered region and one inside *kan*. *kan* elimination was verified with one primer set, each primer flanking the altered region. All deletions in the two final knockout strains AR and ALR were verified by PCR with Pfu and PCR products were sequenced at GATC Biotech. All primers are presented in Table S4. The knockout strains

were transformed with the regulator-reporter plasmids or control plasmids.

Induction assays

The three types of induction assays, based on expression of *leuB*, *kan* or *luxCDABE* had a similar experimental set-up. Two millilitres of precultures were inoculated from agar plates made of the same medium (for adaptation) and grown in 10 mL tubes (Gosselin). The assays were performed in 2-mL 96-well MASTERBLOCKS (Greiner Bio-One) with 500 µL total volume with a range of L-arabinose concentrations and an equal starting OD600 (0.005, 0.0001 and 0.0000625 for the LeuB-, KmR- or LuxCDABE-based assays respectively). After growth, 200 µL per culture was transferred to a transparent 96-well microplate (Greiner Bio-One) for an OD600 measurement with a Synergy MX microplate reader (BioTek). OD600 values were corrected for path length and an average of three blanks. All assays were performed as three independent experiments, being therefore both biological and technical replicates. The data was averaged and the standard deviation was calculated.

For the leucine auxotrophy complementation assays the two system strains expressing *leuB*, the two positive control strains (non-auxotrophs with a frameshift in the plasmid encoded *leuB*) and the two negative control strains (auxotrophs with a frameshift in the plasmid encoded *leuB*) were pre-grown in minimal M9 medium with 18 µg mL^{-1} chloramphenicol, 1x Minimum Essential Medium (MEM) vitamins, 20 mg L^{-1} L-leucine for complementation and with/without inducer (10 mM L-arabinose) for 24 h. In the assays, the leucine concentration was kept below 1 µM to prevent complementation by leucine present in the medium[168]. OD600 was measured after 32 and 48 h.

For the kanamycin resistance assays the two system strains expressing *kan* and the two corresponding negative control strains with a frameshift in *kan* were pre-grown in LB medium with 34 µg mL^{-1} chloramphenicol and with/without inducer (10 mM L-arabinose) for 7h. In the assays, kanamycin concentrations were varied. OD600 was measured after 17 h.

For the bioluminescence assays the four system strains expressing *lux* and the four corresponding negative control strains with a frameshift in *luxA* were pre-grown in LB medium with 34 µg mL^{-1} chloramphenicol for 17 h. In the assays, OD600 and bioluminescence were measured in the microplate reader after 5.5 h. Bioluminescence was measured in white 96-well microplates (Thermo Scientific, Nunc; 200 µL per well) under default settings. The temperature of the plate reader was set at 37 ^{o}C. Bioluminescence values were corrected for the OD600.

Detection of L-arabinose isomerase activity

G. thermodenitrificans T12 AraA (Genbank: KX555561) was compared with. *E. coli* MG1655 AraA (GenBank: AAC73173.1) and *G. thermodenitrificans* CBG-A1 AraA (GenBank: AY302754) by BLASTP 2.3.1+[169,170]. The plasmids expressing the L-arabinose isomerases were made in two steps from pWUR873 (Genbank: KX618638), which contained the low copy p15A origin of replication, the ampicillin marker (*bla* encoding β-lactamase) and the *gpf* gene under control of P_{T7} and a very strong RBS (AAGGAG; -14 to -9 relative to translation start). Firstly, P_{T7} was replaced by the weak to moderate constitutive promoter P_{bla} with KpnI/BcuI, giving pWUR832. The insert was formed by PCR with primers BG4591/BG4304 and pWUR873 as template. Secondly, the *gfp* CDS was replaced by the *araA* CDS from *E. coli* MG1655 or *G. thermodenitrificans* T12 with NdeI/BcuI, giving pWUR833 and pWUR834 respectively. The inserts were formed by PCR in two steps to remove the NdeI site from the CDS (silent mutation, CAT-->CAC). For *araA* of *E. coli*, left and right fragments were created with primers BG6723/BG6726 and BG6725/BG6724 respectively, and combined with primers BG6723/BG6724. For *araA* of *G. thermodenitrificans*, left and right fragments were created with primers BG7219/BG7222 and BG7221/BG7220 respectively, and combined with primers BG7219/BG7220. A negative control plasmid was formed by making the ends of NdeI/BcuI digested pWUR832 blunt with Klenow Fragment and ligating it, giving pWUR917. For verification of the plasmids and the use of enzymes and kits, see section 'Construction of regulator-reporter plasmids and control plasmids'. Dephosphorylation was done here with fastAP (Thermo Scientific). Strain AR was simultaneously transformed with pWUR768/pWUR833, pWUR768/pWUR834, pWUR768/pWUR917, pWUR780/pWUR833 or pWUR780/pWUR834.

The detection assays were performed as described above for the induction assays, except for a few things. Hundred micrograms per millilitre ampicillin was added to maintain the L-arabinose isomerase expressing plasmids. Instead of L-arabinose as inducer of AraC, L-ribulose was added as substrate for the L-arabinose isomerase. The L-ribulose concentration was varied. Cells were not pre-induced, but in the assay 15 μg mL^{-1} kanamycin was added after 1 h of growth to allow induction of *kan*. The bioluminescence values were corrected with the values obtained for the negative control with a frameshift in *luxA* (AR with pWUR780/pWUR833 or pWUR780/pWUR834). The kanamycin resistance assay and the bioluminescence assay were performed as two and three independent experiments respectively.

Enrichment for cells with L-arabinose isomerase activity

Each of the three strains, AR pWUR768 with pWUR833, pWUR834 or pWUR917, was grown separately in 13 mL LB medium with 4 g L^{-1} glycerol, 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol. After 24 h, cells were mixed based on the OD600 in a ratio of $1:1:10^8$ for pWUR833:pWUR834:pWUR917 and grown in 25 ml of the same medium with the addition of 15 µg mL^{-1} kanamycin as selective pressure and 5 mM L-ribulose as substrate for the L-arabinose isomerase. The controls were 1 ml cultures with L-ribulose and with/without kanamycin inoculated with each of the strains separately. After 6 h, dilution series were streaked on three types of LB agar plates with 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol, namely (1) without either L-ribulose or kanamycin, (2) with 15 µg mL^{-1} kanamycin and (3) with both 15 µg mL^{-1} kanamycin and 5 mM L-ribulose. Colonies were counted and 68 individual colonies, originating from the plates with kanamycin and L-ribulose that were inoculated with the mixed culture, were picked for the subsequent bioluminescence-based screen. White 96-well microplates (Thermo Scientific, Nunc) with 200 µl LB medium per well with 15 g L^{-1} agar, 100 µg mL^{-1} ampicillin, 34 µg mL^{-1} chloramphenicol and 0 or 0.5 mM L-ribulose, were inoculated with one colony per well. After 17 h growth, bioluminescence was detected with the lumiglo function of the G:BOX Chemi XT4 (Syngene). Biomass from these plates was used as template in several PCRs to show presence or absence of L-arabinose isomerase genes (primers BG3799/BG6225), identity of L-arabinose isomerase genes (primers BG7642/7643/7644) or occurrence of *araC-kan* CDS exchange (primers BG7009/4588/3652). PCRs were performed with OneTaq (NEB) and primers are presented in Table S4. The *araC-kan* CDS exchange was analysed by sequencing at GATC Biotech.

Acknowledgements

We would like to thank Kirill Datsenko for his advice on making the knockouts and we would like to thank Petra Dersch for providing us with plasmids.

Supplementary information

Table S1. Frameshifts in control plasmids[a]

Construct	Parent[b]	Expected[c]	Obtained[d]	Comments
pWUR778/80/86/88	CCG**CG**G	CCGG	CCGG	as expected[e]
(*luxA*)	G**GC**GCC	GGCC	GGCC	residues 1-160 ok, A161G, A162V, Y163Stop (parent: 361 codons)[f]
pWUR774 (*leuB*)	GACAT**TC**TGTC	GACATCTGTC	GACATTGTC	2 bp instead of 1 bp deletion (stop 5 codons earlier, also 252 codons ok, C253V, L254Stop)[e]
	CTGT**AA**GACAG	CTGTAGACAG	CTGTAACAG	residues 1-252 ok, L253V, S254Stop (parent: 364 codons)[f]
pWUR782 (*leuB*)	GACAT**TC**TGTC	GACATCTGTC	GACACTGTC	2 bp instead of 1 bp deletion (stop 5 codons earlier, 251 instead of 252 codons ok, I252T,
	CTGT**AA**GACAG	CTGTAGACAG	CTGTGACAG	C253V, L254Stop)[e]
				residues 1-251 ok, I252T, L253V, S254Stop (parent: 364 codons)[f]
pWUR776 (*kan*)	CCTG**AT**TCAGG	CCTGATTTCAGG	CCTGATCAGG	1 bp deletion instead of 1 bp insertion (stop 38 codons later, also 115 codons ok)[e]
	GGACT**AA**GTCC	GGACTAAAGTCC	GGACTAGTCC	residues 1-115 ok, S116Q, G117V, E118K, ……., R155E, M156Stop (parent: 272 codons)[f]
pWUR784 (*kan*)	CCTG**AT**TCAGG	CCTGATTTCAGG	CCTGATTTCAGG	as expected[e]
	GGACT**AA**GTCC	GGACTAAAGTCC	GGACTAAAGTCC	residues 1-115 ok, S116F, G117R, E118Stop (parent: 272 codons)[f]

[a]All upper strands are from 5' to 3' in the direction of the gene indicated in brackets. [b]The restriction site in the parent plasmid (Cfr42I, Eam1105I and XagI for *luxA*, *leuB* and *kan* respectively). The cut site is between the two bold nucleotides. [c]Sequence expected after Klenow treatment and ligation. [d]Sequence obtained after Klenow treatment and ligation. [e]Obtained versus expected. For the *luxA* gene, the obtained sequence of the frameshift site conformed to the expected fill in and removal of 5' and 3'overhangs respectively. Surprisingly, the sequence in the *kan* and *leuB* genes differed from this expectation, but this difference can be explained by the GC-content of the termini after digestion. Both *kan* and *leuB* genes had AT-rich termini, whereas *luxA* had GC-rich termini. AT-rich termini make partitioning of the strands easier than GC-rich termini, increasing the preference of the DNA to be bound in the exonuclease domain of Klenow[171]. Increased 3' to 5'exonuclease activity compared to 5' to 3' polymerase activity explained not only the deviation in the sequence, but it also explained the variation between different constructs made with the same restriction enzyme. The overhangs probably mismatched, but were ligated anyway and later repaired by *E. coli* using either one or the other strand as template, giving rise to two different sequences. Despite these differences between controls made with the same restriction enzyme, altered truncated products did not change the signal (growth or bioluminescence) of these negative controls (see characterization section in main text). [f]Obtained versus parent.

Table S2. Relative plasmid copy number of the reporter systems

Strain[a]	10^9 plasmids mL^{-1}[b]		Ratio medium/low[c]
	low copy	medium copy	
ALR + reg.-rep. plasmid (*araC*, *leuB*)	5.7 ± 0.6	24.7 ± 3.7	4.4 ± 0.8
AR + reg.-rep. plasmid (*araC*, *kan*)	5.5 ± 2.0	28.2 ± 1.6	5.1 ± 1.8
ALR + control plasmid (*araC*, *leuB*−)	-	26.0 ± 1.2	-
AR + control plasmid (*araC*, *kan*−)	-	26.4 ± 2.3	-
AR + control plasmid (*araC*, *kan, luxA*−)	-	24.2 ± 2.1	-
DH10B + reg.-rep. plasmid (*araC*, *kan*)	11.9 ± 4.6	55.8 ± 21.4	4.7 ± 2.6
DH10B + pACYC184	20.8 ± 5.0	-	-

[a]The systems vary in the selection reporter (LeuB or KmR) and the copy number of the regulator-reporter plasmid (ColE1 or p15A origins of replications for medium or low copy number respectively). The control plasmids have a frameshift in one of the reporter genes, indicated with a minus. The controls for plasmid isolation are another plasmid with the p15A origin of replication (pACYC184 [172]) and a control strain adapted for cloning (*E. coli* DH10B $T1^R$). reg.-rep. plasmid, regulator-reporter plasmid. [b]The number of plasmid molecules per millilitre culture at an OD600 of 1 was determined by plasmid isolation. All strains were grown overnight in 10 mL LB medium with the appropriate antibiotics. After OD600 measurement, plasmids were isolated with the Plasmid Miniprep kit of Thermo Scientific (#K0503). Plasmid concentration was measured with a ND-1000-Spectrophotometer (NanoDrop Technologies, Inc.). The total amount of isolated plasmid in micrograms was corrected for the amount of supernatant loaded on the column and divided by the OD600 and the culture volume. The negative control value (no plasmid) was subtracted. The number of plasmids per mL culture with OD600 = 1 was obtained by converting this total plasmid weight to the number of plasmid molecules (650 g mol^{-1} bp^{-1}). The data are an average of three independent experiments (standard deviation indicated). The ratio between frameshift control and parent plasmid was 1.0, confirming the expected similarity between the controls and their parent plasmids. pACYC184 is 2.5 times smaller than the regulator-reporter plasmids, but it also has the low copy origin of replication (p15A). pACYC184 had a ~2 higher copy number than the regulator-reporter plasmid with p15A. This inverse relation between plasmid size and copy number has also been described by others [173]. All tested plasmids had a two times higher copy number in the cloning strain *E. coli* DH10B $T1^R$ than in the *E. coli* BW25113 knockout strains. This twofold difference is probably due to the adaptation of *E. coli* DH10B $T1^R$ for cloning and thus for elevated plasmid yields.

Table S3. Sequences.

1. $P_{BAD\text{-}adapt}$ in front of *leuB*/*kan*. Underlined and italic nucleotides indicate restriction sites and operator (half) sites (ecogene.org, Invitrogen pBAD[174,175]) respectively. The sequence contains from 5' to 3' an XmaJI site, an O_2 operator half site (AraC), an O_1 operator half site (AraC), an O_1 operator half site (AraC), a randomized CRP binding site including an NheI site, an I_1 operator half site (AraC), an I_2 operator half site (AraC), the promoter -35 site, the promoter -10 site, and an MreI site.

CCTAGGGCCATTCAGAGAA*GAAACCAATTGTCCATA*TTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCG
CTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCG*TAACAA*
*AAGTGTCTATA*ATCA*CGGCAGAAAAGTCCACA*TTGATTA*ACCTAAGTCGAGATGGAAGCTAGCTCG*CA*TAGCATTTTTATCC*
*ATA*AGAT*TAGCGGATCCTACCTGA*CGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCGCCGGCG

2. $P_{BAD\text{-}adapt}$ in front of *luxCDABE*. Underlined and italic nucleotides indicate restriction sites and operator (half) sites (ecogene.org, Invitrogen pBAD[174,175]) respectively. The sequence contains from 5' to 3' an Acc65I site, an O_2 operator half site (AraC), an O_1 operator half site (AraC), an O_1 operator half site (AraC), a randomized CRP binding site including a PstI site, an I_1 operator half site (AraC), an I_2 operator half site (AraC), the promoter -35 site, the promoter -10 site, and a SalI site.

GGTACCGCCATTCAGAGAA*GAAACCAATTGTCCATA*TTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCG
CTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCG*TAACAA*
*AAGTGTCTATA*ATCA*CGGCAGAAAAGTCCACA*TTGATTA*ACCTAAGTCGAGATGGAACTGCAGTCG*CA*TAGCATTTTTATCC*
*ATA*AGAT*TAGCGGATCCTACCTGA*CGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGGTCGAC

3. Disruption cassette. Underlined and italic nucleotides indicate restriction sites and *lox* sites respectively. The cassette contains from 5'to 3' a SfiI site (used for cloning by GeneArt), a HindIII site, a primer annealing site, a PstI site, *lox*71, a BglII site, *kan*, a SalI site, *lox*66, a NotI site, a primer annealing site, an EcoRI site, and a SfiI site (used for cloning by GeneArt).

GGCCGTCAAGGCCGCATAAGCTTGGTGTCTTTTTTACCTGTTTGACCCTGCAG*TACCGTTCGTATAATGTATGCTATACGAA*
*GTTAT*AGATCTCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTC
AATAATATTGAAAAAGGAAGAGTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATG
GATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAA
GCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAA
ACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGACGCATGGTTACTCACCACTGC
GATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTT
CCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGACCGCGTATTTCGTCTCGCTCAGGCGCAATCA
CGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA
AATGCACAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGG
GGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGC
CTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTT
TCATTTGATGCTCGATGAGTTTTTCTAAGTCGAC*ATAACTTCGTATAATGTATGCTATACGAACGGTA*GCGGCCGCCAACTC
CTTCACCAGAGGTAGGAATTCCTGGGCCTCATGGGCC

Table S4. Primers used in this study

Primer	Annealing location	Sequence (5'-->3')[a]	Features
Construction of plasmids			
BG3691	start *kan*	GCGCGGGATCCCGCCGGCG*AGGAGG*ATACGTATGAGCCATATTCAACGGGAAAC	BamHI, MreI, RBS
BG3692	end *kan*	CGCGCCTCGAGCGATCGTTAGAAAAACTCATCGAGCATCAAATG	XhoI, PvuI
BG3693	start *leuB*	GCGCGGGATCCCGCCGGCG*AGGAGG*ATACGTATGTCGAAGAATTACCATATTGCCG	BamHI, MreI, RBS
BG3694	end *leuB*	CGCGCCTCGAGCGATCGTTACACCCCTTCTGCTACATAGC	XhoI, PvuI
BG3695	inside *leuB*	GAATCTGCTCGCAAGCGTCGCCACAAAGTGACG*AGT*ATCGATAAAGCCAAC	altered serine codon: TCG-->AGT
BG3696	inside *leuB*	GAGGATTGCAGCACGTTGGCTTTATCGAT*ACT*CGTCACTTTGTGGCGACG	altered serine codon: TCG-->AGT
BG3746	start P_{lacI}^{Q}	GCGCGCTCGAGGTTGACACCATCGAATGGTGCAAAACC	XhoI
BG3747	end P_{lacI}^{Q}	GCCGCCGATCGCGGWCCGATTCACCACCCTGAATTGACTCTCTTCC	PvuI, CpoI
BG3940	start *araC*	GCGCGCGGWCCG*AGGAGG*ATACGTATGGCTGAAGCGCAAAATGATC	CpoI, RBS
BG3941	end *araC*	CGCGCCGATCGTTATGACAACTTGACGGCTACATC	PvuI
BG3979	start P_{BAD}	GCCGGGGTACCGCCATTCAGAGAAGAAACC	KpnI
BG3980	end P_{BAD}	CCGGCGTCGACCCCAAAAAAACGGGTATGG	SalI
BG3981	end P_{BAD}	CTCTTCGCCGGCGCCCAAAAAAACGGGTATGG	MreI
BG3982	inside P_{BAD}	*CCTAAGTCGAGATGGAACTGCAGTCG*CATAGCATTTTTATCCATAAGATTAGC	altered CRP binding site, PstI
BG3983	inside P_{BAD}	CGA*CTGCAGTTCCATCTCGACTTAGGT*TAATCAATGTGGACTTTTCTGC	altered CRP binding site, PstI
BG4229	inside P_{BAD_adapt}	G*CGAGCTAGCTTCCATCTCGACTTAGGT*TAATC	NheI, altered CRP binding site
BG4230	inside P_{BAD_adapt}	*GGAAGCTAGCTCG*CATAGCATTTTTATCCATAAG	part altered CRP binding site, NheI
BG4231	start ColE1	TACTGGTACCCATGACCAAAATCCCTTAACGTG	Acc65I
BG4232	end ColE1	TACTCGTACGCCTAGGCGTTCGGCTGC	Pfl23II
BG4304	end *gfp*	TACTACTAGTTTATTTGTAGAGCTCATCCATGCCATGTG	BcuI
BG4368	start P_{BAD}	CTACTCCTAGGGCCATTCAGAGAAGAAACC	XmaJI
BG4591	start *gfp*	TACTGGTACCCCGCTTCGGCGGGGTTTTTTCAAGTTCAAATATGTATCCGCTCATGAGACAATGTGTGGGGA GACCACAACGGTTTCC	KpnI
BG4666	start p15A	TATGTGGTACCTAGCGGAGTGTATACTGGCTTAC	Acc65I
BG4667	end p15A	TACAACCTAGGACAACTTATATCGTATGGGGCTG	XmaJI

Table S4 continued

Primer	Annealing location	Sequence (5'-->3')	Features
Construction of plasmids			
BG6723	start *araA*[b]	CGCGCCATATGACGATTTTTGATAATTATGAAGTGTGG	NdeI
BG6724	end *araA*[b]	GCGCGACTAGTTTAGCGACGAAACCCG	BcuI
BG6725	inside *araA*[b]	GGTGCTCGGCTCCC*CAC*ATGCTGGAAGTCTGC	altered histidine codon: CAT-->CAC
BG6726	inside *araA*[b]	GCAGACTTCCAGCAT*GTG*GGAGCCGAGCACC	altered histidine codon: CAT-->CAC
BG7219	start *araA*[c]	CGCGCCATATGTTATCATTACGTCCTTATGAATTTTGG	NdeI
BG7220	end *araA*[c]	GCGCGACTAGTTTACCTCCCTCGCCAAAATAC	BcuI
BG7221	inside *araA*[c]	GATTCTTGGCGCT*CAC*ATGCTCGAAGTATGC	altered histidine codon: CAT-->CAC
BG7222	inside *araA*[c]	GCATACTTCGAGCAT*GTG*AGCGCCAAGAATC	altered histidine codon: CAT-->CAC
Verification of plasmids			
BG3336	inside ColE1	TTCGCCACCTCTGACTTG	
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3653	inside *kan*	GCCTGTTGAACAAGTCTGGA	
BG3799	inside p15A	CAGAGCAAGAGATTACGCGCAGACC	
BG3857	inside *luxE*	GAAGCGTTTGATAGTTGAGCGG	
BG3858	inside *cat*	CAGGTTCATCATGCCGTCTG	
BG3942	upstream *cat*	CAACGTCTCATTTTCGCCAG	
BG3943	inside *luxC*	CACGAATGTATGTCCTGCG	
BG3977	inside *leuB*	GCACAAATCCTTTCGCTGG	
BG3978	inside *leuB*	GTAATGGCTGGTGGTGATG	
BG4231	start ColE1	TACTGGTACCCATGACCAAAATCCCTTAACGTG	Acc65I
BG4232	end ColE1	TACTCGTACGCCTAGGCGTTCGGCTGC	Pfl23II
BG4627	inside *leuB*	GTATTCCGTGGCGATCTC	
BG4628	inside *leuB*	CGGCATCTATTTCGGTCAG	
BG4629	inside *kan*	CCAGACTTGTTCAACAGGC	
BG4630	inside *kan*	CTCCTGATGACGCATGG	

Table S4 continued

Primer	Annealing location	Sequence (5'-->3')	Features
Verification of plasmids			
BG4630	inside *kan*	CTCCTGATGACGCATGG	
BG4631	inside *luxA*	GAATGGCATGACAGAGGG	
BG4632	inside *luxA*	GTGCCCATATTCTTGAGCC	
	inside T7		
BG6225	terminator	CCTCAAGACCCGTTTAGAGG	
Construction of knockouts			
BG3649	start cassette[d]	*GCTCAACACAACGAAAACAACAAGGAAACCGTGTGA*GGTGTCTTTTTTACCTGTTTGACC	homologous to genome (upstream/start *leuB)*
BG3650	end cassette[d]	*ACGTCTTAGCCATGATTACACCCCTTCTGCTACATA*CTACCTCTGGTGAAGGAGTTG	homologous to genome (end/downstream *leuB*)
BG4490	start cassette[d]	*CAGAACATATTGACTATCCGGTATTACCCGGCATGACAGGAGTAAAAATG*GGTGTCTTTTTTACCTGTTTGACC	homologous to genome (upstream/start *recA*)
BG4491	end cassette[d]	*ATGCGACCCTTGTGTATCAAACAAGACGATTAAAAATCTTCGTTAGTTTC*CTACCTCTGGTGAAGGAGTTG	homologous to genome (end/downstream *recA*)
Verification of knockouts			
BG3651	upstream *leuB*	CAGGTGGATATCGTCGCTAA	
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3653	inside *kan*	GCCTGTTGAACAAGTCTGGA	
BG3654	downstream *leuB*	AACAGTGGGGTTTCGTTTTC	
BG3655	upstream *araC*	GGTTGGGTTAGCGAGAAGAG	
BG3656	downstream *araC*	GGGTAGAATCAAACCGACCA	
BG4190	upstream *recA*	CGTCAGGCTACTGCGTATG	
BG4191	downstream *recA*	GAATACGCGCAGGTCCATAAC	

Table S4 continued

Primer	Annealing location	Sequence (5'-->3')	Features
Analysis of enriched colonies			
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3799	inside p15A	CAGAGCAAGAGATTACGCGCAGACC	
BG4588	inside *araC*	TACTGACAAGCCTCGCGTACCC	
BG6225	inside T7 terminator	CCTCAAGACCCGTTTAGAGG	
BG7009	inside CmR	GGTTATAGGTACATTGAGCAACTG	
BG7642	inside *araA*[b,c]	CCGTGGGACAGCATCGATATGG	
BG7643	inside *araA*[b]	GCCCTGCAGACCGGTTG	
BG7644	inside *araA*[c]	GCTTTGTTCAGAAACATCGCGAATAG	

[a]Underlined nucleotides indicate restriction sites, italic nucleotides indicate other features (for both see last column). [b]*araA* of *E. coli*. [c]*araA* of *G. thermodenitrificans*. [d]Recombination/disruption cassette.

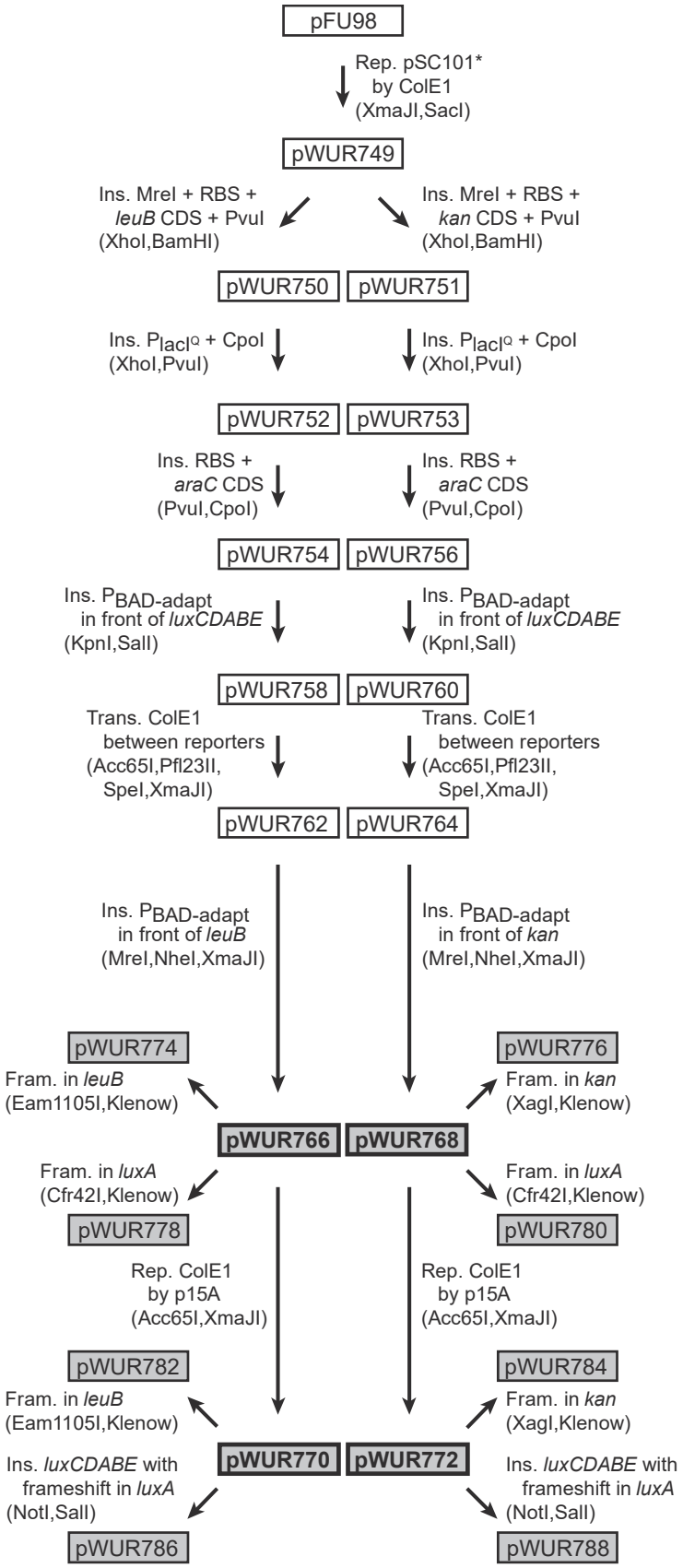

pFU98
Rep. pSC101* by ColE1 (XmaJI,SacI)
pWUR749
Ins. MreI + RBS + leuB CDS + PvuI (XhoI,BamHI)
Ins. MreI + RBS + kan CDS + PvuI (XhoI,BamHI)
pWUR750
pWUR751
Ins. PlacIQ + CpoI (XhoI,PvuI)
Ins. PlacIQ + CpoI (XhoI,PvuI)
pWUR752
pWUR753
Ins. RBS + araC CDS (PvuI,CpoI)
Ins. RBS + araC CDS (PvuI,CpoI)
pWUR754
pWUR756
Ins. PBAD-adapt in front of luxCDABE (KpnI,SalI)
Ins. PBAD-adapt in front of luxCDABE (KpnI,SalI)
pWUR758
pWUR760
Trans. ColE1 between reporters (Acc65I,Pfl23II, SpeI,XmaJI)
Trans. ColE1 between reporters (Acc65I,Pfl23II, SpeI,XmaJI)
pWUR762
pWUR764
Ins. PBAD-adapt in front of leuB (MreI,NheI,XmaJI)
Ins. PBAD-adapt in front of kan (MreI,NheI,XmaJI)
pWUR774
Fram. in leuB (Eam1105I,Klenow)
pWUR776
Fram. in kan (XagI,Klenow)
pWUR766
pWUR768
Fram. in luxA (Cfr42I,Klenow)
Fram. in luxA (Cfr42I,Klenow)
pWUR778
pWUR780
Rep. ColE1 by p15A (Acc65I,XmaJI)
Rep. ColE1 by p15A (Acc65I,XmaJI)
pWUR782
Fram. in leuB (Eam1105I,Klenow)
pWUR784
Fram. in kan (XagI,Klenow)
pWUR770
pWUR772
Ins. luxCDABE with frameshift in luxA (NotI,SalI)
Ins. luxCDABE with frameshift in luxA (NotI,SalI)
pWUR786
pWUR788

Fig. S1. Overview cloning steps. The four regulator-reporter plasmids (depicted in bold with a grey background) were constructed from pFU98[162] in eight subsequent cloning steps. From each of the four plasmids two control plasmids were made (depicted with a grey background) by making a frameshift either in the selection reporter gene (*leuB*/*kan*) or in one gene in the screening reporter operon (*luxA*). For each cloning step a description and the restriction enzymes are included. Rep., replacement; Ins., insertion; Trans., translocation; Fram., introduction frameshift. The origin and formation of the inserts are described here. BG numbers refer to primers (Table S4). pWUR749: digest from pFU168[162]. pWUR750: two step PCR from BW25113 genome to remove AatII and PvuI sites in CDS (first left BG3693/BG3696, first right BG3695/BG3694, second BG3693/BG3694). pWUR751: PCR from recombination cassette[165] (BG3691/BG3692). pWUR752/3: PCR from pET24d (Novagen, BG3746/BG3747). pWUR754/6: PCR from pBAD-TOPO (Invitrogen, BG3940/BG3941). pWUR758/60: two step PCR from pBAD-TOPO with adaptations (first left BG3979/BG3983, first right BG3982/BG3980, second BG3979/BG3980). pWUR762/4: PCR from pWUR758 (BG4232/BG4231). pWUR766/8: two PCRs from pWUR758 with adaptations (BG4368/BG4229, BG4230/BG3981). pWUR770/2: PCR from pACYC184[172] (BG4666/BG4667). pWUR786/8: digest from pWUR778.

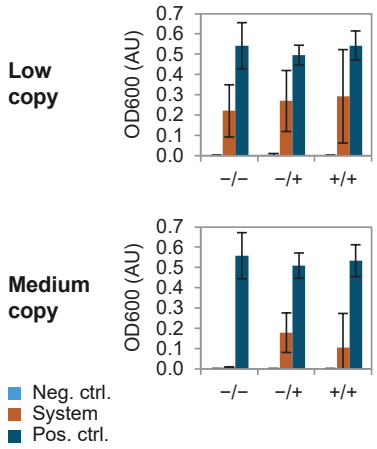


Fig. S2. Effect of the addition of inducer to the preculture on the leucine auxotrophy complementation assay. The plasmid-encoded reporter gene *leuB* was induced or non-induced in the low and medium copy systems with 10 mM (no full induction) of the inducer L-arabinose. The cultures were inoculated from induced or non-induced precultures to see the effect of this treatment. '−/−' non-induced in both precultures and assay cultures, '−/+' induced only in assay cultures, '+/+' induced in both precultures and assay cultures. Bacteria were grown in M9 medium for 32 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*araC* Δ*leuB* Δ*recA* (ALR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph ALR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid with a frameshift in *leuB*.

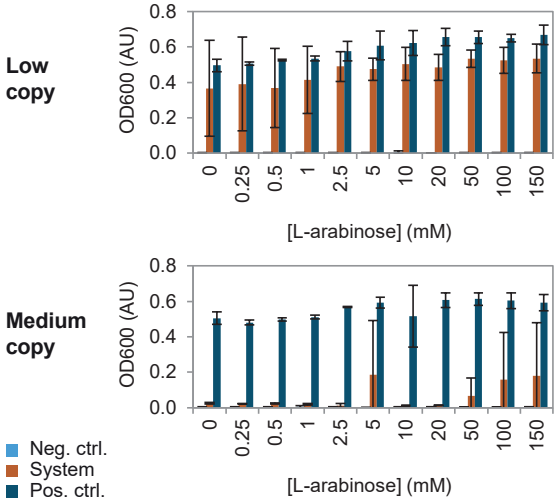


Fig. S3. Selection based on leucine auxotrophy complementation (48 h). The plasmid-encoded reporter gene *leuB* was induced in the low and medium copy systems with the inducer L-arabinose. Bacteria were grown in M9 medium for 48 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*araC* Δ*leuB* Δ*recA* (ALR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph ALR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid with a frameshift in *leuB*.

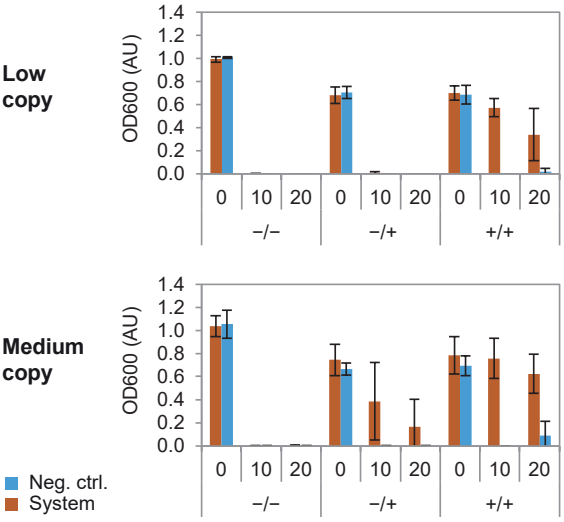


Fig. S4. Effect of the addition of inducer to the preculture on the kanamycin resistance assay. The plasmid-encoded reporter gene *kan* was induced or non-induced in the low and medium copy systems with 10 mM (no full induction) of the inducer L-arabinose. The cultures were inoculated from induced or non-induced precultures to see the effect of this treatment. '−/−'non-induced in both precultures and assay cultures, '−/+' induced only in assay cultures, '+/+' induced in both precultures and assay cultures. Bacteria were grown in LB medium for 17 h in the presence of 0, 10 or 20 μg mL^{-1} kanamycin. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid. Neg. ctrl.: AR with the regulator-reporter plasmid with a frameshift in *kan*.

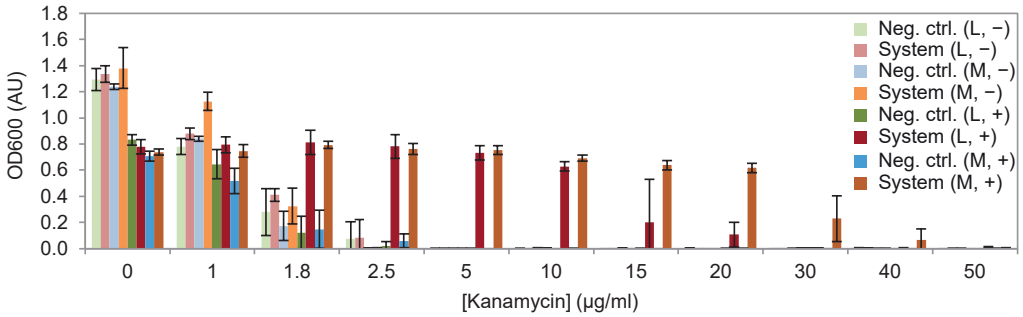


Fig. S5. Kanamycin death curve. The plasmid-encoded reporter gene *kan* was induced '+' or non-induced '–' in the low 'L' and medium 'M' copy systems with 10 mM (no full induction) of the inducer L-arabinose. Bacteria were grown in LB medium for 17 h in presence of different kanamycin concentrations. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid. Neg. ctrl.: AR with the regulator-reporter plasmid with a frameshift in *kan*.

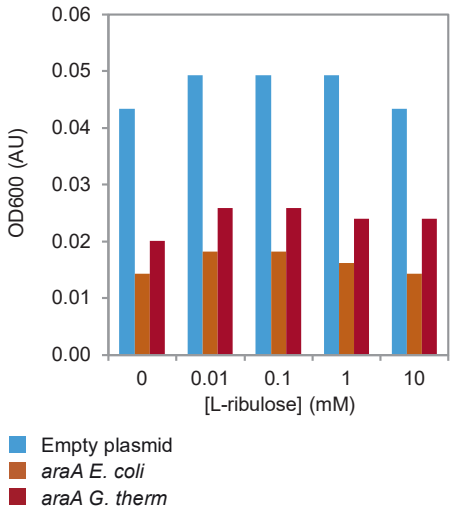


Fig. S6. Growth of cells with or without L-arabinose isomerase (*araA*). Bacteria were grown in LB medium for 5.5 h in presence of a range of L-ribulose (substrate) concentrations. *araA E. coli*: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid and the plasmid expressing *araA* of *E. coli*. *araA G. therm*: AR with the regulator-reporter plasmid and the plasmid expressing *araA* of *G. thermodenitrificans*. Empty plasmid: AR with the regulator-reporter plasmid and the empty plasmid.

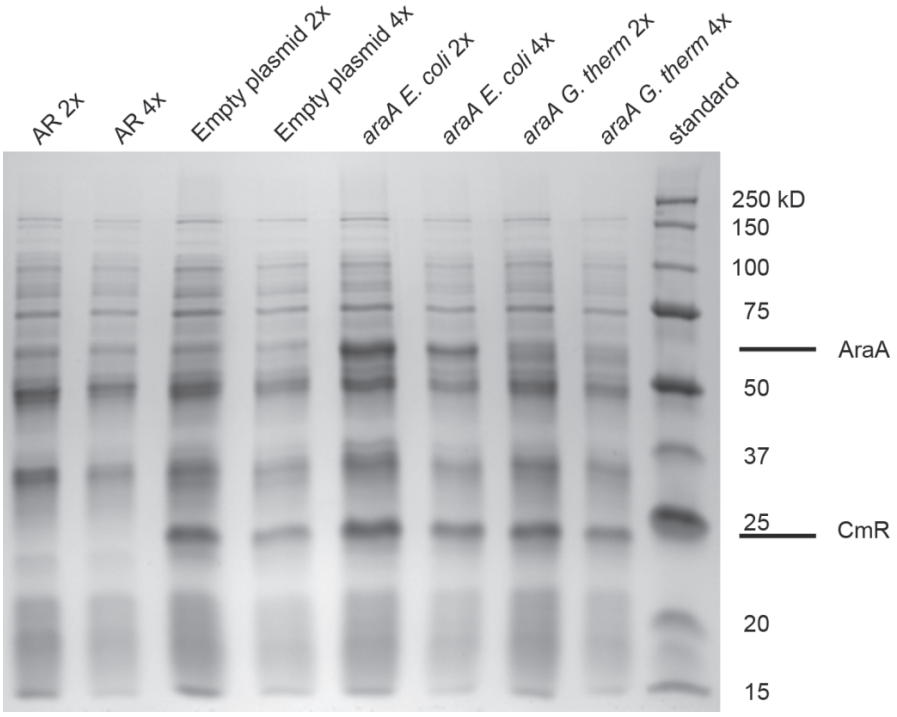


Fig. S7. L-arabinose isomerase (AraA) expression analysis. Three mL LB medium with 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol were inoculated with one of the four strains (strain without plasmids was grown without antibiotics). After 19 h growth, an equivalent of 1 mL of cells of OD600 = 0.8 was centrifuged. The pellet was resuspended in 1x Laemmli Sample Buffer (Biorad) including beta-mercaptoethanol. Two and four times diluted samples were boiled for 15 min at 98°C and, after centrifugation, 10 µl per sample and 10 µl of Precision Plus Protein Unstained Standard (Biorad) were loaded on a 10% SDS-PAGE Mini-PROTEAN TGX Precast Protein Gel (Biorad). The gel was run at 20 mA in 25 mM Tris, 192 mM glycine and 0.1% SDS. Proteins were fixed with 25% isopropanol and 10% acetic acid for 15 minutes and stained with QC Colloidal Coommassie Stain (Biorad) overnight. The gel was destained with milliQ for 3 h. *araA E. coli*: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid (pWUR768) and the plasmid expressing *araA* of *E. coli* (pWUR833). *araA G. therm*: AR with the regulator-reporter plasmid (pWUR768) and the plasmid expressing *araA* of *G. thermodenitrificans* (pWUR834). Empty plasmid: AR with the regulator-reporter plasmid (pWUR768) and the empty plasmid (pWUR917). AR: *E. coli* BW25113 Δ*araC* Δ*recA* without plasmids. 2x and 4x indicate the dilution and CmR indicates the chloramphenicol resistance marker on the regulator-reporter plasmid.

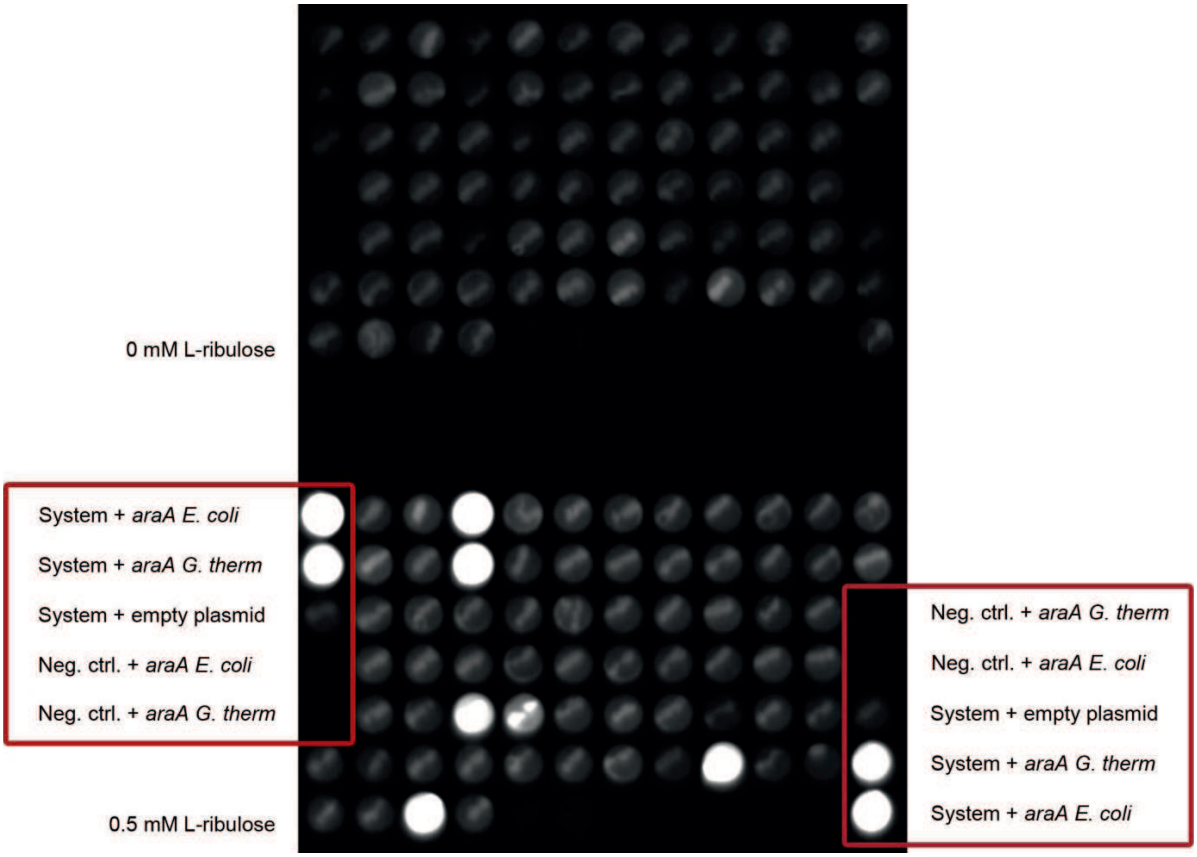
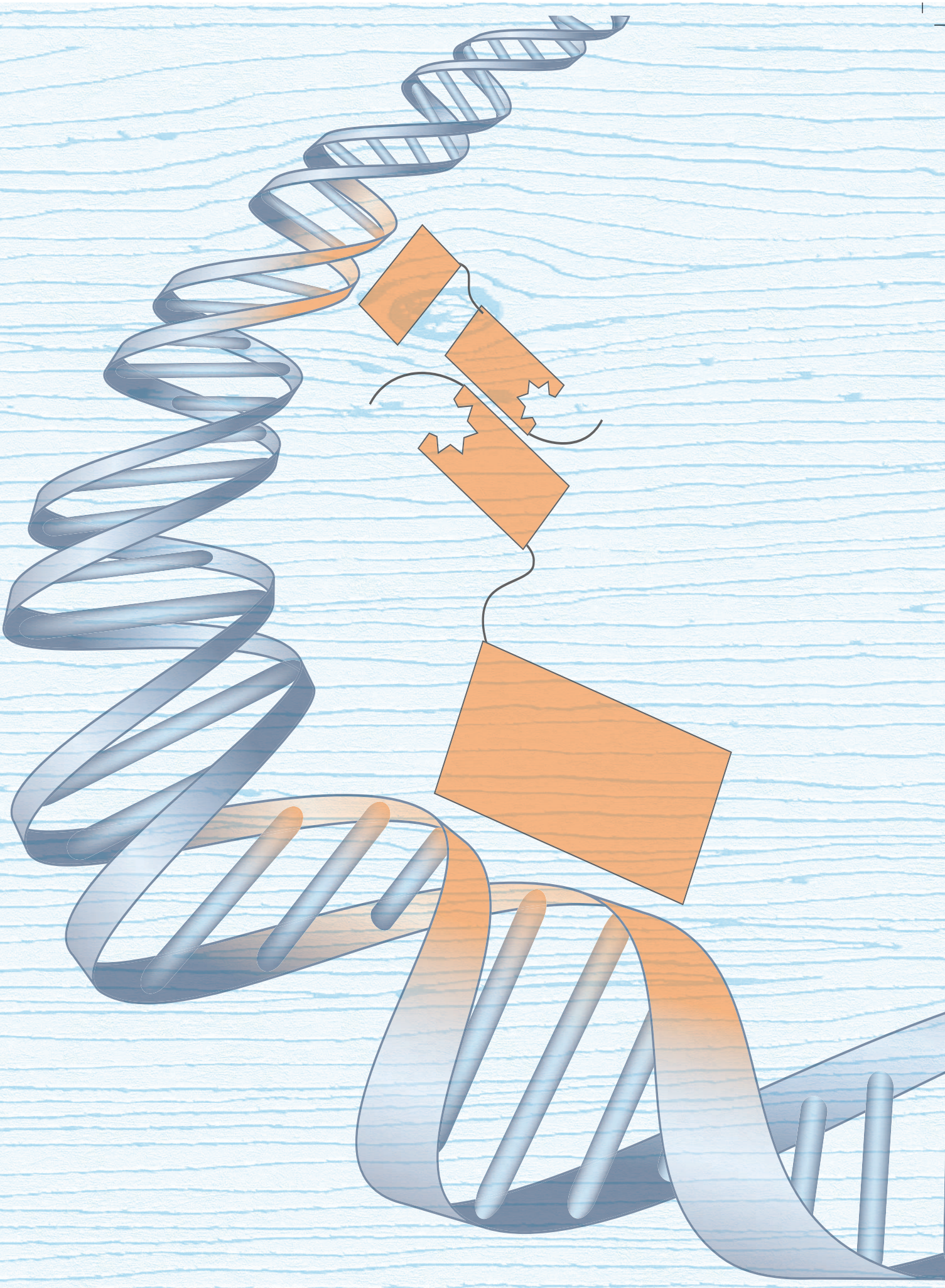


Fig. S8 Bioluminescence-based screening after enrichment of L-arabinose isomerase (*araA*) containing cells. Sixty eight colonies of the enrichment were tested for L-ribulose dependent bioluminescence to confirm presence of *araA*. Conversion of L-ribulose to L-arabinose by the L-arabinose isomerase of *E. coli* or *G. thermodenitrificans* induced the system. Bacteria were grown in a 96-well plate on LB medium with agar with (bottom) or without (top) 0.5 mM L-ribulose for 17 h. The five controls were included twice. System + *araA E. coli*: *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid and the plasmid expressing *araA* of *E. coli*. System + *araA G. therm*: AR with the regulator-reporter plasmid and the plasmid expressing *araA* of *G. thermodenitrificans*. System + empty plasmid: AR with the regulator-reporter plasmid and the empty plasmid. Neg. ctrl. + *araA E. coli*: AR with the regulator-reporter plasmid with a frameshift in *luxA* and the plasmid expressing *araA* of *E. coli*. Neg. ctrl. + *araA G. therm*: AR with the regulator-reporter plasmid with a frameshift in *luxA* and the plasmid expressing *araA* of *G. thermodenitrificans*.

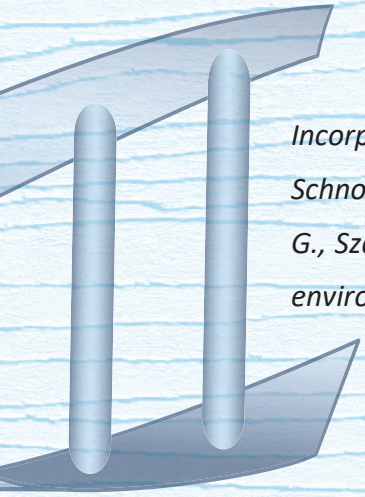


Chapter 4

Modifying the sensor part of a dual bioreporter to broaden its target range

Teunke van Rossum[a], Aleksandra Muras[a], Lisa Röttjers[a], John van der Oost[a], Servé W.M. Kengen[a]

[a]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands



Incorporated in submitted manuscript: Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., van Rossum, T., Siebers, B., Menzel, P., Kublanov, I.V., Rike, A.G., Skretas, G., Szabo, Z., Peng, X., and Young, M.J. (2018) Discovery of novel hydrolases from hot environments. Biotechnology advances.

Abstract

Whole-cell bioreporters are interesting tools for the detection of novel biocatalysts, but their application as high-throughput screens is often still laborious and/or expensive. To simplify detection of novel biocatalysts, a reporter system has previously been developed, based on selection instead of screening. In this approach, only positive cells are selected through the coupling of enzymatic product formation to cell survival. The sensor in this system, the arabinose specific transcriptional regulator AraC, controls the expression of two reporters. The selection reporter KmR allows for rapid reduction of the initially large library size based on growth, whereas the screening reporter LuxCDABE enables exclusion of false positives and quantification of positive variants based on bioluminescence. However, to enlarge the number of target molecules that this system is able to detect, it should be modified to function with other sensor parts, that is, with other transcriptional regulators than AraC. Here, the AraC-based dual selection/screening system was modified to function with LacI as transcriptional regulator. Different versions of the LacI-based system, varying in plasmid copy number (low or medium) and selection reporter (LeuB or KmR), were compared. The best performing system, the low copy system with LeuB as selection reporter, was specific for isopropyl β-D-1-thiogalactopyranoside (IPTG) to reach full induction. In addition, it allowed for detection of previously described weak inducers or anti-inducers. It is concluded that, although the system's characteristics differed from the AraC-based system, the dual reporter system can be modified with different sensor parts, thereby broadening its range of potential target molecules.

Introduction

A whole-cell bioreporter (from here on called bioreporter) is a living microorganism containing a sensor molecule that upon binding of a small molecule of interest switches on a reporter, resulting in a detectable phenotype[116]. These bioreporters are useful tools in a range of applications, such as the detection of pollutants, obtaining novel biocatalysts or strain development[125,176,177]. The main reason for the current interest in bioreporters, is the possibility to use them in high-throughput screening approaches[125,178]. Such efficient methods are required to screen large libraries of variants in projects aiming for enzyme discovery (metagenome libraries), enzyme optimization (gene variant libraries) or production strain development (strain variant libraries). Although enzymatic product formation generally does not provide a growth benefit to the host or production strain, bioreporters with appropriate specificity can couple product formation to growth via a selection reporter. Bioreporters can also dynamically control pathways by fine-tuning metabolic fluxes[177-179]. Another interesting feature is that the reporter can be chosen based on the application[125]. More information about bioreporters can be found in several reviews[125,176-179].

The sensor part of the bioreporter can consist of various kinds of biomolecules, either protein or RNA (riboswitches). A wide range of protein-based systems have been described of which sensor parts include enzymes, transcriptional regulators, extracytoplasmic function (ECF) sigma factors, two-component systems, Periplasmic Binding Proteins (PBPs), or fluorophore-containing proteins (e.g. proteins that allow for Fluorescence Resonance Energy Transfer (FRET) or for Fluorescence Protein Exchange (FBX))[125,177-179]. Bioreporters that continuously express the sensor and the reporter parts have the advantage of a reduced signal response time. It is important to realize that natural sensor parts are evolved to function optimally in a natural setting, generally to improve metabolic efficiency and to enhance host fitness. Hence, most of these sensors do not operate perfectly for unnatural, human-invented applications[177,179], often requiring optimization of sensor parts or design of new sensor parts, which is challenging and time-consuming[125,179]. The design of transcriptional regulator- or riboswitch-based bioreporters is, currently, a bit easier compared to that of other types. Compared to riboswitch-based sensors, the output of transcriptional regulator-based sensors has a higher fold change upon addition of the small molecule and they are applicable for a more diverse range of small molecules due to their more diverse chemistry, consisting of amino acids versus nucleotides[54,179]. Therefore

transcriptional regulator-based bioreporters are employed here.

Transcriptional regulator-based bioreporters are developed for various types of small molecules, including metals[180], amino acids[114], organic acids[87], phenolic compounds[110]. In most cases, the small molecules are very small (< 200 Da), although there are a few examples with slightly bigger molecules such as quercetin (302.24 Da)[115] or NADPH (744.42 Da)[112]. The transcriptional regulators used in bioreporters are often activators, possibly because this was the only available regulator for a certain target molecule or because a repressor that represses upon small molecule binding gives an undesired negative signal. However, such negative signal could be circumvented by letting the repressor control a second repressor that in turn represses the reporter[177,178] or by using a repressor that gives a positive signal by releasing repression upon small molecule binding. In the majority of developed bioreporters, the transcriptional regulators are natural regulators, often originating from soil bacteria like *Pseudomonas putida* or *Corynebacterium glutamicum*. The soil is a complex habitat in which bacteria need to respond to a wide range of signals, requiring a diverse set of regulators[181]. Another often exploited source of transcriptional regulators is the model organism *Escherichia coli*. Not for all small molecules a natural transcriptional regulator is available that has the desired characteristics for a bioreporter, like a dynamic range of a few orders of magnitude. Therefore, quite some bioreporters are based on transcriptional regulators that have been engineered via directed evolution[108,129] or, most recently, via computational design[182,183].

Despite the great interest in bioreporters, the application as high-throughput screens is often still laborious and/or expensive. In order to simplify novel biocatalyst detection, previously, a selection-based reporter system was developed (**Chapter 3**[184]). In this system, the transcriptional regulator AraC is the sensor part that binds to a small molecule of interest, resulting in a conformational change of the regulator, which alters its DNA binding capacity. This allows for expression of two divergently transcribed reporter genes, namely a growth-enabling selection reporter (KmR) and a bioluminescence producing screening reporter (LuxCDABE). These double reporters are the strength of the system, because in the selection step only *E. coli* cells that contain the small molecule or enzymatic product should survive, allowing for a rapid reduction of the initially large metagenomic or mutant library size. The subsequent screening step should exclude false positives, thereby tackling the large false positive rate often encountered for growth-based selection, and makes quantification of positive variants possible. This indeed has been demonstrated to function well, in case of enriching for L-arabinose producing cells (**Chapter 3**[184]). However, the use of AraC limits the

range of potential target molecules, and therefore we want to extend the applicability of the dual bioreporter system by exchanging the sensor part with another transcriptional regulator.

The aim of this study was to show the adaptability of the previously developed AraC-based dual selection/screening system to another transcriptional regulator. LacI, the repressor of lactose metabolism in *E. coli*, was chosen as alternative regulator, because it is well-studied and applied, its crystal structure is available and it has been a subject for protein engineering. Different versions of the LacI-based system, varying in plasmid copy number and selection reporter, were compared in induction assays. The best performing system, the low copy system with LeuB as selection reporter, was specific for isopropyl β-D-1-thiogalactopyranoside (IPTG) to reach full induction. In addition, this system was able to detect previously described weak inducers or anti-inducers.

Results and Discussion

Construction of the system

The lac-repressor (LacI) of *Escherichia coli* down-regulates the expression of genes involved in lactose metabolism in the absence of lactose. LacI forms a tetramer via its C-terminal domain and each dimer of this tetramer binds to one of three operator sites (-82, +11, +412 relative to transcription start), thereby repressing the *lacZYA* operon by blocking the site for RNA polymerase binding and by forming a DNA loop that captures negative supercoils. Although a LacI dimer cannot form a loop, it can repress. When lactose is present, it is hydrolyzed to glucose and galactose by a low background level of the β-galactosidase LacZ, producing allolactose as side product via transglycosylation. Upon binding of allolactose in between the N- and C-terminal subdomains of the core domain of LacI, the N-terminal subdomain changes its conformation relative to the C-terminal subdomain. As a result, the N-terminal DNA-binding domain is released from the DNA and the *lacZYA* operon is transcribed, initiating lactose import and metabolism[185-189]. The lactose operon is also activated by the global regulator CRP (cAMP receptor protein) in response to low glucose levels[132,185,186]. A major advantage of LacI is that it has not only been studied in depth, but it has also been engineered in various aspects, for example to reduce leakiness[190], to tighten regulation[191], to repress in presence instead of absence of IPTG[192,193], or to alter the ligand specificity[194-197]. And most importantly for engineering purposes, high resolution protein structures with and without ligand are available. In addition, LacI is applied for a wide range

of purposes, like regulating protein expression[198], metabolic engineering[199], regulatory circuit design[200] and purification of selected DNAs[201]. Often studies use the gratuitous inducer isopropyl β-D-1-thiogalactopyranoside (IPTG). More information about LacI can be found in the reviews by Lewis and by Davey and Wilson[185,187].

The LacI-based system in this study consisted of the same components as the previously developed AraC-based system (**Chapter 3**[184]), namely a host strain (*E. coli* BW25113 derivatives) and a regulator-reporter plasmid, encoding the transcriptional regulator and both reporters (Fig. 1). The only difference was the identity of the transcriptional regulator, meaning that (1) the host strain had a deletion of the chromosomal *lacI* instead of *araC*, (2) the regulator-reporter plasmid encoded *lacI* instead of *araC*, and (3) the reporters were divergently transcribed from LacI-responsive instead of AraC-responsive promoters. To allow for a good comparison of the AraC- and LacI-based systems, it was decided not to implement additional optimizations, like improving the selection potential of the system by using double positive selection or a combination of positive and negative selection.

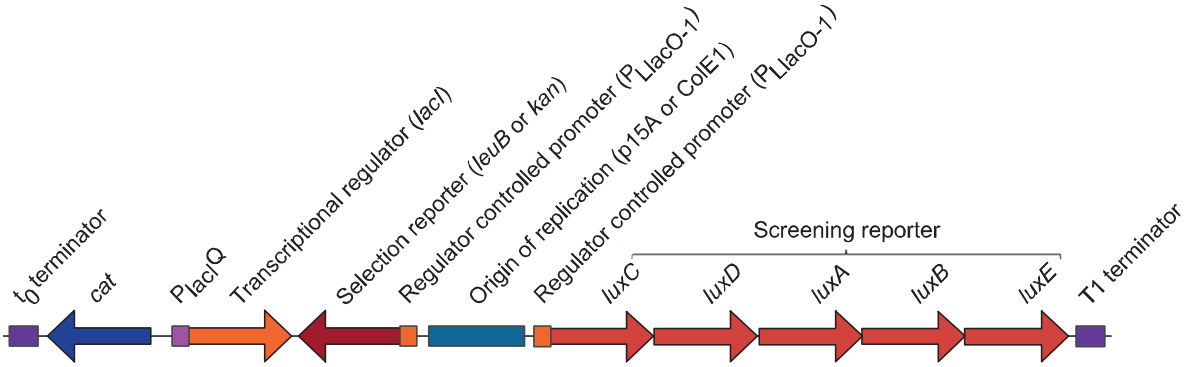


Fig. 1. Linear representation of regulator-reporter plasmid. Different versions of the plasmid vary in the selection reporter (*leuB* or *kan*) and the copy number of the regulator-reporter plasmid (ColE1 or p15A origins of replications for medium or low copy number respectively). The t0 terminator blocks read-through transcription coming from the selection reporter or the chloramphenicol resistance marker (*cat*), whereas the T1 terminator blocks read-through transcription from the screening reporter *luxCDABE*. P_{lacI}^{Q} is a moderate constitutive promoter. $P_{LlacO-1}$ is regulated by LacI. The figure is adapted from Fig. 1 in **Chapter 3**[184].

For the LacI-based system, four versions were constructed to be able to select the best one, as was done for the AraC-based system. These four versions varied in the type of selection reporter and the plasmid copy number. The selection reporters were KmR for kanamycin resistance and LeuB for leucine auxotrophy complementation. The plasmids had replication origins that result in either low (p15A) or medium (ColE1) copy number. As screening reporter genes, all versions had the *luxCDABE* operon, encoding both the luciferase

responsible for bioluminescence and the substrate generating enzymes. $P_{LlacO-1}$ was chosen as LacI-responsive promoter, because it is strong and no longer has the CRP binding site (Table S1)[141]. In principle, now only LacI regulates expression of the reporters. The control plasmids had a designed frameshift (Table S2) either in the selection reporter gene (*kan*/*leuB*) or in one of the screening reporter genes (*luxA*).

The parent of the host strain for the regulator-reporter and control plasmids was *E. coli* BW25113[137]. It has several tandem copies of the *rrnB* terminator inserted in the *lacZ* promoter region[138] and is therefore unable to metabolize lactose. Also its ability to take up lactose/IPTG is reduced[202,203]. In this study, deletions were made of the *lacI*, *leuB* and *recA* genes, to exclude interference from the endogenous regulator, to enable leucine auxotrophy complementation, and to prevent recombination events involving the plasmids respectively. Genes were replaced by a kanamycin resistance marker with Red recombinase and this marker was later removed with FLP or Cre recombinase. The two constructed knockout strains Δ*lacI* Δ*recA* and Δ*lacI* Δ*leuB* Δ*recA* are from now on referred to as LR and LLR respectively. For more details on the system components and their construction see the Materials and methods section, Fig. S1, and **Chapter 3**[184].

The relative plasmid copy numbers in LR and LLR were determined, to show that the medium/low copy number ratio of the regulator-reporter plasmids resembled the reported values of the parent plasmids, the pZ expression vectors. The pZ vectors with either p15A or ColE1 replication origins, had copy numbers of 20-30 and 50-70 respectively[141]. The ratio medium/low copy of the regulator-reporter plasmids was 2-3 (Table S3), which is in good agreement with the described ratios. The previous AraC-based study revealed that the frameshifts in the control plasmids did not influence the copy number, but the large size of the plasmids and the use of a strain not optimized for cloning or expression both reduced the copy number by a factor two compared to the control (**Chapter 3**[184]).

Characterization of the selection (LeuB and KmR) and the screening (LuxCDABE) reporters

All four LacI-based systems were characterized in order to select the best version in both selection and screening and to compare this with the best AraC-based system. A system is qualified as a good system if it has a low leakiness, a high maximal signal, a broad dynamic range and a high sensitivity. However, the relative importance of each of these criteria varies between the selection and the screening step of the system. In the selection step, a high sensitivity and low leakiness are the most important criteria in order to detect even low concentrations of the small molecule of interest without many false positives. Every cell that

survives is interesting and will subsequently be quantified in the screening step, in which all four characteristics are of importance, especially a high sensitivity and a broad dynamic range (**Chapter 3**[184]). To determine these characteristics induction assays were performed in which the systems were induced with a range of IPTG concentrations. As reporter activity or output signal, the optical density (OD600) and/or the bioluminescence were measured. The next paragraphs subsequently describe the selection assay based on leucine auxotrophy complementation by LeuB (Fig. 2), the selection assay based on kanamycin resistance by KmR (Fig. 3), and the screening assay based on bioluminescence by LuxCDABE (Fig. 4).

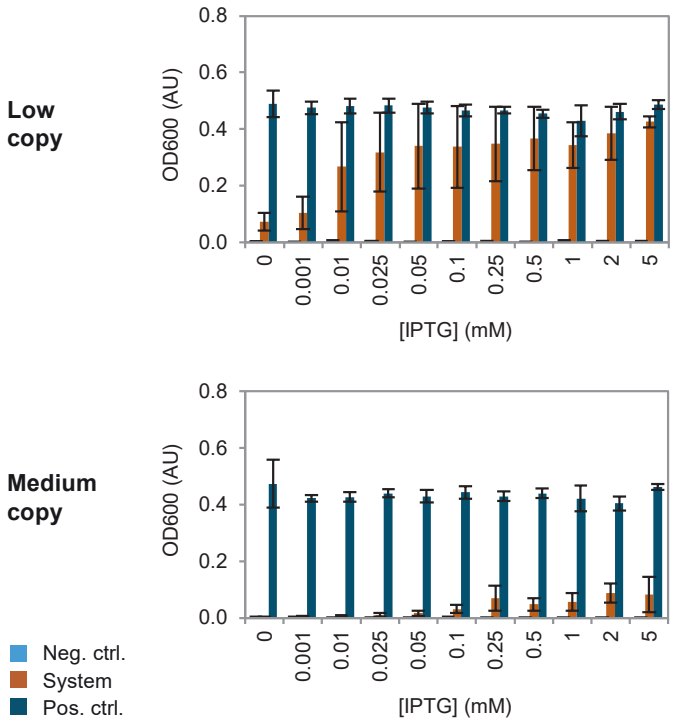


Fig. 2. Selection based on leucine auxotrophy complementation. The plasmid-encoded reporter gene *leuB* was induced in low and medium copy systems by various concentrations of the inducer IPTG. Bacteria were grown in M9 medium for 48 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph LLR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with a frameshift in *leuB*.

The low and medium copy systems with LeuB as selection reporter were grown in minimal M9 medium for 32 h (Fig. S2) and 48 h (Fig. 2). The cells were not pre-induced with IPTG, because this was not required for survival in the assay (Fig. S3). For each system, three strains were analysed: (1) the system itself (auxotroph LLR + regulator-reporter plasmid), (2) a negative control (auxotroph LLR + regulator-reporter plasmid with a frameshift in *leuB*), and (3) a positive control (non-auxotroph LR + regulator-reporter plasmid with a frameshift

in *leuB*). In comparison with the positive controls, which were in stationary phase after 32 h independent of the copy number, the system strains grew slower and their growth rate was dependent on the copy number and the IPTG concentration. The low copy system grown in presence of high IPTG concentrations only reached stationary phase after 48 h, whereas the low copy system in presence of low IPTG concentrations and the medium copy system at all IPTG concentrations had not reached stationary phase at 48 h. As suggested for the AraC-based systems (**Chapter 3**[184]), the difficult complementation in the medium copy system might be caused by the combination of the higher copy number and the dependency on the plasmid encoded LeuB. The frameshift-based controls were once more shown to be good controls (also in KmR- and LuxCDABE- based assays).

The low and medium copy systems with KmR as selection reporter were grown in LB medium for 17 h (stationary phase; Fig. 3). For comparison with the AraC-based system, cells were pre-induced (only non-induced cultures were not pre-induced), although pre-induction was not required for survival in the assay (Fig. S4). For each system, two strains were analysed: (1) the system itself (LR + regulator-reporter plasmid), and (2) a negative control (LR + regulator-reporter plasmid with a frameshift in *kan*). Appropriate kanamycin concentrations (0, 5, 15 and 30 µg mL^{-1} kanamycin) were picked based on death curves obtained at a constant inducer concentration (Fig. S5). In contrast to the negative controls and non-induced system strains, IPTG induced system strains survived above 2.5 µg mL^{-1}. This corresponds to the minimum inhibitory concentration (MIC_{90} = 0.5-2 µg mL^{-1} kanamycin)[143]. The highest kanamycin concentrations required high IPTG concentrations for survival, but even at maximal induction none of the strains could deal with 50 µg mL^{-1} kanamycin, the concentration commonly used for plasmid maintenance. This was probably due to the burden on the cells by maintaining the large plasmids and by expressing their eight genes, resulting in a relative low expression per gene and thus low resistance. This gene dosage effect[144] hypothesis was strengthened by the difference between the low and the medium copy system. When grown with the same kanamycin concentration, the medium copy system grew better. For application of the system, the low range of kanamycin concentrations should not be an obstacle, as long as selections are performed within or in proximity of this range.

All four systems with LuxCDABE as screening reporter, varying in copy number and selection reporter, were grown in LB medium for 4.5 h (Fig. 4). To reduce variation between experiments, measurements were done in late log phase when signal production and wash out due to cell division were about the same. For each system, two strain were analysed: (1)

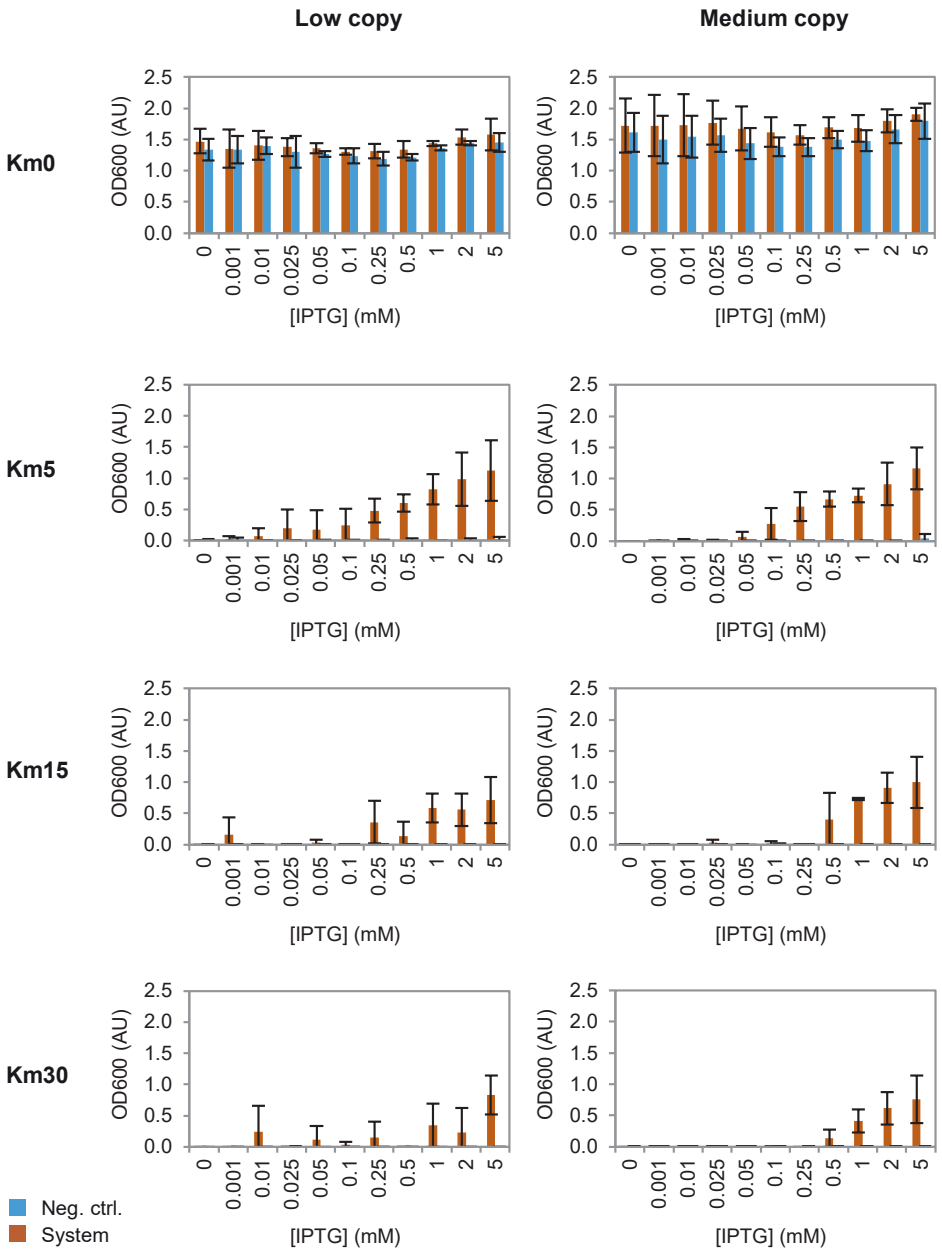


Fig. 3. Selection based on kanamycin resistance. The plasmid-encoded reporter gene *kan* was induced in the low and medium copy systems with the inducer IPTG. Bacteria were grown in LB medium for 17 h in the presence of 0, 5, 15 or 30 µg mL^{-1} kanamycin. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid. Neg. ctrl.: LR with the regulator-reporter plasmid with a frameshift in *kan*.

the system itself (LR or LLR + regulator-reporter plasmid), and (2) a negative control (LR or LLR + regulator-reporter plasmid with a frameshift in *luxA*). The bioluminescence increased with increasing inducer concentration, reaching maximal induction at ~ 3 mM, a value only slightly higher than described in literature (0.1-1 mM[163,190,204]). Medium copy systems had a higher maximal induction compared to lower copy systems, most likely due to a gene dosage effect. In contrast to the previously characterized AraC-based systems (**Chapter 3**[184]), LacI-

based systems differed in maximal induction between the KmR and LeuB versions with the LeuB versions having a higher bioluminescence. The difference was likely related to the coding sequences of the *leuB* and *kan* genes and not to the proteins, because the absence of the functional proteins KmR and LeuB, due to a frameshift in *kan* or *leuB*, did not affect the bioluminescence (Fig. S6). Since AraC stays bound to the DNA upon binding to L-arabinose and therefore prevents possible read-through transcription or transcription from unintended putative promoter regions, AraC-based systems did not show this difference between KmR and LeuB versions.

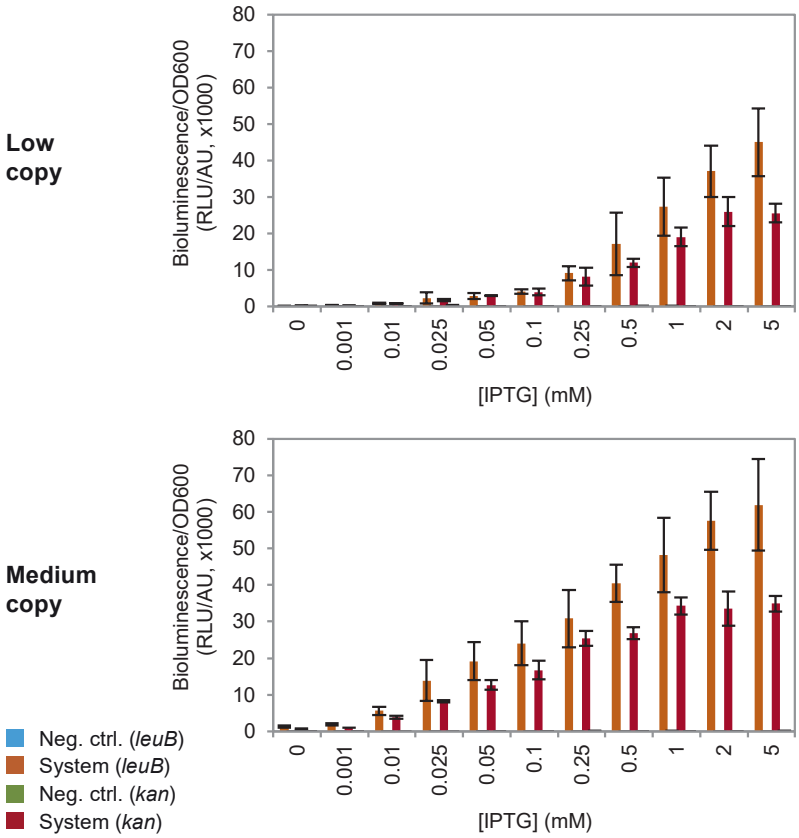


Fig. 4. Screening based on bioluminescence. The plasmid-encoded reporter operon *luxCDABE* was induced in four different systems by various concentrations of the inducer IPTG. The four systems were the low and medium copy systems with either LeuB or KmR as selection reporter. Bacteria were grown in LB medium under non-selective conditions for 4.5 h. The data are an average of three independent experiments (standard deviation indicated). System with LeuB: *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid with *leuB*. Neg. ctrl. with LeuB: LLR with the regulator-reporter plasmid with *leuB* and a frameshift in *luxA*. System with KmR: *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with *kan*. Neg. ctrl. with KmR: LR with the regulator-reporter plasmid with *kan* and a frameshift in *luxA*.

Comparison of the LacI-based systems

After establishment of the response curves, four characteristics were determined (Table 1)

to make a mutual comparison between the LacI-based systems and to compare the best system with the AraC-based system. These characteristics were leakiness, maximal signal, dynamic range and sensitivity. In the next paragraphs, the LacI-based systems are compared regarding low versus medium copy number and LeuB versus KmR as selection reporter.

In the LeuB-based assay, the low copy LacI-based system was slightly leakier, had a higher maximal signal, had a dynamic range shifted to lower concentrations and was more sensitive than the medium copy LacI-based system. The growth rate of the low copy system was much higher than that of the medium copy system (Fig. 2 and Fig. S2), leading to these better characteristics. The difference in growth was probably due to the burden of the medium copy system to the cells as described above. In the KmR-based assay, no leakiness was observed for either copy number by keeping the kanamycin concentration ≥ 5 μg mL^{-1}. Also the maximal signal was similar, but the relative dynamic range and sensitivity varied depending on the kanamycin concentration. In the LuxCDABE-based assay, the medium copy version scored better for all characteristics, most likely due to an overall higher reporter expression (gene dosage effect).

LeuB-based selection was slightly leakier than KmR-based selection, because of the threshold set with a kanamycin concentration ≥ 5 μg mL^{-1}. The maximal signal with LeuB was lower than with KmR due to the lower maximal OD600 reached in minimal medium than in LB medium. The sensitivity was better for LeuB-based selection versus KmR-based selection, and therefore also the dynamic range was shifted to lower concentrations.

In conclusion, all four systems functioned but they differed in their characteristics. Since a good sensitivity is such an important criterion for the selection step, not to miss any positives, LeuB-based selection was favoured over KmR-based selection, despite the longer growth time and the slightly higher leakiness. The low copy system performed better than the medium copy version, making the low copy system with LeuB as selection reporter the best LacI-based system. This system had a low leakiness for both selection and screening, which is important to reduce the number of false positives. In addition, the fold change of maximal signal over leakiness was good for screening (three orders of magnitude) and similar to fold changes in other transcriptional regulator-based systems[111,148]. Also the dynamic range for screening, not for selection, was satisfactory (two orders of magnitude) and comparable to other transcriptional regulator-based systems[107,115,149]. Fortunately, this small dynamic range in the selection is not detrimental, since sensitivity is is the most important here, while the dynamic range is essential during the quantification in the screening step. The sensitivity was good for both selection and screening (0.001-0.01 mM)

Table 1. Characteristics of the reporter systems[a]

Reporter	Copy number	Leakiness (AU)[b]		Maximal signal (AU)[c]		Dynamic range (mM)[d]		Sensitivity (mM)[e]	
LeuB	low	0.07 ± 0.03	− +	0.42 ± 0.02	− +	0.01-0.05	− −	0.001-0.01	+
	medium	0.00 ± 0.00	+ +	0.08 ± 0.06	−	0.025-0.25	−	0.01-0.025	− +
KmR (Km5)	low	0.00 ± 0.02	+ +	1.10 ± 0.49	+ +	0.25-1	−	0.1-0.25	−
	medium	0.00 ± 0.00	+ +	1.12 ± 0.34	+ +	0.1-5	−	0.05-0.1	− +
KmR (Km15)	low	0.00 ± 0.00	+ +	0.71 ± 0.37	+	0.25-1	−	0.1-0.25	−
	medium	0.00 ± 0.01	+ +	0.99 ± 0.41	+ +	1-1	− −	0.5-1	−
KmR (Km30)	low	0.00 ± 0.00	+ +	0.83 ± 0.31	+	5-5	− −	2-5	−
	medium	0.00 ± 0.00	+ +	0.75 ± 0.38	+	1-2	− −	0.5-1	−
LuxCDABE (*leuB*)	low	99 ± 172	+ +	44934 ± 9322	+ +	0.01-5	− +	0.001-0.01	+
	mediu	1360 ± 367	+ +	61897 ± 12507	+ +	0.001-2	+	<0.001	+ +
LuxCDABE (*kan*)	low	116 ± 97	+ +	25643 ± 2627	+ +	0.01-2	− +	0.001-0.01	+
	medium	674 ± 201	+ +	34836 ± 2154	+ +	0.001-1	+	<0.001	+ +

[a]The systems vary in the selection reporter (LeuB or KmR) and the copy number of the regulator-reporter plasmid (medium or low). The KmR-based systems are characterized at three different kanamycin concentrations (5, 15 and 30 μg mL^{-1}). The LuxCDABE-based systems are characterized for both LeuB and KmR containing versions. The standard deviation is included. A qualitative ranking is made (− −, −, − +, +, + +) with − − indicating a poor system and + + a good system. For leakiness, this indication is relative to the maximal signal. Absolute numbers for leakiness and maximal signal cannot be directly compared between the selection reporters LeuB and KmR and the screening reporter LuxCDABE, because they represent growth and bioluminescence respectively. [b]Signal at 0 mM inducer. [c]Signal at saturating inducer concentration. [d]Range of concentrations giving a changeable signal. [e]Lowest detectable inducer concentration.

and biologically relevant, as it agrees with the sensitivity of screening-based bioreporters that have an application in library screening or strain optimization (0.05-10 μM)[110,115,123,149].

Determination of specificity

To test the IPTG specificity of the best system, a LuxCDABE-based assay was performed with a set of compounds resembling IPTG (Fig. 5 and for chemical structures Fig. S7). These compounds were previously described to bind LacI, either as neutral effector (not influencing the DNA binding), weak inducer (decreasing DNA binding) or anti-inducer (increasing DNA binding). All tested compounds except 1,6-hexanediol have the O2 and O3 hydroxyls on the sugar ring that are essential for binding residues R197, N246 and D274 of the LacI core domain via hydrogen bonds[204]. In case of 1,6-hexanediol, water molecules take the places of these hydroxyls[205]. Most tested compounds were assumed not to be metabolized, except for D-galactose and D-melibiose. For the specificity assay, the same set-up was followed as for the induction assay and IPTG was taken along as control. The absolute induction by IPTG deviated slightly from the characterization described above (1 mM, 23268 ± 7788 vs. 27354 ± 7941 RLU/AU; 5 mM, 60045 ± 5559 vs. 45024 ± 9321 RLU/AU), but the values are in the same range. This difference is due to various influences on the metabolism, resulting in deviations in the bioluminescence signal between experiments (**Chapter 3**[184]).

Compared to the inducing effect of IPTG, the effects of the other compounds were very limited. D-lactose, D-melibiose and D-galactose did not influence the bioluminescence. D-lactose was described as inducer based on β-galactosidase and galactoside-transacetylase activities[206], but the actual inducer was probably the D-lactose derivative allolactose since β-galactosidase was present. In a permease activity determination in a β-galactosidase minus strain, D-lactose had no effect on LacI repression[207]. This confirms the bioluminescence-based results in this study, which also come from expression of a reporter gene in a β-galactosidase minus strain. However, *in vitro* operator binding experiments showed D-lactose as anti-inducer[208]. This contradiction might be due to *in vitro* versus *in vivo* measurements. D-melibiose was previously described as (weak) inducer[206,208] and D-galactose as very weak inducer[206,208] or having no effect[207]. That in this study these two sugars had no effect is likely because they are metabolized.

D-fucose, para-nitrophenyl-β-D-galactopyranoside and 1,6-hexanediol had a small inducing effect. The first two were indeed described as very weak inducers[208]. For 1,6-hexanediol, an inducing effect based on β-galactosidase activity was mentioned above 10

mM. Below this concentration the molecule had no effect or was slightly anti-inducing[205]. In this study, the switch was at a slightly lower concentration, namely 5 mM.

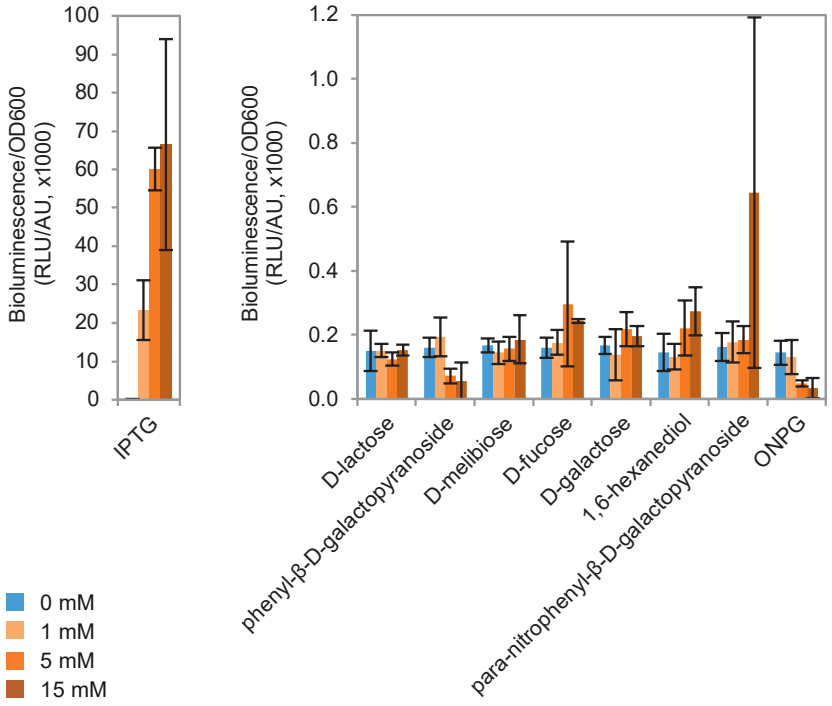


Fig. 5. Specificity determination by screening based on bioluminescence. Induction of the plasmid-encoded reporter operon *luxCDABE* by 0, 1, 5 or 15 mM of IPTG (left) or IPTG resembling molecules (right) was studied in the low copy system with LeuB as selection reporter, namely *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid with *leuB*. Bacteria were grown in LB medium for 4.5 h. The signal of the negative control (LLR with the regulator-reporter plasmid with *leuB* and a frameshift in *luxA*) was subtracted. The data are an average of three independent experiments (standard deviation indicated). The scales of the two y-axes differ.

Phenyl-β-D-galactopyranoside and ortho-nitrophenyl-β-galactopyranoside (ONPG) had a slight anti-inducing effect. They reduced the background expression. Phenyl-β-D-galactopyranoside was described both as inducer based on β-galactosidase and galactoside-transacetylase activities[206] and as anti-inducer based on *in vitro* operator binding experiments[208]and GFP[204]. This discrepancy could not be explained. ONPG was described as neutral effector[208,209], having no effect[207]or as anti-inducer[204]. However, the effect of ONPG can be influenced by pH and the oligomeric state of LacI (dimer or tetramer)[209], maybe explaining the diverse results in different experiments.

Overall, the system was specific, giving high induction only with IPTG. For the other compounds, the majority of the described effects was confirmed, although various experiments contradict one another. The sensitivity of the system was underscored by the system's ability to detect weak inducers as well as anti-inducers, but contaminations of the

tested molecules could not be ruled out. Based on these results, it is hard to draw conclusions on how a molecule should look like to be a neutral effector, an inducer or an anti-inducer. The molecules differ on their C1 and C6 substituent groups, but any common traits among for example inducers could not be found. Daber *et al.* compared crystal structures of LacI bound to IPTG or ONPG and showed that for induction it was essential that the N- and C-terminal subdomains were cross-linked by a water-mediated hydrogen network involving the O6 hydroxyl. However, just the presence of the O6 hydroxyl was not enough; stabilizing interactions between LacI and the rest of the ligand were also required[210]. How D-fucose can induce without O6 hydroxyl was unclear.

Comparison of LacI-based system with the earlier AraC-based system

The performances of the LacI-based systems were compared with those of the previously described AraC-based systems. It appeared that the LacI systems behaved very differently in all three induction assays. (1) The LacI-based systems were less leaky than AraC-based systems. (2) The maximal signal of LacI-based systems was similar to AraC-based systems in the LeuB-based-assay and lower than that of AraC-based systems in the KmR-based assay (only at high Km concentrations), but it was higher in the LuxCDABE-based assay. (3) The dynamic range of LacI-based systems was not as good as that of AraC-based systems in the LeuB-based assay and at high kanamycin concentration in the KmR-based assay, but it was better at low kanamycin concentrations and in the LuxCDABE-based assay. (4) The sensitivity of the LacI-based systems was lower than that of AraC-based systems in the KmR-based assay, but it was higher in the LeuB- and LuxCDABE-based assays. Although it is hard to pinpoint exactly which characteristic of the LacI and AraC regulation was responsible for each individual behaviour in the assays, there are a few things that play a role. (1) The mechanism of the two regulators is very different. LacI is a repressor that dissociates from the DNA upon induction, whereas AraC turns from repressor to activator upon induction and stays attached to the DNA. (2) The binding kinetics of the regulators to the DNA and to their inducers are dissimilar. For example, LacI and AraC bind IPTG and L-arabinose *in vitro* with dissociation constants of 2.5-2.8 µM[209,211] and 0.3-3 mM[212,213] respectively. (3) For LacI, the observations were reflecting the average level of induction per cell. For AraC, the observations reflected the number of cells that were fully induced as a result of induction of *araE* by L-arabinose-bound-AraC; the *araE* gene encodes the low affinity L-arabinose transport system[142]. (4) The transcription and translation rates of the two regulators probably slightly deviate. Although they have the same promoter and the same ribosomal

binding site (RBS), the coding sequence is different and could influence the rate of transcription, translation and folding, for example via codon usage[214]. The number of transcriptional regulators could change the response to the inducer. For example, a high number of repressors gives a less steep response curve than a low number of repressors[117]. (5) The transcription of the reporters probably varied between LacI- and AraC-based systems, because the LacI and AraC responsive promoters were not the same. The LacI responsive promoter was closer to the consensus *E. coli* promoter sequence TTGACA(N)^{17}TATAAT and most likely stronger than the AraC responsive promoter. Since the final observation in the assays was the sum of all these factors, it was not surprising that the behaviour of the LacI and AraC-based systems was different.

Focusing only on the best LacI- or AraC- based systems, each with different characteristics, brings up the question which system to use in which situation. The advantages of the low copy LacI-based system with LeuB is two orders of magnitude more sensitive than the AraC-based system and for screening its fold change of maximal signal over leakiness is two orders of magnitude higher. On the other hand, for the medium copy AraC-based system with KmR the stringency of selection could be easily adapted by changing the kanamycin concentration, and cells can be grown on rich LB medium instead of minimal medium, reducing the selection time. Although these factors are important for the choice of one system over the other, the main factor is the target molecule. The natural ligands of LacI and AraC are a disaccharide and a monosaccharide respectively, and LacI is therefore suited for slightly bigger molecules than AraC. Also the adaptability of the regulators towards the target molecule is crucial for the decision.

Conclusion

In this study, LacI-based versions were constructed of the previously developed AraC-based bioreporter (**Chapter 3**[184]). These systems express two reporters under control of LacI. The selection reporter allows for a rapid reduction of the initially large multi-gene metagenomic or single-gene mutant library size based on growth, whereas the screening reporter enables exclusion of false positives and quantification of the positive variants based on bioluminescence. The different versions vary in plasmid copy number (low/medium) or selection reporter (LeuB/KmR) and were compared in terms of leakiness, maximal signal, dynamic range and sensitivity. The best LacI-based system had a low copy number and LeuB as selection reporter. Although the system was specific only for IPTG to reach full induction, it was able to detect previously described weak inducers or anti-inducers.

As predicted upfront (**Chapter 3**[184]), a different regulator indeed demands for some optimization followed by characterization. The LacI- and AraC-based systems have different characteristics and require other system components like medium versus low copy number and KmR versus LeuB as selection reporter for the best performance. However, this study does show that this double reporter system can be modified with a different transcriptional regulator. The availability of transcriptional regulators for small molecules targets is then the major determinant to make this system widely applicable in novel biocatalyst detection or small molecule detection in general. Fortunately, more and more efforts are being made to identify and characterize new transcriptional regulators. Next to regulator engineering, one approach is to pick up promoters responsive to a target molecule from promoter libraries[215,216]. An alternative is the use of helper enzymes to convert the target enzymatic product to a molecule for which a transcriptional regulator is known[217]. Other issues that could hamper the applicability of bioreporters are of more general nature, like problems with heterologous expression or uptake of substrate or product by the cell. In short, efforts like increasing the number of available sensors, expanding the host and reporter repertoires and improving general issues like heterologous expression, should make it possible in the future to detect a wide range of enzymatic products or other small molecules with bioreporters.

In conclusion, this study shows the adaptability of the dual bioreporter to another transcriptional regulator, broadening the bioreporter's range of target molecules. It is therefore an important step towards an improved detection method for small molecules and thereby for finding novel biocatalysts.

Materials and methods

Bacterial strains and media

E. coli DH10B T1^{R} (Invitrogen, catalog number C6400-03) was used for plasmid propagation and was grown and transformed by standard methods[160]. *E. coli* BW25113 JW0336-1 and JW0063-1 of the KEIO-collection[161] were the parent strain for the constructed knockouts strains and the origin of the *leuB* gene respectively. The knockout strains hosted the regulator-reporter plasmids or their controls. Transformations and cell growth were done as described in **Chapter 3**[184].

Construction of the systems

The medium copy regulator-reporter plasmids pWUR767 and pWUR769 (~10 kb each) were constructed in seven succeeding cloning steps from pFU98[162] (kindly provided by Petra Dersch), following the same methods as described for the previously constructed AraC-based plasmids (**Chapter 3**[184]). Also the low copy regulator-reporter plasmids pWUR771 and pWUR773 and the control plasmids with a frameshift in one of the reporter genes were created as described for the AraC-based plasmids. The individual cloning steps and the primers are presented in more detail in Fig. S1 and Table S4 respectively.

For the host strains, the kanamycin resistance gene *kan* in *E. coli* BW25113 JW0336-1 (Δ*lacI::kan*) of the KEIO-collection[161] was eliminated by FLP recombinase encoded on pCP20[164] as described by Datsenko and Wanner[137]. The Δ*lacI* Δ*leuB* double knockout was constructed according to Datsenko and Wanner[137], except for the disruption cassette. This cassette with the flanking FLP recognition target (FRT) sites flanking *kan* (same homologous regions as in Baba *et al.*[161]) was created by PCR with the Geneart plasmid containing the recombination cassette minus homologous regions from Westra *et al.*[165] as template. The product was directly treated with DpnI and subsequently purified with the PCR purification kit of Thermo Scientific (#K0702). The *kan* gene was eliminated by FLP recombinase (as above). The Δ*lacI* Δ*recA* double knockout and the Δ*lacI* Δ*leuB* Δ*recA* triple knockout (designated LR and LLR respectively) were constructed as described for Δ*araC* Δ*recA* and Δ*araC* Δ*leuB* Δ*recA* in **Chapter 3**[184], using Cre recombinase instead of FLP recombinase for *kan* elimination. Each recombination event was verified by PCR and sequencing according to the methods in **Chapter 3**[184]. All primers are presented in Table S4.

The knockout strains were transformed with the regulator-reporter plasmids or control plasmids. For all four system strains, the relative plasmid copy number was determined based on plasmid isolation as described in **Chapter 3**[184].

Induction and specificity assays

The three types of induction assays, based on expression of *leuB*, *kan* or *luxCDABE* had a similar experimental set-up, which is described in **Chapter 3**[184]. The only exceptions to this protocol were that the growth time in the bioluminescence assay was 4.5 h instead of 5.5 h and that the inducer was IPTG instead of L-arabinose.

The specificity assays were performed in the same way as the induction assays. Concentrations of 0, 1, 5 and 15 mM of D-lactose, phenyl-β-D-galactospyranoside, D-melibiose, D-fucose, D-galactose, 1,6-hexanediol, para-nitrophenyl-β-D-galactopyranoside,

ONPG and IPTG were used. The OD600-corrected bioluminescence of the negative control (LLR with the regulator-reporter plasmid pWUR787 with *leuB* and a frameshift in *luxA*) was subtracted from the OD600-corrected bioluminescence of the system (LLR with the regulator-reporter plasmid pWUR771).

Acknowledgements

We would like to thank Petra Dersch for providing us with plasmids and Tjerk Sminia for input from an organic chemistry view.

Supplementary information

Table S1. Sequences.

1. $P_{LlacO\text{-}1}^{141}$ in front of *leuB*/*kan*. Underlined and italic nucleotides indicate restriction sites and operator sites respectively. The sequence contains from 5'to 3' a KpnI site, an O1 operator site (LacI), the promoter -35 site, an O1 operator site (LacI), the promoter -10 site, and an MreI site.
*GGTACCAATTGTGAGCGGATAACAA***TTGACA***TTGTGAGCGGATAACAA***GATACT**GAGCACATCAGCAGGACGCACTGACC*CGCCGGCG*

2. $P_{LlacO\text{-}1}^{141}$ in front of *luxCDABE*. Underlined and italic nucleotides indicate restriction sites and operator sites respectively. The sequence contains from 5'to 3' a KpnI site, an O1 operator site (LacI), the promoter -35 site, an O1 operator site (LacI), the promoter -10 site, and a SalI site.
*GGTACCAATTGTGAGCGGATAACAA***TTGACA***TTGTGAGCGGATAACAA***GATACT**GAGCACATCAGCAGGACGCACTGACC*GTCGAC*

Table S2. Frameshifts in control plasmids[a]

Construct	Parent[b]	Expected[c]	Obtained[d]	Comments
pWUR779/81/87/89	CCG**CG**G	CCGG	CCGG	as expected[e]
(*luxA*)	G**GC**GCC	GGCC	GGCC	residues 1-160 ok, A161G, A162V, Y163Stop (parent: 361 codons)[f]
pWUR775 (*leuB*)	GACAT**TC**TGTC	GACATCTGTC	GACACTGTC	2 bp instead of 1 bp deletion (stop 5 codons earlier, 251 instead of 252 codons ok, I252T,
	CTGT**AA**GACAG	CTGTAGACAG	CTGTGACAG	C253V, L254Stop)[e]
				residues 1-251 ok, I252T, L253V, S254Stop (parent: 364 codons)[f]
pWUR783 (*leuB*)	GACAT**TC**TGTC	GACATCTGTC	GACATTGTC	2 bp instead of 1 bp deletion (stop 5 codons earlier, also 252 codons ok, C253V, L254Stop)[e]
	CTGT**AA**GACAG	CTGTAGACAG	CTGTAACAG	residues 1-252 ok, L253V, S254Stop (parent: 364 codons)[f]
pWUR777 (*kan*)	CCTG**AT**TCAGG	CCTGATTTCAGG	CCTGATCAGG	1 bp deletion instead of 1 bp insertion (stop 38 codons later, also 115 codons ok)[e]
	GGACT**AA**GTCC	GGACTAAAGTCC	GGACTAGTCC	residues 1-115 ok, S116Q, G117V, E118K, ……., R155E, M156Stop (parent: 272 codons)[f]
pWUR785 (*kan*)	CCTG**AT**TCAGG	CCTGATTTCAGG	CCTGTTCAGG	1 bp deletion instead of 1 bp insertion (stop 38 codons later, 114 instead of 115 codons ok)[e]
	GGACT**AA**GTCC	GGACTAAAGTCC	GGACAAGTCC	residues 1-114 ok, D115V, S116Q, G117V, ……., R155E, M156Stop (parent: 272 codons)[f]

[a]All upper strands are from 5' to 3' in the direction of the gene indicated in brackets. [b]The restriction site in the parent plasmid (Cfr42I, Eam1105I and XagI for *luxA*, *leuB* and *kan* respectively). The cut site is between the two bold nucleotides. [c]Sequence expected after Klenow treatment and ligation. [d]Sequence obtained after Klenow treatment and ligation. [e]Obtained versus expected. The differences were due to AT-rich termini after digestion, which allow for mismatching; this was also observed for the AraC-based control plasmids in **Chapter 3**[184]. [f]Obtained versus parent.

Table S3. Relative plasmid copy number of the reporter systems

Strain[a]	10^9 plasmids mL^{-1}[b]		Ratio medium/low
	low copy	medium copy	
LLR + reg.-rep. plasmid (*lacI*, *leuB*)	5.2 ± 1.0	17.4 ± 1.6	3.3 ± 0.7
LR + reg.-rep. plasmid (*lacI*, *kan*)	7.5 ± 2.6	16.6 ± 2.3	2.2 ± 0.8

[a]The systems vary in the selection reporter (LeuB or KmR) and the copy number of the regulator-reporter plasmid (ColE1 or p15A origins of replications for medium or low copy number respectively). reg.-rep. plasmid, regulator-reporter plasmid. [b]The number of plasmid molecules per millilitre culture at an OD600 of 1 was determined by plasmid isolation for the four different systems. The data are an average of three independent experiments (standard deviation indicated).

Table S4. Primers used in this study

Primer	Annealing location	Sequence (5'-->3')[a]	Features
Construction of plasmids			
BG3691	start *kan*	GCGCGGGATCCCGCCGGCG*AGGAGG*ATACGTATGAGCCATATTCAACGGGAAAC	BamHI, MreI, RBS
BG3692	end *kan*	CGCGCCTCGAGCGATCGTTAGAAAAACTCATCGAGCATCAAATG	XhoI, PvuI
BG3693	start *leuB*	GCGCGGGATCCCGCCGGCG*AGGAGG*ATACGTATGTCGAAGAATTACCATATTGCCG	BamHI, MreI, RBS
BG3694	end *leuB*	CGCGCCTCGAGCGATCGTTACACCCCTTCTGCTACATAGC	XhoI, PvuI
BG3695	inside *leuB*	GAATCTGCTCGCAAGCGTCGCCACAAAGTGACG*AGT*ATCGATAAAGCCAAC	altered serine codon: TCG-->AGT
BG3696	inside *leuB*	GAGGATTGCAGCACGTTGGCTTTATCGAT*ACT*CGTCACTTTGTGGCGACG	altered serine codon: TCG-->AGT
BG3746	start P_{lacI}^{Q}	GCGCGCTCGAGGTTGACACCATCGAATGGTGCAAAACC	XhoI
BG3747	end P_{lacI}^{Q}	GCCGCCGATCGCGGWCCGATTCACCACCCTGAATTGACTCTCTTCC	PvuI, CpoI
BG3938	start *lacI*	GCGCGCGGWCCG*AGGAGG*ATACGTGTGAAACCAGTAACGTTATACGATGTCGCAG	CpoI, RBS
BG3939	end *lacI*	CGCGCCGATCGTCACTGCCCGCTTTCCAGTCG	PvuI
BG4231	start ColE1	TACTGGTACCCATGACCAAAATCCCTTAACGTG	Acc65I
BG4232	end ColE1	TACTCGTACGCCTAGGCGTTCGGCTGC	Pfl23II
BG4666	start p15A	TATGTGGTACCTAGCGGAGTGTATACTGGCTTAC	Acc65I
BG4667	end p15A	TACAACCTAGGACAACTTATATCGTATGGGGCTG	XmaJI
Verification of plasmids			
BG3336	inside ColE1	TTCGCCACCTCTGACTTG	
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3653	inside *kan*	GCCTGTTGAACAAGTCTGGA	
BG3857	inside *luxE*	GAAGCGTTTGATAGTTGAGCGG	
BG3858	inside *cat*	CAGGTTCATCATGCCGTCTG	
BG3942	upstream *cat*	CAACGTCTCATTTTCGCCAG	
BG3943	inside *luxC*	CACGAATGTATGTCCTGCG	
BG3977	inside *leuB*	GCACAAATCCTTTCGCTGG	
BG3978	inside *leuB*	GTAATGGCTGGTGGTGATG	
BG4231	start ColE1	TACTGGTACCCATGACCAAAATCCCTTAACGTG	Acc65

Table S4 continued

Primer	Annealing location	Sequence (5'-->3')[a]	Features
Verification of plasmids			
BG4232	end ColE1	TACTCGTACGCCTAGGCGTTCGGCTGC	Pfl23II
BG4627	inside *leuB*	GTATTCCGTGGCGATCTC	
BG4628	inside *leuB*	CGGCATCTATTTCGGTCAG	
BG4629	inside *kan*	CCAGACTTGTTCAACAGGC	
BG4630	inside *kan*	CTCCTGATGACGCATGG	
BG4631	inside *luxA*	GAATGGCATGACAGAGGG	
BG4632	inside *luxA*	GTGCCCATATTCTTGAGCC	
Construction of knockouts			
BG3649	start cassette[b]	*GCTCAACACAACGAAAACAACAAGGAAACCGTGTGA*GGTGTCTTTTTTACCTGTTTGACC	homologous to genome (upstream/start *leuB)*
BG3650	end cassette[b]	*ACGTCTTAGCCATGATTACACCCCTTCTGCTACATA*CTACCTCTGGTGAAGGAGTTG	homologous to genome (end/downstream *leuB*)
BG4490	start cassette[b]	*CAGAACATATTGACTATCCGGTATTACCCGGCATGACAGGAGTAAAAATG*GGTGTCTTTTTTA CCTGTTTGACC	homologous to genome (upstream/start *recA*)
BG4491	end cassette[b]	*ATGCGACCCTTGTGTATCAAACAAGACGATTAAAAATCTTCGTTAGTTTC*CTACCTCTGGTGAA GGAGTTG	homologous to genome (end/downstream *recA*)
Verification of knockouts			
BG3651	upstream *leuB*	CAGGTGGATATCGTCGCTAA	
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3653	inside *kan*	GCCTGTTGAACAAGTCTGGA	
BG3654	downstream *leuB*	AACAGTGGGGTTTCGTTTTC	
BG3657	upstream *lacI*	ACAACGGGTAGCAAAACAGA	
BG3751	downstream *lacI*	CACTTCAACATCAACGGTAATCG	
BG4190	upstream *recA*	CGTCAGGCTACTGCGTATG	
BG4191	downstream *recA*	GAATACGCGCAGGTCCATAAC	

[a]Underlined nucleotides indicate restriction sites, italic nucleotides indicate other features (for both see last column). [b]Recombination/disruption cassette.

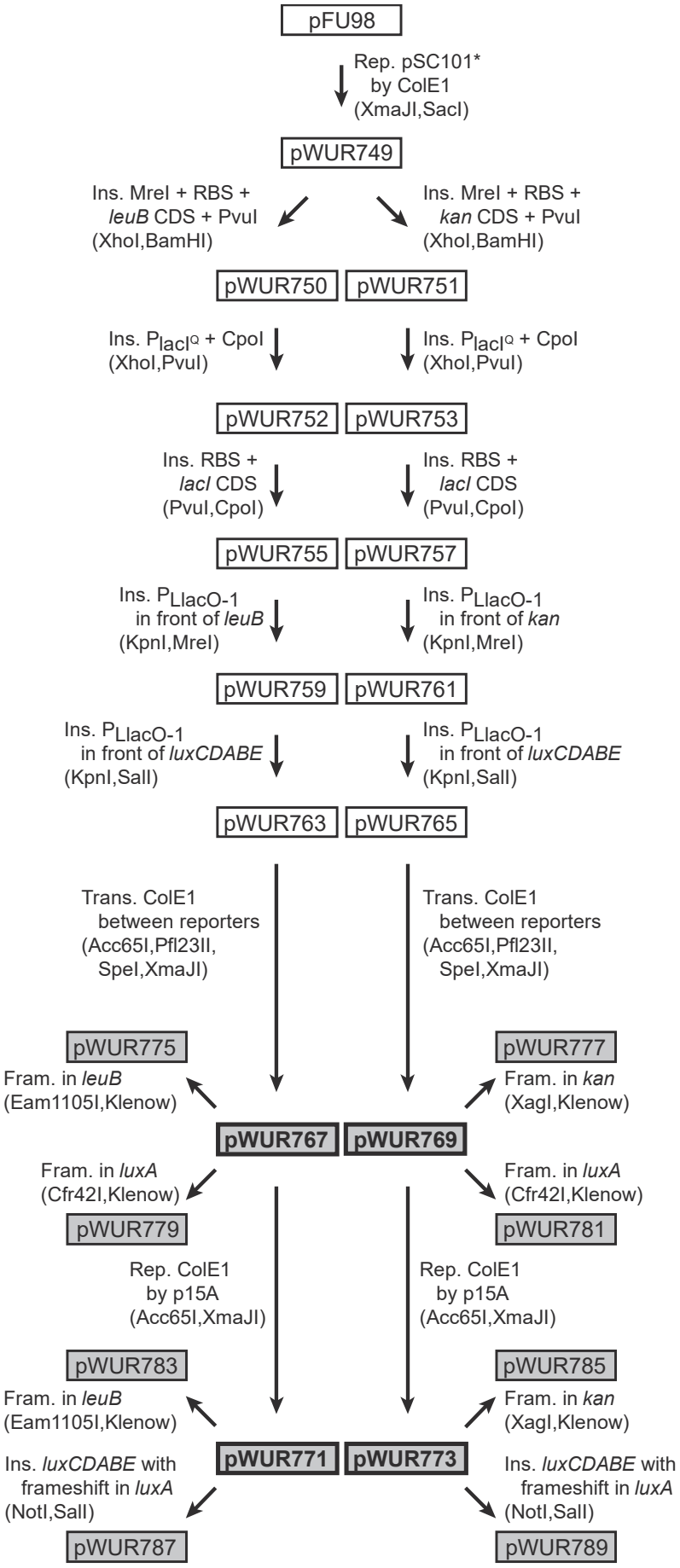

pFU98
Rep. pSC101* by ColE1 (XmaJI,SacI)
pWUR749
Ins. MreI + RBS + leuB CDS + PvuI (XhoI,BamHI)
Ins. MreI + RBS + kan CDS + PvuI (XhoI,BamHI)
pWUR750
pWUR751
Ins. PlacIQ + CpoI (XhoI,PvuI)
Ins. PlacIQ + CpoI (XhoI,PvuI)
pWUR752
pWUR753
Ins. RBS + lacI CDS (PvuI,CpoI)
Ins. RBS + lacI CDS (PvuI,CpoI)
pWUR755
pWUR757
Ins. PLlacO-1 in front of leuB (KpnI,MreI)
Ins. PLlacO-1 in front of kan (KpnI,MreI)
pWUR759
pWUR761
Ins. PLlacO-1 in front of luxCDABE (KpnI,SalI)
Ins. PLlacO-1 in front of luxCDABE (KpnI,SalI)
pWUR763
pWUR765
Trans. ColE1 between reporters (Acc65I,Pfl23II, SpeI,XmaJI)
Trans. ColE1 between reporters (Acc65I,Pfl23II, SpeI,XmaJI)
pWUR775
Fram. in leuB (Eam1105I,Klenow)
pWUR777
Fram. in kan (XagI,Klenow)
pWUR767
pWUR769
Fram. in luxA (Cfr42I,Klenow)
Fram. in luxA (Cfr42I,Klenow)
pWUR779
pWUR781
Rep. ColE1 by p15A (Acc65I,XmaJI)
Rep. ColE1 by p15A (Acc65I,XmaJI)
pWUR783
Fram. in leuB (Eam1105I,Klenow)
pWUR785
Fram. in kan (XagI,Klenow)
pWUR771
pWUR773
Ins. luxCDABE with frameshift in luxA (NotI,SalI)
Ins. luxCDABE with frameshift in luxA (NotI,SalI)
pWUR787
pWUR789

Fig. S1. Overview cloning steps. The four regulator-reporter plasmids (depicted in bold with a grey background) were constructed from pFU98[162] in eight subsequent cloning steps. From each of the four plasmids two control plasmids were made (depicted with a grey background) by making a frameshift either in the selection reporter gene (*leuB*/*kan*) or in one gene in the screening reporter operon (*luxA*). For each cloning step a description and the restriction enzymes are included. Rep., replacement; Ins., insertion; Trans., translocation; Fram., introduction frameshift. The origin and formation of the inserts are described here. BG numbers refer to primers (Table S3). pWUR749: digest from pFU168[162]. pWUR750: two step PCR from BW25113 genome to remove AatII and PvuI sites in CDS (first left BG3693/BG3696, first right BG3695/BG3694, second BG3693/BG3694). pWUR751: PCR from recombination cassette[165] (BG3691/BG3692). pWUR752/3: PCR from pET24d (Novagen, BG3746/BG3747). pWUR755/7: PCR from pET24d (BG3938/BG3939). pWUR759/61: digest from pMA-RQ P_{LlacO1}-KM (Geneart). pWUR763/5: digest from pMA_RQ $P_{LlacO-1}$-KS (Geneart). pWUR767/9: PCR from pWUR758[184] (BG4232/BG4231). pWUR771/3: PCR from pACYC184[172] (BG4666/BG4667). pWUR787/9: digest from pWUR778[184].

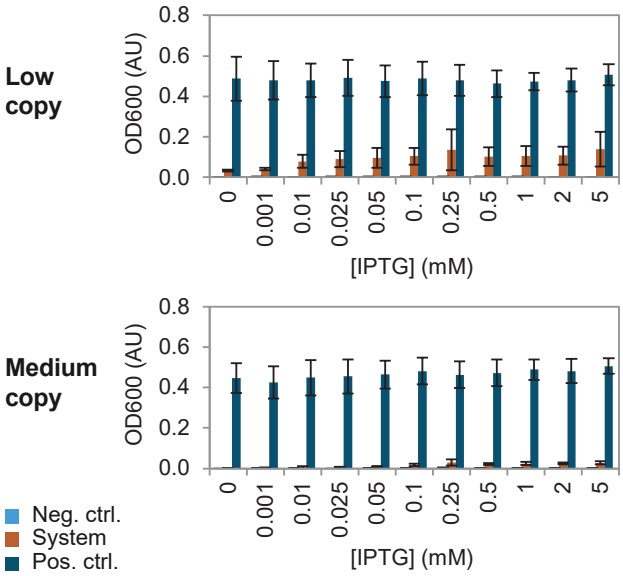


Fig. S2. Selection based on leucine auxotrophy complementation (32 h). The plasmid-encoded reporter gene *leuB* was induced in the low and medium copy systems with the inducer IPTG. Bacteria were grown in M9 medium for 32 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph LLR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with a frameshift in *leuB*.

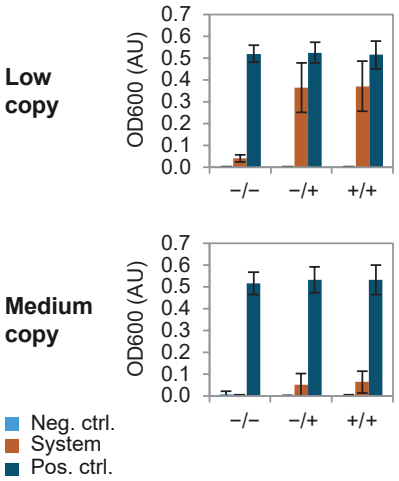


Fig. S3. Effect of the addition of inducer to the preculture on the leucine auxotrophy complementation assay. The plasmid-encoded reporter gene *leuB* was induced or non-induced in the low and medium copy systems with 0.5 mM of the inducer IPTG. The cultures were inoculated from induced or non-induced precultures to see the effect of this treatment. '−/−' non-induced in both precultures and assay cultures, '−/+' induced only in assay cultures, '+/+' induced in both precultures and assay cultures. Bacteria were grown in M9 medium for 32 h. The data are an average of three independent experiments (standard deviation indicated). System: auxotroph *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph LLR with the regulator-reporter plasmid with a frameshift in *leuB*. Pos. ctrl.: non-auxotroph *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with a frameshift in *leuB*.

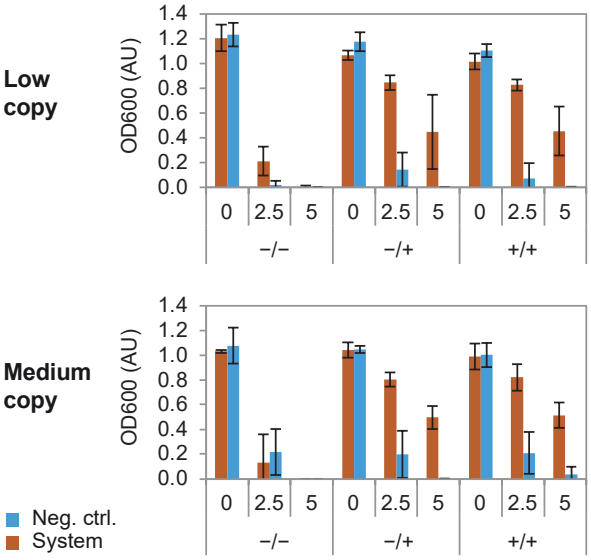


Fig. S4. Effect of the addition of inducer to the preculture on the kanamycin resistance assay. The plasmid-encoded reporter gene *kan* was induced or non-induced in the low and medium copy systems with 0.5 mM of the inducer IPTG. The cultures were inoculated from induced or non-induced precultures to see the effect of this treatment. '−/−' non-induced in both precultures and assay cultures, '−/+' induced only in assay cultures, '+/+' induced in both precultures and assay cultures. Bacteria were grown in LB medium for 17 h in presence of 0, 5, or 10 μg mL^{-1} kanamycin. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid. Neg. ctrl.: LR with the regulator-reporter plasmid with a frameshift in *kan*.

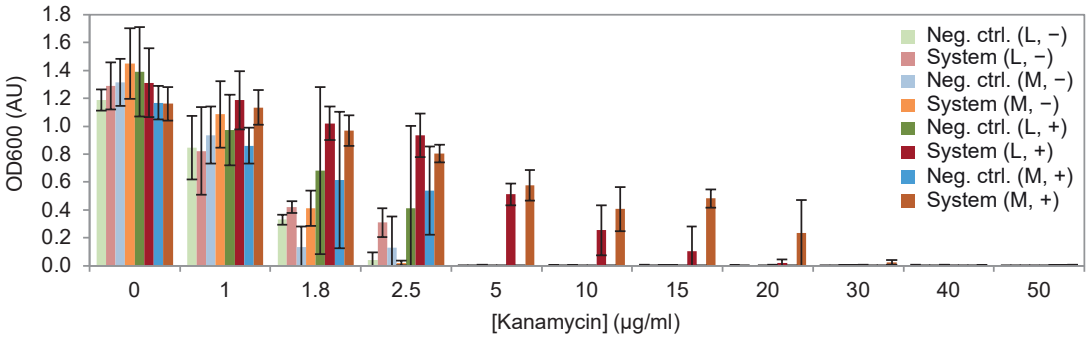


Fig. S5. Kanamycin death curve. The plasmid-encoded reporter gene *kan* was induced '+' or non-induced '–' in the low 'L' and medium 'M' copy systems with 0.5 mM of the inducer IPTG. Bacteria were grown in LB medium for 17 h in presence of different kanamycin concentrations. The data are an average of three independent experiments (standard deviation indicated). System: *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid. Neg. ctrl.: LR with the regulator-reporter plasmid with a frameshift in *kan*.

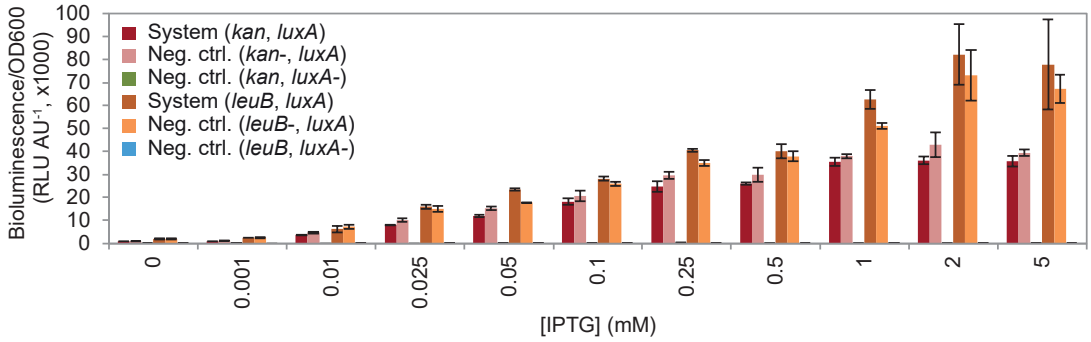
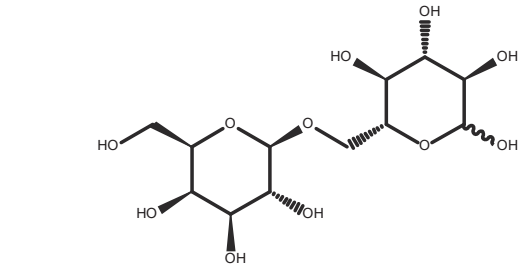
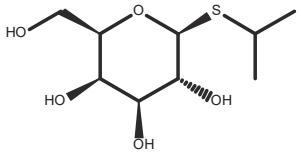


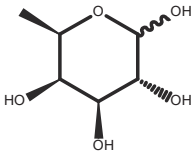
Fig. S6. Screening based on bioluminescence for LeuB- versus KmR-based systems. The plasmid-encoded reporter operon *luxCDABE* was induced in the medium copy system with either LeuB or KmR as selection reporter by various concentrations of the inducer IPTG. Bacteria were grown in LB medium for 4.5 h. The data are an average of three independent experiments (standard deviation indicated). System with LeuB: *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid with *leuB*. Neg. ctrl. without LeuB: LLR with the regulator-reporter plasmid with a frameshift in *leuB*. Neg. ctrl. with LeuB: LLR with the regulator-reporter plasmid with *leuB* and with a frameshift in *luxA*. System with KmR: *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with *kan*. Neg. ctrl. without KmR: LR with the regulator-reporter plasmid with a frameshift in *kan*. Neg. ctrl. with KmR: LR with the regulator-reporter plasmid with *kan* and with a frameshift in *luxA*.



6-O-β-D-galactopyranosyl-β-D-glucopyranose (allolactose)
natural inducer

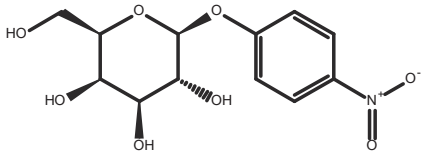


isopropyl β-D-1-thiogalactopyranoside (IPTG)
gratuitous inducer

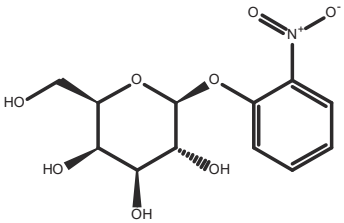


D-fucose

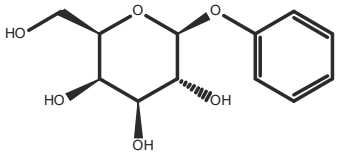
1,6-hexanediol



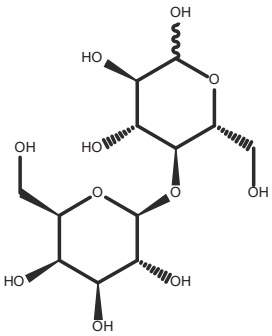
para-nitrophenyl-β-D-galactopyranoside



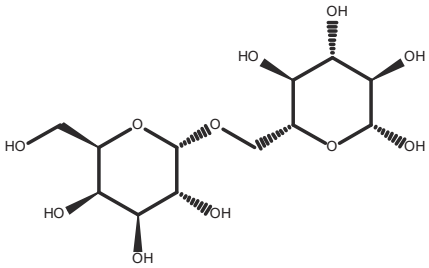
ortho-nitrophenyl-β-galactopyranoside (ONPG)



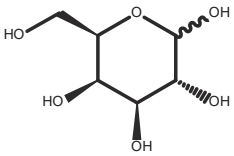
phenyl-β-D-galactospyranoside



4-O-β-D-galactopyranosyl-β-D-glucopyranose (D-lactose)

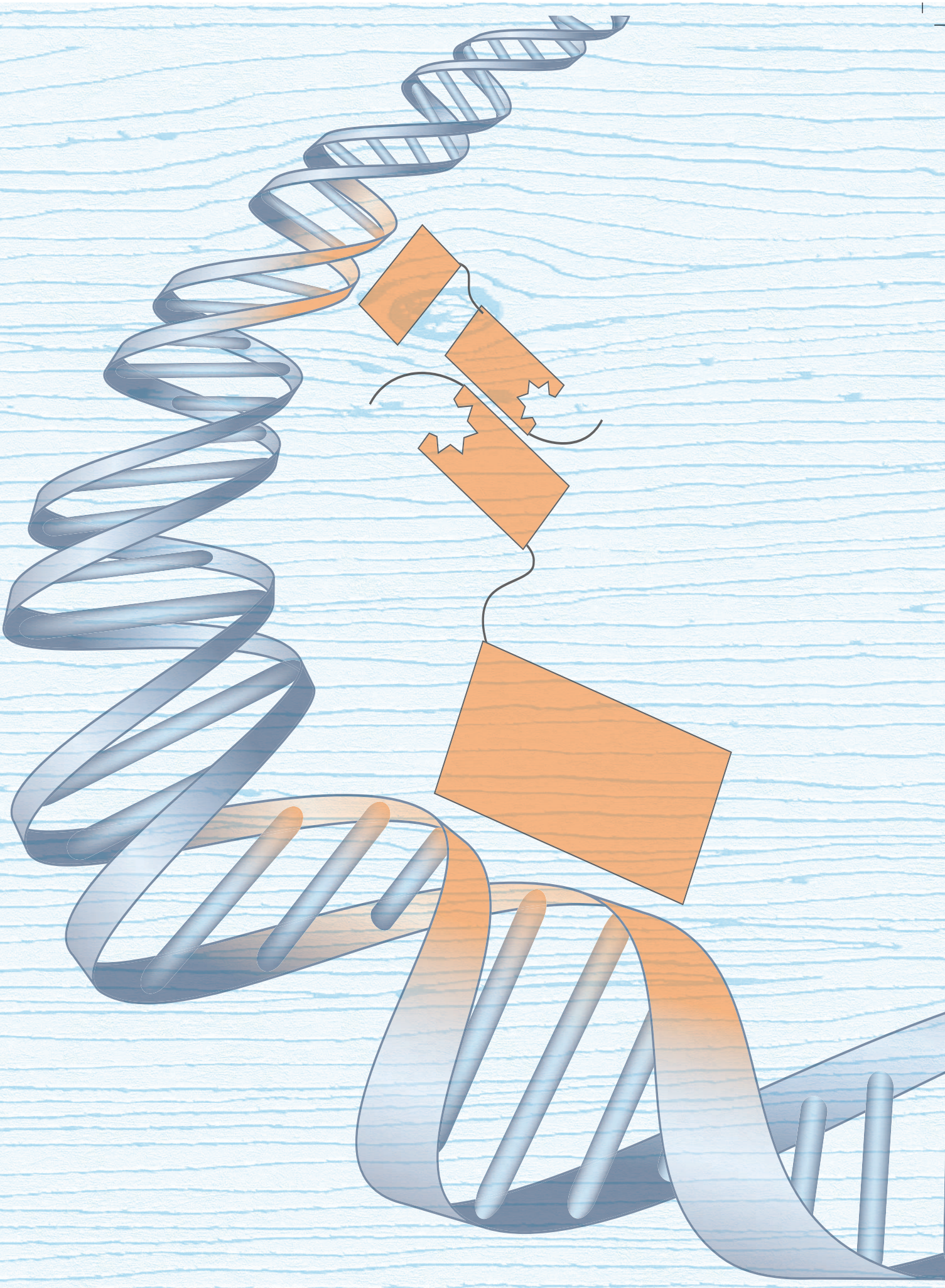


6-O-α-D-galactopyranosyl-β-D-glucopyranose (D-melibiose)



D-galactose

Fig. S7. Chemical structures of compounds used in specificity determination. Note the galactose moiety in most structures.



Chapter 5

Engineering the ligand specificity of the transcriptional regulator AraC and enrichment of desired variants by combined selection and screening

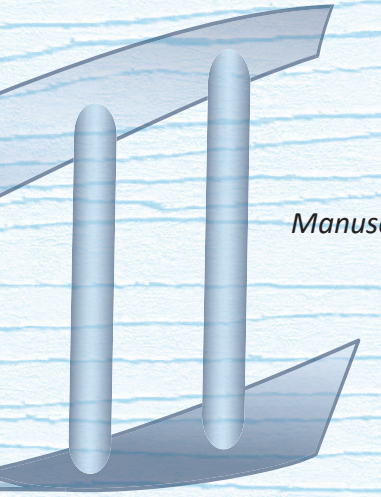
Teunke van Rossum[a], Andre C. Stiel[b,c], Lisa Röttjers[a], Fadi Al Makindji[a], Birte Höcker[b,d], John van der Oost[a], and Servé W. M. Kengen[a]

[a]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[b]Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

[c]Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

[d]Lehrstuhl für Biochemie, Universität Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany



Manuscript in preparation

Abstract

Bioreporters can be used in high-throughput screening methods for finding better biocatalysts and improved cell strains. The specificity determinant of a bioreporter is the sensor part, often a transcriptional regulator, which needs to be modified for every small molecule that is to be detected. In most specificity engineering projects, regulator variant libraries are screened with fluorescent activated cell sorting (FACS) using GFP as reporter, a rather complex and expensive method. This study aimed at simplifying the process by using growth-based selection rather than screening to obtain desired variants of the transcriptional regulator AraC. As a proof of principle, a previously developed AraC-based dual selection and screening system was used to obtain AraC variants with specificity towards D-xylose instead of its natural ligand L-arabinose. The dual reporter system divergently transcribes two reporters under AraC control, one for selection (*kan*) and one for screening (*luxCDABE*). In order to find D-xylose-specific AraC variants, two combinatorial site-saturation mutagenesis libraries of *araC* were designed, each having six codons that encode residues in the ligand binding pocket changed to NNK. Variants were selected based on kanamycin resistance in the presence of D-xylose. Further selection and screening assays allowed for exclusion of false positives, quantification and selection of three final variants. Although selection was done on D-xylose, best performing AraC variants showed only a very low level of induction (bioluminescence assay). Moreover, these variants were also responsive to monosaccharides other than D-xylose. This suggests that generalists rather than specialists were obtained, which is a commonly observed result in early screening rounds. Subsequent selection and screening rounds are needed to obtain better performing AraC variants. Although the current set of selected variants should be investigated in more depth to determine whether their ligand specificity is truly modified, after optimization of the selection and screening protocol, this same set-up could be used to select not only AraC variants with a better response to D-xylose, but also variants specific to other target molecules.

Introduction

Bioreporters are genetically engineered microbial cells that report the presence of a specific compound by producing a measurable signal. They can be used in high-throughput screening methods for finding, for example, novel biocatalysts[110,112] or better strains[218,219]. The sensor part of the bioreporter binds a small molecule such as the product of an enzyme or a desired metabolite. As a result, the reporter is switched on, giving a distinguishable phenotype[116]. The specificity determinant of this method is the sensor part, which thus needs to be modified for every small molecule to be detected. The most commonly used sensor part is a ligand-binding transcriptional regulator[33,34]. To get a transcriptional regulator with the desired specificity, there are two options. Firstly, known regulators can be implemented. If no regulator with the desired specificity is known from literature or from one of the dedicated databases[220,221], one could try to find a regulator and/or the promoter it regulates by screening genomic or metagenomic libraries or by transcriptomics[75,221,222]. An alternative is the use of helper enzymes to convert the target molecule into a molecule for which a regulator is known[217,221]. Secondly, regulators can be engineered by changing their specificity via directed evolution[108,219,221,223] or computational design[182,183,196,221]. Another engineering approach makes use of combining parts of different proteins, like in QUerying for EnzymeS using the Three-hybrid system (QUEST)[88], chemical complementation[224], chimeric transcription regulators[225] or chimeric two component regulatory systems[226].

Directed evolution and computational design each have advantages and disadvantages. In directed evolution, diversity is generated based on random or semi-random methods. Random methods like DNA shuffling and error-prone PCR (epPCR) have the benefit that no knowledge about the structure is required. However, DNA shuffling is only possible when multiple gene variants or natural homologues are available, and epPCR has a bias towards transitions (pyrimidine to pyrimidine or purine to purine) and an unequal distribution of mutations along the gene. The method 'Sequence saturation mutagenesis' (SeSAM[227]) or the upgraded version SeSaM-Tv P/P[228], deals with these issues and thereby greatly improves epPCR. Nevertheless, changes in specificity often require a combination of mutations[229] which, because of size limitations, is not likely covered in an epPCR library[230]. This combination is more probable in a semi-random approach like site-saturation mutagenesis in which multiple specific target residues are chosen and replaced by all or a subset of residues. This method therefore does require prior knowledge on the structure of the protein[230]. Computational design goes even further by making and screening libraries *in*

silico. The major advantage of this approach is the large sequence space that can be covered. This makes computational design suitable when more drastic engineering steps are required, for example when changing the specificity of the regulator to a small molecule that does not resemble the natural ligand[220]. However, detailed prior knowledge is required, and both energy functions and sampling methods are inaccurate[83,231,232]. Since both directed evolution and computational design have some drawbacks, the most powerful approach is the combination of the two[231]. In the initial computational design step, a large sequence space can be covered and diversity can be focused, whereas in a subsequent directed evolution step the necessary optimization and fine-tuning can be achieved[182,220,231,232]. In this respect, many methods have been developed for designing libraries or oligomers to create libraries[232-237]. These include approaches to limit the number of possible amino acids or codons[232,238-240].

A transcriptional regulator that has been the subject of many engineering studies is AraC, the regulator of the L-arabinose metabolism in *E. coli*. It is a dimer with 292 amino acids per monomer. In the absence of L-arabinose, each monomer binds to one of two operators, O_2 and I_1 (210 bp apart), with its C-terminal DNA binding domains (Fig. 1A). This binding results in looping the DNA, blocking the binding of RNA polymerase and thereby repressing transcription of the downstream genes involved in L-arabinose transport and conversion. Upon binding of L-arabinose to the N-terminal domain (Fig. 1B), the N-terminal arm of AraC folds over L-arabinose and no longer holds the C-terminal DNA-binding domains in a restricted conformation. The DNA-binding domains can then reorient to bind the adjacent I_1 and I_2 operators, thereby activating transcription of the genes involved in L-arabinose metabolism and transport. How the conformational change exactly takes place remains to be elucidated, but the N-terminal arm as well as the inter-domain linker are involved in this process, and the way the arm is folded in the presence of L-arabinose is crucial for inducibility. No direct binding of the N-terminal arm and the DNA-binding domains have been demonstrated.[130,241-243] Since AraC is studied in great detail, and because a crystal structure is available with and without its ligand, it has been an interesting subject for engineering. Error-prone PCR and screening with fluorescent activated cell sorting (FACS), has previously provided an AraC variant with reduced inhibition by isopropyl β-D-1-thiogalactopyranoside (IPTG)[244]. The combination of the DNA-binding domain of AraC with an enzyme that could bind the target ligand, resulted in chimeras specific to 2,3-dihydro-2,5-dihydroxy-4H-benzopyran-4-one (DDBO)[88] or isopentenyl pyrophosphate (IPP)[225]. Combinatorial site-saturation mutagenesis and screening with FACS gave AraC variants that

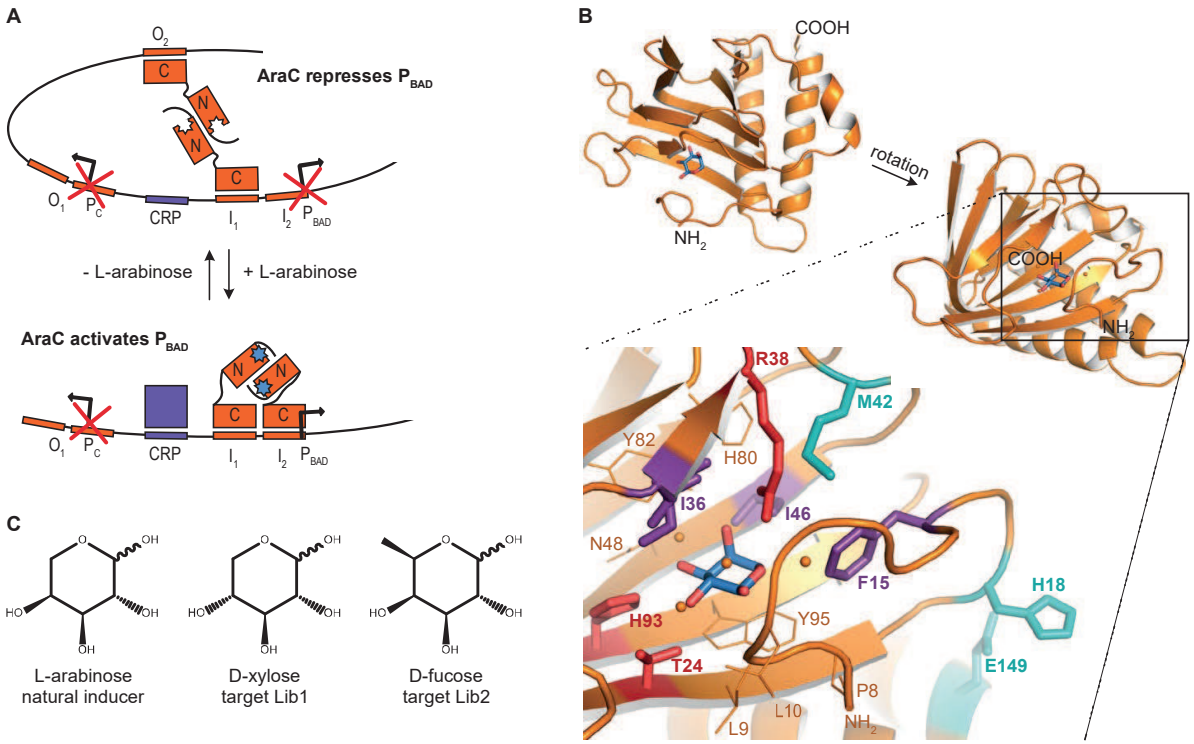


Fig. 1. AraC mechanism, structure and library design. (A) In the absence of L-arabinose, AraC represses genes involved in L-arabinose metabolism and transport by binding two operators, O_2 and I_1 (210 bp apart), with its C-terminal DNA binding domains. This binding loops the DNA, blocking binding of RNA polymerase. Upon binding of L-arabinose to the N-terminal domain, the N-terminal arm of AraC folds over L-arabinose and no longer holds the C-terminal DNA-binding domains in a restricted conformation. The DNA-binding domains can then reorient to bind the more closely located I_1 and I_2 operators, thereby activating transcription from P_{BAD}. Full activation is only reached when also Catabolite Repression Protein (CPR) binds at low glucose concentrations. AraC also represses its own P_C promoter, which is only transiently expressed upon L-arabinose binding[130,241,243]. Adapted from[242]. (B) Crystal structure (PDB 2ARC[245]) of the N-terminal domain (orange) bound to L-arabinose (blue). In the enlarged representation of the L-arabinose binding pocket, the target residues for libraries 1 (Lib1) and 2 (Lib2) are indicated as sticks and water molecules as spheres. Red, target residues of Lib1; cyan, target residues of Lib2; purple, target residues of Lib1 and Lib2. (C) Chemical structures of L-arabinose, D-xylose and D-fucose.

responded to D-arabinose[128], mevalonate[87], triacetic acid lactone (TAL)[129] or ectoine[246].

In many of these engineering projects, variant libraries were screened with FACS using GFP as reporter. Although the results obtained with this approach were impressive, the screening method required complex and expensive equipment. The present study set out to simplify the process to obtain AraC variants with altered ligand specificity by using growth-based selection rather than screening. As a proof of principle of the previously developed AraC-based dual selection and screening system[184], this system was used to select for AraC variants with a specificity towards D-xylose. In contrast to the natural ligand L-arabinose, D-xylose only elicits very low transcription[128] despite the similarity of the two

monosaccharides. The dual reporter system divergently transcribes two reporters under AraC control, one for selection (*kan*) and one for screening (*luxCDABE*). This allows for a rapid reduction of the initially large library size by selection based on kanamycin resistance, followed by exclusion of false positives and quantification of the positive variants based on bioluminescence. Two combinatorial site-saturation mutagenesis libraries of *araC* were prepared. In each library, six codons that code for residues in the ligand binding pocket were changed to NNK. Based on kanamycin resistance in the presence of D-xylose, variants with an altered response, albeit small, to several monosaccharides were selected. Subsequent selection and screening assays were performed to exclude false positives.

Results

Library design and construction

To show that combined selection and screening could be used to obtain a transcriptional regulator with altered ligand specificity, *araC* libraries were designed, constructed and screened. These libraries required four components, (1) a target molecule, (2) a host strain, (3) a vector and (4) an insert of *araC* variants. The target molecule D-xylose only differs from the natural ligand L-arabinose in the orientation of the O4 hydroxyl (Fig. 1C). The host strain was the same as for the dual reporter system, namely *E. coli* BW25113 Δ*araC* Δ*recA* (AR)[184]. To prevent D-xylose from being metabolized by the host strain AR, the *xylA* gene was deleted. This gene codes for D-xylose isomerase, which converts D-xylose to D-xylulose, the first step in D-xylose catabolism.[247] The new strain was designated ARX. The vector was the regulator-reporter plasmid of the screening/selection system, pWUR768[184]. This medium copy plasmid encodes *araC* from a constitutive promoter and two divergently oriented reporters from AraC-controlled promoters. The selection reporter KmR allows for selection based on kanamycin resistance, whereas the screening reporter LuxCDABE allows for screening based on bioluminescence. To prevent the wildtype *araC* from the vector ending up in the library, it was replaced by *sacB* from *B. subtilis*, giving pWUR947. *SacB* encodes levansucrase, which converts sucrose into compounds that are lethal to *E. coli*[248] and therefore allowed counter-selection on sucrose to remove vectors that were not properly digested and still contained *sacB* instead of an *araC* variant. The ability to counter-select on 5% sucrose was confirmed to be functional (Table S1). Before designing the library, the manageable library size needed to be determined. This size is dependent on the transformation efficiency of the host strain ARX with the large regulator-reporter plasmid. It

was assumed that this efficiency was the same as for the AR strain. AR cells that were made electrocompetent from a 100-mL culture with an OD600 of 0.4 gave $4.4 \cdot 10^7$ colonies when transformed with 2 µg of regulator-reporter plasmid containing the wildtype *araC* (pWUR768). This was the optimal ratio of cells and DNA (Fig. S1). Although this efficiency was rather low, scaling up twenty times could give a maximal theoretical number of transformants of $8.8 \cdot 10^8$. To prevent an even lower number of transformants when transforming ARX with ligation mixtures instead of a plasmid, *E. coli* DH10B was used to make the libraries before transferring them to ARX.

For generating *araC* variants, combinatorial site-saturation mutagenesis was chosen, because enough information about ligand binding residues was available. To make optimal use of the selection power of the dual reporter system, a large library was preferred over an iterative approach with small libraries. The codons for six residues in the ligand binding pocket were changed to NNK with N being any nucleotide and K being a G or a T, encoding all twenty amino acids with some redundancy and just a single stop codon. A similar approach previously turned out to be successful for other specificity changes of AraC in which screening instead of selection was used[87,128,129,246]. NNK at six positions gives 32^6 or $1.07 \cdot 10^9$ possibilities, just above the predicted maximum number of transformants. Although it was not expected that all variants would be covered, still six residues were chosen to optimally exploit the selection power of the system. To choose the target residues, the crystal structure of the N-terminal domain of AraC in complex with L-arabinose (PDB 2ARC[245], Fig. 1B) was inspected for interactions between the protein and the ligand. The N-terminal arm is folded over L-arabinose, with two main interactions between the arm and L-arabinose being the electrostatic interaction with the P8 backbone carbonyl and the hydrophobic interaction with F15. In between the arm and L-arabinose, there are water molecules enclosed, forming a water-mediated hydrogen bonding network. Apart from the N-terminal arm, T24, R38 and H93 interact with the L-arabinose's oxygen atoms, I46 forms hydrophobic interactions, and W95 stacks with the ligand's ring. To see which residues would be good targets for changing the specificity of AraC to D-xylose, D-xylose was aligned with L-arabinose in the pocket. In addition, conservation scores were calculated. These scores are a relative measure of evolutionary conservation per position in the primary sequence. Based on this, we excluded Y82 due to high conservation scores. Moreover, W95 was excluded due to the highly stabilizing stacking interaction. Hence, residues T24, H93, I46 and I36 were chosen as first targets, because they have a low conservation and a potential influence on the proximal monosaccharide oxygens. F15 and R38 were also included, since

they are positioned at the entrance of the pocket. A second design was made with D-fucose as target molecule for changing specificity. D-fucose is very similar to L-arabinose, but it functions as anti-inducer instead of inducer[249]. Although selection on D-fucose is not covered in this study, the D-fucose based library was also used for selection of D-xylose specific variants due to the similarity of the molecules. The only difference between D-fucose and L-arabinose is an additional methyl group attached to the ring C5 (Fig. 1C) and the two monosaccharides are bound by AraC in a very similar fashion[249]. The focus should therefore be on the residues that make van der Waals interactions with the methyl group. W95 and R38 were left untouched. This left F15, M42 and I46. I36 was added, because of the non-conserved residues it is nearest to the methyl group. To stabilize the arm for D-fucose instead of L-arabinose, residues H18 and E149 were included, because of their described effect on arm stabilization[250]. These two residues are the only targets of both designs that are not located in the first shell of the ligand binding pocket. E149 lies at the end of an α-helix of the α-helical subdomain for dimerization and it is almost contacting H18 in the N-terminal arm[250]. Altogether, two libraries were made, each with six target residues (Fig. 1B): one based on D-xylose (Lib1: F15, T24, I36, R38, I46, H93) and one based on D-fucose (Lib2: F15, H18, I36, M42, I46, E149).

To test the quality of the libraries 48 clones per library were sequenced (Tables S2 and S3). As expected for such large libraries, none of the sequences were the same, none had the complete wildtype sequence and per position, only a small number of clones had the wildtype codon (<5%) or amino acid (<10%). Per position, the diversity of codons or amino acids was high (Lib1: 63-78% of possible codons, 71-95% of possible amino acids; Lib2: 72-88% of possible codons, 67-90% of possible amino acids). However, the number of sequenced clones was too small to identify a possible bias towards particular amino acids, except for the bias created by the degenerate NNK codon. The amino acids leucine, serine or arginine, which are each encoded by three different codons when using NNK, indeed were present more than the other codons, and amino acids that were encoded by two codons were present more than those encoded by one codon. Also the percentage of clones having at least one stop codon corresponded to the expected value of 17% (Lib1: 24% of clones; Lib2: 17% of clones). The size and the degeneracy of the libraries (Table 1) were determined as the number of transformants (T) and the number of different members among the transformants ($D = D_{max} \cdot (1-e^{-T/Dmax})$ with D_{max} being the maximal number of possible variants[251]) respectively. Since 26% of Lib1 variants and 64% of Lib2 variants had insertions or deletions on off-target positions and thus no complete *araC* sequence, the library size was

corrected for these numbers. Off-target point mutations were thought of as extra variation. Although Lib2 had more transformants than Lib1, the final number of variants was lower due to the higher number of off-target mutations. Lib1 and Lib2 contained 45% and 32% of the maximal number of possible variants respectively.

Table 1. Quality control of the libraries Lib1 and Lib2

	Lib1		Lib2	
	No correction	Correction after sequencing (74%)[d]	No correction	Correction after sequencing (36%)[d]
Number of possible variants[a] = D_{max}	$1.07 \cdot 10^9$	$1.07 \cdot 10^9$	$1.07 \cdot 10^9$	$1.07 \cdot 10^9$
Library size[b] = T	$8.5 \cdot 10^8$	$6.3 \cdot 10^8$	$1.2 \cdot 10^9$	$4.1 \cdot 10^8$
% of possible variants = T/D_{max}	79%	59%	108%	39%
Degeneracy[c] = $D = D_{max} \cdot (1-e^{-T/Dmax})$	$5.9 \cdot 10^8$	$4.8 \cdot 10^8$	$7.1 \cdot 10^8$	$3.4 \cdot 10^8$
% of possible variants = D/D_{max}	55%	45%	66%	32%

[a]Six codons per library were changed to the degenerate codon NNK (N= A, G, C or T and K = G or T), giving 32^6 possible variants. [b]Number of transformants based on dilution series. [c]Number of different members among the transformants[251]. [d]74% of Lib1 and 36% of Lib2 transformants did not contain insertions or deletions.

KmR-based library selection in the presence of D-xylose

To enrich for AraC variants that are induced by D-xylose, a series of selection and screening steps were performed (Fig. 2). In the first step, the selection reporter of the dual reporter system was used. Only when a variant turns on kanamycin resistance in the presence of D-xylose, the cell should be able to survive. In this way, the large number of variants should be reduced quickly. Cells were first selected for 5 h in liquid medium, followed by selection on agar plates. This method has previously been used successfully for the enrichment of cells with L-arabinose isomerase activity[184]. For both selection steps, a high D-xylose concentration of 100 mM was chosen, because desired variants would probably not be as sensitive to D-xylose as wildtype AraC is to L-arabinose. D-xylose responsive variants were selected on 15 μg mL^{-1} kanamycin to have sufficient selective pressure. A kanamycin concentration of 5 μg mL^{-1} was proven insufficient and 30 μg mL^{-1} was assumed to be too stringent, since wildtype AraC gave insufficient kanamycin resistance up to 10 mM L-arabinose[184]. The coverage was estimated to be 10, based on the OD600 and the maximal number of possible variants (10x $1.1 \cdot 10^9$). However, the total number of variants was lower than the maximum number due to the transformation efficiency, the degeneracy and insertions or deletions on off-target positions (Table 1). The coverages were therefore 23 and 32, for Lib1 and Lib2 respectively.

Used reporter	KmR	KmR	KmR	LuxCDABE	KmR	KmR	LuxCDABE
Number of variants	dilution series	100 (50 L + 50 S)		8	4	2	2
[Km] (μg mL^{-1})	15	15	15, 20	0	10	10	0
[D-xylose] (mM)	100	100	20, 100	5, 20, 50, 100	20, 100	20, 100	5, 20, 100
Incubation time	5 h	O/N	O/N	O/N (equilibrium log)	O/N	O/N	O/N (equilibrium log)

Fig. 2. Overview of the selection and screening steps to enrich for D-xylose inducible AraC variants. The selection and screening were based on kanamycin resistance (KmR) and bioluminescence (LuxCDABE) respectively. Each step was done for both libraries. For screening, the data points were taken during the late log phase at a point for which signal production and wash out due to cell division were about equal (equilibrium log phase). The number of variants that were transferred to the next step are indicated. L, large colony; S, small colony; Km, kanamycin; O/N, overnight.

After selection in liquid medium, dilution series were not only plated in the presence of both kanamycin and D-xylose, but also without either of the two and with only kanamycin to get an idea of the total number of cells and the number of false positives (Table S4). Some general trends were seen. In Lib1 cultures with or without kanamycin and Lib2 cultures without kanamycin, the total number of cells had doubled compared to the estimated inoculum, based on both OD600 and colony count. In Lib2 cultures with kanamycin, the total number of cells was similar to the estimated start value based on colony count but doubled based on OD600. Only 0.001% Lib1 and 0.5% Lib2 cells were able to survive the kanamycin in the presence of D-xylose. Together these observations indicated that cells did not die from the kanamycin in the selection cultures, but they were unable to form colonies on the agar plates unless they were resistant to kanamycin. For both libraries, more cells were resistant to kanamycin in the presence of D-xylose than in the absence of D-xylose. These were the cells of interest. However, they should still be separated from the false positives. Lib1 and Lib2 varied in the percentage of false positives; 82% and 3% respectively. Similar differences between the libraries in the number of false positives and the total number of cells surviving kanamycin in the presence of the target molecule, were also obtained for the variants selected on L-arabinose. Whether Lib1 had a higher number of constitutive AraC variants (variants that are impaired in repression and thus bind to the I_1 and I_2 sites even in the absence of L-arabinose[242]) or more escape mutants than Lib2 could not be concluded.

Selection assay on agar plates

Ideally, the selection should be followed by screening based on bioluminescence to exclude the false positives, as was done for the enrichment of cells with L-arabinose isomerase

activity in the previous study[184]. Unfortunately, bioluminescence-based screening on agar plates was not possible, because D-xylose had a negative effect on the growth of the cells, resulting in less biomass. This difference in biomass and thus bioluminescence made a fair comparison between plates with and without D-xylose impossible. For the earlier L-arabinose isomerase experiment, this was not an issue because the differences between negative and positive cells were much larger compared to the D-xylose non-responsive and responsive AraC variants in this study. Bioluminescence-based screening in liquid medium was also not feasible at this point, because the starting OD600 should be controlled and this was unpractical for many colonies. Therefore, a KmR-based selection assay on agar plates was performed first (Fig. 3 and Fig. S2). A hundred colonies per library were picked from the selection plates with D-xylose. Since these colonies varied in size and all sizes could be of interest, fifty small (1-50) and fifty large (51-100) colonies were taken. After pre-induction, cultures were spotted on plates with 15 or 20 µg mL^{-1} kanamycin and 20 or 100 mM D-xylose. To check for false positives, non-pre-induced cultures were spotted on plates with only kanamycin. Plates with kanamycin and L-arabinose or L-rhamnose were taken as controls; L-arabinose was included to check if the variants could still respond to L-arabinose, and the non-metabolizable L-rhamnose to check the effect of just a high sugar concentration. In addition, some control strains were included that have the regulator-reporter plasmid with the wildtype *araC* (ARX pWUR768), a frameshift in *kan* (ARX pWUR776) or *sacB* instead of *araC* (ARX pWUR947). The last two were negative controls for kanamycin resistance and for leaky expression in the absence of AraC respectively. The variants of interest should be more resistant to kanamycin in the presence than in the absence of D-xylose. This phenotype was found for 85 (39 small + 46 large) Lib1 and 75 (46 small + 29 large) Lib2 variants; of these 16 (2 small + 14 large) Lib1 and 7 (6 small + 1 large) Lib2 variants were responding only to D-xylose, and not to L-arabinose or L-rhamnose. Since D-xylose only elicits very low transcription for wildtype AraC, it was not surprising that wildtype was only resistant to 15 µg mL^{-1} kanamycin in the presence of 100 mM D-xylose and not resistant to 20 µg mL^{-1} kanamycin. The negative controls could not survive the kanamycin at all. All variants grew without kanamycin, and the selection with 20 µg mL^{-1} kanamycin was more stringent than with 15 µg mL^{-1} kanamycin, leading to less false positives (variants that survive kanamycin also without any of the monosaccharides). Variants resulting from small colonies gave less false positives than variants from large colonies. For Lib1, variants from small colonies were less specific in their response to the monosaccharides, whereas variants from large colonies were more specific towards D-

xylose. For Lib2, this was the other way around. For follow-up analysis, eight Lib1 and seven Lib2 variants were chosen that did not survive Km20 without any of the monosaccharides and grew better on Km20 with D-xylose than with L-arabinose or L-rhamnose. These were Lib1-Xyl23, 33, 52, 56, 58, 65, 66 and 72 and Lib2-Xyl1, 8, 9, 31, 37, 38 and 45.

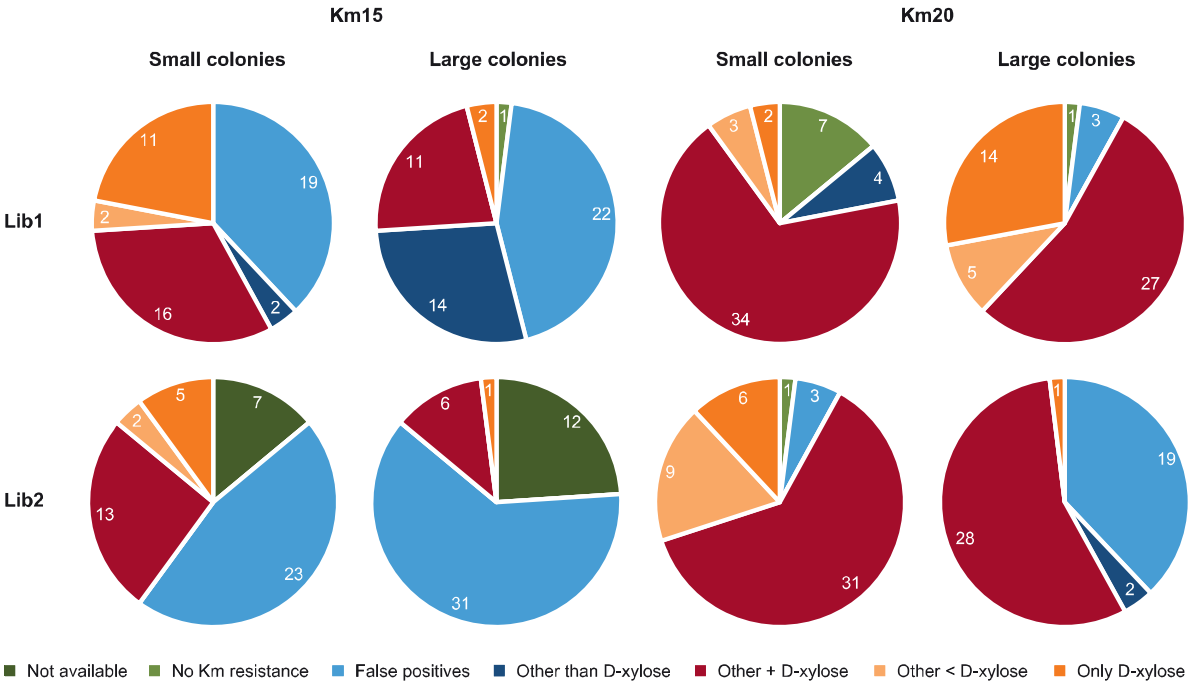


Fig 3. Analysis of D-xylose response of *araC* library variants by kanamycin resistance-based selection. The response of *araC* library 1 (Lib1) and 2 (Lib2) variants towards D-xylose was analysed by growing cells on LB medium overnight in the presence of 0, 15 or 20 µg mL^{-1} kanamycin and 0, 20 or 100 mM D-xylose. 20 mM L-arabinose and 100 mM L-rhamnose were controls. An aliquot of 5 µL cell suspension was spotted on each agar plate (Fig. S2). Each pie diagram represents 50 variants, originating from small or large colonies. The different phenotypes are presented as different colours. False positives were resistant to kanamycin independent of the presence of monosaccharides. Monosaccharide dependent resistance is indicated by 'Other than D-xylose' (only response to L-arabinose and/or L-rhamnose), 'Other + D-xylose' (similar response to D-xylose and L-arabinose and/or L-rhamnose), 'Other < D-xylose' (response to D-xylose is the strongest), 'Only D-xylose' (only response to D-xylose). Km, kanamycin.

Selection and screening assays in liquid medium

To verify the observed D-xylose response of the selected variants, a screening assay based on bioluminescence was performed (Fig. S3). For both libraries, seven or eight variants were grown in the presence of various D-xylose concentrations for 17 h. The controls were L-arabinose and L-rhamnose and the strains with wildtype *araC*, a frameshift in *luxA*, or *araC* replaced by *sacB*. The OD600 and bioluminescence were measured every 15 min. The 96-well plate was not covered with a breathable film, because this would make the measurements impossible. Instead a clear transparent hard cover was used. To see if the

bioluminescence was affected by evaporation in the outer wells of the 96-well plate, all 0 mM monosaccharide samples were included twice, one in the outer wells and one in the inner wells. Unfortunately, the evaporation in the outer wells did affect the bioluminescence that much that no definite verification of the D-xylose response was possible. Nevertheless, some things could be observed. The leakiness, the signal in the absence of any monosaccharide, differed per strain. The negative control with a frameshift in *luxA* had a negligible leakiness, as expected. The negative control with *sacB* instead of *araC* had a low leakiness, resulting from the low expression in the absence of AraC. Most variants had a higher leakiness compared to wildtype, to various extents, whereas some had a similar leakiness as wildtype and one (Lib2-Xyl37) even had a lower leakiness. None of the variants responded to any of the monosaccharides as strongly as wildtype to L-arabinose. However, the variants did seem to moderately respond to the monosaccharides, each variant in a slightly different way. The variants that were most promising in terms of D-xylose response and/or low leakiness were chosen to analyse further in a selection assay in liquid medium. These were Lib1-Xyl23, 33, 56 and 65 and Lib2-Xyl8, 9, 31 and 37.

For both libraries, the four selected variants were grown in the presence of D-xylose, L-arabinose or L-rhamnose and 0 or 10 μg mL^{-1} kanamycin for 17 h (Fig. S4). The kanamycin concentration was lower than the concentration used in the selection assay on agar plates, because the differences between the variants were more pronounced in that way. The controls were the strains with wildtype *araC*, a frameshift in *kan*, or *araC* replaced by *sacB*. L-arabinose and D-xylose, but not L-rhamnose, had a negative effect on growth (see Km0). Wildtype only responded to L-arabinose and the two negative controls to none of the monosaccharides. All variants except Lib2-Xyl37 responded to both the target molecule D-xylose and to L-rhamnose. The D-xylose response was concentration dependent. The magnitude of the response for these monosaccharides differed per variant. Lib1-Xyl56, Lib1-Xyl65 and Lib2-Xyl31 also responded to L-arabinose and for the first two, this response was lower than for the other monosaccharides. Some background was seen for 0 mM, which originated mostly from only one of the three triplicates. The variants that had more background or higher overall signal in this assay, also had more leakiness in the above described screening assay (Fig. S3).

For each library, two variants were tested in more depth for their response to the target molecule D-xylose and for the specificity of this response. To this end, the response of the chosen variants Lib1-Xyl23, Lib1-Xyl33, Lib2-Xyl8 and Lib2-Xyl9 to different L-arabinose-resembling monosaccharides (Fig. S5 for chemical structures) was analysed in both a

selection (Fig. 4) and a screening assay (Fig. 5). The tested monosaccharides D-lyxose, D-fucose and L-rhamnose cannot be metabolized by ARX[138,252,253] and D-arabinose can only be metabolized in case a mutation makes expression of L-fucose metabolic genes constitutive[254]. Uptake of D-fucose, L-rhamnose and D-arabinose should be possible, although the latter with low efficiency[128,138,252,255,256]. Uptake of D-lyxose is unknown.

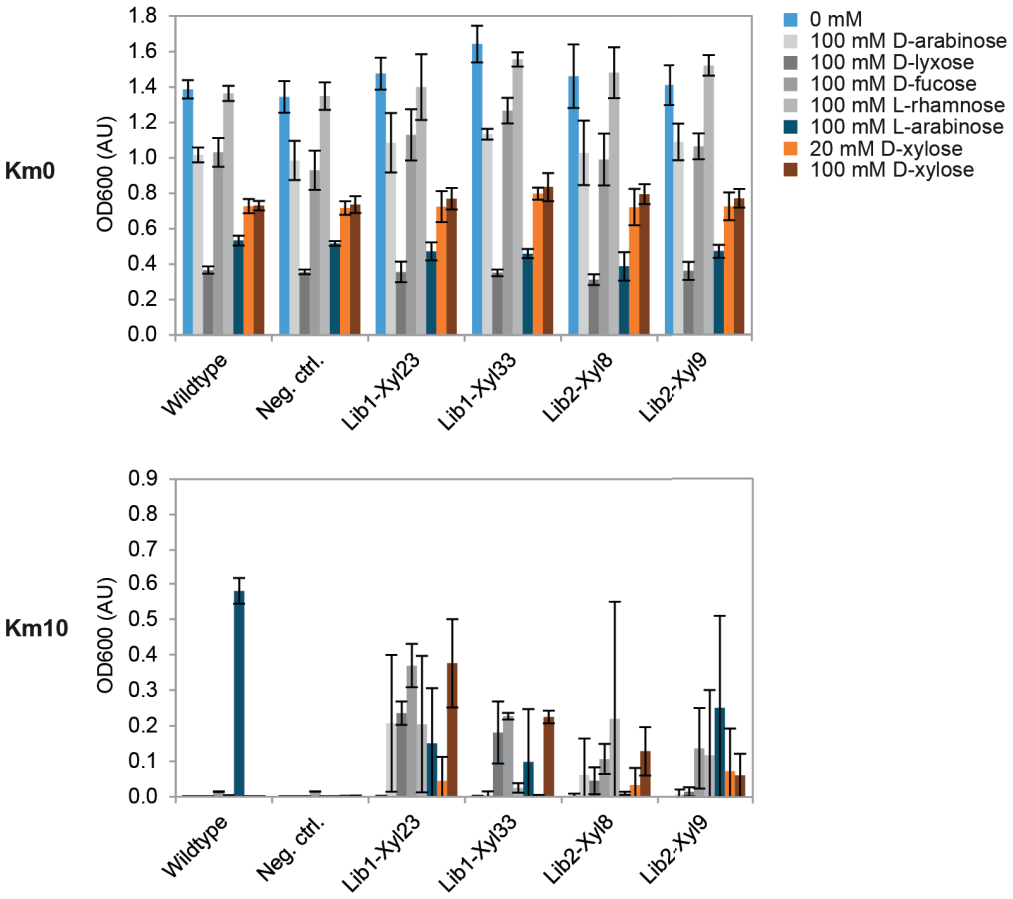


Fig. 4. Specificity determination of *araC* library variants by kanamycin resistance-based selection. The response of *araC* library 1 (Lib1) and 2 (Lib2) variants towards various monosaccharides was determined by growing cells in LB medium in the presence of 0 or 10 µg mL^{-1} kanamycin for 17 h. For each library, two independent experiments were performed in triplicate and values were averaged. Also Wildtype and Neg. Ctrl. values of Lib1 and Lib2 experiments were averaged. The final standard deviation is indicated. Wildtype: *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC*. Neg. ctrl.: ARX with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *kan*. The two graphs have a different y-axis scale.

This selection assay had the same set-up as the selection assay with four variants per library. Most of the monosaccharides had a negative effect on growth (Km0). Wildtype only responded to L-arabinose, whereas the variants responded to several monosaccharides, but to various degrees. Lib1-Xyl23 and Lib1-Xyl33 responded strongly to D-lyxose, D-fucose and

D-xylose and very little to D-arabinose and L-rhamnose. Lib2-Xyl8 responded strongly to D-fucose and D-xylose and little to D-lyxose, whereas Lib2-Xyl9 only responded slightly to D-fucose. Surprisingly, the strong response to L-rhamnose of the previous selection assay was now gone as well as the response of Lib2-Xyl9 to the target molecule D-xylose.

The screening assay (Fig. 5A) was performed as the screening assay with eight variants per library. The only exception was that the outer two rings of the 96-wells plates were not used, because evaporation from these wells was too high. Wildtype responded positively to D-fucose and D-xylose and negatively to D-lyxose. As in the other screening assay, the variants had a different leakiness than wildtype, the variants did not respond to the monosaccharides in the same way as wildtype and the responses differed from variant to variant. A response was defined as significant when the fold change of the values at 100 mM over 0 mM monosaccharide was higher than 1.1 or smaller than 0.9 (Fig. 5B).

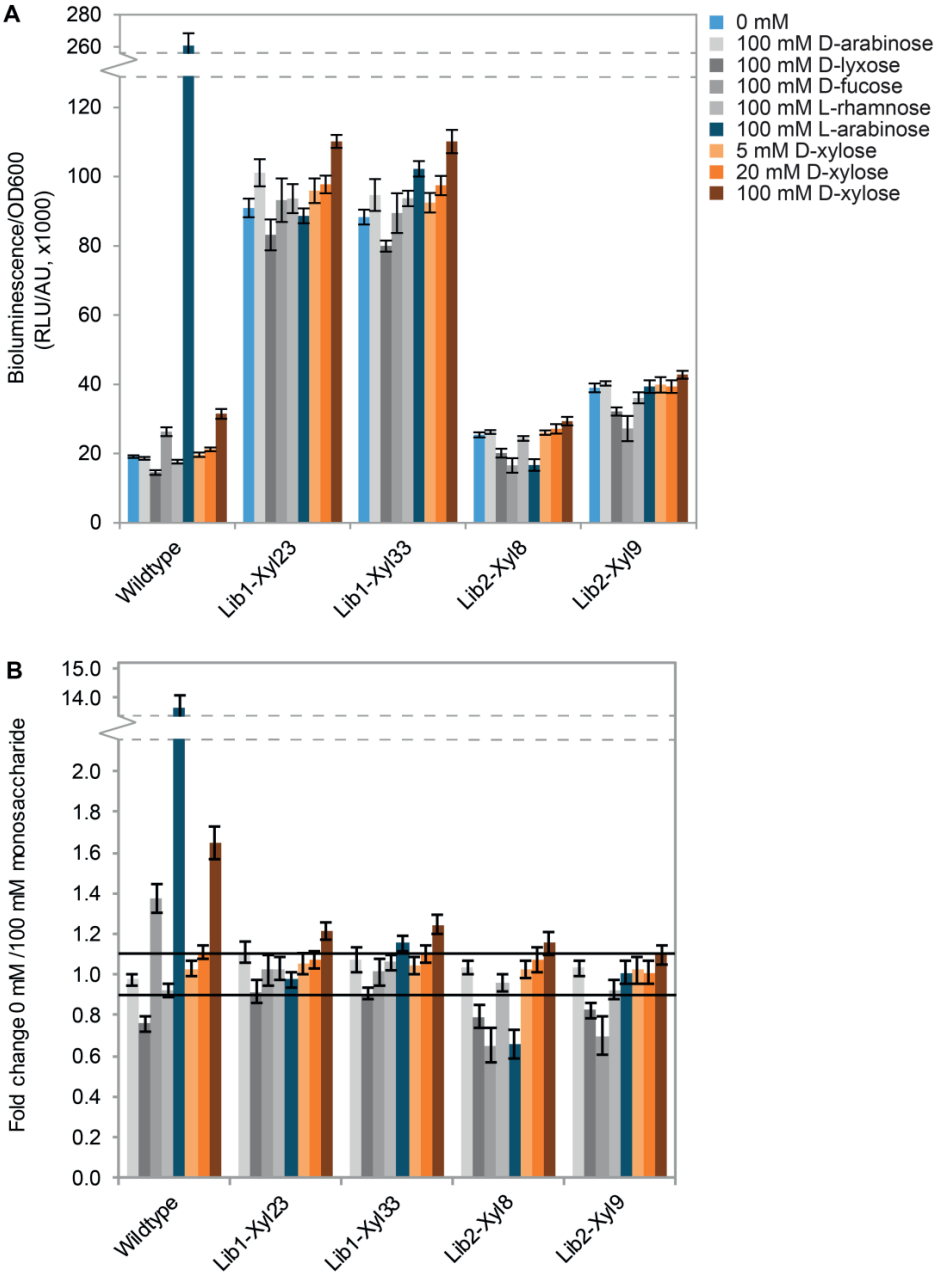


Fig. 5. Specificity determination of *araC* library variants by bioluminescence-based screening. (A) The response of *araC* library 1 (Lib1) and 2 (Lib2) variants towards various monosaccharides was determined by growing cells in LB medium for 17 h, while OD600 and bioluminescence were measured every 15 min. An average was taken of three subsequent time points in the late log phase for which the signal production and wash out due to cell division were about equal. The signal of the negative control (*E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *luxA*) was subtracted and the two 0 mM values were averaged. For each library, two independent experiments were performed and values were averaged. Also wildtype values of Lib1 and Lib2 experiments were averaged. The final standard deviation is indicated. (B) The fold change of the values at 100 mM over 0 mM monosaccharide. Boundaries were set at 1.1 and 0.9 for positive and negative responses respectively. Wildtype: ARX with the regulator-reporter plasmid with wildtype *araC*.

Sequencing of the obtained variants

To link the observed phenotypes to the structure of the variants, their *araC* genes were sequenced (Table 2). Lib1-Xyl23 and Lib1-Xyl33 showed the same nucleotide sequence, which corresponds to their very similar phenotype in the assays. Both variants had an extra mutation at an off-target position, N177S. This residue is located at the end of the linker between the two domains and at the start of the C-terminal domain. The two Lib2 variants did differ, leaving three different AraC variants. To rule out that the binding of AraC to the operators was altered by mutations in the operator regions, these regions were sequenced. Only Lib2-Xyl8 had a point mutation in between the O_1 and O_2 operators.

Table 2. Sequences of Lib1-Xyl23, Lib1-Xyl33, Lib2-Xyl8 and Lib2-Xyl9

Wildtype	F15[a]	H18[b]	T24[c]	I36[a]	R38[c]	M42[b]	I46[a]	H93[c]	E149[b]	N177[d]
Lib1-Xyl23	S	H	V	R	D	M	R	L	E	S
Lib1-Xyl33	S	H	V	R	D	M	R	L	E	S
Lib2-Xyl8	W	W	T	A	R	N	A	H	W	N
Lib2-Xyl9	N	R	T	N	R	D	A	H	A	N

[a]Purple, target residues of Lib1 and Lib2. [b]Cyan, target residues of Lib2. [c]Red, target residues of Lib1. [d]N177 off-target mutation (not present in any of 48 clones sequenced after library formation).

Discussion

In this study, growth-based selection was used to select AraC variants with specificity towards D-xylose instead of the natural ligand L-arabinose. Two combinatorial site-saturation mutagenesis libraries of *araC* were made, each having six codons that code for residues in the ligand binding pocket changed to NNK. Variants were selected based on kanamycin resistance in the presence of D-xylose. Further growth-based selection assays and bioluminescence-based screening assays allowed for exclusion of false positives, quantification and selection of three final variants, which indeed showed an altered response, albeit small, to D-xylose and to some other monosaccharides.

The two libraries differed not only in the target residues, but also with respect to the results of each step, from construction to selection and screening: (1) In the library construction, Lib1 had less transformants than Lib2, but due to less insertions/deletions on off-target positions, this library had the highest number of variants with intact genes. (2) During the initial selection on kanamycin in the presence of D-xylose, about 500 times less Lib1 clones survived than Lib2 clones. Of the survivors, 82% (Lib1) and 3% (Lib2) were false

positives. (3) In the selection assay on agar plates, less false positives were obtained from Lib1 (41% and 3% for 15 and 20 μg mL^{-1} kanamycin respectively) than from Lib2 (54% and 22% for 15 and 20 μg mL^{-1} kanamycin respectively) Why these numbers are so different from the initial selection is not clear. Perhaps it is because the conditions varied in the two assays; cells grew either in a colony or at a 5 μl spot, encountering different local conditions, like cell density. Another possibility is a difference in the type and the rate of arising escape mutants between the two libraries. That escape mutants can arise with this method was shown in previous work in which escape mutants had *kan* under the constitutive promoter instead of under the AraC-controlled promoter.

Over time the fraction of the plasmids in the cell with this mutation increased.[184] Lib1 variants that grew relatively slow on selective plates (small colonies) were less specific in their response to the target molecule D-xylose or the other tested monosaccharides, L-arabinose or L-rhamnose, whereas faster growing variants (large colonies) appeared more specific towards D-xylose. For Lib2, this was the other way around. The most promising variants that were selected for further assays mostly were obtained from large colonies for Lib1 and small colonies for Lib2. (4) In the liquid selection and screening assays, most Lib1 variants and only half of Lib2 variants had a higher leakiness than wildtype. (5) Lib1-Xyl23/33 was better induced by the target molecule D-xylose thanLib2-Xyl8 and Lib2-Xyl9, but less than wildtype. However, Lib1-Xyl23/33 was more specific towards D-xylose than wildtype, since it no longer responded to L-arabinose.Lib2-Xyl8 and Lib2-Xyl9 had a reverse response to D-fucose compared to wildtype and Lib1-Xyl23/33. Both results correspond to the monosaccharides on which the library design was based, D-xylose for Lib1 and D-fucose for Lib2, but the changes are small and only a few variants were studied in detail. Although all these differences between the libraries are likely a result of the different composition of the two libraries, variation between experiments cannot be ruled out, since all experiments including the construction of the libraries were performed only once, except for the selection and screening assays to determine the specificity of the three final variants.

The fifteen variants that where selected for further selection and screening responded differently to the tested monosaccharides than the wildtype, but the various assays gave diverse outcomes. For example, in the first selection assay (Fig. S4) all variants responded well to L-rhamnose, whereas in the second selection assay (Fig. 4) this response was less pronounced. In contrast to the earlier analyses, the assays for specificity determination (Fig. 4 and Fig. 5) were performed twice and were thus more accurately describing the responses, and hence were used as basis to compose a general overview (Fig.

6). From this overview, it is clear that the observed responses differed between the selection and the screening assays. Most striking are the relative high responses of the variants to some of the monosaccharides in the selection assay in comparison to the screening assay and the opposite responses in the two assays. The first probably originates from the chosen kanamycin concentration (10 µg mL^{-1}). At this low concentration only a low *kan* expression is necessary to confer resistance[184] and with the high leakiness of especially Lib1-Xyl23/33 (4.5x) even less extra expression is necessary than for the wildtype. The high standard deviations are also likely resulting from this low kanamycin concentration. Since the death curve is rather steep around this concentration, only a very minor difference in concentration could already result in growth. Increasing the concentration was tried, but then the small responses did not lead to any resistance. How the responses to for example D-lyxose could be positive in the selection assay and negative in the screening assay, cannot be explained. Evaluating the results in a simpler system with only one reporter would be a next step. Despite these inconsistencies, what can we say about the responses of the wildtype and the three final variants? The variants had an altered response to the tested monosaccharides compared to the wildtype and a higher leakiness). The wildtype AraC was very well induced by the natural ligand L-arabinose, only very little by D-fucose and D-xylose, and was repressed by D-lyxose. The slight inducing effect of D-xylose and no response to D-arabinose have been observed before, but the responses to D-fucose and D-lyxose have previously been described as anti-inducing and very low inducing respectively[128,249]. As no contamination with L-arabinose could be detected neither for D-fucose nor for D-xylose (not shown), the discrepancy of the wildtype responses with literature data remains elusive. Lib1-Xyl23/33 was slightly induced by the target molecule D-xylose and by D-arabinose. Since the response to L-arabinose was gone, its relative specificity towards D-xylose over L-arabinose was higher than that of wildtype, as was aimed for. The additional response to D-arabinose is also interesting to develop a D-arabinose responsive variant. Lib2-Xyl8 was slightly induced by the target molecule D-xylose and repressed by L-arabinose and D-fucose. Lib2-Xyl9 was repressed by D-fucose. The obtained Lib2 variants are therefore less interesting in respect to D-xylose, but the inverse reaction towards D-fucose is interesting and also promising for a future selection of Lib2 on D-fucose. The possible effects of the other monosaccharides on the variants was less clear, but in general the variants seemed to respond to more monosaccharides than wildtype. This phenomenon of creating generalists instead of specialists is commonly observed in directed evolution approaches and enrichment of specialists often requires at least one more round of library formation[220,257]. However, it is

not impossible to find a specialist in the first library[183]. To cover both possibilities, two types of variants should be selected from the selection assay on agar plates; variants that show a better response to D-xylose than to the other monosaccharides and variants that respond similarly to different monosaccharides. Maybe some of the latter would have a stronger response to D-xylose or a lower leakiness than the final variants.

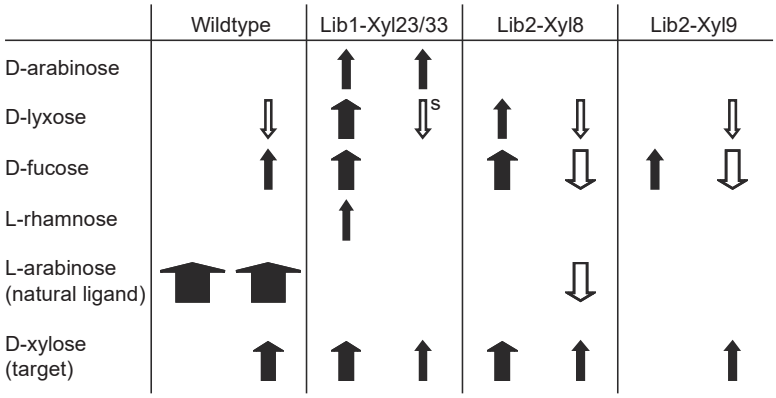


Fig. 6. Overview of the monosaccharide specificity of *araC* variants Lib1-Xyl23/33, Lib2-Xyl8 and Lib2-Xyl9. This is an overview of the data presented in Fig. 4 and Fig. 5. For each strain, the left arrow indicates the response in selection and the right arrow the response in screening. The size of the arrow represents the strength of the response (not to scale). Black upward and white downward directed arrows indicate positive or negative responses respectively. s: smaller response than wildtype. Wildtype: *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC*.

Although the monosaccharide response profile of the final variants is different from that of the wildtype, no definite conclusions on altered specificity could be drawn. The differences are only small, except for the reduced response to L-arabinose and the increase in leakiness. Does this mean that the variants have instead of an altered specificity merely an impaired repression and a disrupted response to L-arabinose? If this were the case, it should be due to mutations inside AraC, since the operators do not contain mutations. Variants that are impaired in repression and thus bind to the I_1 and I_2 sites even in the absence of L-arabinose are called constitutive, also when further induction with L-arabinose is possible. Most of these variants have mutations in the N-terminal arm (e.g. H18R), but mutations at other locations have been found as well, including mutations in the linker between the two domains and within the dimerization domain[242,250,258]. Although the leakiness of the selected variants is indeed increased compared to that of the wildtype, no definite statement can be made on impaired repression or altered ligand specificity. Some mutations that are found in the final AraC variants disrupt the ligand induction (F15S[243], H18W[258], I46A[241]) when present as single mutations, but in these variants the mutations are combined with others and it is therefore impossible to conclude what the combinatorial influence is. It is possible that in

certain combinations the mutations may have the opposite effect[87]. Also the combinatorial effect on expressibility, solubility, overall folding or local structure should be considered. Nothing conclusive can be said about the mutations in the variants in this study compared to those in other engineered variants, like the D-arabinose responsive variants described by Tang *et al.*[128,129,246,259].

To determine the nature and relevance of the obtained variants it would be good to perform further experiments. A first verification could be a dose-dependent response to D-xylose using a simpler *in vivo* method or even an *in vitro* transcription-translation assay with GFP as single reporter[260]. The individual contribution of each of the six mutations could be investigated by making single mutations or by reverting mutations back one by one to wildtype. To proof altered ligand specificity instead of changed dynamics of the N-terminal arm, several *in vitro* assays like binding assays, isothermal titration calorimetry, tryptophan fluorescence quenching[259] or anisotropy[242] might proof useful. Unfortunately, only the N-terminal domain could be used for this due to the poor solubility of the whole protein[242]. Once interaction between the ligand and AraC will be established, the variants could be investigated computationally to get an idea on what the effect of the mutations is with and without ligands. Methods like Molecular dynamics (MD) and variants thereof such as Self-guided Langevin dynamics (SGLD)[241] are interesting in this respect. In addition, protein structures of the variants would be useful to determine the interactions of these variants with their ligand. The structure could also be a good starting point for a new round of library generation and selection/screening to further improve the strength of the response as well as the specificity towards D-xylose versus other monosaccharides. In this respect, especially a structure of Lib1-Xyl23/33 would be useful.

Although our dual selection and screening method resulted in some potentially interesting variants, some comments regarding the library design and selection can be made. The choice to use the degenerate NNK codon at the target positions reduced the maximum number of possible variants on the protein level, possibly leading to more screening costs to find the optimal variant. However, a more focused non-degenerate library, for instance designed with DYNAMCC_0 (algorithm to design a specific set of degenerate codons for a defined collection of amino acids without redundancy) or made by ProxiMAX (synthesis method that adds one codon per cycle of ligation, amplification and digestion; each amino acid is encoded by one codon only), is often more expensive by requiring more oligonucleotides, especially when multiple codons are targeted[239,261]. It is a trade-off between library costs and screening costs, but to simply show what our dual

selection and screening system is capable of, the NNK method sufficed. For the same reason no iterative approach with multiple small libraries was used. Since sufficient literature is available on the AraC ligand binding pocket, a rational choice of target positions was possible, alleviating the need for using methods like alanine scanning. The chosen positions indeed resulted in some interesting variants, Lib1 for D-xylose and Lib2 for D-fucose, confirming that the used design approach was appropriate, but one cannot rule out that a different set would have given better variants. For further fine-tuning based on minute secondary structure rearrangements, it would be good to address the sequence neighbouring residues of promising residues. The quality of the libraries was sufficient for the method used. With this strain and plasmid, a higher transformation efficiency is not possible and the relatively high number of off-target mutations is as expected for commercial libraries created for *de novo* DNA synthesis. Although Lib1 and Lib2 contained only 32% and 45% of the maximum number of possible variants respectively, the sizes were still large enough to obtain some interesting variants.

The selection on kanamycin in the presence of D-xylose does require some optimization. Before starting the selection, it would be good to do dilution series without kanamycin and D-xylose, with only kanamycin and with both. This would allow determination of some more information about the library, enabling calculations of the enrichment during selection of the number of leaky and/or responsive variants over the number of non-leaky and/or non-responsive variants. The D-xylose concentration of 100 mM might be too high, increasing the risk of false positives, especially since the wildtype was also slightly induced by D-xylose[259]. In the selection assay on agar plates, most variants responded to 20 mM D-xylose, but were more resistant to 100 mM. Using 20 mM in the initial selection might be a good way to reduce false positives, despite the fact that some variants with only a small response to D-xylose will be missed. The kanamycin concentration of 15 μg mL^{-1} is probably good for the selection in liquid, but for the subsequent selection on agar plates a concentration of 20 μg mL^{-1} is better to reduce the number of false positives, based on the selection assay on agar plates. A further increase of the kanamycin concentration is not recommended, because of the risk of missing the variants with low sensitivity. Since in agar plates local concentration differences occur due to cell growth and colony formation, slight differences in resistance between liquid and solid cultures are to be expected. For the selections and screenings, cells were grown in LB medium. As negative control for induction, no ligand was added, but to be absolutely sure that AraC is not responding to any of the medium components, minimal medium should be used, at least

once for verification. The faster growth rate in LB medium, however, is a big advantage during all selection and screening steps. In addition, this regulator-reporter plasmid was chosen for use in LB medium. Going to minimal medium might require optimization of the plasmid and for sure characterization of the monosaccharide responses. For further testing of the selected variants, it is best to pick the small colonies, because the larger colonies are often false positives.

The selection and screening assays that were performed to exclude false positives can be done more straightforward based on the options explored in this study. The selection assay on agar plates should be done on 20 μg mL^{-1} kanamycin only, since this gave less false positives compared to 15 μg mL^{-1} kanamycin. Based on this assay, variants should be picked for a screening assay in liquid medium in 96-wells plates and grown in the plate reader, using only the inner wells to get the best reliability. More variants could be included when only testing with and without the target molecule and no other molecules, except for the positive control of wildtype with L-arabinose. It could be that after all optimizations of the protocol still too many false positives are obtained, for example when the number of constitutive mutants in the libraries is so high that these mutants dominate in the initial selection. If this happens, a combination of positive and negative selection might proof useful, as recently used in other studies[257,262]. In that case, another selection reporter should be used that also allows for negative selection.

Conclusions

The dual selection and screening system employed here, is a promising method for the selection of AraC variants with an altered ligand specificity. Selection based on growth allows for a rapid reduction of the initial large library, and subsequent screening based on bioluminescence excludes false positives and makes quantification possible. Indeed some variants were obtained that have an altered, albeit small, response to several monosaccharides. Lib1-Xyl23/33 has a higher specificity for D-xylose or D-arabinose over L-arabinose than wildtype, whereas Lib2-Xyl8 and Lib2-Xyl9 have an inverse response to D-fucose compared to wildtype. However, these variants should be investigated in more depth to determine whether their ligand specificity is truly modified. The rational target choice seemed to target the right positions, but to get better variants, the selection and screening set-up should be optimized, for example by increasing the stringency of selection. After optimization, this same set-up could be used to select not only AraC variants with a better

response to D-xylose, but also variants specific to other target molecules.

Materials and methods

Bacterial strains and media

E. coli DH10B $T1^R$ (C6400-03, Invitrogen) was used for plasmid propagation and was grown and transformed by standard methods[160]. *E. coli* BW25113 Δ*araC* Δ*recA* (AR) was the parent strain for *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX), which hosted the regulator-reporter plasmid, the control plasmids or the library plasmids. Cells were grown at 37°C in LB medium with 34 µg mL^{-1} chloramphenicol (indicated as LB-Cm), unless stated otherwise. Transformations were done as described in **Chapter 3**[184]. OD600 was measured with a Biowave Cell Density Meter CO8000 (Biochrom). D-arabinose (10850, Fluka Analytical), D-lyxose (220477, Sigma-Aldrich), D-fucose (2256300050, Acros Organics), L-rhamnose (W373011, Sigma-Aldrich), L-arabinose (5118.3, Carl Roth) and D-xylose (W360600, Sigma-Aldrich).

Host strain construction

The host strain *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) was constructed from AR and verified as described for AR in **Chapter 3**[184]. The only exceptions were the L-arabinose concentration (20 µM instead of 10 mM) for induction of the Red system genes (*γ*, *β*, and *exo*), the sequence of the disruption cassette and the polymerase that was used to create the fragments for sequencing (Phusion HF instead of Pfu; both Thermo Scientific). The disruption cassette was made by PCR with as template the linear fragment presented in Table S5, which has *cat* instead of *kan* flanked with *lox*71(left)/*lox*66(right) sites[140]. The homologous regions that were introduced during PCR were 50 nucleotides long and inside the *xylA* CDS, leaving, after recombination, the first 54 and last 78 nucleotides. All primers are presented in Table S6. The inability of ARX to metabolize D-xylose was verified by growth in 2 mL M9 medium, with or without 22.2 mM D-xylose or D-glucose. AR and *E. coli* BW25113 JW3537-1 (*ΔxylA::kan*) were taken as controls. The OD600 was measured after 22 h growth.

Vector construction

The vector pWUR947 was made from the regulator-reporter plasmid pWUR768[184] by replacing the *araC* gene by *sacB* (inclusive AscI site downstream of *sacB*) with CpoI/PvuI. The

sacB insert was made by two subsequent PCRs with Pfu to remove the KpnI site from the CDS (silent mutation, TAC --> TAT) from *B. subtilis* subsp. *subtilis* str. 168[263]. Left and right fragments were formed with primers BG5913/BG5916 and BG5915/BG5914 respectively, and combined with primers BG5913/BG5914 (Table S6). Products were purified with the GeneJET PCR Purification Kit (K0702). The RsrI/PvuI digested vector was treated with Antarctic Phosphatase (NEB), according to the protocol of NEB, and extracted from an agarose gel with the Zymoclean Gel DNA recovery Kit (D4002, Zymoresearch). The fragments were ligated for 1 h at room temperature with T4 ligase. pWUR947 was verified by PCR with DreamTaq, by restriction analysis and by sequencing at GATC Biotech. Plasmids were isolated with the Plasmid Miniprep Kit (K0503). All enzymes and kits came from Thermo Scientific, unless stated otherwise.

Library design and construction

AraC protein structures were visualized with the PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC. Conservation scores were calculated using the ConSurf server (http://consurf.tau.ac.il/2016/). For library construction, pWUR947 was isolated with the JETstar Plasmid Maxiprep Kit (220020, Geneprice) and send to Baseclear. The library fragments were made using an in-house *de novo* gene synthesis approach, based on Ligase Chain Reaction, and using degenerate NNK codons at the target positions (library 1 (Lib1): F15, T24, I36, R38, I46, H93 and library 2 (Lib2): F15, H18, I36, M42, I46, E149). After PCR amplification of the library fragments and subsequent CpoI/SgsI digestion, library amplicons and digested vector were purified from an agarose gel. The vector was dephosphorylated prior to overnight ligation at 16°C. ElectroMAX DH10B electrocompetent *E. coli* was transformed with the ligation mixtures and transformants were plated on LB-Cm agar plates. The results were approved only if the empty vector control yielded <1% of the number of colonies of the library ligation plates. Per library plate, 48 clones were picked, grown overnight in LB-Cm medium, and plasmids were isolated for Quality Control Sanger sequencing. The libraries were sequenced using two short Sanger runs per clone, covering the entire cloned library fragments. Sanger traces were analysed using SeqScape software. About half of the clones (20 of 48 Lib1 clones and 26 of 48 Lib2 clones) had no additional non-silent mutations on positions other than the target positions. Upon passing quality control, ElectroMAX DH10B electrocompetent *E. coli* was transformed with the remaining ligation mixtures for both libraries at a scaled up level, recovered after transformation in LB-Cm medium, pelleted by centrifugation and resuspended in LB medium with 12.5% glycerol.

The glycerol stocks were verified to yield at least $1 \cdot 10^{10}$ CFU per library.

To transfer the libraries to the host strain ARX, for each library, 400 mL LB-Cm in a 2-L erlenmeyer was inoculated with halve of the library glycerol stock. After 23 h growth, plasmids were isolated (JETstar Plasmid Maxiprep Kit, 220020, Geneprice) from 100 mL of the culture. This was repeated for the other halve of the glycerol stock. DNA was verified by restriction analysis. ARX cells were grown in 2100 mL 2xYP medium to an OD600 of 0.4 and made electrocompetent, resulting in a final suspension of 8 mL ARX electrocompetent cells. These were transformed with ~40 µg of library plasmids (~20 µg from each isolation) in twenty individual electroporations of 400 µL. Directly after each electroporation, 1 mL of LB was added. All transformations were pooled and 12 mL LB was used to wash out the cuvettes, resulting in a recovery culture of 40 mL total. After 1 h 37°C, 1960 mL LB-Cm medium with 5% sucrose in a 5-L erlenmeyer was inoculated with the recovery culture to amplify the library. Directly after inoculation, a dilution series on LB-Cm agar plates was performed in duplicate, starting from two times 5 mL of the 2-L culture. Dilutions were added to the culture, not to lose too many variants. After overnight growth, OD600 was measured and again dilution series were performed in duplicate, starting from two times 10 mL of the culture. The rest of the culture was centrifuged and resuspended to get 50 mL with 20% glycerol, aliquoted per 1 mL and frozen at -80°C. Dilution series were performed in duplicate from the frozen aliquot. Per non-amplified library, 48 colonies were sent to GATC Biotech for sequencing with primers BG3942 and BG8211 (Table S6).

Library selection

Per library, 100 mL LB-Cm medium with 4 g L^{-1} glycerol in a 1-L erlenmeyer was inoculated with a 1 mL aliquot of the amplified library. After overnight growth, the OD600 was measured and pre-cultures were diluted into the selection cultures to a starting OD600 of 2.2, which should contain ~$1.1 \cdot 10^{10}$ (10x the theoretical number of possible variants). Cells were grown for 6 h in 25 mL LB-Cm with 4 g L^{-1} glycerol and 100 mM D-xylose in a 250-mL erlenmeyer. After 1 h, 15 µg mL^{-1} kanamycin was added to start the selection. Control cultures of 25 mL with MQ or 100 mM L-arabinose instead of D-xylose were taken along, as well as control cultures of 1 mL without kanamycin with either MQ, 100 mM L-arabinose or 100 mM D-xylose. After growth, OD600 was measured and the selection was continued by dilution series on LB-Cm agar plates, LB-Cm agar plates with 15 µg mL^{-1} kanamycin, and LB-Cm agar plates with 15 µg mL^{-1} kanamycin and 100 mM D-xylose or L-arabinose, depending on the monosaccharide in the selection culture. Non- and pre-induced controls (ARX with

pWUR768[184] (regulator-reporter plasmid with wildtype *araC*), pWUR776[184] (regulator-reporter plasmid with wildtype *araC* and a frameshift in *kan*) or pWUR947 (regulator-reporter plasmid with *sacB* instead of *araC*)) were streaked on agar plates from the same batch to verify the selective conditions. Of the cultures with kanamycin, 12.5 mL were centrifuged and each pellet was resuspended to a final volume of 750 µL incl. 20% glycerol, which was frozen at -80°C. From these glycerol stocks, selection on LB-Cm agar plates with 15 µg mL^{-1} kanamycin, and LB-Cm agar plates with 15 µg mL^{-1} kanamycin and 100 mM D-xylose or L-arabinose was repeated with different dilutions to get separate colonies.

Selection assay on agar plates

Hundred colonies, fifty small and fifty large colonies, were picked from the selection plates (LB-Cm agar plates with 15 µg mL^{-1} kanamycin and 100 mM D-xylose). Together with the controls ARX pWUR768, ARX pWUR776 and ARX pWUR947, they were pre-grown in 500 µL LB-Cm medium in 2-mL 96-well MASTERBLOCKS (780271, Greiner Bio-One), overnight. After 100x dilution (twice 10x) in 96-well plates (655161, Greiner Bio-One), they were grown in these plates for 1 h in 200 µL LB-Cm medium with or without 20 mM D-xylose, L-arabinose or L-rhamnose. 5 µL of the cultures with MQ were spotted on LB-Cm agar plates with 0, 15 or 20 µg mL^{-1} kanamycin, whereas 5 µL of the cultures with D-xylose, L-arabinose or L-rhamnose were spotted on LB-Cm agar plates with 15 or 20 µg mL^{-1} kanamycin and with 20 or 100 mM D-xylose, 20 mM L-arabinose or 100 mM L-rhamnose respectively. After overnight growth, photos were made with a G:BOX Chemi XT4 (Syngene).

Selection and screening assays in liquid medium

Before starting the assays, monoclonal glycerol stocks were made. The selection assays were performed as the KmR-based induction assays in **Chapter 3**[184] with a few alterations. Instead of pre-induction, 10 µg mL^{-1} kanamycin was added after 1 h of growth to allow induction of *kan*. Next to 100 mM L-arabinose, more monosaccharides were included, namely 100 mM D-arabinose, D-lyxose, D-fucose, L-rhamnose and 20 and 100 mM D-xylose. The assays were performed in triplicate within one experiment.

For the screening assays, cells were first pre-grown on LB-Cm agar plates overnight and then in 2 mL LB-Cm medium in 10 mL tubes (TP10-01, Gosselin) for 6 h. After OD600 measurements, cells were diluted to have an equal starting OD600 of 0.0001 in the assay. They were grown in a white 96-well plate with transparent bottom (6005181, Perkin Elmer) in the inner wells (C3-C11, D3-D11, E3-E11 and F3-F11) only. Each inner well contained 200

µL LB-Cm with 0 (in duplicate), 5, 20 or 100 mM D-xylose or 100 mM D-arabinose, D-lyxose, D-fucose, L-rhamnose or L-arabinose. The other wells were not used due to excessive evaporation and were therefore filled with MQ. Cells were grown in a Synergy MX microplate reader (BioTek) with continuously medium shaking for 17 h. Every 15 min bioluminescence (gain 185, sensor distance 4.5 mm) and OD600 were measured. Both bioluminescence and OD600 values were corrected for an average of four blanks and the bioluminescence was corrected for the OD600. An average was taken of three subsequent time points in the late log phase for which the signal production and wash out due to cell division were about equal. The signal of the negative control ARX pWUR780 was subtracted and the two 0 mM values were averaged. The experiment was performed twice and values were averaged.

Sequencing of obtained variants

Plasmids from ARX Lib1-Xyl23, Lib1-Xyl33, Lib2-Xyl8 and Lib2-Xyl9 were isolated with the Plasmid Miniprep Kit (K0503, Thermo Scientific) and sequenced at GATC Biotech with primers BG3652, BG3943 and BG8211 (Table S6).

Acknowledgements

We would like to thank Tjerk Sminia for input from an organic chemistry viewpoint.

Supplementary information

Table S1. SacB-based counter selection

	OD600 (AU)[a]									
Sucrose (%)	0	0.1	0.25	0.5	1	2.5	5	6	7.5	10
AR pWUR768 (*araC*)[b]	1.059	1.096	1.086	1.096	1.098	1.076	1.002	0.964	0.905	0.710
AR pWUR947 (*sacB*)[c]	1.083	0.786	0.002[d]	0.000[d]	0.000[d]	0.000[d]	0.000[d]	0.001[d]	0.000[d]	0.000[d]

[a]Cells were first pre-grown on LB-Cm agar plates overnight and then in 2 mL LB-Cm medium in 10 mL tubes (Gosselin) for 8 h. After OD600 measurements, cells were diluted to have an equal starting OD600 of 0.0001 in the assay. They were grown in 200 µL LB-Cm in a 96-well plate (Greiner Bio-one, 655101) in a Synergy MX microplate reader (BioTek) with continuously medium shaking for 17 h. Every 10 min OD600 was measured. OD600 values were corrected for an average of four blanks. Similar results were obtained when using BB-Cm medium[264], except for very limited growth of AR pWUR947 (OD600 = 0.011) in presence of 0.1 % sucrose. [b]AR pWUR768 (*araC*): *E. coli* BW25113 Δ*araC* Δ*recA* with the regulator-reporter plasmid with *araC*. [c]AR pWUR947 (*sacB*): AR with the regulator-reporter plasmid with *sacB*. [d]These samples were transferred to fresh LB-Cm medium and no growth was observed after 17 h.

Table S2. Target and off-target mutations in the *araC* coding sequence of 48 clones of Lib1

Clone	Target mutations[b]						Off-target mutations[c]
	F15	T24	I36	R38	I46	H93	
1	R	S	T	P	T	I	H18P, T50N
2	V	M	T	V	G	W	
3				V			
4	C						
5[a]	M	G	L	Q	T	A	A102S
6	M	L	W	Q	T		ins. 1 nt 279-280
7	S	W	D	N	H	Y	
8	R	R	L	L	R	G	
9	L	N	T	T	Q	R	
10[a]	H		W				
11[a]	H	H	P		R	W	A17G, V20G; sil. nt 381; del. nt 46
12	P	T	E	P	K	C	sil. nt 168; ins. nt 381-2
13	W	L	E	L	I	I	
14[a]		Y		V	K	A	N16K, V20G, Q60R, E63K, R99P, L133V; sil. nt 184, 444; ins. 1 nt 38-9
15	*	G	L	T	G	K	
16	L	M	A	L	R	N	V57L, V65I; sil. nt 120; del. nt 111, 483; ins. 1 nt 252-3, 461-2, 470-1
17	L	L	L	T	S	W	C280S
18	L	G	R	W	L	E	G118E; del. 6 nt 799-804
19	V	I	*	T	S	T	
20	G	*	*	F	Q	P	
21		S	M	G	P	P	H203N; del. nt 391
22	R	Q	S	L	*	L	
23	D	R	A	V	N	*	ins. 1 nt 27-28
24	G	V		S			N59I, R67G; del. nt 236
25[a]					P	D	F265I
26	C	S	F	K	R	Q	del. nt 506
27	G	C	T	F	L	C	
28	R	D	S	C	T	L	
29					M	G	T50N
30	E	V	L	C	H	D	
31	T		T	R	T	P	del. nt 520
32	K	T	L	R	R	V	
33	S	E	S	C	A	L	sil. nt 204
34	M	F	*	S	E	A	
35	M	F	*	S	E	A	

Table S2 continued

Clone	Target mutations[b]						Off-target mutations[c]
	F15	T24	I36	R38	I46	H93	
36	V	T	E	S	D	F	del. nt 373
37				F			
38	V	C	S	F	P	R	A185G
39	L	A	Q	L	R	W	
40	A	P	M	L	S	P	del. nt 66
41	G	*	R	V	S	L	A101P
42	R	T	V	R	P	I	G33S
43	T	I	S	K	A	L	E165K
44	G	L	S	E	S	L	
45	V		Y	V	F	G	del. nt 68-69, 73, 684; ins. 1 nt 221-2
46	S	S	Q	G	E	V	
47	*	P	V	S	L	G	D176N
48	L	M	G	H	I	R	
Possible codons/aa (%)[d]	69/76	75/95	69/76	63/71	78/81	78/86	
Wildtype codons/aa (%)[e]	0/0	3/10	0/0	0/7	5/5	0/0	

[a]Incomplete sequence. [b]At empty spots, incomplete sequence or partial or complete deletion of codon. *, stop codon. [c]nt, nucleotides that have a silent mutation (sil.), are deleted (del.) or in between which an insertion is located (ins.). [d]The percentage of the total number of possible codons (32) or amino acids (21 incl. stop codon) per position. [e]The percentage of the clones that have the wildtype codon or amino acid.

Table S3. Target and off-target mutations in the *araC* coding sequence of 48 clones of Lib2

Clone	Target mutations[b]						Off-target mutations[c]
	F15	H18	I36	M42	I46	E149	
1	C	R	A	A	Q	L	R231L; del. nt 162, 478
2	M	L	R	F	S	T	del. nt 263
3	E	L	R	*	I	H	
4	K	K	*	S	T	T	
5[a]	L	G	*	T	I	H	
6	K	S	T	K	D	T	del. nt 27
7				S	T	*	R62Q; del. nt 1-108, 733-7
8	W	T	R	N	V	A	N29I; sil. nt 709; del. nt 470
9	Q	Y	K	T	Q	E	S199R
10	R		L	P	R	D	del. nt 53-4, 172, 579
11	L	R	R	Y	H	T	ins. 1 nt 483-4
12				D	S	C	del. nt 1-108
13	F	S	T	L	S	S	del. nt 190
14	T	L	Y	S	V	S	E165K
15	N	L	L		A	L	G41R; del. nt 230-1, 249
16	L	V	S	W	L	Y	del. nt 400; ins. 1 nt 516-7

Table S3 continued

Clone	Target mutations[b]						Off-target mutations[c]
	F15	H18	I36	M42	I46	E149	
17				T	A	R	del. nt 1-108, 702; ins. 1 nt 831-2
18	G	R	L	V	R	W	T50A, R231F
19[a]	C	S	C	W	H	C	S14P; del. nt 33
20	G	H	L	S	M	C	sil. nt 540; ins. 1 nt 821-2
21				P	H	I	sil. 453; del. nt 1-108
22				G	K	S	del. nt 1-108
23	R	L	D	L	P	Y	V200D
24	H	P	A	P	P	H	sil. nt 585; del. nt 403
25				C	*	L	del. nt 1-108, 589
26[a]	R	F	V	V	N	R	ins. 1 nt 171-2
27	L	H	L	T	A	R	Q230L
28	*	G	*	C	T	T	
29	V	T	K	E	L	G	
30	I	L	*	I	G	L	R146H; del. nt 248-250; ins. 1 nt 814-5
31				P	L	F	del. nt 1-108, 243
32	E	N	G	T	C	H	
33	C	L		K	G	D	del. nt 212, 39-40, 483; ins. 1 nt 721-2
34	I	D	N	P	S	G	G83C; del. nt 157
35		D	R	T	A	C	del. nt 149
36	I	D	S	T	A	P	ins. 1 nt 21-22, 718-9
37	A	Y	S	V	G	T	
38	M	L	N	R	Q	*	Q94K; del. nt 478
39	H	S	G	P	S	A	
40	A	M	P	G	V	G	P273S
41				I	W	G	G44N; sil. nt 627; del. nt 1-108
42	S	V	L	R	L	V	del. nt 427-8
43				R	P	G	P100L; del. nt 1-108
44	S	A	R	F	S	A	
45	P	G	L	T	L	A	H213R; del. nt 111
46	L	A	A	V	S	R	sil. nt 149
47	V	S	R	M	V	S	
48	P	Q	C	S	M	S	S131N; del. nt 433
Possible codons/aa[d]	81/90	75/76	72/67	81/90	88/86	78/81	
Wildtype codons/aa[e]	3/3	5/5	0/0	2/2	4/4	0/2	

[a]Incomplete sequence. [b]At empty spots, incomplete sequence or partial or complete deletion of codon. *, stop codon. [c]nt, nucleotides that have a silent mutation (sil.), are deleted (del.) or in between which an insertion is located (ins.). [d]The percentage of the total number of possible codons (32) or amino acids (21 incl. stop codon) per position. [e]The percentage of the clones that have the wildtype codon or amino acid.

Table S4. Cell count after selection

Culture[a]	Cells mL^{-1} of selection culture after colony count on different agar plates[b]				
	Km0[c]	Km15[c]	Km15[d]	Km15 + L-arabinose[d]	Km15 + D-xylose[d]
Lib1, Km0 + MQ	$8.9 \cdot 10^8$	$8.6 \cdot 10^3$	NA	NA	NA
Lib1, Km0 + L-arabinose	$6.8 \cdot 10^8$	$1.3 \cdot 10^4$	NA	NA	NA
Lib1, Km0 + D-xylose	$6.0 \cdot 10^8$	$8.6 \cdot 10^3$	NA	NA	NA
Lib1, Km15 + MQ	$9.4 \cdot 10^8$	$8.8 \cdot 10^3$	NA	NA	NA
Lib1, Km15 + L-arabinose	$1.1 \cdot 10^9$	$4.2 \cdot 10^4$	$8.7 \cdot 10^3$	$9.1 \cdot 10^3$	NA
Lib1, Km15 + D-xylose	$1.6 \cdot 10^9$	$1.5 \cdot 10^4$	$1.2 \cdot 10^4$	NA	$1.4 \cdot 10^4$
Lib2, Km0 + MQ	$6.4 \cdot 10^8$	$7.0 \cdot 10^4$	NA	NA	NA
Lib2, Km0 + L-arabinose	$1.2 \cdot 10^{10}$	$8.8 \cdot 10^4$	NA	NA	NA
Lib2, Km0 + D-xylose	$1.6 \cdot 10^9$	$4.2 \cdot 10^4$	NA	NA	NA
Lib2, Km15 + MQ	$7.1 \cdot 10^8$	$9.4 \cdot 10^4$	NA	NA	NA
Lib2, Km15 + L-arabinose	$2.5 \cdot 10^8$	$8.6 \cdot 10^4$	$6.0 \cdot 10^4$	$6.0 \cdot 10^5$	NA
Lib2, Km15 + D-xylose	$4.1 \cdot 10^8$	$6.6 \cdot 10^4$	$3.1 \cdot 10^4$	NA	$9.6 \cdot 10^5$

[a]*araC* variants from libraries Lib1 and Lib2 were selected in liquid LB with 4 g L^{-1} glycerol based on kanamycin resistance in presence of 100 mM D-xylose, 100 mM L-arabinose or MQ. [b]After selection in liquid medium, cells were plated on LB agar plates under various conditions (Km0, without kanamycin; Km15, with 15 µg mL^{-1} kanamycin; L-arabinose or D-xylose, 100 mM) to get the cell count per mL of selection culture. [c]Plated directly after selection cultures. [d]Plated from glycerol stocks.

Table S5. Sequence of disruption cassette. Underlined and italic nucleotides indicate restriction sites and *lox* sites respectively. The cassette contains from 5' to 3' a BspTI site (used to make linear), a primer annealing site, a BglII site, *lox*71, *cat*, *lox*66, a NotI site, a primer annealing site and an Eco32I site (used to make linear).

CTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTGCGCTGT
GCATTGGGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCT
CGGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGATGCACACAAACAGGAGTCTGGTATCGC
TCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGCACACGAGAAGTACGGAATCGAAT
CTTTTGCACTGATTCACGACTCCTTCGGTACGATTCCGGCTGACGCTGCGAACCTGTTCAAAGCAGTGCGCGAAACTATGG
TTGACACTTATGAGTCTTGTGATGTACTGGCTGATTTCTACGACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAA
AATGCCAGCACTTCCGGCTAAAGGTAACTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCGTTCGCGTAAAGATCT*TA*
*CCGTTCGTATAATGTATGCTATACGAAGTTAT*GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGGCAATGAGC
TTGCACTGCAGAACTTTCTCGAGGATATACCATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCA
TCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTT
TTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGG
AATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAA
CTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGT
GTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTT
CACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACTATGGGCAAATATTATACGCAAGGC
GACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAA
TTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAA*ATAACTTCGTATAATGTATGCTATACGAACGGTA*GCGGCCG
CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACT
ATATCCGGGTAACGAATTCAAGCTTGATATC

Table S6. Primers used in this study

Primer	Annealing location	Sequence (5'-->3')[a]	Features
Construction/verification of plasmids/libraries			
BG3652	inside *kan*	AGTAACCATGCGTCATCAGG	
BG3653	inside *kan*	GCCTGTTGAACAAGTCTGGA	
BG3942	upstream *cat*	CAACGTCTCATTTTCGCCAG	
BG3943	inside *luxC*	CACGAATGTATGTCCTGCG	
BG5913	start *sacB*	GCGCGCGGACCGAGGAGGATACGTATGAACATCAAAAAGTTTGCAAAACAAG	RsrII
BG5914	end *sacB*	GCCGCACGATCGGGCGCGCCTTATTTGTTAACTGTTAATTGTCCTTGTTC	PvuI, AscI
BG5915	inside *sacB*	GAACGGCAAATGG*TAT*CTGTTCACTGACTCCCGCGGATC	altered tyrosine codon: TAC-->TAT
BG5916	inside *sacB*	GATCCGCGGGAGTCAGTGAACAG*ATA*CCATTTGCCG	altered tyrosine codon: TAC-->TAT
BG8211	inside *kan*	CTCCTTCATTACAGAAACGGC	
Construction/verification of* xylA *knockout			
BG3655	upstream *araC*	GGTTGGGTTAGCGAGAAGAG	
BG3656	downstream *araC*	GGGTAGAATCAAACCGACCA	
BG4190	upstream *recA*	CGTCAGGCTACTGCGTATG	
BG4191	downstream *recA*	GAATACGCGCAGGTCCATAAC	
BG5870	upstream *xylA*	GTGAGGTGCAAAGGGTGAG	
BG5871	downstream *xylA*	AGGTGCCAAGATCTATCCCG	
BG6221	start casette[b]	*CAAGCCTATTTTGACCAGCTCGATCGCGTTCGTTATGAAGGCTCAAAATCC*GAGTCGGACTTCGCGTTCGC	homologous to genome (nt[c] 4-54 of CDS of *xylA*)
BG6222	end cassette[b]	*CAGATTTTCCAGTTGTTCCTGGCGACCACTCTGATGCACCGGAGACAAATG*GCTAGTTATTGCTCAGCGGTGGC	homologous to genome (nt[c] 1246-1296 of CDS of *xylA*)

[a]Underlined nucleotides indicate restriction sites, italic nucleotides indicate other features (for both see last column). [b]Recombination/disruption cassette. [c]nt, nucleotides.

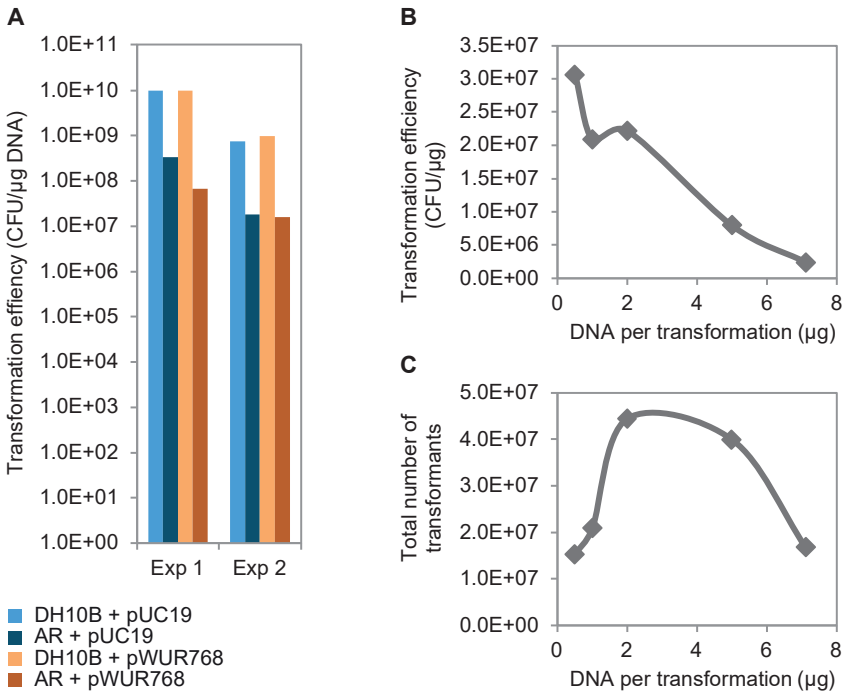
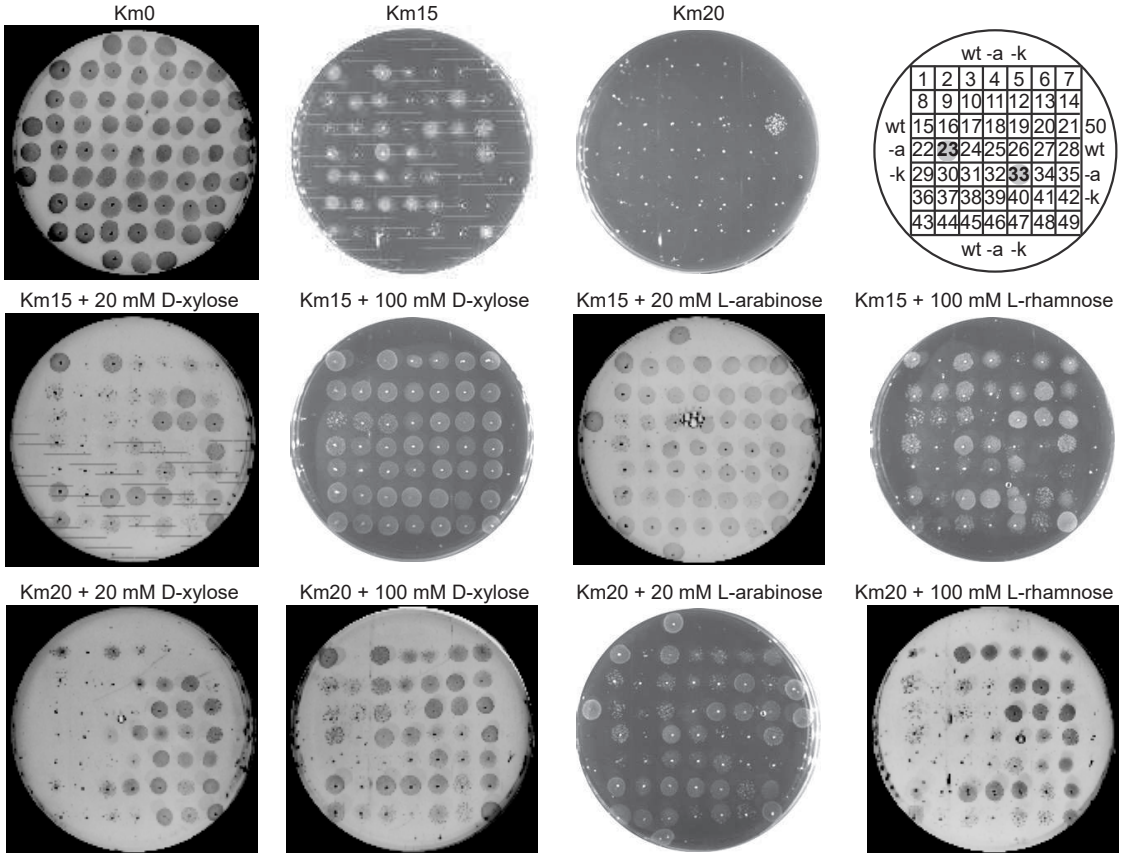
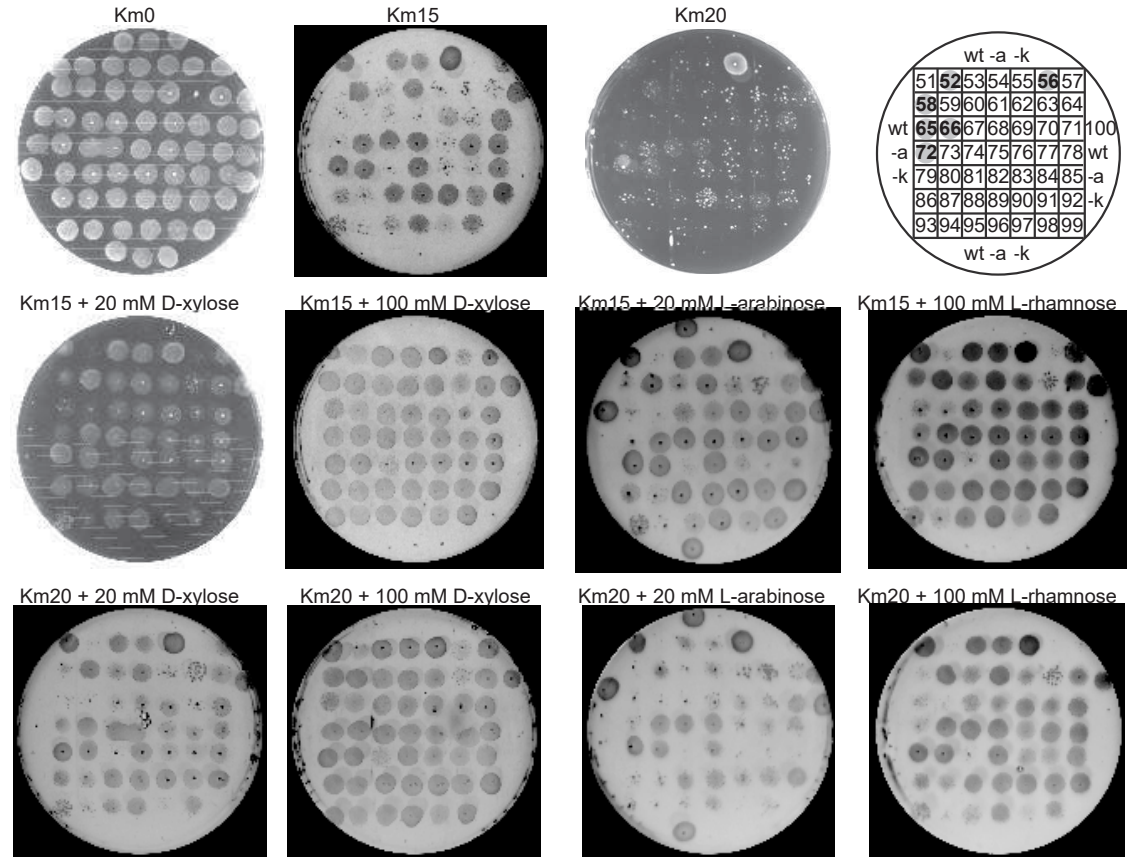


Fig. S1. Transformation efficiency of *E. coli* BW25113 AR with pWUR768. A) Small scale transformations to compare the transformation efficiency of AR + pWUR768 with commonly used control strain *E. coli* DH10B $T1^R$ and control plasmid pUC19[265]. Cells were grown in 100 mL 2xYP medium to an OD600 of 0.4 and made electrocompetent, resulting in a final suspension of 400 µL electrocompetent cells. These were transformed with 20 pg pUC19 or 600 pg pWUR768 by electroporation (2 mm cuvettes, 2.5 kV, 25 µF, 200 Ω, 40 µL cells, 2 µL DNA, recovery in LB in a total volume of 1 mL for 1 h). Cells were plated on LB agar plates with 100 µg mL^{-1} ampicillin for cells with pUC19 and 34 µg mL^{-1} chloramphenicol for cells with pWUR768. Exp1 and Exp2 indicate two independent experiments. B) and C) Large scale transformations to determine the optimal amount of DNA per transformation based on the transformation efficiency and the total number of transformants. Same method used as in A), but AR cells were grown in 500 mL 2xYP, the final cell suspension was 2000 µL and 400 µL cells were transformed with 10 µL pWUR768 (total recovery volume of 3 mL).

A


Km0
Km15
Km20
wt -a -k
1 2 3 4 5 6 7
8 9 10 11 12 13 14
wt 15 16 17 18 19 20 21 50
-a 22 23 24 25 26 27 28 wt
-k 29 30 31 32 33 34 35 -a
36 37 38 39 40 41 42 -k
43 44 45 46 47 48 49
wt -a -k
Km15 + 20 mM D-xylose
Km15 + 100 mM D-xylose
Km15 + 20 mM L-arabinose
Km15 + 100 mM L-rhamnose
Km20 + 20 mM D-xylose
Km20 + 100 mM D-xylose
Km20 + 20 mM L-arabinose
Km20 + 100 mM L-rhamnose

B


Km0
Km15
Km20
wt -a -k
51 52 53 54 55 56 57
58 59 60 61 62 63 64
wt 65 66 67 68 69 70 71 100
-a 72 73 74 75 76 77 78 wt
-k 79 80 81 82 83 84 85 -a
86 87 88 89 90 91 92 -k
93 94 95 96 97 98 99
wt -a -k
Km15 + 20 mM D-xylose
Km15 + 100 mM D-xylose
Km15 + 20 mM L-arabinose
Km15 + 100 mM L-rhamnose
Km20 + 20 mM D-xylose
Km20 + 100 mM D-xylose
Km20 + 20 mM L-arabinose
Km20 + 100 mM L-rhamnose

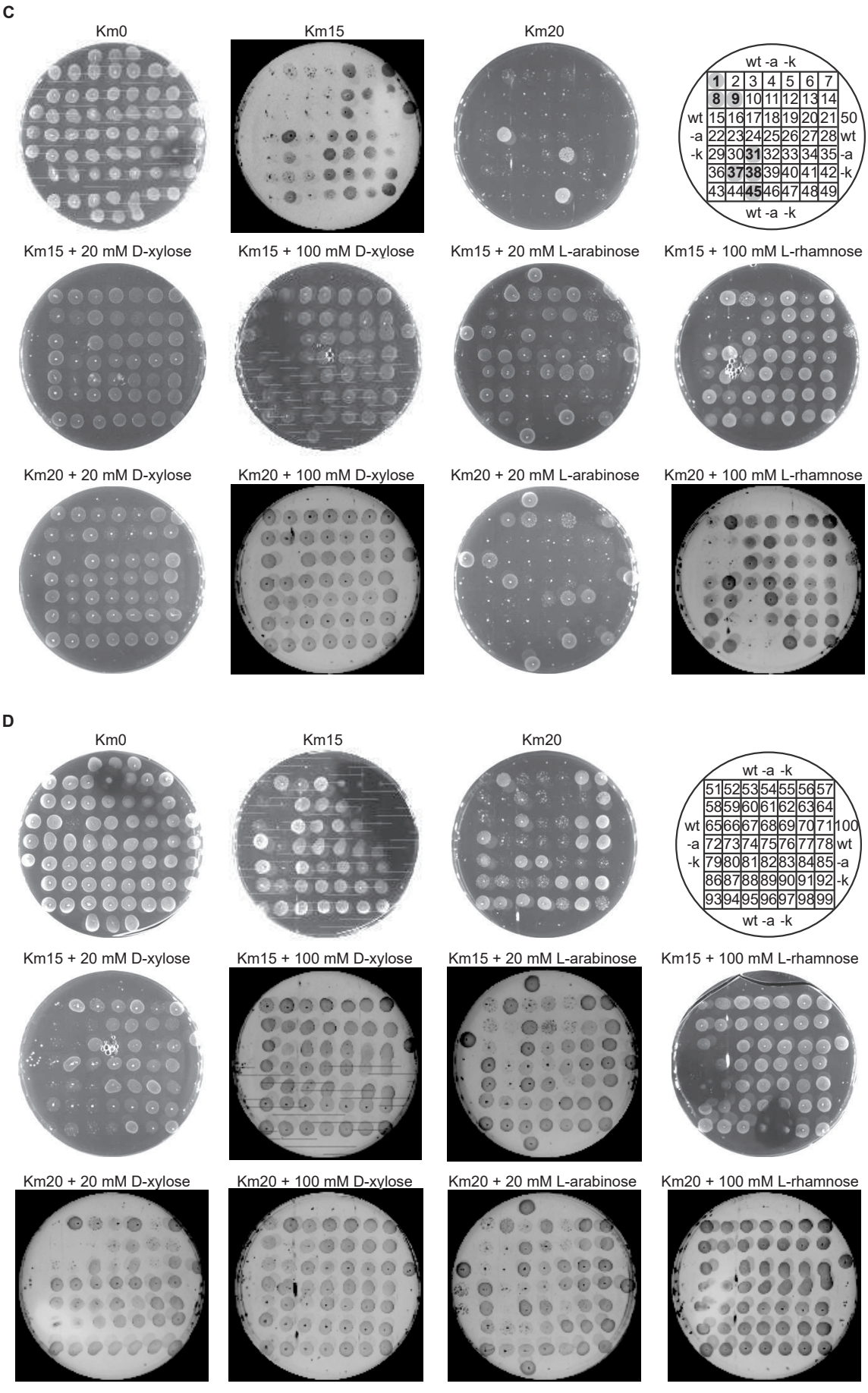

C
Km0
Km15
Km20
wt -a -k
1 2 3 4 5 6 7
8 9 10 11 12 13 14
wt 15 16 17 18 19 20 21 50
-a 22 23 24 25 26 27 28 wt
-k 29 30 31 32 33 34 35 -a
36 37 38 39 40 41 42 -k
43 44 45 46 47 48 49
wt -a -k
Km15 + 20 mM D-xylose
Km15 + 100 mM D-xylose
Km15 + 20 mM L-arabinose
Km15 + 100 mM L-rhamnose
Km20 + 20 mM D-xylose
Km20 + 100 mM D-xylose
Km20 + 20 mM L-arabinose
Km20 + 100 mM L-rhamnose
D
Km0
Km15
Km20
wt -a -k
51 52 53 54 55 56 57
58 59 60 61 62 63 64
wt 65 66 67 68 69 70 71 100
-a 72 73 74 75 76 77 78 wt
-k 79 80 81 82 83 84 85 -a
86 87 88 89 90 91 92 -k
93 94 95 96 97 98 99
wt -a -k
Km15 + 20 mM D-xylose
Km15 + 100 mM D-xylose
Km15 + 20 mM L-arabinose
Km15 + 100 mM L-rhamnose
Km20 + 20 mM D-xylose
Km20 + 100 mM D-xylose
Km20 + 20 mM L-arabinose
Km20 + 100 mM L-rhamnose

Fig. S2. Analysis of D-xylose response of *araC* library variants by kanamycin resistance-based selection. The response of *araC* library 1 (A, Lib1-Xyl1 to Lib1-Xyl50; B, Lib1-Xyl51 to Lib1-Xyl100) and library 2 (C, Lib2-Xyl1 to Lib2-Xyl50; D, Lib2-Xyl51 to Lib2-Xyl100) variants towards D-xylose was analysed by growing cells on LB medium overnight in presence of 0, 15 or 20 µg mL^{-1} kanamycin and 0, 20 or 100 mM D-xylose. 20 mM L-arabinose and 100 mM L-rhamnose were controls. 5 µL cell suspension was spotted on each agar plate. The variants that were used for follow up experiments are indicated in bold with a grey background. wt, wildtype (*E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC*); -a, neg. ctrl (ARX with the regulator–reporter plasmid with *sacB* instead of *araC*); -k, neg. ctrl. (ARX with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *kan*); Km, kanamycin.

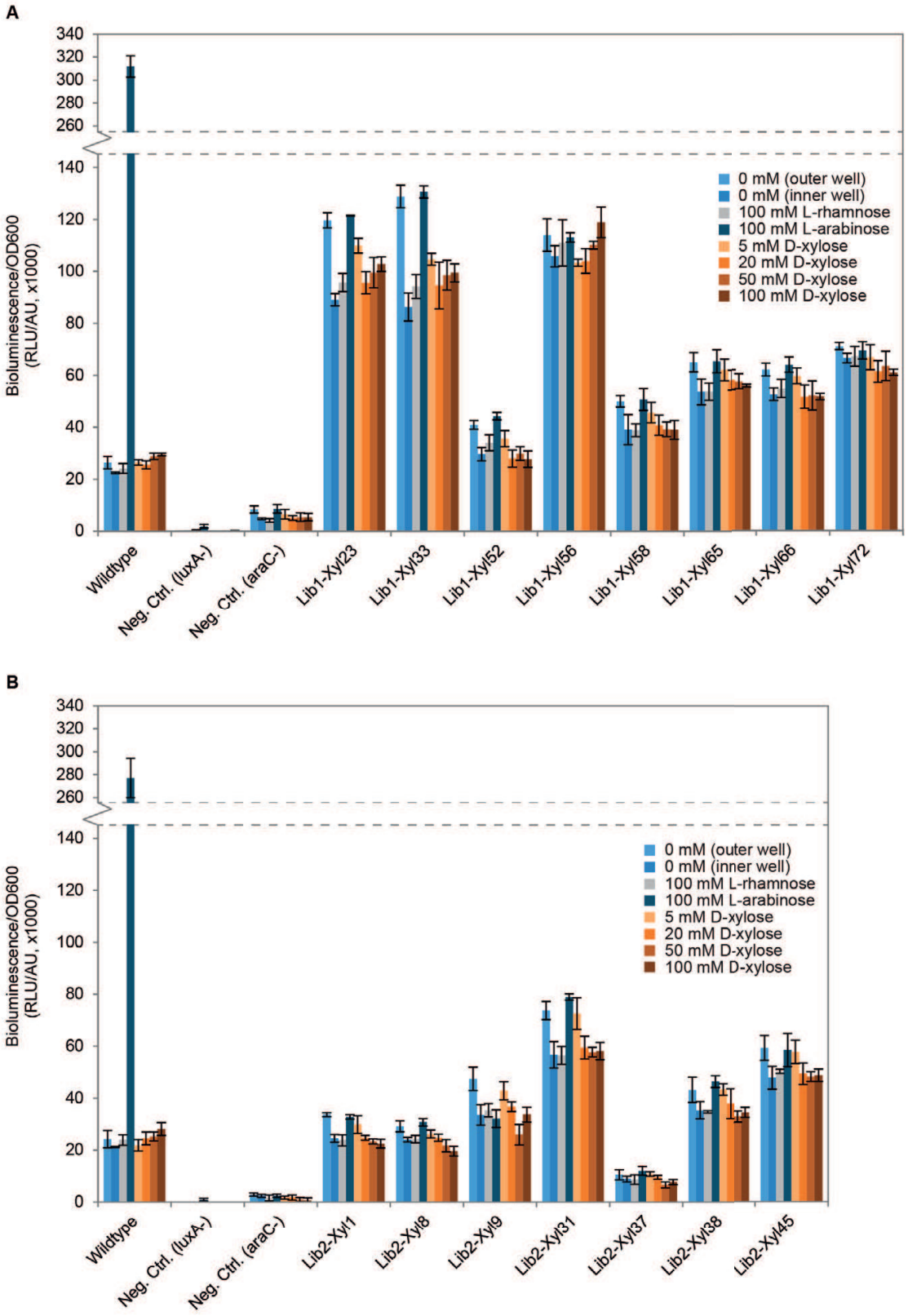

A
B
Bioluminescence/OD600 (RLU/AU, x1000)
0 mM (outer well)
0 mM (inner well)
100 mM L-rhamnose
100 mM L-arabinose
5 mM D-xylose
20 mM D-xylose
50 mM D-xylose
100 mM D-xylose
Wildtype
Neg. Ctrl. (luxA-)
Neg. Ctrl. (araC-)
Lib1-Xyl23
Lib1-Xyl33
Lib1-Xyl52
Lib1-Xyl56
Lib1-Xyl58
Lib1-Xyl65
Lib1-Xyl66
Lib1-Xyl72
Lib2-Xyl1
Lib2-Xyl8
Lib2-Xyl9
Lib2-Xyl31
Lib2-Xyl37
Lib2-Xyl38
Lib2-Xyl45

Fig. S3. Determination of D-xylose response of *araC* library variants by bioluminescence-based screening. The D-xylose response of the *araC* library 1 (Lib1, A) and 2 (Lib2, B) variants were measured in two independent experiments. Cells were grown in LB medium in 96-wells plates for 17 h, while OD600 and bioluminescence were measured every 15 min. An average was taken of three subsequent time points in the late log phase for which the signal production and wash out due to cell division were about equal (standard deviation indicated). L-arabinose and L-rhamnose were controls. To show evaporation effects, 0 mM was included twice, in an outer and in an inner well. Lib1-Xyl56 was measured together with Lib2 variants. Wildtype: *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC*. Neg. ctrl. (luxA-): ARX with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *luxA*. Neg. ctrl. (araC-): ARX with the regulator–reporter plasmid with *sacB* instead of *araC*.

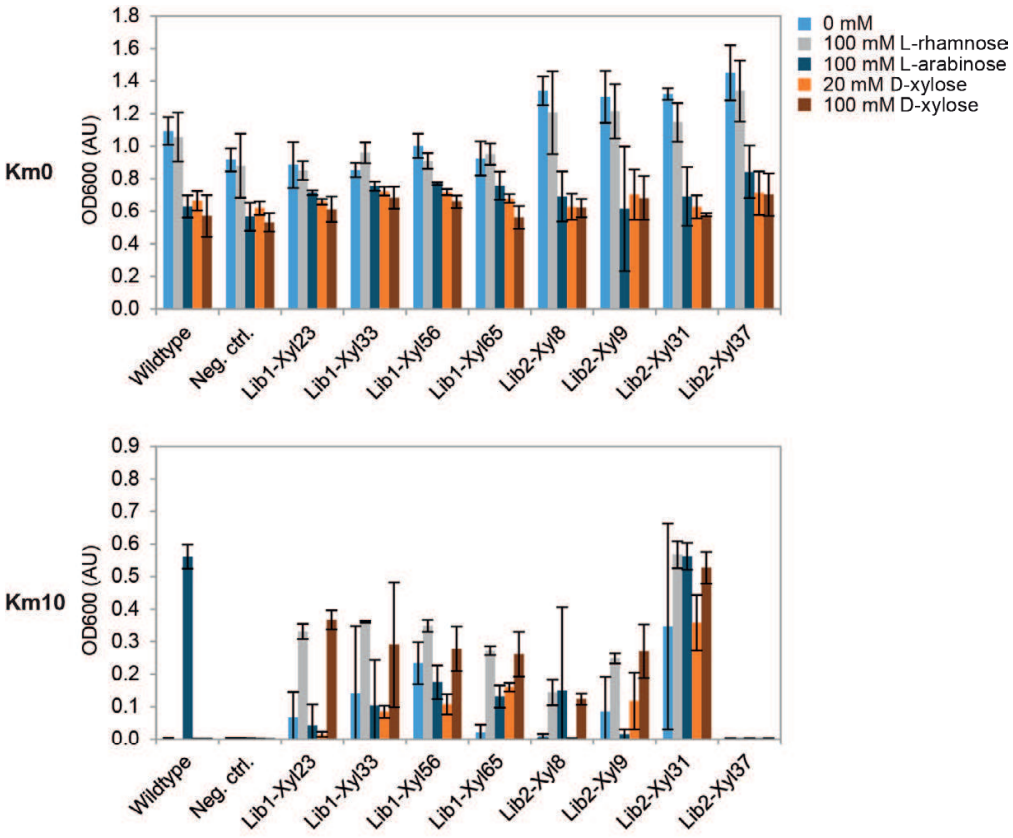


Fig. S4. Determination of D-xylose response of *araC* library variants by kanamycin resistance-based selection. The D-xylose response of *araC* library 1 (Lib1) and 2 (Lib2) variants was determined by growing cells in LB medium in presence of 0 or 10 µg mL^{-1} kanamycin for 17 h. L-arabinose and L-rhamnose were controls. For each library, one experiment was performed in triplicate and values were averaged. Also Wildtype and Neg. Ctrl. values of Lib1 and Lib2 experiments were averaged. The final standard deviation is indicated. Wildtype: *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX) with the regulator–reporter plasmid with wildtype *araC*. Neg. ctrl.: ARX with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *kan*. The two graphs have a different y-axis scale.

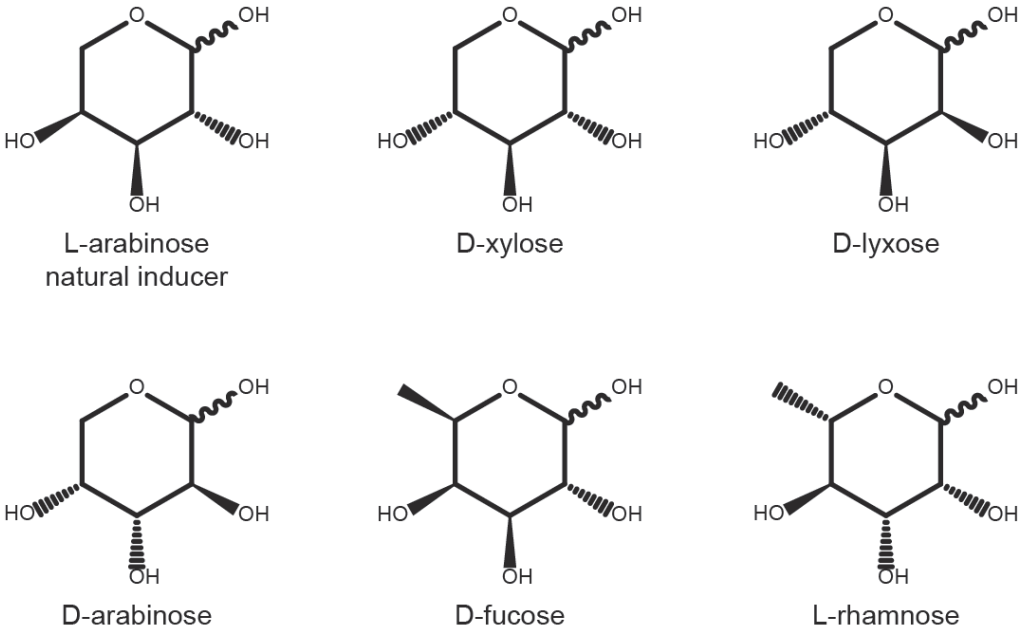
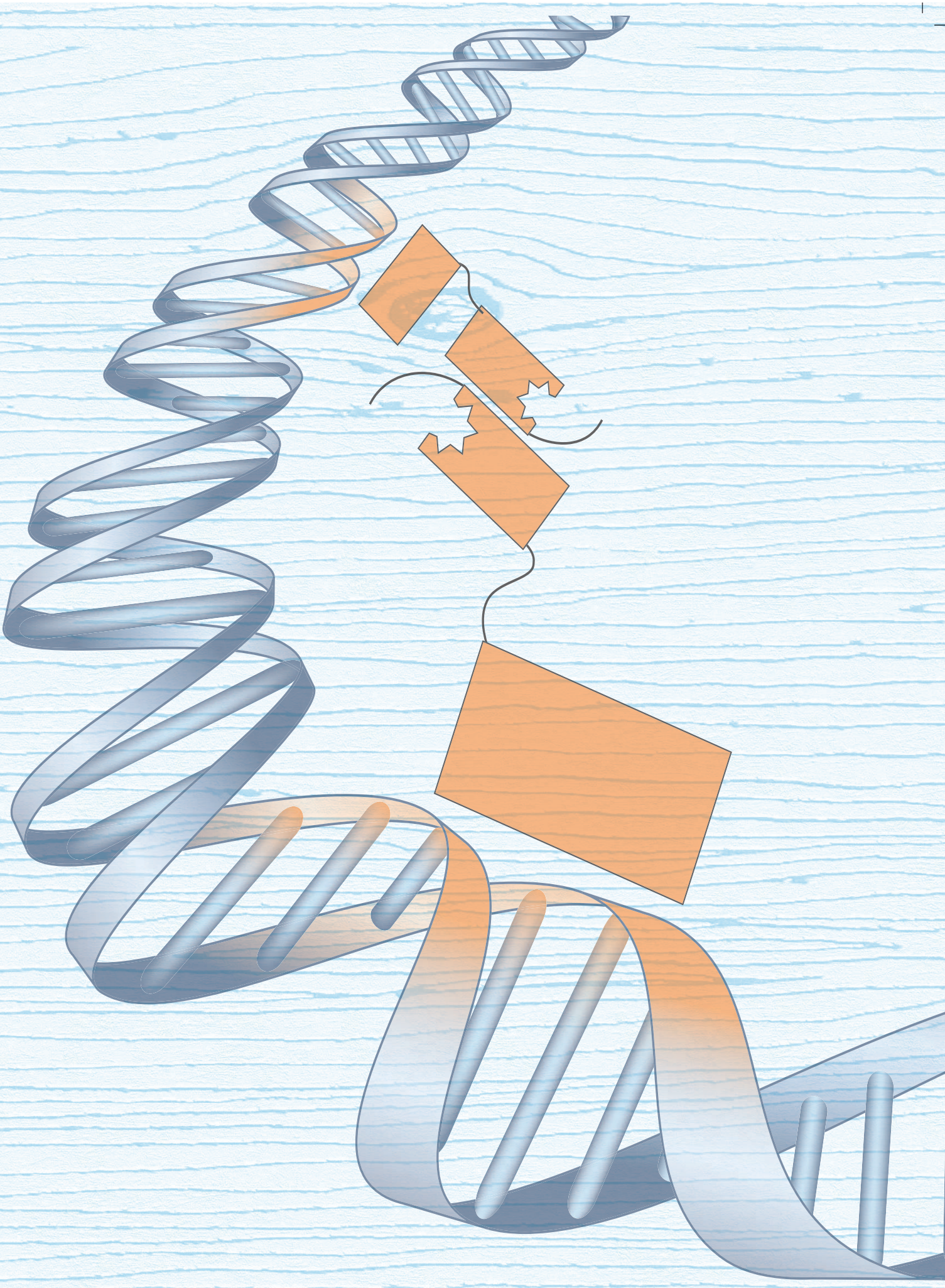


Fig. S5. Chemical structures of monosaccharides used in specificity determination.

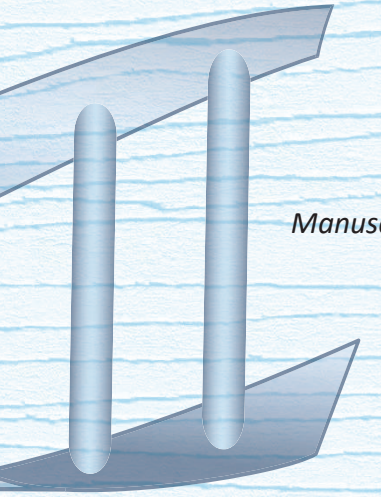


Chapter 6

Inhibitory and stimulatory effects of L-arabinose on growth of *Escherichia coli* BW25113

Teunke van Rossum[a], John van der Oost[a], Servé W.M. Kengen[a]

[a]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands



Manuscript in preparation

Abstract

Escherichia coli preferentially grows on sugars, like mono- and disaccharides. To be able to switch its metabolism to convert the energetically most favourable carbon source available, *E. coli* uses regulatory mechanisms called carbon catabolic repression (CCR). For example, the expression of genes involved in L-arabinose catabolism and transport are repressed by D-glucose and activated by L-arabinose via the transcriptional regulators CRP and AraC respectively. These regulatory phenomena are understandable as these carbon sources all may act as growth substrate. However, literature data are available that suggest that even in *E. coli* strains that are incapable of growth on L-arabinose, both stimulatory and inhibitory growth effects occur in presence of this sugar. In this study, these growth effects were investigated in more detail to understand the underlying regulatory mechanisms. In LB medium, L-arabinose negatively affected growth of wildtype strain BW25113 (lower final OD600), but not of Δ*araC* and Δ*crp* strains. In addition, the effect was stronger for Δ*xylA*, encoding the first enzyme in D-xylose catabolism. Growth of strains in which wildtype *araC* was replaced by *sacB* or by *araC* variants that encode L-arabinose unresponsive AraC mutants was still inhibited by L-arabinose as well as some other monosaccharides to various extent. In M9 minimal medium, L-arabinose stimulated growth of both BW25113 and Δ*araC* strains in an early phase of growth, but reduced the final OD600 of only BW25113. Hypothetical regulatory mechanisms are discussed that may explain the L-arabinose effects on growth of *E. coli*.

Introduction

The Gram-negative bacterium *Escherichia coli* is a facultative anaerobe that thrives in the mucus layer of the animal intestine. Here, it grows on mucus-derived nutrients, preferably sugars like mono- and disaccharides, which are released through the action of extracellular hydrolases of anaerobes. *E. coli* itself lacks the hydrolases to degrade complex polysaccharides[266,267]. The sugar metabolism of *E. coli* has been studied predominantly in laboratory mono-cultures, where different growth phases reflect the operation of different metabolic pathways. In aerobic batch cultures on rich LB medium, *E. coli* initially uses the carbon sources in a sequential manner starting with D-glucose, its preferred substrate, and followed by maltose and maltodextrins. When the preferred carbon sources are depleted, the growth rate slows down, amino acids and nucleotides are consumed and *E. coli* switches to simultaneous use of multiple carbon sources like carbohydrates (e.g. L-arabinose), polyols, and organic acids, while it produces acetate. In the next phase, growth slows down by 40%, the excreted acetate is assimilated again, oligopeptides are used and the cells produce most of the biomass precursors themselves. Before reaching stationary phase, cells grow very slowly while maximally expressing most catabolic pathways. When carbon sources are depleted, growth ceases[168,268-270]. The switch from acetate production to acetate assimilation also occurs for aerobic growth in minimal medium going from high (exponential phase) to low D-glucose concentrations (glucose-limited fed batch conditions). At first, energy mainly comes from the glycolysis, but when D-glucose levels are low, enzymes of the TCA cycle are upregulated for both energy production and the formation of intermediates for macromolecular biosynthesis. Under these poor conditions, cells use a wide range of carbon sources and they switch from D-glucose import via the phosphotransferase system (PTS) to high affinity D-glucose import via ABC transporters[271].

To be able to switch its metabolism to convert the energetically most favourable carbon source available, *E. coli* uses regulatory mechanisms called carbon catabolic repression (CCR) at various levels, e.g. global- or operon specific regulation on transcription level or post-transcriptional regulation via sRNA. In this way, the preferred sugar represses the catabolism of less favourable sugars. In case of D-glucose, its import via PTS is coupled to repression of genes involved in the transport and the catabolism of secondary sugars via the signal molecule cAMP. When D-glucose levels are high, less cAMP is available to bind and activate the global regulator CRP (cAMP receptor protein), which therefore cannot activate the expression of the secondary sugar genes. Also a hierarchy between non-PTS sugars has

been described: lactose, arabinose, xylose, sorbitol, rhamnose and ribose. This order can be quantitatively explained by differential CRP-cAMP activation of the promoters of the genes involved in the transport and catabolism of these sugars[272]. However, it is likely that this is not a strict order, but rather an order that is dependent on the relative concentrations as is described for L-arabinose and D-xylose catabolism. Each of these two sugars bound to its respective transcriptional activator represses expression of the catabolic genes of the other sugar, either directly in case of L-arabinose or indirectly by repressing expression of the other transcriptional activator in case of D-xylose[273,274].

The genes involved in L-arabinose catabolism and transport are regulated by the transcriptional regulators AraC and CRP and their respective inducers L-arabinose and cAMP[130,132]. The way AraC regulates expression depends per gene[131]. For example, the *araBAD* operon, which encodes the first three enzymes in L-arabinose catabolism (AraA, L-arabinose isomerase; AraB, L-ribulosekinase; AraD, L-ribulose-5-phosphate 4-epimerase), is repressed by AraC in the absence of L-arabinose and activated in the presence of L-arabinose (Fig. 1 of **Chapter 5**). In the repressive state, the dimer AraC is bound with the C-terminal domain of each monomer to one of two operators upstream the *araBAD* operon, O_2 and I_1 (210 bp apart), and thereby represses expression of these genes by looping the DNA. Once AraC binds L-arabinose, it undergoes a conformational change, allowing the C-terminal domains to reorient and bind the two adjacent I_1 and I_2 operators. This binding activates expression of the *araBAD* operon[130,131]. Full expression is only possible when both L-arabinose and cAMP are present, relieving the CCR at low D-glucose concentrations[147]. When little L-arabinose is present, it is taken up by basal levels of the transporter proteins, but once AraC binds L-arabinose it activates expression from the transporter genes *araE* (low affinity, high capacity L-arabinose/proton symporter using electrochemical potential[130]) and *araFGH* (high affinity, low capacity ABC L-arabinose transporter using ATP[130]), leading to more transport proteins, more inducer uptake and thus providing a positive feedback loop on expression. This loop causes an all-or-nothing behaviour of induction in which at low L-arabinose concentrations the cells within a population are heterogeneously induced and only at high L-arabinose concentrations homogenous induction occurs. Expression of *araE* from a constitutive promoter allows for homogeneous expression for which the expression level depends on the extracellular inducer concentration[142,275]. Next to these genes that have a clear function in L-arabinose catabolism and transport, AraC directly regulates its own gene *araC* (repression and L-arabinose dependent transient expression[130]), *araJ* (putative transporter; L-arabinose dependent activation[131,276]), *ytfQ* (galactofuranose ABC transporter

periplasmic binding protein; L-arabinose independent repression[131,277]), *ydeNM* (predicted sulfatase and a predicted sulfatase maturase respectively; L-arabinose dependent repression[131]), and *xylA* (D-xylose isomerase, L-arabinose dependent repression[273]). AraC binding sites are present upstream of all these genes, but also in the coding sequence of *dcp* (dipeptidyl carboxypeptidase). The function of this site is unknown. [131] Read through transcription of *araE* and *araBAD* leads to co-regulation of *ygeA* (putative racemase, UniProt) and *polB* (DNA polymerase II[278]) respectively[131]. Next to these direct targets of AraC, various genes are regulated indirectly by AraC and/or L-arabinose. Most of these genes are repressed and include genes involved in maltose metabolism (*malE*, *malF*, *malG*, *malK*, *malM*, and *lamB*), threonine metabolism (*tdcA*, *tdcB*, *tdcC*, *tdcD*, and *tdcE*), D-glucarate/D-galactarate metabolism (*garD*, *garL*, *garP*, and *garR*), and tryptophan metabolism (*tnaA*, *tnaB*, and *tnaL*). *IsrB* is the only gene that is indirectly activated by AraC and L-arabinose[131].

Knowledge on how L-arabinose is involved in the regulation of carbon catabolism and on how L-arabinose affects cell growth could be valuable for various fields. L-arabinose is a significant component of plant material, which is used as substrate for the production of green chemicals[274] and it is a substrate for bacterial growth in the animal intestine where *E. coli* grows on the L-arabinose released by other bacteria[267]. In the field of protein production, L-arabinose is used as inducer of the *araBAD* promoter that controls the corresponding gene[279]. Knowledge on how L-arabinose affects *E. coli* cells could therefore contribute to an improved production of green chemicals and proteins and could help to understand the role of *E. coli* in the intestinal microbiome and in disease. One aspect of L-arabinose metabolism in *E. coli* that is not yet understood, was first described almost sixty years ago by Gross and Englesberg. They observed that the growth of mutants of *E. coli* B/r that lack the enzymes for the first or second step of L-arabinose catabolism, Δ*araA* or Δ*araB* respectively, were inhibited by L-arabinose when grown in casein hydrolysate, whereas a mutant that lacks the regulator, Δ*araC*, was unaffected by L-arabinose under these conditions[280]. Recently, the inhibitory effect of L-arabinose was also described for growth of the strain *E. coli* BW25113 Δ*araC* Δ*recA* with AraC encoded on a plasmid, in LB medium. This strain has a deletion of the *araBAD* operon (**Chapters 3**[184] and **5**). In contrast, a stimulatory effect of L-arabinose was observed during growth in M9 minimal medium (**Chapter 3**[184]). The aim of the current study was to try to understand how L-arabinose brings about these inhibitory and stimulatory effects on growth. To that end, the data of **Chapters 3**[184] and **5** were studied in more detail and the growth of various knockout strains in the presence and the absence of L-arabinose was analysed and compared to the growth of parent strain

BW25113. In LB medium, L-arabinose negatively affected growth of BW25113 (lower final OD600), but not of Δ*araC* and Δ*crp* and the effect was stronger for Δ*xylA*. Strains in which wildtype *araC* was replaced by *sacB* or by *araC* variants that encode L-arabinose unresponsive AraC mutants were still inhibited by L-arabinose as well as some other monosaccharides. In M9 minimal medium, L-arabinose stimulated growth of both BW25113 and Δ*araC* in an early phase of growth, but reduced the final OD600 of only BW25113.

Results

Inhibitory effect of L-arabinose during growth in LB medium

The previous observations described in **Chapter 3** that L-arabinose negatively affects growth of *Escherichia coli* were made with *E. coli* BW25113 Δ*araC* Δ*recA* (AR) with the regulator-reporter plasmid pWUR772 or pWUR768, having a low (p15A) or medium (ColE1) plasmid copy number respectively[184]. The strain AR has a deletion of the *araBAD* operon (Δ(*araD-araB*)*567*), ranging from 24 nt upstream *araB* to 18 nt into the *araD* coding sequence (CDS). This leaves the regulatory binding sites for AraC and CRP intact, but deletes *araB*, *araA* and the first 18 nt of *araD*. In addition, this strain lacks the CDSs of *araC* and *recA*. The plasmid pWUR768 contains the chloramphenicol resistance marker *cat* for plasmid maintenance, *araC* expressed from the constitutive promoter P_{lacI}^{Q}, and two reporter genes, which are divergently expressed from adapted P_{BAD} promoters that no longer have the CRP binding site. These reporters are the kanamycin marker *kan* and the luciferase operon *luxCDABE*. The negative growth effect was observed when bacteria were grown aerobically in LB medium containing various concentrations of L-arabinose (Fig. 1A). After 17 h, the final OD600 (stationary phase) was measured. Growth of both strains was negatively affected by L-arabinose, with a higher concentration having a more severe effect. However, AR with the low copy plasmid pWUR772 was more sensitive to L-arabinose than AR with the medium copy plasmid pWUR768. At 0.5 mM, AR pWUR722 reached maximum inhibition (0.58 ± 0.04 of final OD600 without L-arabinose), whereas AR pWUR768 reached this maximum only at 5-10 mM (0.57 ± 0.04 of final OD600 without L-arabinose).

L-arabinose also negatively affected the growth of AR pWUR768 when a second plasmid was added that expressed the L-arabinose isomerase AraA under the constitutive P_{bla} promoter (Fig. 1B). This enzyme catalyses the reversible conversion of L-arabinose to L-ribulose, the first step in L-arabinose catabolism. In the presence of AraA, the inhibition by 10 mM L-arabinose was similar as in absence of AraA (0.61 ± 0.17 vs. 0.55 ± 0.14 of final

OD600 without L-arabinose). Interestingly, when L-ribulose instead of L-arabinose was present, only growth of the strain with AraA, which could convert L-ribulose to L-arabinose, was inhibited by L-ribulose, confirming the specific inhibitory effect of L-arabinose (Fig. 1B; 0.66 ± 0.12 of final OD600 without L-ribulose).

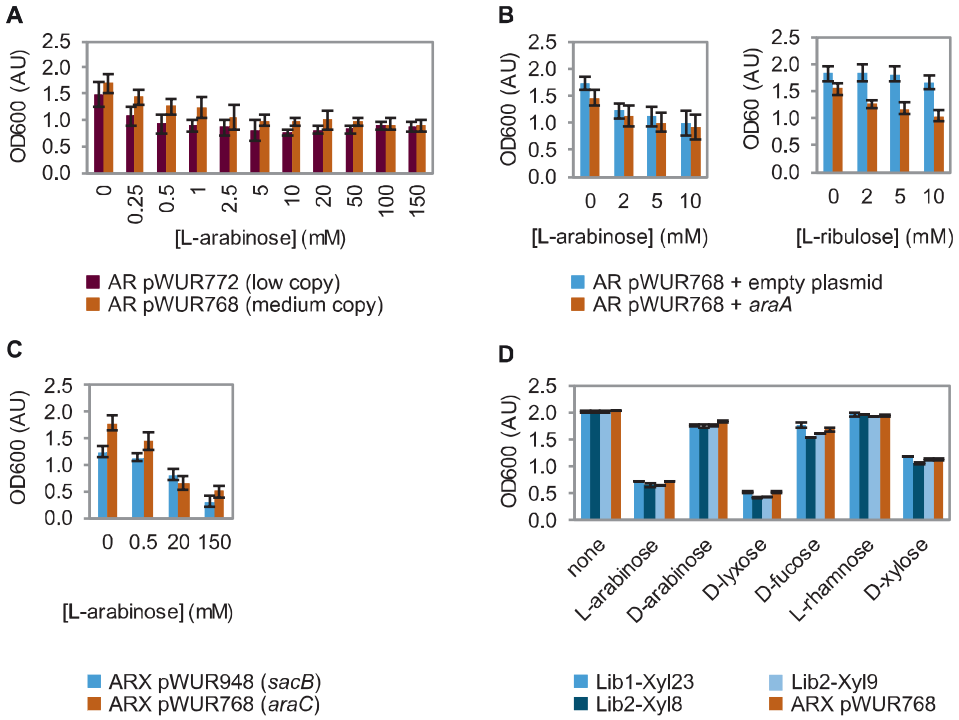


Fig 1: Negative effect of L-arabinose on growth in LB medium. (A) Effect over a range of L-arabinose concentrations for strains with a low or medium copy plasmid that encodes the transcriptional regulator AraC. (B) Effect of L-arabinose and L-ribulose in the absence (empty plasmid) and the presence (*araA*) of L-arabinose isomerase AraA. (C) Effect when plasmid-encoded *araC* is replaced by *sacB*, which encodes levansucrase. (D) Effect of 100 mM of various monosaccharides for strains with wildtype or variant *araC* (**Chapter 5**). In all assays, cells were grown in 500 µL LB medium in a 2-mL 96-well MASTERBLOCK for 17 h, except for the assay in (D) in which cells were grown in 200 µL LB medium in a 96-well plate in a platereader for 16 h. Data are an average of three (A), four (B), or two (D) independent experiments or an average of two replicates within one experiment (C). Error bars represent standard deviations. AR(X) pWUR772 or pWUR768: *E. coli* BW25113 Δ*araC* Δ*recA* (Δ*xylA*) with respectively the low or the medium copy regulator–reporter plasmid with wildtype *araC*. ARX pWUR947: ARX with the medium copy regulator–reporter plasmid with *sacB* instead of *araC*. Lib1-Xyl23, Lib2-Xyl8, Lib2-Xyl9: ARX with the medium copy regulator–reporter plasmid with an *araC* variant that encodes L-arabinose unresponsive AraC.

In a different context, the CDS of *araC* was replaced by *sacB*, encoding the levansucrase of *B. subtilis* (**Chapter 5**). Levansucrase has two activities, the hydrolysis of sucrose to D-glucose and D-fructose and the transfructosylation, in which the fructose moiety of sucrose is added to an acceptor molecule. Once expressed in *E. coli*, SacB is translocated to the periplasm, in which the transfructosylation activity is lethal[248,281,282]. The

sacB encoding plasmid pWUR947 was introduced in *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* (ARX), a strain that lacks XylA (the first enzyme in D-xylose catabolism), and is therefore unable to grow on D-xylose. In this context however, the strain ARX pWUR947 was used as control for absence of *araC* (Fig. 1C). Just the replacement itself reduced the growth (0.70 ± 0.05 of final OD600 with *araC*). Although the growth was further reduced in the presence of L-arabinose, this strain was less sensitive to 20 mM L-arabinose than the strain with *araC*, ARX pWUR768 (0.65 ± 0.03 vs. 0.37 ± 0.03 of final OD600 without L-arabinose). That the negative effect of L-arabinose, and also D-xylose, was more severe for strain ARX pWUR768 than for AR pWUR768 was verified in another experiment (Fig. S1; with 20 mM L-arabinose 0.41 ± 0.02 vs. 0.61 ± 0.15 of final OD600 without L-arabinose and with 20 mM D-xylose 0.41 ± 0.03 vs. 0.63 ± 0.15 of final OD600 without D-xylose).

Next to the replacement of *araC* by *sacB*, *araC* was replaced by *araC* variants that encode AraC mutants that no longer respond to L-arabinose and only very limited respond to the other tested monosaccharides (**Chapter 5**). However, these mutants probably retain some DNA binding activity at the adjacent operator sites, which are bound for activation, because their leakiness (reporter expression in absence of inducer) was not only higher than that of wildtype, but also higher than that of ARX pWUR947, which has no *araC* (Fig. S2). They were thus capable of slight enhancement of background reporter expression, which required binding to the two adjacent operator sites. The strains with wildtype or variant *araC* were grown for 16 h in LB medium in the presence of various monosaccharides (Fig. 1D). Not only was the growth of these mutants inhibited by L-arabinose to the same extent as the strain with wildtype *araC*, all strains were also inhibited equally by the other monosaccharides. The severity was different per monosaccharide: D-lyxose (0.24 ± 0.00 of final OD600 without D-lyxose) > L-arabinose (0.34 ± 0.00) > D-xylose (0.56 ± 0.00) > D-fucose (0.82 ± 0.01) > D-arabinose (0.88 ± 0.01) > L-rhamnose (0.97 ± 0.01) of which L-rhamnose had barely any effect.

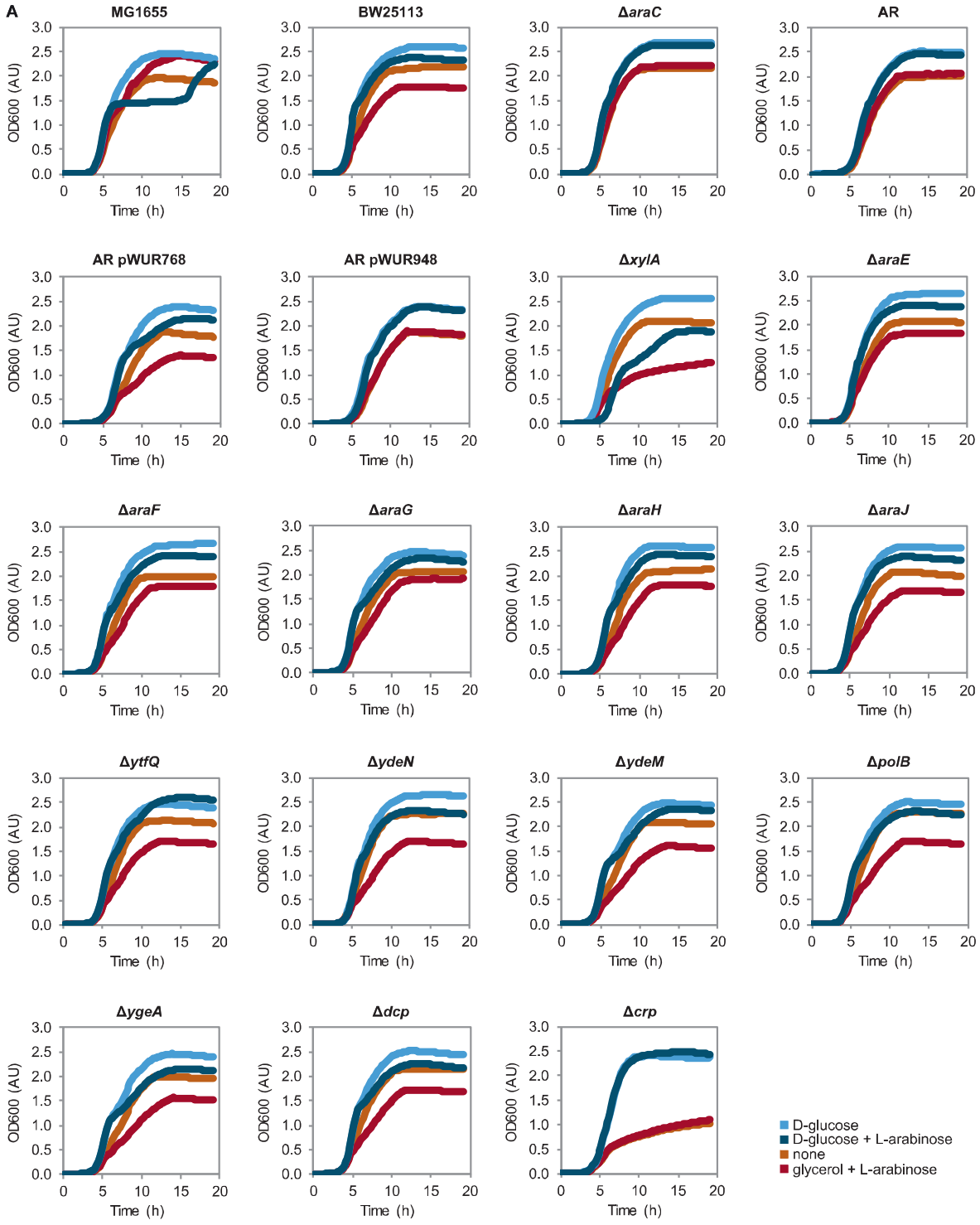
Inhibitory effect of L-arabinose on growth of various strains in LB medium

The above described negative effect on growth by L-arabinose was observed in a complex system, namely in strains with plasmids and several deleted genes. In addition, only the effect on the final OD600 was measured. To see if L-arabinose also affects the growth in other growth phases and to study the negative effect in a simpler system, growth curves were made for various strains. *E. coli* MG1655, wildtype *E. coli*, was included, because it could, in contrast to AR and its parent strain BW25113, grow on L-arabinose (Fig. S3A). It

does contain the *araBAD* operon, encoding the first three genes for L-arabinose catabolism. To determine the possible involvement of the transcriptional regulator of L-arabinose metabolism, both BW25113 and its derivative Δ*araC* were included as well as AR pWUR768 and AR pWUR948. The latter two have no *araC* gene on the genome, but they do have a functional and non-functional *araC* gene under a constitutive promoter on a plasmid respectively. The non-functional *araC* gene was made by introducing a frameshift. Also BW25113 derivatives were included that have a single knockout of a gene directly regulated by AraC. These genes were *xylA*[273], *araE*, *araF*, *araG*, *araH*, *araJ*, *ytfQ*, *ydeN*, *ydeM* and *dcp*[131]. *Dcp* had an AraC binding site within the gene, but no regulation of *dcp* transcription was observed. The genes *ygeA* and *polB*, of which transcription is regulated by AraC via read-through transcription of *araE* and *araBAD* respectively, were also included[131]. To see what the influence was of CCR, BW25113 Δ*crp* was included. All strains were grown in LB medium with 0 or 1 g L^{-1} D-glucose and in the presence of 0 or 10 mM L-arabinose for 19 h. The OD600 was measured every 10 minutes.

The growth of the various strains differed among strains and per condition (Fig. 2A), but all strains grew faster on D-glucose (Fig. 2B) and most strains had a shorter lag phase when grown on D-glucose (Fig. 2C), the preferred carbon source. Growth of Δ*crp*, AR and AR with plasmids was slower with a longer lag phase and a lower final OD600 in all conditions, as is described for Δ*crp* and Δ*recA*[161]. MG1655 was able to grow on L-arabinose and reached a higher OD600 than the other strains. In the presence of D-glucose, its growth was arrested at an OD600 of 1.4 for 9 h. Growth of BW25113 was inhibited by L-arabinose, starting at an OD600 of 0.5, but in the presence of D-glucose this inhibition was delayed to an OD600 of 1.4. Most other strains grew similar as BW25113, except for a few. The inhibitory effect of L-arabinose was gone for Δ*araC*, as observed by Gross and Englesberg[280]. Also the growth of strain AR was no longer negatively affected by L-arabinose, but once *araC* was encoded on a plasmid (AR pWUR768), the inhibition was restored. The onset of inhibition was at the same OD600 as for BW25113, but it was more abrupt. The final OD600 difference in the absence and the presence of L-arabinose was similar as for BW25113. Inactivation of the plasmid-encoded *araC* (AR pWUR748), removed the inhibition again. The growth of Δ*xylA* was more severely negatively affected by L-arabinose. It had a 1.5 h longer lag phase and a slower growth rate in the presence of both D-glucose and L-arabinose than BW25113 and a lower final OD600. The inhibition by L-arabinose was reduced for Δ*ytfQ* only without D-glucose being present and gone for Δ*crp*. Δ*araE* was more gradually inhibited by L-arabinose, instead of having a clear onset of inhibition like BW25113. Δ*ydeN* grew slower in LB medium with

and without D-glucose than BW25113, but only in the absence of L-arabinose. Most strains that were negatively affected by L-arabinose in the final OD600, also grew slower in the presence of L-arabinose, but only in the absence of D-glucose.



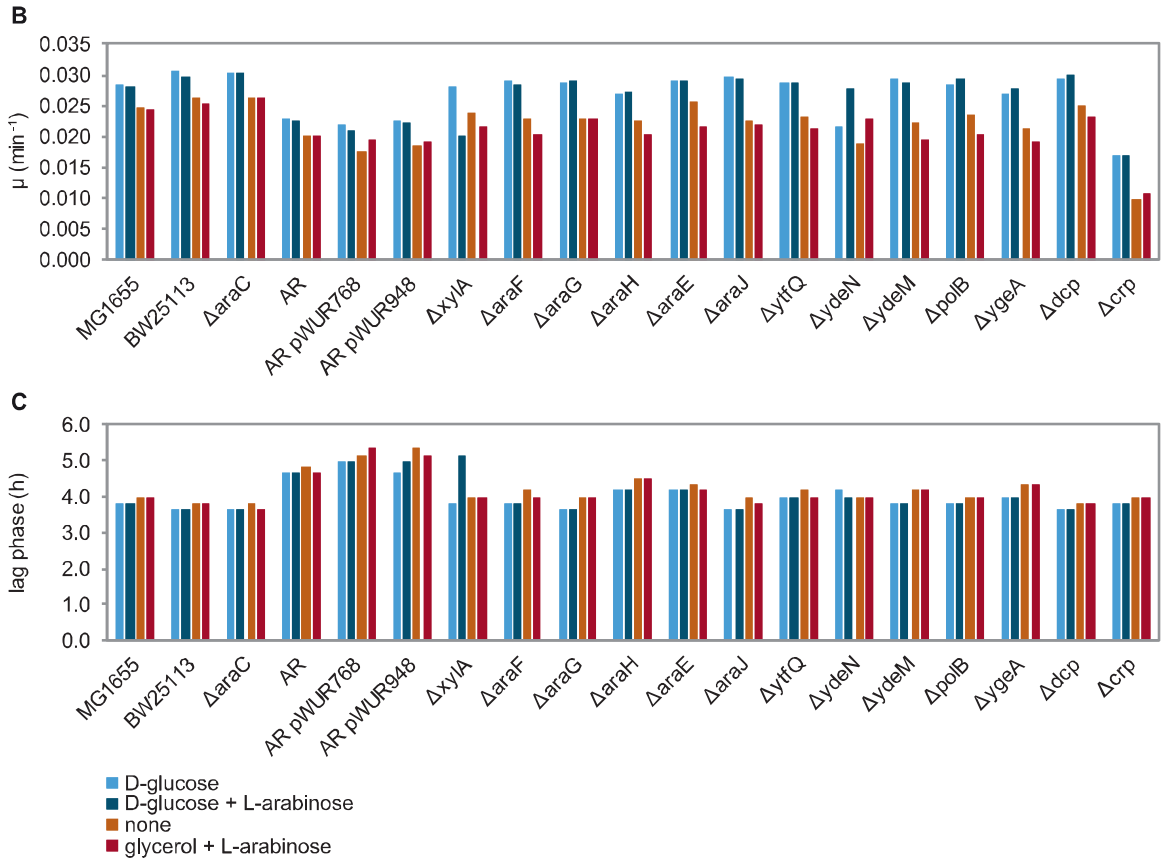


Fig 2: Negative effect of L-arabinose on growth of various strains in LB medium. (A) Growth curves. (B) Growth rate. (C) Lag phase. The lag phase was defined as the time needed to reach an OD600 of 0.1. Cells were grown in 200 µL LB medium with 0 or 1 g L^{-1} D-glucose and 0 or 10 mM L-arabinose in a 96-well plate in a platereader for 19 h. OD600 was measured every 10 min. AR pWUR768: *E. coli* BW25113 Δ*araC* Δ*recA* (Δ*xylA*) with the medium copy regulator–reporter plasmid with wildtype *araC*. AR pWUR948: AR with the medium copy regulator–reporter plasmid with *araC* with a frameshfit. All knockout strains are derivatives of BW25113. MG1655 but not BW25113 can grow on L-arabinose.

Stimulatory and inhibitory effects of L-arabinose during growth in M9 medium

In contrast to the negative effect of L-arabinose on growth in LB medium, a positive effect by L-arabinose was previously observed for growth in M9 minimal medium. These observations were made with the strains AR pWUR782 and AR pWUR774. These strains are similar to the above described AR pWUR768, but instead of the intact reporter *kan*, they have *leuB* (essential in L-leucine biosynthesis[134]) with a frameshift in the *leuB* gene. pWUR782 and pWUR774 have a low and medium copy number respectively. Bacteria were grown in M9 medium with 4 g L^{-1} D-glucose as sole carbon source and various concentrations of L-arabinose. After 48 h, the final OD600 (stationary phase) was measured (Fig. 3A). Growth of both strains was positively affected by L-arabinose, with a higher concentration having a more severe effect. In general, AR with the low copy plasmid pWUR782 grew a bit better,

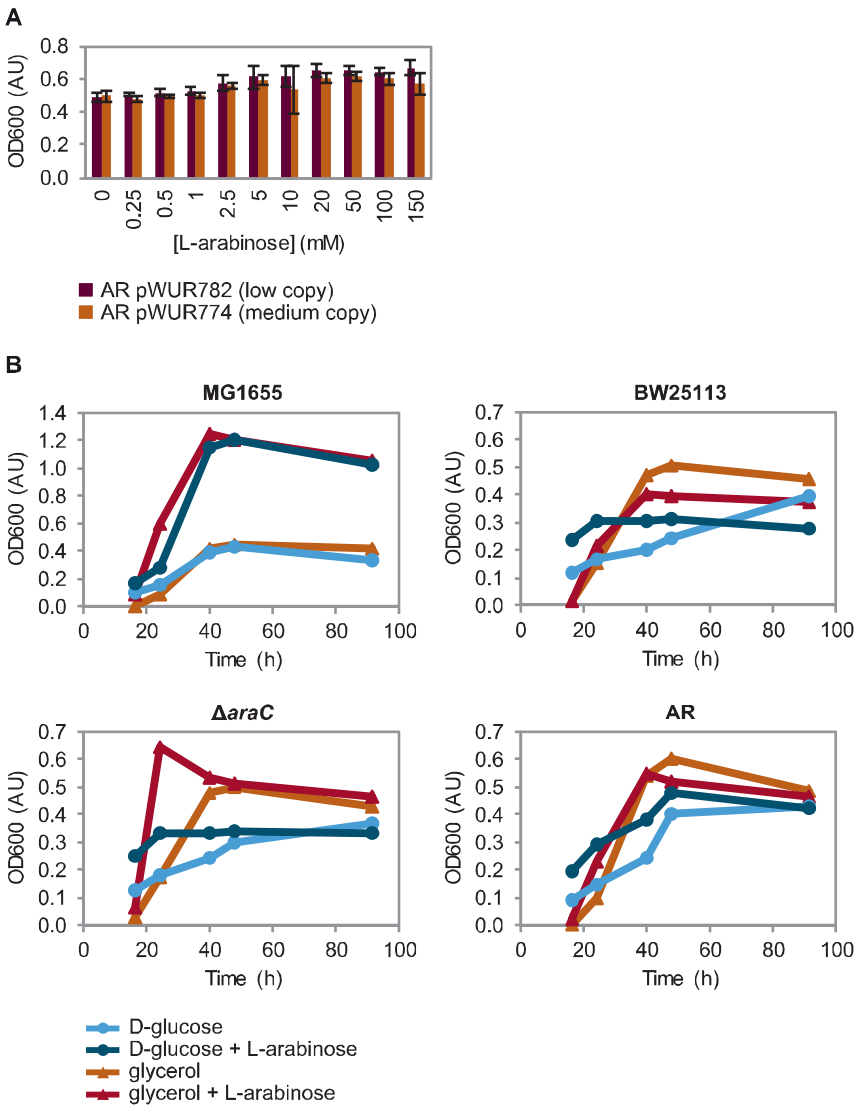


Fig 3: Positive and negative effects of L-arabinose on growth in M9 minimal medium. (A) Positive effect over a range of L-arabinose concentrations for strains with a low or medium copy plasmid that encodes the transcriptional regulator AraC. Cells were grown in 500 µL M9 minimal medium in a 2-mL 96-well MASTERBLOCK for 48 h. Data are an average of three independent experiments (standard deviations indicated). AR pWUR782 or pWUR774: *E. coli* BW25113 Δ*araC* Δ*recA* with respectively the low or medium copy regulator–reporter plasmids with wildtype *araC* and a frameshift in the reporter *leuB*. (B) Comparison of growth of MG1655 (with *araBAD* operon), BW25113 (without *araBAD* operon), Δ*araC* and AR (both without *araBAD* operon and without *araC*) in the presence and the absence of L-arabinose. Cells were grown in 2 mL M9 minimal medium in 10-mL tubes with 1 g L^{-1} D-glucose or glycerol as sole carbon source and with or without 10 mM (= 1.5 g L^{-1}) L-arabinose for 91.5 h. The graphs have a different y-axis.

probably due to less burden of the plasmid. Maximum stimulation was reached at 20 mM L-arabinose for both AR pWUR782 (1.33 ± 0.13 of final OD600 without L-arabinose) and AR pWUR774 (1.22 ± 0.11 of final OD600 without L-arabinose).

The stimulatory effect of L-arabinose on growth was unlikely due to growth on L-

arabinose, because of the *araBAD* operon deletion. To definitely rule out this possibility, cells were grown in M9 minimal medium with L-arabinose as only carbon source (2-mL cultures in 2-mL Eppendorfs; Fig. S3A). Only MG1655 cells, which have an intact *araBAD* operon, could grow on L-arabinose and AR cells indeed could not. Even after 13 days, no growth was observed. Since growth of AR cells without plasmid was also positively affected by L-arabinose, the effect was plasmid independent. The stimulatory effect seen after 24 h and 48 h, was gone after 72 h. To verify that L-arabinose was not consumed, cells were grown at larger scale (50-mL cultures in 250-mL erlenmeyers). HPLC analysis confirmed that L-arabinose itself was not consumed. However, when AR cells were grown in the presence of L-arabinose with D-glucose or glycerol as carbon source, D-glucose and glycerol were consumed faster than without L-arabinose (Fig. S3B). Here, the stimulatory effect was gone after 39 h, probably due to the better growth in erlenmeyers than in Eppendorfs.

To see if L-arabinose also positively affected growth of a wildtype *E. coli* and to see which phases of growth were affected by L-arabinose, growth curves of MG1655, BW25113, AR and Δ*araC* were made (Fig. 3B). Cells were grown in 2 mL M9 minimal medium with 1 g L^{-1} D-glucose or glycerol as sole carbon source and with or without 10 mM (= 1.5 g L^{-1}) L-arabinose. Based on Fig. 3A this L-arabinose concentration should give a close to maximum stimulation. MG1655 could grow on L-arabinose and therefore reached a higher OD600 in the presence than in the absence of L-arabinose. In case L-arabinose affected growth in another way, this was shielded by the growth on L-arabinose. Growth of BW25113, Δ*araC* and AR was all positively affected by L-arabinose in the early time points (24 h point for Δ*araC* probably a measurement error) for both carbon sources. However, the effect was prolonged when grown on D-glucose (16-47.5 h) compared to growth on glycerol (24 h). Surprisingly, after 91.5 h, only growth of BW25113 was negatively affected by L-arabinose when grown on either D-glucose (0.7 of OD600 without L-arabinose) or glycerol (0.8 of OD600 without L-arabinose).

Discussion

This study presents new experimental data of earlier observations that L-arabinose affects growth of *E. coli* strains that are incapable of L-arabinose catabolism (**Chapters 3**[22] and **5**)[184,280], in order to better understand these effects. In addition, the growth of various knockout strains in the presence and the absence of L-arabinose was analysed. An overview of the obtained results is presented in Table 1. In LB medium, L-arabinose inhibited growth

of *E. coli* BW25113 (lower final OD600). This effect was stronger in strains lacking *xylA*, encoding the first enzyme in D-xylose catabolism, but the effect was absent in knockout strains of *araC*, encoding the transcriptional regulator of L-arabinose metabolism, or *crp*, encoding the global regulator involved in CCR. However, Δ*araC* strains in which plasmid-encoded *araC* was replaced by *sacB* or by *araC* variants that encode L-arabinose unresponsive AraC mutants, were still inhibited by L-arabinose as well as some other monosaccharides to various extent: D-lyxose > L-arabinose > D-xylose > D-fucose > D-arabinose > L-rhamnose of which L-rhamnose had barely any effect. In M9 minimal medium with either D-glucose or glycerol as sole carbon source, L-arabinose stimulated growth of both BW25113 and Δ*araC* in an early phase of growth, but reduced the final OD600 of only BW25113.

Three potential causes could be excluded as reason of the negative effect of L-arabinose on growth of BW25113 in LB medium and in late phases of growth in M9 medium. (1) Toxicity by L-ribulose-5-phosphate accumulation. Growth of Δ*araD* strains, which lack L-ribulose-5-phosphate 4-epimerase, is inhibited by L-arabinose due to accumulation of L-ribulose 5-phosphate. This inhibitory effect can be resolved by disrupting the formation of L-ribulose 5-phosphate, for example by additionally mutating *araB*, encoding L-ribulosekinase[283]. Since BW25113 lacks the whole *araBAD* operon, accumulation of L-ribulose-5-phosphate is not possible for this strain. (2) Toxicity by methylglyoxal accumulation. When the rates of L-arabinose catabolism are high, glycolysis is overburdened. This leads to a build-up of dihydroxyacetone phosphate, stimulating methylglyoxal formation. The accumulation of toxic concentrations of methylglyoxal inhibits growth[284,285]. Since BW25113 is unable to catabolize L-arabinose, L-arabinose resultant accumulation of methylglyoxal is not possible for this strain. (3) Osmotic stress by too high sugar concentration. However, an osmotic effect for L-arabinose has been observed over a range from 50-250 mM with a stronger effect at higher concentrations[286], whereas the negative effect of L-arabinose described in this study reached a maximum at concentrations below 50 mM. In both cases, an *E. coli* strain unable to catabolize L-arabinose was used. In addition, the growth inhibitory effect of L-arabinose is barely changed when trehalose, an osmoprotectant, is overproduced and the L-arabinose concentration at which the growth is half of the growth without L-arabinose (IC_{50}) is 325 mM for *E. coli* W3110, which can use L-arabinose, grown in M9 medium with 20 g L^{-1} D-glucose[287]. Together this suggests that the negative effect on growth by L-arabinose described here, has a different underlying reason than osmotic stress.

Table 1: Overview of positive and negative effects on growth by L-arabinose

		Effect by L-arabinose[a]			
Genome	**Plasmid**	**LB**	**LB + D-glucose**	**M9 + glycerol[b]**	**M9 + D-glucose[b]**
BW25113[c]		-	- (later)[d]	+ (early)	+ (early)
				- (late)	- (late)
Δ*araC*		none	none	+ (early)	+ (early)
				none (late)	none (late)
Δ*araC*	*araC*	-	-	ND (early)	+ (early)
				ND (late)	ND (late)
Δ*araC*	*araC* with frameshift	none	none	ND	ND
Δ*araC*	*sacB*	- (less sensitive)[d]	ND	ND	ND
Δ*araC*	*araC* variants	-	ND	ND	ND
Δ*araC*	*araC*, *araA*	-	ND	ND	ND
Δ*araC* Δ*xylA*	*araC*	- -	ND	ND	ND
Δ*xylA*		- -	- -	ND	ND
Δ*crp*		none	none	ND	ND
Δ*araE*		- (more gradual)[d]	- (more gradual)[d]	ND	ND
Δ*araF*		-	-	ND	ND
Δ*araG*		-	-	ND	ND
Δ*araH*		-	-	ND	ND
Δ*araJ*		-	-	ND	ND
Δ*ytfQ*		-	- (very minor)[d]	ND	ND
		-	-		
Δ*ydeN*		needs for normal μ[e]	needs for normal μ[e]	ND	ND
Δ*ydeM*		-	-	ND	ND
Δ*polB*		-	-	ND	ND
Δ*ygeA*		-	-	ND	ND
Δ*dcp*		-	-	ND	ND

[a]Negative and positive growth effects of L-arabinose (lower of higher OD600 than without L-arabinose) are indicated with - or + respectively, with - - indicating a stronger negative effect than for BW25113. ND indicates that data are not determined. [b]In M9, L-arabinose affected growth differently in early and in late growth. [c]All knockout strains are derivatives of BW25113, which has a deletion of the *araBAD* operon. [d]Compared to BW25113 in LB medium. [e] Without L-arabinose Δ*ydeN* had a lower growth rate μ than BW25113.

Then what is the reason for this negative effect of L-arabinose on growth? Although no definitive conclusion can be drawn, two hypotheses can be put forward. Either the intracellular L-arabinose itself is burdening the cells, or the cells are fooled into switching to L-arabinose catabolism, while they cannot use this carbon source. Let us first discuss the former hypothesis. Since BW25113 is unable to catabolize L-arabinose, the negative feedback loop on L-arabinose concentration is missing. Meanwhile the positive feedback

loop of switching on transcription of L-arabinose transporters AraE and AraFGH is in place, resulting in more and more intracellular L-arabinose in the presence of extracellular L-arabinose. This high concentration could lead to catalysis by enzymes that can use L-arabinose only as side reaction and thereby produce possible toxic compounds. Another option is the inhibition of enzymes by L-arabinose. On www.brenda-enzymes.org a multitude of enzymes in carbon metabolism can be found, which have activity with L-arabinose as substrate or are inhibited by L-arabinose. This hypothesis fits with the onset of inhibition at the point that transcription of L-arabinose transporter genes is switched on. Although transcriptomics should be used for verification, is it likely that these genes were expressed at an OD600 of 0.5 or 1.4 in LB medium without or with D-glucose respectively (Fig. 2A). D-glucose exerts CCR and thereby leads to delayed expression of the L-arabinose transporter genes. In MG1655, this was visible as diauxie, in which cells temporarily stopped growing (OD600 = 1.4) while switching to a different carbon source. However, the growth arrest took inexplicably long for *E. coli*. For example, the diauxic lag phase for the shift from D-glucose to lactose metabolism in minimal medium takes 30-60 min, but can be extended for a couple of hours by the addition of L-valine[288,289]. Maybe the complex composition of the LB medium extended the growth arrest of MG1655. For BW25113, inhibition suddenly started at the indicated OD600 values, whereas for Δ*araE* inhibition was more gradual. This timing also fits with the onset of L-arabinose catabolism as observed by transcriptomics[268-270] and of the reduction of the growth rate when preferable carbon sources are depleted[168]. In the three strains that were unable to switch on this transport (Δ*araC*, Δ*araC* with a frameshift in the plasmid-encoded *araC* and Δ*crp*) L-arabinose did not inhibit growth. Strains with a high *araC* expression (medium copy versus low copy number plasmid) were less sensitive to L-arabinose, probably because more transcriptional regulator proteins led to a less steep response curve[117] and therefore L-arabinose transport was switched on at higher L-arabinose concentrations. However, if it was indeed the L-arabinose accumulation itself that caused the problems, why could the cells not solve this by switching on efflux transporters? L-arabinose can induce the expression of various transporters, but whether these export L-arabinose and, if so, under which conditions they operate, is not yet known[285]. Whether this hypothesis also fits with the growth inhibition in M9 minimal medium is hard to say, because the current data do not indicate the onset of L-arabinose transport in this medium.

The second hypothesis is that cells are fooled into switching to L-arabinose catabolism, while they cannot use this carbon source. This hypothesis has two sides. On the

one hand, cells invest ATP in the production of proteins that are involved in L-arabinose catabolism and transport and they invest ATP in the import of L-arabinose, while not obtaining energy from L-arabinose catabolism. Therefore less energy is available for growth. On the other hand, AraC and L-arabinose indirectly repress genes involved in the catabolism of other carbon sources when grown in LB medium, like maltose (*malE*, *malF*, *malG*, *malK*, *malM*, and *lamB*), threonine (*tdcA*, *tdcB*, *tdcC*, *tdcD*, and *tdcE*), D-glucarate/D-galactarate (*garD*, *garL*, *garP*, and *garR*), tryptophan (*tnaA*, *tnaB*, and *tnaL*), and L-fucose (*fucI*)[131]. Most of these genes are also regulated by CRP and their protein levels are upregulated in minimal medium in a D-glucose-limited fed batch phase, a phase in which the cells try to use alternative carbon sources[271]. In the presence of L-arabinose, these genes are repressed and cells would have more trouble using alternative carbon sources, while not being able to grow on L-arabinose, possibly explaining the reduced growth. This hypothesis also fits with the onset of L-arabinose import, with the delayed inhibition in the presence of D-glucose, with the higher sensitivity in the presence of less AraC and with the absence of inhibition in Δ*araC*, Δ*araC* with a frameshift in the plasmid-encoded *araC* and Δ*crp*. In addition, growth inhibition by L-arabinose is observed both in LB medium and in M9 minimal medium, presumably when the preferred carbon sources run out.

Both hypotheses seem plausible, but they do not explain the phenotype of strains in which the plasmid-encoded *araC* is replaced by *sacB* or by *araC* variants that encode L-arabinose unresponsive AraC mutants. In these strains, *araE* and *araFGH* are either not switched on or switched on very little (some DNA binding activity, probably to two adjacent operators, remained in the AraC variants) and they should not be able to import much L-arabinose[290]. Still, the *sacB* expressing strain was only a bit less sensitive to L-arabinose and the *araC* variants expressing strains were similarly inhibited by L-arabinose as the strain with wildtype *araC*. Apart from the inexplicable import of L-arabinose, these strains lack the wildtype *araC*, making the AraC dependent second hypothesis less likely. However, since growth was the only output, the inhibitory effect in *sacB* expressing strains could have a different underlying cause. SacB can use L-arabinose as acceptor, but it is not described whether it is able to use another substrate than sucrose, which is probably not present in LB medium or only in very little amounts. Since the AraC variants retained some DNA binding activity, probably to two adjacent operators, perhaps some functions of the wildtype AraC necessary for the inhibitory growth effect were still intact.

A few other observations regarding the negative effect on growth by L-arabinose are also not yet explainable. The effect was reduced in Δ*ytfQ* in the presence of D-glucose; it was

similar in Δ*ydeN* but with a lower growth rate in the absence of L-arabinose; it was stronger in Δ*xylA*; and the growth of ARX pWUR768 was also inhibited by other monosaccharides than L-arabinose. Based on the current data, nothing conclusive can be said about the growth effects of L-arabinose on Δ*ytfQ* and Δ*ydeN*. The stronger growth inhibiting effect of L-arabinose for Δ*xylA* than for BW25113 occurred both in LB medium and in LB medium with D-glucose. In the latter, the lag phase of Δ*xylA* was 1.5 h longer than that of BW25113 in the presence of L-arabinose. Since the effect was also stronger for ARX pWUR768 compared to AR pWUR768, it would be interesting to see the effect of L-arabinose on the growth of ARX without the plasmid-encoded *araC* to discover more about the possible dependence on AraC. Although XylA has very low activity with L-arabinose as substrate[291,292] and thereby could reduce the intracellular L-arabinose concentration, it is unlikely that this would cause such a large effect on growth. A more reasonable explanation would be the cross-talk between the D-xylose and the L-arabinose catabolism, but further analysis is required to demonstrate this.

L-arabinose was not the only monosaccharide that negatively inhibited the growth of ARX pWUR768 and its derivatives in which the plasmid-encoded *araC* was replaced by *araC* variants that encode L-arabinose unresponsive AraC mutants or by *sacB*. The severity of the effect differed per monosaccharide (D-lyxose > L-arabinose > D-xylose > D-fucose > D-arabinose > L-rhamnose of which L-rhamnose had barely any effect), but cannot be directly linked to the ability for monosaccharide uptake. All monosaccharides were probably taken up by this strain, but the uptake of D-arabinose has a low efficiency (**Chapter 5**)[128,138,252,255,256] and the uptake of D-lyxose is unknown. Also the inability to catabolize the monosaccharide, as is the case for L-arabinose, cannot be linked to the severity of the inhibiting effect. None of these monosaccharides can be catabolized by ARX (**Chapter 5**)[138,252,253]; only D-arabinose can be catabolized in case a mutation makes expression of the L-fucose catabolic genes constitutive[254]. It looks like the pentoses (D-lyxose, L-arabinose, D-xylose) had a more severe effect than the hexoses (D-fucose, L-rhamnose), but no real statement can be made. D-arabinose was left out here because of less efficient uptake. Based on these limited data, it is hard to say whether the inhibition by these different monosaccharides had some common ground or whether the underlying causes were different in each case.

Next to the negative effect on growth in both LB medium and M9 minimal medium, L-arabinose had a positive effect on growth in M9 minimal medium in an early phase of growth (Table 1). This effect was not caused by growth on L-arabinose as carbon source and

since growth of both BW25113 and Δ*araC* was stimulated by L-arabinose, the effect was likely AraC independent. Although the effect was prolonged when growing on D-glucose compared to growth on glycerol, it is hard to say if this was the result of CCR. The molar amounts of these carbon sources were not the same. However, both glycerol and D-glucose were consumed faster in the presence of L-arabinose and also the growth rate was higher. This increased growth rate has also been observed by Afroz *et al.*[293] in a strain lacking *araBAD* when grown in M9 minimal medium with glycerol and 0.2% casamino acids. This effect is reduced when next to the *araBAD* deletion the expression of either *araFGH* or *araE* was made constitutive with the latter having a stronger effect or when *araE* expression is made constitutive in a Δ*araBAD* Δ*araFGH* background. The effect is absent in a Δ*araBAD* Δ*araE* strain with constitutive *araFGH* expression. It is not known how it is possible that uncoupling of expression of the transporter genes and the import of L-arabinose reduced the stimulatory effect of L-arabinose, whereas a deletion of *araC* did not influence the effect. And how to explain this stimulatory effect of L-arabinose? Do cells somehow detect that they are on richer medium with multiple carbon sources and therefore grow better? Unfortunately, these data do not yet provide an answer.

In order to explain the growth inhibitory and stimulatory effects of L-arabinose and to verify the growth curves of the knockout strains (experiments that were performed only once) further experiments are required. Growth curves with more time points over a longer time frame and in triplicate would give a lot of information, certainly when including more strains and comparing them in LB medium with or without D-glucose and in M9 minimal medium with glycerol or D-glucose. Sugar analysis by HPLC would further enhance the understanding. Although is it unlikely influencing the results, it would be best to use BW25113 instead of BW25113 (DE3)[294], which was used as wildtype in this study. BW25113 (DE3) has an integration of the λDE3 prophage in the promoter region of *ybhC*, encoding a putative acyl-CoA thioester hydrolase (Uniprot). Since Δ*ybhC* grows properly in LB medium[161] and most knockout strains in this study, which had BW25113 without λDE3 as parent, were similarly affected by L-arabinose as BW25113 (DE3), it is unlikely that this integration changed the outcome. Other interesting strains to include for the growth assays are (1) ARX and Δ*xylR* to see the effect of cross-talk between L-arabinose and D-xylose catabolism, (2) BW25113, ARX and AR with plasmid-encoded *araC* replaced by *sacB* or some other gene to elucidate the role of *sacB*, (3) ARX and AR with plasmid-encoded *araC* variants that are not responding to L-arabinose or that are constitutive ON or OFF to study the role of AraC, (4) strains that have the L-arabinose transporters constitutively expressed or expressed from an

inducible promoter that is controlled by another regulator than AraC (possibly in an Δ*araC* background) to determine whether the negative growth effect seems dependent on AraC because it switches on the transporters or because of some other reason, (5) Δ*araE* Δ*araFGH* to see if L-arabinose still effects the cells when it can barely be imported, and (6) strains constitutively expressing genes involved in the catabolism of alternative carbon sources that are indirectly regulated by AraC and L-arabinose, like *tnaA* or *tdcB*, to see if possible repression of these genes by L-arabinose or AraC can be removed. Although these growth assays would provide a lot of information, preparing all these strains is rather laborious. A better approach would be to start with a transcriptomic, a proteomic and/or a metabolomic analysis of BW25113 and Δ*araC* with and without L-arabinose. Such approach would give a clue in which direction to look and to give focus to the growth assays.

Conclusions

The work performed in this study extended our knowledge on the inhibitory and stimulatory effects of L-arabinose on growth of *E. coli* BW25113, but more work is needed to unravel the underlying mechanism of these effects. Nevertheless, the obtained insights provide a basis for lines of research to elucidate this mechanism. Although the effects have only been described for *E. coli* strains that are unable to catabolize L-arabinose, unravelling of the mechanism could possibly reveal a common biology of various *E. coli* strains. Moreover, the knowledge on how L-arabinose is involved in the regulation of carbon catabolism and on how L-arabinose affects cell growth could contribute to various application fields in which L-arabinose functions as substrate or inducer. Knowledge on how L-arabinose influences *E. coli* cells could contribute to an improved production of green chemicals, to an improved production or expression of proteins and to a better understanding of the role of *E. coli* in the intestinal microbiome and in disease.

Materials and methods

Bacterial strains and media

The used *Escherichia coli* strains are presented in Table 2. For LB medium, 10 g tryptone, 5 g yeast extract and 5 g NaCl per litre of medium were used. For M9 medium, 1x M9 minimal salts, 2 mM $MgSO_4$, 4 g L^{-1} D-glucose and 0.1 mM $CaCl_2$ were used, unless stated otherwise.

Strains with plasmids were grown in the presence of 34 µg mL^{-1} or 18 µg mL^{-1} chloramphenicol for plasmid maintenance in LB medium or M9 minimal medium respectively. L-arabinose (5118.3, Carl Roth).

Construction of pWUR948

To introduce a frameshift in *araC* in pWUR768, the plasmid was digested inside *araC* at the unique restriction site PauI (Thermo Scientific). PauI was inactivated for 20 min at 80°C and DNA ends were made blunt by Klenow (Thermo Scientific) according to the protocol of Thermo Scientific. After gel extraction with the Zymoclean Gel DNA Recovery Kit (D4002, Zymo Research), the ends were ligated with T4 ligase for 1 h at room temperature. Plasmids were propagated in *E. coli* DH10B $T1^R$, which was grown and transformed by standard methods[160]. Transformants were plated on LB agar plates with 10 mM L-arabinose. DH10B pWUR768 was used as control. After overnight growth, bioluminescence was measured with the lumiglo function of the G:BOX Chemi XT4 (Syngene) to select colonies that were no longer able to produce bioluminescence in the presence of L-arabinose. To confirm that these colonies had a frameshift in *araC*, plasmids were isolated with the Plasmid Miniprep kit of Thermo Scientific (#K0503) and sequenced at GATC Biotech with primers BG3942 (CAACGTCTCATTTTCGCCAG) and BG3653 (GCCTGTTGAACAAGTCTGGA).

Growth assays in LB medium

The growth assay with a range of L-arabinose concentrations was performed as the KmR-based induction assay described in **Chapter 3**[184]. All cultures with L-arabinose were inoculated from pre-cultures with 10 mM L-arabinose. The growth assay with L-arabinose isomerase AraA was performed as the detection assay described by **Chapter 3**[184]. The growth assay with *araC* replaced by *sacB* was performed as the KmR-based selection assay described in **Chapter 5**, except that the samples were included in duplicate. The growth assay with other monosaccharides was performed as the bioluminescence-based screening assay described in **Chapter 5**. In none of the assays, kanamycin was added and in all assays, cells were grown in 500 µL LB medium in a 2-mL 96-well MASTERBLOCK for 17 h, except for the assay with the other monosaccharides in which cells were grown in 200 µL LB medium in a 96-well plate in a platereader for 16 h. Values were an average of three (L-arabinose concentration range assay), four (AraA and ARX pWUR768 with other monosaccharides assays) or two (*araC* variants with other monosaccharides assay) independent experiments or an average of two replicates within one experiment (*sacB* assay). For the assay with other

monosaccharides, the 0 mM samples were included in duplicate.

For the growth assay with various strains, strains were pre-grown on LB agar plates overnight, followed by pre-growth in 500 µL LB medium with 0 or 1 g L^{-1} D-glucose in a 2-mL 96-well MASTERBLOCK (780271, Greiner Bio-One) for 7 h. After OD600 measurement, cells were diluted to have a starting OD600 of 0.0001 in the assay. Cells were grown in 200 µL LB medium with 0 or 1 g L^{-1} D-glucose and 0 or 10 mM L-arabinose in a transparent 96-well plate (655101, Greiner Bio-One) in a Synergy MX microplate reader (BioTek) for 19 h. Every 10 minutes the OD600 was measured. For each sample, the value of the first time point was subtracted from the values of the other time points and all values were corrected for the path length with the average path length of eight blanks.

Growth assays in minimal M9 medium

The growth assay with a range of L-arabinose concentrations was performed as the LeuB-based induction assay described in **Chapter 3**[184]. The medium was supplemented with 1x minimum essential medium (MEM) vitamins. For the growth assay with various strains, cells were pre-grown on M9 minimal medium plates with 1 g L^{-1} glycerol as sole carbon source and in 2 mL M9 minimal medium in 10-mL tubes (Gosselin) with 1 g L^{-1} glycerol or D-glucose as sole carbon source (a higher glycerol concentration did not give good growth). From here 2 mL M9 minimal medium in 10-mL tubes (Gosselin) with 1 g L^{-1} glycerol or D-glucose and 0 or 10 (= 1.5 g L^{-1}) mM L-arabinose were inoculated at a starting OD600 of 0.0001. After 16, 24, 40, 47.5 and 91.5 h, the OD600 was measured with a Synergy MX microplate reader (BioTek) by transferring 200 µL culture to a transparent 96-well plate (655101, Greiner Bio-one). The values were corrected for path length and an average of three blanks.

Table 2: Bacterial strains used in this study

Strain	Plasmid	Description
DH10B T1^{R} (Invitrogen)	-	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*), Φ80*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(*ara, leu*)7697 *galU galK* λ- *rpsL nupG tonA*
MG1655 (ATCC)	-	F-λ- *ilvG- rfb-50 rph-1*
BW25113[a 294]	-	BW25113 (DE3): Δ(*araD-araB*)*567* Δ(*rhaD-rhaB*)*568* Δ*lacZ4787*(::*rrnB-3*) *hsdR514 rph-1 attB*::λ*DE3*
AR (**Chapter 3**[184])	-	BW25113 Δ*araC* Δ*recA*
AR(**Chapter 3**[184])	pWUR768(**Chapter 3**[184])	AR + reg.-rep. plasmid[b] (*araC* under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
AR (**Chapter 3**[184])	pWUR772 (**Chapter 3**[184])	AR + reg.-rep. plasmid[b] (*araC* under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, p15A)
AR (**Chapter 3**[184])	pWUR948 (this study)	AR + reg.-rep. plasmid[b] (*araC* with frameshift under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
AR (**Chapter 3**[184])	pWUR768 + pWUR833 (**Chapter 3**[184])	AR + pWUR768 + *araA* expression plasmid (*araA* under constitutive P_{bla}, p15A)
AR (**Chapter 3**[184])	pWUR768 + pWUR917 (**Chapter 3**[184])	AR + pWUR768 + empty plasmid (no *araA* cds, p15A)
AR (**Chapter 3**[184])	pWUR774 (**Chapter 3**[184])	AR + reg.-rep. plasmid[b] (*araC* under constitutive P_{lacI}^{Q}, *leuB* with frameshift and *luxCDABE* under AraC control, ColE1)
AR (**Chapter 3**[184])	pWUR782 (**Chapter 3**[184])	AR + reg.-rep. plasmid[b] (*araC* under constitutive P_{lacI}^{Q}, *leuB* with frameshift and *luxCDABE* under AraC control, p15A)
ARX (**Chapter 5**)	-	BW25113 Δ*araC* Δ*recA* Δ*xylA*
ARX (**Chapter 5**)	pWUR768 (**Chapter 3**[184])	ARX + reg.-rep. plasmid[b] (*araC* under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
ARX (**Chapter 5**)	pWUR947 (**Chapter 5**)	ARX + reg.-rep. plasmid[b] (*sacB* under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
ARX (**Chapter 5**)	Lib1-Xyl23 (**Chapter 5**)	ARX + reg.-rep. plasmid[b] (*araC* variant Lib1-Xyl23 under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
ARX (**Chapter 5**)	Lib2-Xyl8 (**Chapter 5**)	ARX + reg.-rep. plasmid[b] (*araC* variant Lib2-Xyl8 under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
ARX (**Chapter 5**)	Lib2-Xyl9 (**Chapter 5**)	ARX + reg.-rep. plasmid[b] (*araC* variant Lib2-Xyl9 under constitutive P_{lacI}^{Q}, *kan* and *luxCDABE* under AraC control, ColE1)
Δ*araC* (KEIO[161])	-	BW25113 JW0063-1 Δ*araC771*::*kan*
Δ*araF* (KEIO[161])	-	BW25113 JW1889-1 Δ*araF751*::*kan*
Δ*araG* (KEIO[161])	-	BW25113 JW1888-1 Δ*araG750*::*kan*
Δ*araH* (KEIO[161])	-	BW25113 JW1887-1 Δ*araH749*::*kan*
Δ*araE* (KEIO[161])	-	BW25113 JW2809-1 Δ*araE766*::*kan*
Δ*araJ* (KEIO[161])	-	BW25113 JW0386-1 Δ*araJ760*::*kan*

Table 2 continued

Strain	Plasmid	Description
Δ*xylA* (KEIO[161])	-	BW25113 JW3537-1 Δ*xylA748*::*kan*
Δy*tfQ* (KEIO[161])	-	BW25113 JW4186-1 Δ*ytfQ775*::*kan*
Δ*ydeN* (KEIO[161])	-	BW25113 JW5243-2 Δ*ydeN720*::*kan*
Δ*ydeM* (KEIO[161])	-	BW25113 JW1492-1 Δ*ydeM790*::*kan*
Δ*polB* (KEIO[161])	-	BW25113 JW0059-1 Δ*polB770*::*kan*
Δ*ygeA* (KEIO[161])	-	BW25113 JW2808-1 Δ*ygeA765*::*kan*
Δ*dcp* (KEIO[161])	-	BW25113 JW1531-1 Δ*dcp-759*::*kan*
Δ*crp* (KEIO[161])	-	BW25113 JW5702-4 Δ*crp-765*::*kan*

[a]Only strain BW25113 has the integration of λDE3. Strains AR, ARX and the strains of the KEIO collection do not have this integration. [b]reg.-rep., regulator-reporter plasmid.

Supplementary information

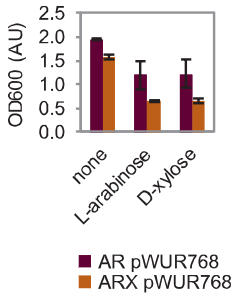


Fig S1: Negative effect of L-arabinose and D-xylose on growth of ARX pWUR768 and AR pWUR768. The assay was performed as the KmR-based selection assay described in **Chapter 5**. In short, cells were grown in 500 µL LB medium with 0 or 20 mM L-arabinose or D-xylose in a 2-mL 96-well MASTERBLOCK for 17 h (no kanamycin was used). Data are an average of four independent experiments (standard deviations indicated). AR(X) pWUR768: *E. coli* BW25113 Δ*araC* Δ*recA* (Δ*xylA*) with the medium copy regulator–reporter plasmid with wildtype *araC*.

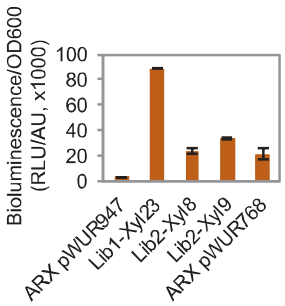


Fig S2: Leakiness of the luciferase operon under control of wildtype AraC or one of three non L-arabinose responsive AraC variants. The assay was performed as the LuxCDABE-based screening assay described in **Chapter 5**, but the outer wells were used as well. In short, cells were grown in 200 µL LB medium without inducer in a white 96-well plate with transparent bottom for 16 h. The bioluminescence and OD600 were measured every 15 min. Bioluminescence over OD600 values were averaged over three subsequent time points in the late log phase for which the signal production and wash out due to cell division were about equal. The signal of the negative control ARX pWUR780 was subtracted. Values of ARX pWUR768 and ARX pWUR947 are an average of two independent experiments (overall standard deviation indicated). Although this specific experiment was only performed once (or twice for two of the strains), the trend that leakiness of *araC* variants > ARX pWUR768 > ARX pWUR947 was observed in various experiments. ARX pWUR768: *E. coli* BW25113 Δ*araC* Δ*recA* Δ*xylA* with the medium copy regulator–reporter plasmid with wildtype *araC*. ARX pWUR947: ARX with the medium copy regulator–reporter plasmid with *sacB* instead of *araC*. ARX pWUR780: ARX with the regulator–reporter plasmid with wildtype *araC* and a frameshift in *luxA*.

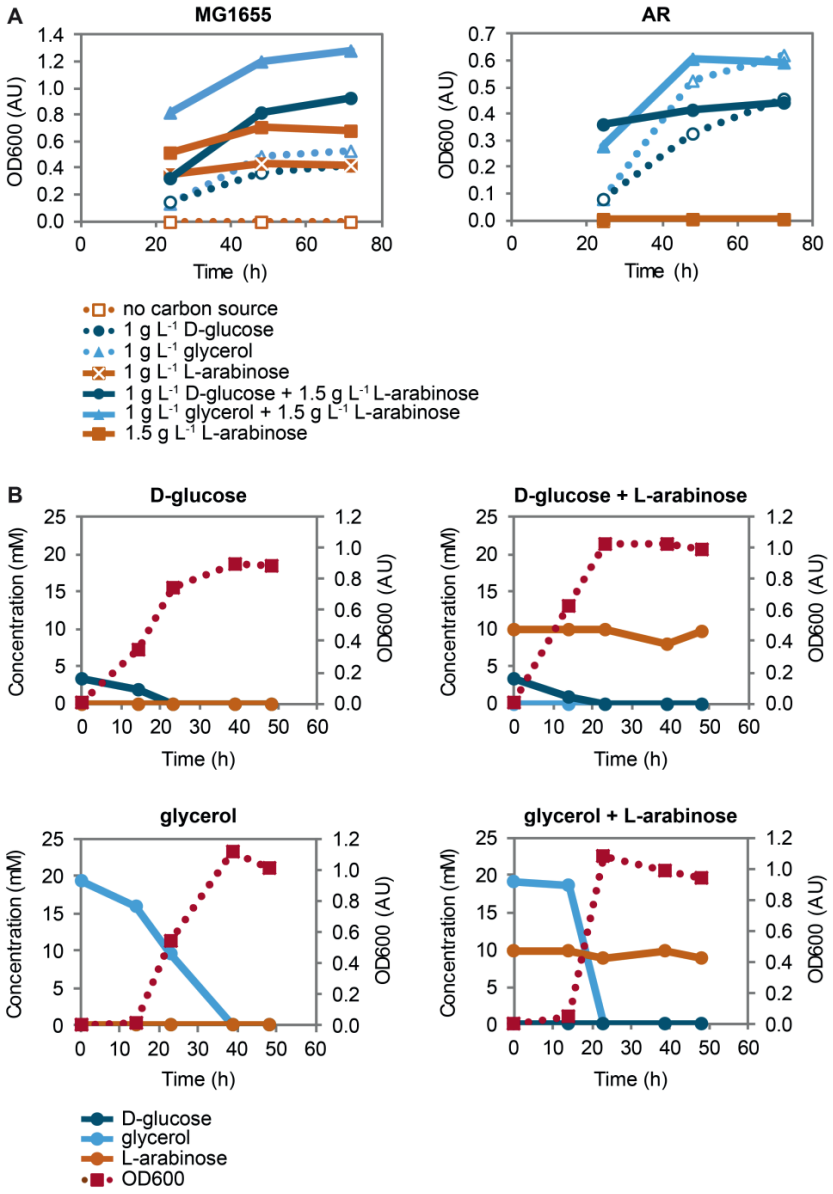
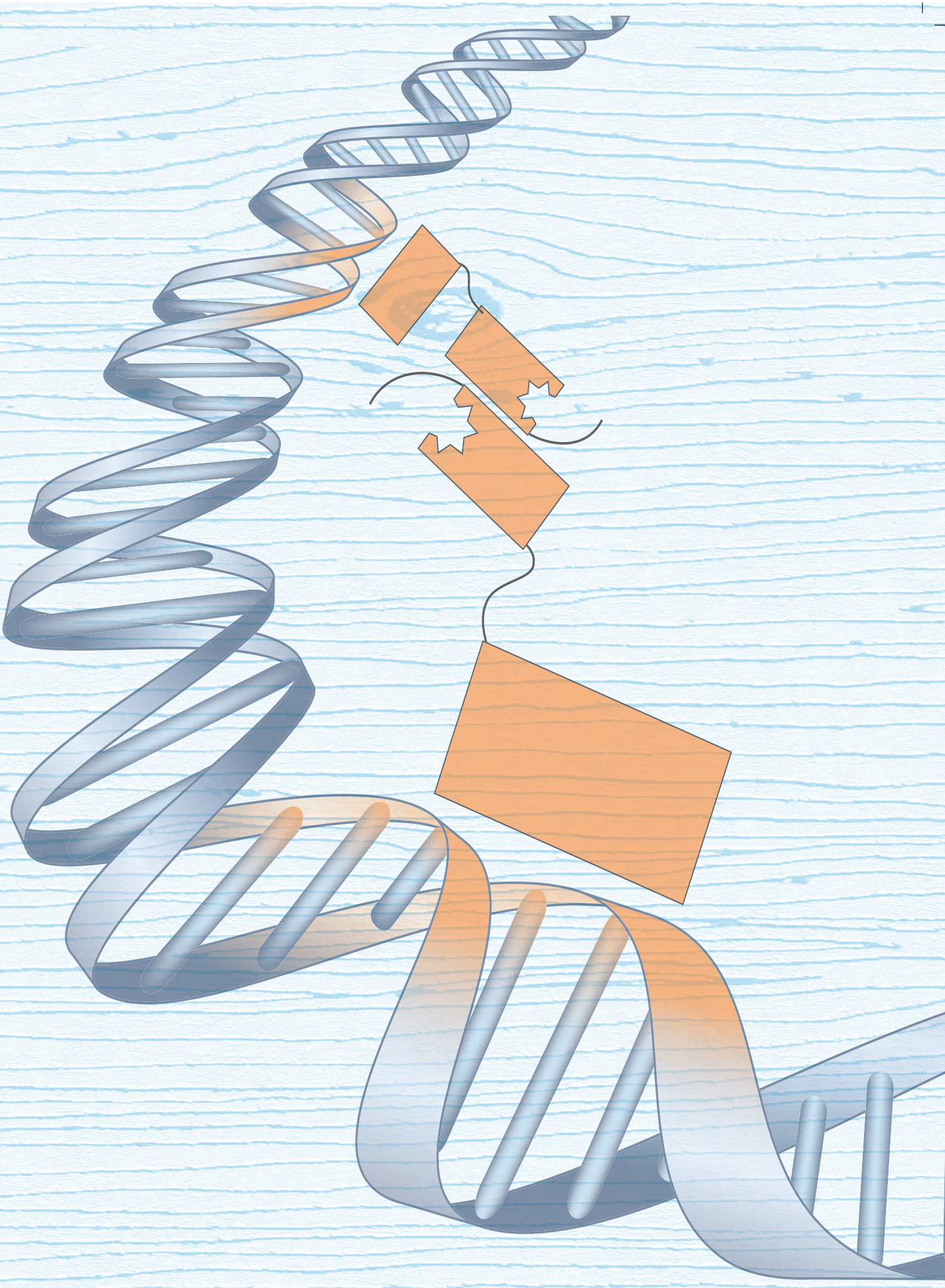
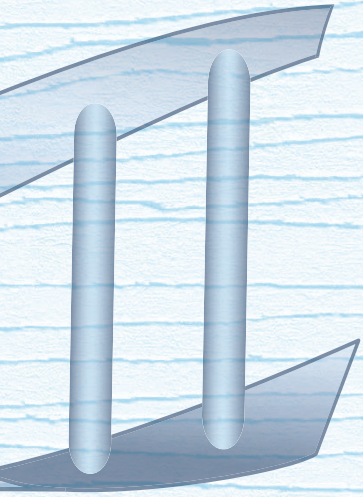


Fig S3: Positive effect of L-arabinose on growth in M9 minimal medium. (A) Growth of *E. coli* MG1655 (with *araBAD* operon) and *E. coli* BW25113 Δ*araC* Δ*recA* (AR; without *araBAD*) on various single or mixed carbon sources. Cells were grown in 2 mL M9 minimal medium with the indicated amounts of carbon sources in 2-mL Eppendorfs. Cultures were inoculated from biomass grown on LB agar plates and resuspended in 10x diluted M9 minimal medium without carbon source (MG1655 and AR had different starting OD600). After 24, 48, 72 h and 13 days of growth, the OD600 was measured (200 µl samples in a transparent 96-well plate (655101, Greiner Bio-one).using a Synergy MX microplate reader (BioTek)).(B) Consumption of carbon sources by AR cells. Cells were pre-grown on LB agar plates and in 2 mL M9 minimal medium in 10-mL tubes (Gosselin) with 1 g L^{-1} glucose or glycerol as sole carbon source for 24 h. From here, 50 mL M9 minimal medium in 250-mL erlenmeyers were inoculated (starting OD600 = 0.0001). The carbon source was 1 g L^{-1} glucose or glycerol with 0 or 10 mM (=1.5 g L^{-1}) L-arabinose. After 14, 23, 39 and 48 h of growth, 1.5 mL samples were taken for OD600 measurement and HPLC. For HPLC, 160 µL sample or glycerol, glucose or L-arabinose standard, were mixed with 40 µL internal standard (10 mM DMSO in 0.02 M H_2SO_4). Samples were analysed on a Dionex ICS-5000+ DP (Thermo Scientific) HPLC system with an Aminex HPX-87H column (0.8 mL min^{-1}, 60 °C), a mobile phase of 8 mM H_2SO_4, UV detector (210 nm) and RI detector (35 °C).Data were analysed with Chromeleon version 7.



Chapter 7

Summary and General discussion



Summary

Enzymes are proteins that catalyse chemical reactions. They accelerate the reaction by lowering the activation energy, thereby allowing the equilibrium to be reached more quickly. The molecules upon which enzymes may act are called substrates and the enzyme converts the substrates into molecules known as products. Nearly all metabolic processes in a living cell need enzyme catalysis in order to proceed at rates fast enough to sustain life. In **Chapter 1**, the relevance of enzymes for society is described. This dates back to more than 8000 years ago, when people unknowingly already made use of enzymatic conversions via fermentation by whole-cell microorganisms to make early forms of bread and beer**.** In the last two centuries, our knowledge on how enzymes are functioning increased tremendously. They have remarkable features that make them interesting for industrial applications, such as a high specificity and selectivity, and they can make processes 'greener' by replacing often polluting or toxic chemical reactions. This resulted in the large scale industrial production and application of enzymes in diverse areas ranging from food industry and detergents to pharma and DNA technology. Currently, there is both the need and the room to increase the number of enzyme applications. A more extensive implementation of enzymes in industry is very important because, sooner or later, we have to make the shift from a fossil fuel based economy to a biobased economy. At this moment, only 5% of all chemical products are produced biologically. For the latter, only 150-170 of the 3000 different types of known enzymes are being applied. It is estimated that only 1% of enzymes is known and, therefore, there is much room to extend the number of industrial enzyme applications. But how to find novel enzymes for these applications? Enzymes can be obtained from nature (natural evolution, screening of metagenomics libraries) or by enzyme engineering via laboratory evolution (screening of random or semi-random enzyme variant libraries) or computational design (*in silico* generation and screening of enzyme variant libraries followed by experimental verification screens). Together these approaches comprise a multitude of methods to find or generate an enormous amount of genetic diversity. To obtain the desired variants from these large libraries it is essential to have an efficient screening method, in which phenotype and genotype are linked. However, since screening methods are often time-consuming, complicated and/or require expensive equipment, this screening step is the main bottleneck in obtaining novel enzymes.

Reporter-based *in vivo* screening and selection is a new development in order to deal

with the large numbers in screening for novel enzymes. This approach is extensively discussed in **Chapter 2**, including a comparison of other *in vivo* screening and selection strategies and the various reporter-based mechanisms. In reporter-based approaches, it is not the enzymatic conversion or its product that result in a measurable property, but rather a genetically encoded reporter that gives a discriminating phenotype. Since the enzyme activity is measured indirectly, this approach is independent of the reaction and thus general. A sensor part couples the enzyme activity to the reporter and, hence the reporter choice determines the detection output, e.g. bioluminescence or fluorescence. The cell containing these sensor and reporter components, and functions as a reporter, is called a whole-cell bioreporter or simply bioreporter. As sensor part, various biomolecules are possible, either protein- or RNA-based (riboswitches), but the most popular sensor is a transcriptional regulator. This transcriptional regulator very specifically binds the enzymatic product, which results in a conformational change that modifies its DNA binding capacity and switches on expression of a reporter gene. Consequently, the specificity of the sensor has to be modified for each product. Although the development of this and other reporter-based strategies takes time and effort, the many advantages make this a very powerful screening method. These advantages include: wide applicability, screening for enantioselectivity, signal enhancement, no need for artificial substrates and high-throughput screening.

The most commonly used reporter-based screening method combines the reporter GFP with screening by fluorescence activated cell sorting (FACS). Though this is a high-throughput method that has proven very successful in obtaining new and/or improved enzymes, it does entail expensive equipment and experienced people operating this equipment. The aim of this thesis is to make this technology easier by developing a generic and high-throughput *in vivo* reporter-based system that involves selection instead of screening. The advantage of selection over screening is that only positive cells, containing the active enzyme, stay in the library pool, which allows for a quick reduction of the initially large library size. Although this thesis is not describing the first reporter-based selection system, other systems often are not applicable for a wide range of enzymes. In this thesis, the modular set-up of the system should make it more generic. To show that the developed system can be used in finding novel enzymes, a proof of principle is required. This consist of three aspects: (1) detect a product of an enzymatic activity, (2) apply the system at library scale, and (3) change the specificity of the system to make it applicable for a wide range of enzymatic products and thus different enzymes. In **Chapter 3**, the first two aspects are

demonstrated, while **Chapters 4** and **5** each focus on a different approach for the third aspect.

Our bioreporter is based on the most common reporter-based strategy, namely the transcriptional regulatory-based strategy, and couples enzymatic activity to growth of the bacterium *Escherichia coli*. **Chapter 3** covers the development of this *in vivo* transcriptional regulator-based selection system. A high false positive rate is a returning problem for growth-based selection and, therefore, our system was designed with dual reporters, both a selection and a screening reporter. The sensor part of the bioreporter is based on the transcriptional regulator AraC, which is involved in L-arabinose metabolism in *E. coli*, because AraC has been well studied. Furthermore, protein structures of AraC with and without ligand are available and it has been a topic of several engineering studies. In our system, the AraC sensor binds the product of the enzymatic reaction and switches on transcription of both a selection reporter (LeuB or KmR; enabling growth), for rapid reduction of the initially large library size, and a screening reporter (LuxCDABE; causing bioluminescence), for exclusion of false positives and quantification of positive variants. The characteristics of four different systems, differing in the selection reporter (LeuB or KmR) and in the plasmid origin of replication (low or medium copy number), are compared. The medium copy number system with KmR as selection reporter was found to be the best performing system based on leakiness, maximal signal, dynamic range and sensitivity in both selection and screening. Most importantly, a proof of principle of this system was provided by selecting cells expressing an L-arabinose isomerase derived from mesophilic *E. coli* or thermophilic *Geobacillus thermodenitrificans*. A more than a millionfold enrichment of cells with L-arabinose isomerase activity was established by selection and exclusion of false positives by screening. This shows the value of the dual selection and screening system for the detection of both mesophilic and thermophilic enzymes at library scale.

However, in order to demonstrate that our bioreporter is generic and can be applied for a wide range of enzymes, its specificity needs to be adaptable towards the product of any enzyme. In **Chapters 4** and **5**, two different approaches to change the specificity of the bioreporter are presented. In **Chapter 4**, the replacement of the transcriptional regulator AraC by LacI (the regulator of lactose metabolism in *E. coli*), is described. The characteristics of four different systems, all having LacI as transcriptional regulator, but varying in the selection reporter (LeuB or KmR) and in the plasmid origin of replication (low or medium copy number), were compared. The low copy system with LeuB as selection reporter was selected as best performing system and using this system, it was demonstrated that

previously described weak inducers or anti-inducers can be detected. The newly developed LacI-based system was compared with the original AraC-based system. The LacI-based system has a better sensitivity and a higher fold change of maximal signal over leakiness, but its dynamic range for selection is lower than that of the AraC-based system. Although some optimization is required, the replacement of the transcriptional regulator is rather straightforward due to the modularity of the system. It is a good approach to alter the specificity of the dual selection and screening system and thereby to broaden its range of potential target molecules.

A second approach to change the system's specificity is described in **Chapter 5**. This approach is based on engineering the ligand specificity of AraC, from L-arabinose to D-xylose, by targeting residues in the ligand binding pocket with combinatorial site-saturation mutagenesis. Others have already successfully modified the specificity of AraC using a GFP- and FACS-based screening of transcriptional-regulator variants. The aim here was to offer a simpler and alternative method by using growth-based selection instead. To this end, the dual reporter system itself was applied for selection and screening of transcriptional-regulator variants. The complete process is described, starting from library design and construction up to kanamycin resistance-based selection and bioluminescence-based screening of these libraries in the presence of D-xylose. Some of the developed AraC variants showed an altered, albeit small, response to D-xylose and several other tested monosaccharides. The selected variants yet have to be investigated in more depth to verify whether their ligand specificity is truly modified. Nonetheless, these variants will be interesting starting points for further engineering and indicate that the right positions in the protein were targeted. However, to obtain variants that give a better response, the selection and screening set-up needs to be optimized. After this optimization step, the same set-up could be used to select not only AraC variants with a better response to D-xylose, but also variants specific to other target molecules.

Inhibitory and stimulatory effects of L-arabinose on growth of *E. coli* were observed during the experimental work with the AraC-based dual reporter system (**Chapters 3** and **5**). In **Chapter 6**, these observations are supplemented with follow-up experiments to understand the underlying regulatory mechanisms of these effects. The growth effects caused by L-arabinose are described for the system strain, its parent strain *E. coli* BW25113 and various single gene knockout strains derived from BW25113. In LB medium, L-arabinose negatively affects growth of wildtype strain BW25113 (lower final OD600), but not of Δ*araC* or Δ*crp* strains. In addition, the effect is stronger in strains with Δ*xylA*, encoding the first

enzyme in D-xylose catabolism. Growth of strains in which wildtype *araC* is replaced by *sacB* or by *araC* variants that encode L-arabinose unresponsive AraC mutants, is still inhibited by L-arabinose. Other related monosaccharides show to various extent also inhibition. In M9 minimal medium, L-arabinose stimulates growth of both BW25113 and Δ*araC* strains in the early phase of growth, but ultimately reduces the final OD600 of only BW25113. Based on the different genotypes and phenotypes of the various tested strains, hypothetical regulatory mechanisms that may explain the effects of L-arabinose on growth of *E. coli* are discussed.

To critically question the work presented in this thesis, a general discussion on the developed bioreporter is provided below. The bioreporter is compared to other screening and selection methods and suggestions for further improvements are outlined.

General discussion

Since the first use of enzymes for making bread and bear thousands of years ago, a lot has changed. Knowledge on enzymes' functions, structures and mechanisms allowed enzymes to be applied in a wide range of industries during the last century. Advances in the fields of recombinant DNA technology, omics and computational tools further increased this number tremendously since the 1970s. Enzymes allowed us, humans, to do many things more efficiently and at larger scale, such as the treatment of textile or the use of detergents, but enzymes also allowed us to do things which were previously very difficult, like DNA engineering. However, we have not reached the top yet. To be able to make the switch from a fossil fuel based economy to a biobased economy, we need to expand the implementation of enzymes in industry extensively. Currently, only 5% of chemical products is produced biologically, but fortunately only 150-170 of the 3000 known enzyme types are being applied and only 1% of nature's repertoire of enzymes is known[1]. This leaves ample space to extend the number of industrial enzyme applications to be able to produce more chemicals biologically. To this end, the number of available enzymes should be increased. This can be done via metagenomics, directed evolution or computational design, all of which have led to a multitude of successes. Regardless of the approach, a high-throughput screening method is required to be able to screen the enormous amount of library clones. However, screening methods are often time-consuming, complicated and/or require expensive equipment.

An interesting approach to handle the large numbers in screening for novel enzymes is reporter-based *in vivo* screening or selection. This is a very powerful screening method

due to its many advantages: wide applicability, screening for enantioselectivity, signal enhancement, no need for artificial substrates and high-throughput screening. The most common reporter-based screening method combines the reporter GFP with screening by FACS. Although this method is high-throughput and has proven very successful in obtaining new and/or improved enzymes, it does require expensive equipment and experienced people to operate this equipment. The aim of this thesis is to simplify this technology by providing a generic and high-throughput *in vivo* reporter-based selection system or bioreporter. By using selection rather than screening the initial large library size can be rapidly reduced. To this purpose, a transcriptional regulator-based dual reporter system, has been developed and characterized, containing both a selection and a screening reporter under control of the transcriptional regulator AraC (**Chapter 3**). Moreover, a proof of principle is provided, by detecting L-arabinose isomerase activity *in vivo* both in assay format and at library scale; cells with this activity were enriched more than a millionfold (**Chapter 3**). In addition, the changeability of the specificity to make the system applicable for a wide range of enzymatic products and thus enzymes has been shown by replacing the transcriptional regulator AraC for LacI (**Chapter 4**) and possibly, although more experiments are needed to really conclude this, by engineering the specificity of AraC (**Chapter 5**). But how does the system's performance compare to other screening and selection methods and is there room for improvement? In this chapter, a discussion of these questions is provided.

Comparison of this thesis' bioreporter with other screening and selection methods

To compare the bioreporter developed in this thesis with other types of screening methods, some comparison criteria need to be established. Since the aim of this thesis is to provide a generic and high-throughput *in vivo* reporter-based selection system or bioreporter as a simpler and alternative method for the currently available enzyme screening methods, the components of this aim should be verified and thus give some comparison criteria. These are universality, throughput, complexity and labour intensiveness. In addition, it is important to look at the development of the method and at the number of true and false positives the method gives. These different criteria will function as the stepping stones of the comparison, going from *design* to *handling* and *output* of the screening and selection methods. The other methods included in this comparison do by far not cover all existing methods, but they are frequently used and give a good context for a critical discussion of our bioreporter. An overview of the comparison is provided in Table 1.

Design: universality

The first criterion in the design part is the *principle of the method* and thus the *universality*. If a method is universal or generic it means that it can be used in any enzyme screen, regardless of the enzyme type and origin. The dual reporter system developed here is based on the detection of the enzymatic product via a genetic reporter that gives a discriminating phenotype and not directly the product or the conversion itself that give a measurable property. In theory, this should make our system as well as other transcriptional regulator-based systems universal for all enzyme classes. Only very large products, like polymers, are very unlikely to be bound by a transcriptional regulator. The same is true for riboswitch-based systems, but due to more diverse chemical building blocks of transcriptional regulators, amino acids versus nucleic acids in riboswitches, the range of products to be detected is broader for systems based on transcriptional regulators than those based on riboswitches. Compared to these two screening methods, other types of reporter-based and also non-reporter-based screening are more restricted in their use. Posttranslational-modification-based screening is often applied for proteases, since they could directly act on the reporter protein to activate it. When no reporter is involved in the screening, the product or the conversion itself should have a measurable property. This narrows down the application range of the screening method to enzymes providing such a property. In the case of agar plate screening, microtiter plate screening and *in vitro* compartmentalization coupled to microfluidics (µIVC), multiple properties could be measured, such as colour or pH on agar plates and colour or fluorescence in microtiter plates and µIVC. This still allows multiple enzyme types to be screened for. In case of FACS or growth-based selection, the application range is even narrower, because only one type of output can be measured. The products should give fluorescence for the first or growth for the latter.

Another aspect of universality is the *origin of the enzyme*, either the library type or the host organism. All methods that are compared here can be used for metagenomic and enzyme variant libraries. Also no difference exists between methods regarding the original organism from which the enzyme is obtained. However, all *in vivo* screening methods, including ours, might run into problems with heterologous expression. Whether an enzyme is properly expressed in the screening host depends on things like the temperature, the codon usage and the presence of cofactors and chaperones. In *in vitro* screening, such as µIVC, this is less of a problem since one could more easily play around with the expression conditions.

Table 1: Comparison of common screening and selection methods with or without reporter[a]

Screening/selection method		Design		Handling		
		Universality	**Development**	**Throughput**	**Complexity**	**Labour intensiveness**
No reporter	Growth-based selection	-	+	-/+	+	+
	Agar plates	-/+	+	-	+	-
	Microtiter plates	-/+	+	-	-/+	-
	FACS[b]	-	-/+	-/+	-	-/+
	µIVC[c]	-/+	-	+	-	-/+
Reporter	**This thesis' bioreporter[d]**	**+**	**-**	**-/+**	**+**	**+**
	Growth-based selection	+	-	-/+	+	+
	FACS[b]	+	-	-/+	-	-/+
	µIVC[c]	+	-	+	-	-/+

[a]The methods are scored for the different criteria from + to -/+ to – with + indicating that the method is the best (e.g. a + at universality, development and complexity means the method is widely applicable, easy to adapt for another enzymatic reaction and simple to handle). [b]Fluorescent-activated cell sorting. [c]*In vitro* compartmentalization coupled to microfluidics.

The universality is also determined by the *ability of the substrate and the enzyme to come together*. In this respect, the developed bioreporter has the same limitation as other *in vivo* methods, namely the cell membrane. Preferably, the substrate should pass the membrane into the cell and the product should stay inside to ensure a genotype-phenotype linkage. For some methods the enzyme passing the membrane to the outside of the cell is an option, as for example in agar plate screening in which an extracellular enzyme forms a halo around the colony or in FACS coupled to cell display of the enzyme and product. Another option to measure the enzymatic conversion outside the cell is the use of a two-component system to transduce extracellular product formation to expression of a reporter gene, but it is important to have spatial separation of variants to avoid cross talk and thus prevent loss of the genotype-phenotype linkage. Compared to all discussed methods, only µIVC, lacking a cell membrane, is more universal regarding the ability of enzyme and substrate to come together, but, as for the other methods it is important that the product is retained.

Although no method is entirely universal, the developed bioreporter seems to do better in the universality area than most other methods due to the reaction independent measuring via a reporter. However, one major issue needs to be resolved before really stating this. This issue is *the changeability of the specificity of the bioreporter*, which is in

more detail discussed in **Chapters 4** and **5**. To change this specificity, one could replace the regulator with a known or engineered regulator with a different specificity of which the latter could also be a synthetic regulator consisting of parts from several proteins[295]. At the start of this project, computational design of transcriptional regulators seemed within reach over the course of the project, but unfortunately, this approach is not yet straightforward, despite great advances that have been made in this area[182,183,196] This limitation is also reflected in other projects; more and more groups turn to an alternative approach in which the enzymatic product is converted by other enzymes to a product for which a regulator is available[217,296]. Thus, the envisioned modularity of the bioreporter in which the specificity could simply be changed, is not yet achieved. Nevertheless, once available, replacing the regulator is rather simple, as for other transcriptional regulator-based systems, due to the modularity of the system, though some optimization of conditions or expression levels is required (**Chapter 4**). In short, if there would be a transcriptional regulator available or could be easily obtained for any enzymatic product, this bioreporter would be widely applicable. Unfortunately, screening for novel regulators or engineering them still takes a lot of time or effort. Once advances in bioinformatics and computational design allow regulator screening and engineering respectively, to be more quick/straightforward, the bioreporter can be widely applied.

Design: development

Next to the universality of the screening method, the *development* is an important criterion in the design part. Here, development means the adaptation of an existing method for another enzyme and not the development from scratch as was done in this thesis. It entails the whole procedure from figuring out the details of the set-up, the construction of new components where needed, and the determination of the proper conditions. For our bioreporter, this development is time consuming and laborious, mainly due to obtaining a transcriptional regulator with the right specificity as already discussed above. Subsequent implementation of this new regulator is straightforward due to the modularity of the system, but it still requires a construction step and some condition optimization. The latter is necessary due to the extra components that are present because it is a reporter-based system (the reporter and the sensor). For a good interplay between these components, the right conditions need to be determined. In this thesis, that was done by making different system variants with varying selection reporter and plasmid copy number and by varying the kanamycin and product concentrations (**Chapters 3** and **4**). Most other transcriptional

regulator-based methods encounter the same problems in adapting the method to another enzyme, because this is inherent to their specificity. However, some synthetic regulators are easier to adapt. A good example here is Chemical Complementation, a yeast-three hybrid system in which the formation or breakage of a bond by the enzyme brings together the activation and the DNA binding domains of the transcriptional regulator, thereby allowing or disrupting the transcription of the reporter gene respectively. The only element that needs to be changed is the chemical linkage between the substrate and the two protein domains[89,90].

Other reporter-based systems have the same issue of a long and laborious development. Whether it functions on transcriptional, translational or posttranslational level, for all these systems a specific sensor needs to be found or engineered and subsequently implemented. The advantage of riboswitches is that initial screening for novel aptamers, the RNA parts that confer specificity, can be done with a very high-throughput *in vitro* method such as Systematic evolution of ligands by exponential enrichment (SELEX). Unfortunately, this does not always translate one to one with *in vivo*. A promising new way to obtain riboswitches is via computational design, but this field is still in its infancy[297,298]. For non-reporter based methods like agar plate and microtiter plate screening, the development is much less time consuming and laborious. It mostly involves finding a natural or artificial substrate that has a measurable property and optimization of the assay conditions. For FACS and µIVC on the other hand, the development involves more extensive condition optimization, for example to establish proper gating for sorting and to enhance stability of droplets[299]. This also requires experienced people to do so, for µIVC even more than for FACS.

Handling: *throughput*

The design of the method is followed by its execution or handling. Here, an important criterion is the *throughput* of the method. This is the number of variants that can be screened in a reasonable time frame and determines how large the library to be screened can be or is covered. Our bioreporter is high-throughput because it uses growth-based selection. This means that the initially large library size, only limited by the transformation efficiency and thus ~10^9, can be rapidly reduced by simply growing the cells. The subsequent screening step in microtiter plates has much lower throughput, but since the library pool is already reduced to hundreds, this is not a problem. For other reporter-based methods, the throughput depends on the reporter choice. With GFP as reporter and thus fluorescence as

signal, FACS can be used. FACS is like growth-based selection a high-throughput method that is only limited by the transformation efficiency. LacZ on the other hand gives a colour, which can be detected on agar plates or in microtiter plates, both having a much lower throughput (library sizes of 10^5 and 10^4 respectively), even when coupled to robotics. µIVC is as an *in vitro* method not limited by the transformation efficiency and has therefore an even higher throughput (library size >10^9).

Handling: complexity

Another important criterion in handling is the *complexity of the experiments*. Selection and screening with our bioreporter is rather simple, once developed. It is just growing the cells using appropriate selection conditions and screening the ones that survive in microtiter plates. Also no specific or expensive equipment is necessary, except for a medium expensive plate reader, and no very experienced or trained personnel is required. For other reporter-based methods, the complexity depends on the reporter choice. If it is coupled to FACS or microfluidics, expensive equipment is necessary and very experienced people are required to adequately operate this equipment and thus control the conditions. Microtiter and agar plate screening are rather simple in handling, but microtiter plate screening requires medium expensive equipment.

Handling: labour intensiveness

The last criterion in handling is the *labour intensiveness* of the experiments. Selection and screening with our bioreporter is not laborious, certainly not in respect to the throughput. Growing the cells does not cost much effort and screening is done only at a small scale. For other reporter-based systems, it depends on the reporter choice how much work has to be done. Agar plate and microtiter plate screening are quite laborious, especially considering the low throughput. They require a lot of handling time for pipetting and in case of the agar plates, for making the plates. FACS and microfluidics are not so laborious as long as the conditions are properly set-up. Non-reporter based methods differ in throughput, complexity and labour intensiveness as already described for the reporter-based methods.

Output: number of true positives

Once the experiments are executed, the performance or output of the screen should be determined. This concerns the total number of *true positive hits*, but also the *fraction of false positives*. For a high number of true positive hits, two things are very important. The first is the *design, quality and coverage of the library*. Even if a screening method is excellent,

if no enzymes with the target activity or improved target activity are present, such a method is worthless. The higher the coverage is, the higher the likelihood of finding a certain unique hit or even several copies of this unique hit. For a high coverage, also a high throughput is required in case of large libraries. The second thing is *the characteristics of the method*. A high sensitivity and a low leakiness are important to obtain the true positives from the library. To distinguish enzymes with high activity from the ones with low activity, either during the screening or in the subsequent quantification and mutual ranking of the hits, a large dynamic range, a high maximal signal and a high fold change of maximal signal over leakiness are important. In this thesis, the system was tested with a mock library consisting of cells with L-arabinose isomerase from *E. coli* and *G. thermodenitrificans* and cells without L-arabinose isomerase in a ratio of $1:1:10^8$ (**Chapter 3**). With a coverage of 100, enough positive cells were present. That the characteristics were sufficient to detect an enzyme at library scale was shown by the successful enrichment of the cells with L-arabinose isomerase activity from both species in this proof of principle. An enrichment of more than 10^6 was achieved in a single round of selection and screening, whereas other transcriptional regulator-based systems needed either one selection round or at least two FACS rounds to get an enrichment of only more than 10^5. These screens or selections were started with (mock) library sizes mostly up to 10^6, one 10^7, instead of 10^8 as done here[84,111,154,219,223,300,301]. Although not for all published transcriptional regulator-based bioreporters, enrichment numbers are provided and some good performing bioreporters might be missing, our bioreporter seems to do better than most others. Since the characteristics are similar to those of many other transcriptional regulator-based systems that were tested with different enzyme types (**Chapter 3**), it is likely that this bioreporter's working range is also suitable for various enzyme types. To fully test the bioreporter, enzymes should be selected from a real metagenomic or enzyme variant library and when problems are encountered, characteristics can be further improved as discussed below in the section 'Other improvements'. In the previous paragraph, some attempt is made to compare our bioreporter with other transcriptional regulator-based bioreporters, but a proper comparison between different screening methods can only be made by using the same library and coverage for all methods, or even better, multiple libraries. With the large influence the library has on the output, this is the ideal way, but it is also very laborious. Even a less ideal approach, comparing the number of true positive hits over many publications of one method is rather difficult. Most publications either do not include these numbers at all or present them differently. In enzyme variant screens for example, some

state that hits are the ones that have a more than 10% higher activity than the parent enzyme, others take the top 10 hits and mostly the numbers of the rest of the hits are not shown. Also the library, providing the context of these numbers, is described differently. Either the total number of clones that were screened are mentioned or the library size, often without coverage. Thus trying to compare all these methods based on these numbers would not give a correct account of their mutual performance. A comparison based on the methods' characteristics is slightly easier to do, because more information is included in the publications, either the numbers themselves are mentioned or they could be deducted from the response curves depending on the axes' scales. However, attention needs to be paid to the used definitions of these characteristics. Alternative definitions for the same term might be used in different publications. For example, dynamic range is either defined as the range of substrate concentration giving a changeable signal (used in this thesis) or as the signal range that can be measured upon addition of substrate. Sensitivity also has multiple definitions: the lowest substrate concentration that can be measured (used in this thesis) or the slope of the response curve. When making a comparison of various reporter-based methods using these characteristics, one should keep in mind that these characteristics are largely influenced by the sensor kinetics, the expression levels of the various bioreporter components and the reporter choice. For instance, the characteristics of bioreporters using transcriptional regulators or riboswitches coupled to FACS both cover a wide and overlapping range of values depending on the above mentioned criteria and not simply on whether the sensor consists of protein or RNA. Also for non-reporter based methods, characteristics are often determined by the specifics of the methods such as which spectrophotometer with which settings is used for microtiter plate screening. Nevertheless, about the fold change of maximal signal over leakiness some things can be mentioned. Selection has a lower fold change than screening[46]. Of the screening methods, agar plate screening has the lowest fold change[37] and the fold change is higher for µIVC than for microtiter plate screening, which makes µIVC very suitable for improving the activity of enzymes that already have a high activity[302].

Output: number of false positives

The other criterion that determines the output of the system is the *number of false positives*. Growth-based selection has a high risk for false positives. To reduce this risk our bioreporter involves a subsequent bioluminescence-based screening step. That this indeed reduces the false positive risk is shown in the proof of principle with L-arabinose isomerase (**Chapter 3**).

After the selection step, 90% of the surviving cells were false positives, but after the subsequent screening step this was reduced to 0% as verified by PCR. The combination of selection and screening is thus very valuable in dealing with this high false positive risk in growth-based selection. For engineering of the ligand specificity of AraC, these numbers differ per library and per step in the selection and screening process (**Chapter 5**). They also differ from the numbers in the proof of principle with L-arabinose isomerase. This makes sense since another type of false positives is introduced, namely the regulator variants that constitutively activate reporter gene expression. In addition, the whole set-up is not yet optimized. The nature of the false positives was not determined, except for one large group (60% of the false positives in the proof of principle with L-arabinose), which turned out to have an exchange of the RBS region of the *kan* and *araC* genes, placing *kan* under a constitutive promoter instead of the AraC-controlled promoter, allowing it to survive in the presence of kanamycin even in the absence of active enzyme. By keeping the kanamycin concentration sufficiently high, leakiness of the reporter promoter cannot result in false positives. However, library selection took place at a boundary concentration because the signal of the AraC variants was low. This allowed some false positives due to leakiness.

Since false positives are a large risk for growth-based selection, various solutions are being developed, ranging from toggled negative and positive selection[262] to using genes involved in catabolism[303] or in anabolism[304] as reporter. Although these tricks certainly help, some false positives will always be present as it is inherent to the method. With a selective pressure evolution will always lead to some escape mutants. To compare the number of false positives obtained with our bioreporter with those obtained by other screening or selection methods is difficult, because also in this respect publications often lack numbers. Nevertheless, what is clear is that the nature of the false positives differs per method. For example, false positives in µIVC occur due to co-encapsulation of multiple droplets into one or due to cross talk between droplets, e.g. by diffusion of product[299]. Cross talk between cells is possible for *in vivo* methods, when the enzyme and/or product are not contained within the cell or at the cell surface. In growth-based selection, cross talk can also occur in a different form; especially under low stringency, neighbouring cells can create locally lower selective pressure, e.g. by removing antibiotics, and thereby helping other cells survive. Another general cause of false positives is the difference in expression levels, leading to a difference in signal which is not necessarily related to the amount of enzyme activity. Also the difference between growing cells on solid media and in liquid media should be kept in mind. Conditions might need to be adapted depending on the way of growth. Whatever

method is used, it is good to optimize the conditions and the set-up to reduce the number of false positives, especially in the first round. This will reduce the amount of work needed to later discriminate between true and false positives.

Although for a more accurate comparison of performance it would be great to have a standard way to describe performance criteria of the different screening and selection methods[305,306], our bioreporter can compete with other screening and selection methods (Table 1). But is the aim reached to provide a generic and high-throughput *in vivo* reporter-based selection system or bioreporter as a simpler and alternative method for the currently available enzyme screening methods? Although the throughput is not as high as for µIVC, it is a high-throughput method and it is much simpler and less laborious than many other methods just like other growth-based selection methods. The advantage this bioreporter has over the latter is the combination of a selection and screening reporter, making further quantification, ranking and reduction of false positives possible, though other ways to reduce the false positives are also imaginable as discussed below. The other aspect of the envisioned system, the genericity, is not yet achieved.

Suggested improvements for this thesis' bioreporter

The developed bioreporter is able to detect an enzyme at library scale as shown as proof of principle for L-arabinose isomerase (**Chapter 3**). However, the use of this bioreporter for selecting AraC variants with another specificity, showed that the system could be further improved (**Chapter 5**). This would simultaneously make the system better for obtaining novel enzymes. In this section, multiple improvements are suggested for the reporters, the construction method and various other points. An overview of these suggested improvements is presented in Fig. 1.

Selection and screening reporters

As selection reporter, two different options were used in this thesis, either KmR for kanamycin resistance or LeuB for L-leucine auxotrophy complementation. For both reporters, the cell growth was dependent on the presence of the inducer with higher concentrations giving more growth (**Chapters 3** and **4**), which are essential characteristics for a growth-based selection reporter. Depending on which transcription regulator controlled their expression, one or the other reporter performed best; KmR for AraC and LeuB for LacI (**Chapter 4**). This is likely explained by different regulation dynamics and thus different

reporter expression levels. Only KmR was tested in a proof of principle and this reporter allowed the enrichment of cells with L-arabinose isomerase activity over a millionfold (in combination with the screening reporter), making it well suited for enzyme detection. However, for selecting transcriptional-regulator variants this reporter has not yet proven its value. First, the kanamycin concentrations need to be optimized and a *counter selection* reporter should be added to remove false positives, e.g. constitutive regulator variants, already in an early stage. To prevent further increasing the plasmid size by adding an additional component, it would be better to have one selection reporter that enables both *positive and negative selection*. In that way, variants that are constitutively active or only active in the absence of the ligand can be reduced during negative selection. The most common reporter with this ability is PyrF, an orotidine 5'-phosphate decarboxylase. PyrF allows positive selection based on uracil auxotrophy complementation in a Δ*pyrF* background and negative selection based on sensitivity to 5-fluoroorotic acid (5-FOA)[156]. Other options include ThyA (thymidylate synthase; positive and negative selection based on thymine auxotrophy complementation in a Δ*thyA* background and sensitivity to the antibiotic trimethoprim respectively[307]) and TolC (outer membrane protein; positive and negative selection based on sensitivity to small toxic molecules like sodium dodecyl sulfate (SDS) in the absence of TolC and sensitivity to the bacteriocin colicin E1 in the presence of TolC respectively[308]). In any case, it should be tested which of them has the best working range in combination with the current system. A very interesting alternative selection method is described by Liu *et al.*[303]. They used MalQ as reporter, an enzyme required for maltose utilization in a Δ*malQ* background. This allows selection based on the ability to utilize maltose as sole carbon source and was shown to be very robust compared to antibiotic resistance; escape mutants only arose after many transfers. However, it would not be able to deal with constitutive regulators, because these would always turn on transcription of MalQ, enabling growth on maltose irrespective of the presence or the absence of inducer.

As screening reporter, LuxCDABE was chosen, because it is more sensitive than the other common reporters GFP and LacZ[309,310], it does not require addition of a substrate in contrast to for example Luc and LacZ, it has less background noise in comparison to GFP[311] and for this bioreporter coupling to FACS was not necessary because the growth-based selection provides high throughput. No experimental comparison was made in this study between LuxCDABE and other screening reporters in regards to sensitivity, but the sensitivity is indeed good (< 1 µM for the medium copy system with LacI; **Chapter 4**). However, working

with LuxCDABE is not robust due to the dependence on the cell's metabolism; the intracellular redox pool and the oxygen concentration strongly affect the luciferase activity[312]. Slight changes in the metabolism can already alter the signal (**Chapter 3**). This makes the detection of small differences between samples such as different regulator variants difficult due to relative large standard deviations (**Chapter 5**). Also the comparison between experiments in absolute values is impossible. In this respect, it would be good to test the system with *GFP*, which is only oxygen dependent[311], as screening reporter to see what sensitivity would be conceivable and to possibly enable detection of small differences.

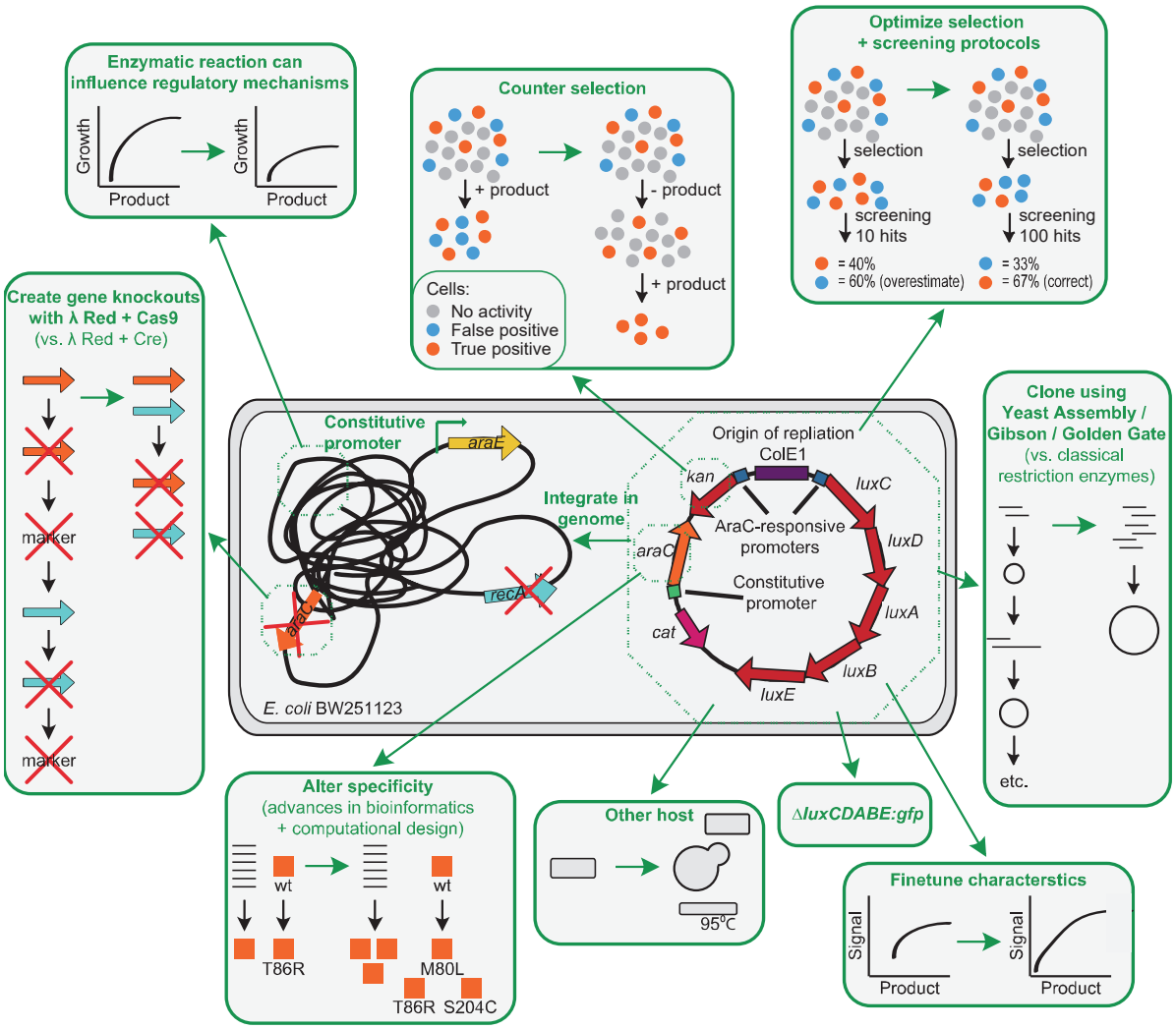


Fig. 1: Overview of suggested improvements for the bioreporter developed in this thesis. The bioreporter is presented in the centre and the improvements are indicated in green. The legend in the box 'Counter selection' is also the legend for the box 'Optimize selection + screening protocols'.

Construction methods

To make our bioreporter, we constructed a set of regulator-reporter plasmids as well as a

series of knockout strains. The *cloning* was done with classical restriction enzymes and *making the knockout strains* by replacing the gene of interest by an antibiotic marker using λ Red recombinase and removal of this marker by the recombinases FLP or Cre. In retrospect, where these the best methods? In this section various alternatives are discussed.

The low and medium copy versions of the regulator-reporter plasmids were cloned in eight subsequent steps using restriction enzymes. Seven of these steps were done in parallel for each of the selection reporters. In the final plasmids, the restriction sites between the different components make it modular and allow the easy replacement of such a component. Although the plasmids function as required, the construction process could have been done more efficiently by using *another cloning method*. While cloning with restriction enzymes followed by ligation allows the combination of three DNA fragments in a single step (**Chapter 3**), three other methods allow much more DNA fragments. In Gibson Assembly, up to ca. eight DNA fragments with an overlap of at least 40 bp are assembled *in vitro* into products up to several hundred kilobases. An exonuclease removes nucleotides from the 5' ends, allowing annealing of the overlapping regions. Gaps are filled in by a polymerase and fragments are ligated by a ligase. Except for the simultaneous and seamless assembly of many fragments in a short time, no restriction enzymes are required, which are often limited in choice due to the sequences of the fragments[313,314]. In Golden Gate cloning, up to ca. nine DNA fragments are seamlessly cloned in a short time, making use of a four nucleotide sequence identical in both adjacent DNA fragments and a type IIS restriction enzyme site right next to each of them. This enzyme cuts outside its recognition site, thereby creating overhangs of the identical sites and removing the recognition sites. The DNA fragments are ligated *in vitro*. In Yeast Assembly, yeast is transformed with up to ca. ten DNA fragments that have 20-60 bp overlap and of which one contains the origin of replication and one the selection marker for yeast. If transfer to *E. coli* is required, these elements for plasmid replication and maintenance in *E. coli* are also needed. Yeast ligates the fragments[315-317]. Using one of these three methods would have reduced the cloning effort to only about two steps instead of eight. The future replacement of the transcriptional regulator for changing the bioreporter's specificity requires a few cloning steps or many PCRs to make one large insert, when done by restriction enzymes, since the CDS of the regulator gene as well as the two promoter-operator regions in front of the reporter genes have to be replaced. However, using either one of the three above described cloning methods, this can be done more efficiently.

The knockout strains were initially created by replacing the gene of interest by an

antibiotic marker using λ Red recombinase and removal of this marker by FLP recombinase[137]. The scar left after this procedure is still being recognized by FLP recombinase, resulting in undesired recombination and thereby the undesired removal of genome parts when subsequent genes are removed in the same strain. Since FRT sites that leave an unrecognizable scar were not available, the switch was made to Cre recombinase and *lox71*/*lox66* sites, which do leave an unrecognizable scar after recombination[140]. The here developed protocol integrates aspects of three other protocols: recombination by λ Red recombinase to replace the gene of interest by the antibiotic marker[137] followed by removal of the antibiotic marker by Cre recombinase[166] via recombination of the *lox71*/*lox66* sites[140]. The protocol was made more efficient by placing the λ Red recombinase and the Cre recombinase on one plasmid, removing the intermediate plasmid clearing and transformation steps (S.C.A Creutzburg, personal communication). Unfortunately, this more efficient protocol was not successful for the gene knockouts created in this study, but Song and Lee 2013[318] showed the efficiency of such a set-up both for making in the same strain successive gene knockouts or two gene knockouts simultaneously. Although in this thesis a good protocol was developed for making gene knockouts, the removal of multiple genes in one strain had to be done in sequence and included the removal of the marker for each individual gene. Currently, an *alternative genome engineering method* is widely applied that tackles these issues. This new method still uses λ Red for recombination, but includes a counter selection based on the CRISPR-Cas system (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated system), a bacterial adaptive immune system against foreign genetic elements. For genome engineering, the type II CRISPR-Cas system from *Streptococcus pyogenes* is most frequently used. In this system, a RNA-guided nuclease Cas9 cleaves the invading DNA. The RNA, of which a part is complementary to the invader's DNA, is transcribed from a memory or CRISPR array on the bacterial genome, where it was incorporated during a previous encounter with the same invader. To create a gene knockout in *E. coli* the target strain is first transformed with a plasmid that encodes λ Red and Cas9. In the second step, expression of λ Red is induced and this strain is provided with the homologous flanks of the target gene and the synthetic single guide sgRNA complementary to the target gene. Expression of the sgRNA is induced to allow sgRNA guided cleavage of non-recombined target gene by Cas9. This method does not require marker removal and includes a counter selection. It therefore allows efficient sequential gene deletions. Also multiplexing, the simultaneous deletion of multiple genes, is possible by providing for each gene the complementary sgRNA and flanks[319].

Other improvements

Next to the above suggested improvements, some other improvements might optimize the bioreporter's performance. Firstly, the reporters and the regulator can be *integrated into the genome*. This has the advantages of a higher stability, no need for an antibiotic marker for plasmid maintenance and a smaller library plasmid. Having the system on a plasmid, there is the risk of recombination, as was indeed observed in **Chapter 3**. Genome integration strongly reduces this risk, because on the genome less recombination takes place and only one copy is present, reducing the risk further (still some sequence similarity in the promoter regions of the three components; regulator and two reporters). In addition, the lack of a selection marker reduces the strain on the cells[320] and the library plasmid containing the enzyme variants can be smaller, allowing a higher transformation efficiency and thus a larger library. Despite these advantages, integration was not done yet, because in the characterization of the AraC-based system the medium copy system functioned better than the low copy system (e.g. better sensitivity, **Chapter 3**). This means that integration, resulting in a single copy of the system genes, will require further optimization of expression, like stronger promoters, to ensure a high enough signal. A relative low signal in a single copy system was also observed by others[321], but the signal is very much dependent on the integration location on the genome[322,323]. Integration will also require an extra step when changing the system's specificity, since the regulator and the promoter regions upfront the reporters need to be altered on the genome instead of the plasmid. For engineering the specificity of a regulator using the selection and screening capability of the system itself, a strain with only the reporters integrated is necessary, allowing a plasmid-based library of regulator variants. Although the advantages of integration make it worth to try this approach, it is recommended to only integrate the system components as a final step when other improvements such as the implementation of a counter selection are done. This would allow optimization of the system in a more efficient way, on plasmids first, to prevent unnecessary work.

Whether genome-integrated or plasmid-encoded, some *fine tuning of the bioreporter's characteristics* (leakiness, dynamic range, sensitivity, maximal signal, fold change of maximal signal over leakiness) might help to improve its performance. Although the system functioned very well for the enrichment of cells with L-arabinose isomerase activity, this is not necessarily the case for all enzymes. Also when the specificity is changed, some optimization of the characteristics is likely required as shown when AraC was replaced by LacI (**Chapter 4**). Next to varying the selection reporter and the plasmid copy number, a

possible way to do this, is by playing with the expression levels of the individual system components, for example by altering the promoter or the RBS strengths[324]. The ratio of transcriptional regulator over operator influences the response curve and thus the characteristics. For instance, a relative high number of repressors results in a lower sensitivity, since more inducer is required to occupy all binding sites and thus to de-repress expression[117,325]. For activators, this is the other way around[325]. Further things that can be changed to alter the bioreporter's characteristics are the number, the location and the sequence of the operators[326], the promoter in front of the reporter[218], the medium contents[326], the biosynthetic and degradation capacity of the cell for the target molecule[326], the transport capacity of the cell for the target molecule[326,327], the kinetics of the transcriptional regulator[327] and reducing the burden to the cell[328]. With all these options, it is important to keep in mind that focusing on changing one characteristic, e.g. dynamic range, could simultaneously alter another characteristic, e.g. the amount of inducer necessary for half of maximal induction[329]. It might also be interesting to make several system variants with different characteristics, e.g. with different sensitivities, allowing the selection/screening of enzyme variants or regulator variants with different activity or affinity. Although this fine tuning can already improve the current bioreporter, it is recommended to only do this when the system does not perform well enough for other enzymes or with another regulator.

Another improvement is the *optimization of the selection and screening protocols*. This protocol was successful in enriching cells with L-arabinose isomerase activity, but should of course be tested for various other enzymes. Unfortunately, the protocol was not yet optimal for obtaining regulator variants with a changed specificity. The first step here is to implement a counter selection to filter out the variants that are constitutively activating reporter gene expression irrespective of the presence of the target molecule. Next, a more stringent selection by increasing the kanamycin concentration and reducing the target molecule concentration (in this work D-xylose) would help in selecting the best variants. In addition, a subsequent screening step incorporating more colonies would help to better distinguish true from false positives. For a more detailed discussion on this topic, please see the discussion in **Chapter 5**.

Although *E. coli* is a model organism that is easy to work with, not all enzymes will be heterologously expressed very well in this organism. For that reason, it would be interesting to *transfer the bioreporter's principle and components to a set of other microorganisms*, e.g. a Gram-positive, a thermophile, and an eukaryote. This would further enhance the

applicability for a wide range of enzymes, once modifying the bioreporter's specificity is more straightforward in the future. However, transfer to other organisms is likely an extensive project, since different organisms require different promoters, RBSs, origins of replications, etcetera. Indeed, Jha *et al.*[330] showed that transfer of a reporter system from *E. coli* to *Pseudomonas putida* required a plasmid suitable for replication in this new host as well as engineering of the promoter, the operator and the regulator. In addition, the choice of reporters might be influenced by the target organism. For example, in thermophiles these reporters need to be heat resistant. That bacterial regulators can be transferred to eukaryotes and be used as sensor of a bioreporter, either for screening or for selection, is shown recently for yeast, allowing the use of a similar set-up in this eukaryote except for some adaptations in the promoters and operators[331-333]. Due to the time-consuming adaptations, transfer to other organisms is recommended only when running into expression problems in *E. coli*.

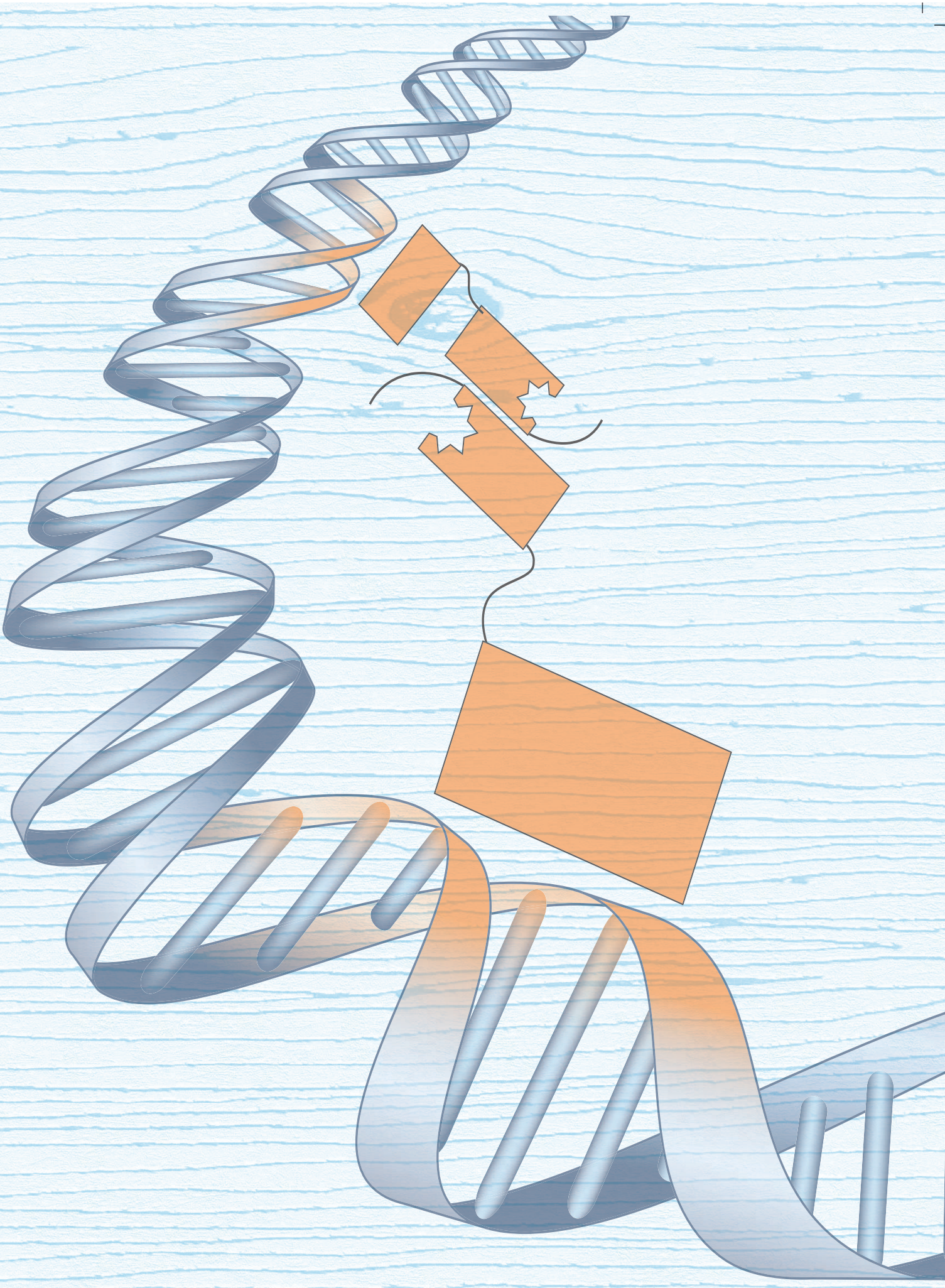
The next point is more a thing to keep in mind than an actual improvement. Depending on the target molecule and the transcriptional regulator of choice, different *regulatory mechanisms taking place inside the cell are affected*. These may influence the growth of the cell and thus interfere or at least alter the outcome of selection or screening. The AraC-based system is a good example of this. Due to the way that expression of the L-arabinose transporter AraE is under control of AraC, an all or nothing induction takes place in which cells are either fully induced or not induced. By placing *araE* under a constitutive promoter, the induction level becomes gradual with the level depending on the L-arabinose concentration[142]. This was not done so far, because with the current bioreporter a proof of principle with L-arabinose isomerase was successful. However, it might help for other L-arabinose producing enzymes. Another regulatory effect that one has to consider when working with AraC and L-arabinose, is the stimulation and inhibition of *E. coli* cell growth by L-arabinose. These effects were observed during the characterization of the AraC-based system and during the modification of the AraC specificity (**Chapters 3** and **5**). The mechanism of these effects is not yet understood (**Chapter 6**), but they do influence the selection and screening. During selection, growth is the output and although kanamycin resistance is the first requirement for growth, the level of growth is also influenced by the presence of L-arabinose, possibly altering the characteristics of the system, e.g. sensitivity. During screening, cells that grow in the presence of L-arabinose grow differently than cells that grow in the absence of L-arabinose and thereby L-arabinose influences the bioluminescence. Although this is not a problem during exponential phase and thus not

during the bioluminescence-based assays in microtiter plates, this prevented comparison of spotted cells on agar plates after overnight growth. These growth effects by L-arabinose were an interesting observation during this work, the kind that a researcher is happily stumbling upon and is worthwhile to further study. However, together with the AraE-influenced induction they are a good example of regulatory effects that one has to deal with when choosing a certain target molecule and corresponding regulator. When working with such a system, it is advisable to keep this in mind.

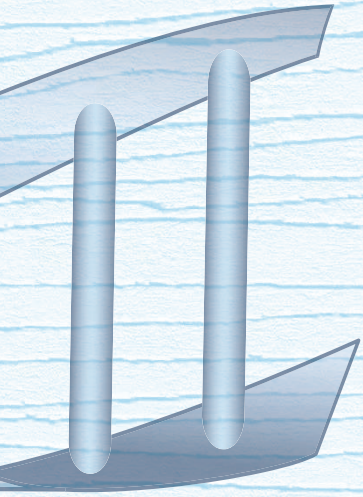
Concluding remarks and perspectives

The aim of this thesis was to provide a simpler alternative to the current screening and selection methods for enzymes by developing a generic and high-throughput *in vivo* reporter-based selection system or bioreporter. For the most part, this was indeed achieved. A high-throughput *in vivo* reporter-based selection or bioreporter, which is simpler than many screening methods, was provided. In a proof of principle, the ability to detect an enzyme activity (mesophilic or thermophilic L-arabinose isomerase) both in assay format and at library scale was shown. Here, the dual reporter system was very important for a rapid reduction of the initially large library size by growth-based selection and the subsequent exclusion of false positives by bioluminescence-based screening. That the specificity of the bioreporter can be changed, was demonstrated by the successful replacement of the transcriptional regulator AraC with another one that was readily available (LacI). Unfortunately, altering the system's specificity by engineering the specificity of AraC was not yet equally successful. To make the bioreporter better equipped to select transcriptional-regulator variants, the implementation of a counter selection and some condition optimization are needed. Even with these improvements, adaptation of the specificity will, in the near future, stay the major limitation of transcriptional regulator-based bioreporters becoming more widely applicable. For these kind of bioreporters to become generic, further advances in the fields of computational design and/or bioinformatics are required to make more transcriptional regulators available, thereby reducing the development time and labour in changing the system's specificity for each enzymatic product. Meanwhile, for each enzyme one should consider what the best approach will be when no transcriptional regulator specific for the target product is available. In some cases engineering might be faster, whereas in other cases metagenomics or a conversion by enzymes to a product for which a transcriptional regulator is available might be faster[221].

Despite these issues regarding the modification of the sensor's specificity, transcriptional regulator-based bioreporters will continue to play an important role in finding new and/or improved enzymes. Their many advantages (reaction independent, reporter choice, screening for enantioselectivity, signal enhancement, no need for artificial substrates and high-throughput screening), make the investment to develop such a bioreporter worthwhile. Also the field of metabolic engineering frequently applies these bioreporters, in that case for strain improvement, and will thus be a large contributor to the number of bioreporters and sensors available for enzyme screening[221]. Growth-based selection and microfluidics are likely to occupy the enzyme screening landscape in coming years; the first simple to handle and the other with even higher throughput and better suited to further improve enzymes' activities. Overall, it will be exciting to look back in ten or twenty years from now to see the full extent of the contribution that bioreporters can make to our society and economy by enhancing the number of available enzymes.



Appendices



References

1. Singhania, R. R., Patel, A. K., Thomas, L., Goswami, M., Giri, B. S., and Pandey, A. (2015) Industrial enzymes. In *Industrial biorefineries & white biotechnology*. Pandey, A., Höfer, R., Taherzadeh, M., Nampoothiri, K. M. and Larroche, C., (eds). pp. 473-497, Elsevier, Amsterdam.
2. (2010) Medical and health sciences - Volume II. EOLSS Publications, United Kingdom.
3. Fox, P. F. (1993) Cheese: An overview. In *Cheese: chemistry, physics and microbiology*. Fox, P. F., (ed.), Springer, Boston.
4. Newell-McGloughlin, M., and Re, E. (2006) The evolution of biotechnology. Springer, Dordrecht.
5. Elleuche, S., Schröder, C., Sahm, K., and Antranikian, G. (2014) Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. *Curr Opin Biotech*. **29**, 116-123.
6. Norus, J. (2006) Building sustainable competitive advantage from knowledge in the region: the industrial enzymes industry. *Eur Plan Stud*. **14**, 681-696.
7. Koshland, D. E. (1995) The key–lock theory and the induced fit theory. *Angew Chem Int Ed Engl*. **33**, 2375-2378.
8. Michaelis, L., and Menten, M. L. (1913) Die Kinetik der Invertinwirkung. *Biochem Z*. **49**, 333-369.
9. Röhm, O. (1913) Verfahren zum Reinigen von haeuslichen und gewerblichen Gegenstaenden und Stoffen aller Art. **DE 283 923**
10. Avery, O. T., Macleod, C. M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. *J Exp Med*. **79**, 137-158.
11. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. **171**, 737-738.
12. Nelson, D. L., and Cox, M. M. (2005) Lehninger principles of biochemistry (4th ed.). W. H. Freeman and Company, New York.
13. Keller, M. A., Piedrafita, G., and Ralser, M. (2015) The widespread role of non-enzymatic reactions in cellular metabolism. *Curr Opin Biotechnol*. **34**, 153-161.
14. Maurer, K. H. (2004) Detergent proteases. *Curr Opin Biotech*. **15**, 330-334.
15. Jemli, S., Ayadi-Zouari, D., Hlima, H. B., and Bejar, S. (2016) Biocatalysts: application and engineering for industrial purposes. *Crit Rev Biotechnol*. **36**, 246-258.
16. Yang, G., and Ding, Y. (2014) Recent advances in biocatalyst discovery, development and applications. *Bioorg Med Chem*. **22**, 5604-5612.
17. Porter, J. L., Rusli, R. A., and Ollis, D. L. (2016) Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem*. **17**, 197-203.
18. Ni, Y., Holtmann, D., and Hollmann, F. (2014) How green is biocatalysis? To calculate is to know. *ChemCatChem*. **6**, 930-943.
19. de Regil, R., and Sandoval, G. (2013) Biocatalysis for biobased chemicals. *Biomolecules*. **3**, 812-847.
20. Uchiyama, T., and Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotech*. **20**, 616-622.
21. Kaul, P., and Asano, Y. (2012) Strategies for discovery and improvement of enzyme function: state of the art and opportunities. *Microb Biotechnol*. **5**, 18-33.
22. Madhavan, A., Sindhu, R., Binod, P., Sukumaran, R. K., and Pandey, A. (2017) Strategies for design of improved biocatalysts for industrial applications. *Bioresour Technol*. **245**, 1304-1313.
23. Littlechild, J. A. (2017) Improving the 'tool box' for robust industrial enzymes. *J Ind Microbiol Biot*. **44**, 711-720.
24. Krüger, A., Schäfers, C., Schröder, C., and Antranikian, G. (2018) Towards a sustainable biobased industry - Highlighting the impact of extremophiles. *N Biotechnol*. **40**, 144-153.
25. Reetz, M. T. (2016) What are the limitations of enzymes in synthetic organic chemistry? *Chem Rec*. **16**, 2449-2459.
26. Cheng, F., Zhu, L., and Schwaneberg, U. (2015) Directed evolution 2.0: improving and deciphering enzyme properties. *Chem Commun*. **51**, 9760-9772.
27. Schmidt-Dannert, C., and Arnold, F. H. (1999) Directed evolution of industrial enzymes. *Trends Biotechnol*. **17**, 135-136.
28. Barrozo, A., Borstnar, R., Marloie, G., and Kamerlin, S. C. L. (2012) Computational protein engineering: bridging the gap between rational design and laboratory evolution. *Int J Mol Sci*. **13**, 12428-12460.

29. Xiao, H., Bao, Z., and Zhao, H. (2015) High throughput screening and selection methods for directed enzyme evolution. *Ind Eng Chem Res*. **54**, 4011-4020.
30. Jacques, P., Béchet, M., Bigan, M., Caly, D., Chataigné, G., Coutte, F., Flahaut, C., Heuson, E., Leclère, V., Lecouturier, D., Phalip, V., Ravallec, R., Dhulster, P., and Froidevaux, R. (2017) High-throughput strategies for the discovery and engineering of enzymes for biocatalysis. *Bioprocess Biosyst Eng*. **40**, 161-180.
31. Desobgo, S. C. Z., Mishra, S. S., Behera, S. K., and Panda, S. K. (2017) Scaling-up and modelling applications of solid-state fermentation and demonstration in microbial enzyme production related to food industries. In *Microbial enzyme technology in food applications*. Ray, R. C. and Rosell, C. M., (eds). pp. 452-468, CRC Press, Boca Raton.
32. Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem*. **79**, 563-590.
33. Bott, M. (2015) Need for speed - finding productive mutations using transcription factor-based biosensors, fluorescence-activated cell sorting and recombineering. *Microb Biotechnol*. **8**, 8-10.
34. Wang, Q. Z., Tang, S. Y., and Yang, S. (2017) Genetic biosensors for small-molecule products: Design and applications in high-throughput screening. *Front Chem Sci Eng*. **11**, 15-26.
35. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature*. **409**, 258-268.
36. Sanchez, S., and Demain, A. L. (2011) Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Org Process Res Dev*. **15**, 224-230.
37. Leemhuis, H., Kelly, R. M., and Dijkhuizen, L. (2009) Directed evolution of enzymes: library screening strategies. *IUBMB Life*. **61**, 222-228.
38. Liszka, M. J., Clark, M. E., Schneider, E., and Clark, D. S. (2012) Nature versus nurture: developing enzymes that function under extreme conditions. *Annu Rev Chem Biomol Eng*. **3**, 77-102.
39. Illanes, A., Cauerhff, A., Wilson, L., and Castro, G. R. (2012) Recent trends in biocatalysis engineering. *Bioresour Technol*. **115**, 48-57.
40. Davids, T., Schmidt, M., Böttcher, D., and Bornscheuer, U. T. (2013) Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr Opin Chem Biol*
41. Huang, X. F., Santhanam, N., Badri, D. V., Hunter, W. J., Manter, D. K., Decker, S. R., Vivanco, J. M., and Reardon, K. F. (2013) Isolation and characterization of lignin-degrading bacteria from rainforest soils. *Biotechnol Bioeng*, in press.
42. Kennedy, J., O'Leary, N. D., Kiran, G. S., Morrissey, J. P., O'Gara, F., Selvin, J., and Dobson, A. D. W. (2011) Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J Appl Microbiol*. **111**, 787-799.
43. Timmis, K., Carreño, C. A., and de Lorenzo, V. (2010) Genetic traps for surveying new catalysts in (meta) genomic DNA. In *Handbook of Hydrocarbon and Lipid Microbiology*. pp. 4563-4579, Springer Berlin Heidelberg
44. Fernández-Arrojo, L., Guazzaroni, M. E., López-Cortés, N., Beloqui, A., and Ferrer, M. (2010) Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol*. **21**, 725-733.
45. Boersma, Y. L., Dröge, M. J., and Quax, W. J. (2007) Selection strategies for improved biocatalysts. *FEBS J*. **274**, 2181-2195.
46. Goldsmith, M., and Tawfik, D. S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol*. **22**, 406-412.
47. Brouns, S. J. J., Wu, H., Akerboom, J., Turnbull, A. P., de Vos, W. M., and van der Oost, J. (2005) Engineering a selectable marker for hyperthermophiles. *J Biol Chem*. **280**, 11422-11431.
48. Heselpoth, R. D., and Nelson, D. C. (2012) A new screening method for the directed evolution of thermostable bacteriolytic enzymes. *J Vis Exp*, e4216.
49. Acharya, P., Rajakumara, E., Sankaranarayanan, R., and Rao, N. M. (2004) Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase. *J Mol Biol*. **341**, 1271-1281.
50. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., and Houk, K. N. (2013) Computational enzyme design. *Angew Chem Int Ed Engl*
51. Kipnis, Y., and Baker, D. (2012) Comparison of designed and randomly generated catalysts for simple chemical reactions. *Protein Sci*. **21**, 1388-1395.
52. Khare, S. D., and Fleishman, S. J. (2012) Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett*, in press.
53. Hediger, M. R., De Vico, L., Svendsen, A., Besenmatter, W., and Jensen, J. H. (2012) A computational methodology to screen activities of enzyme variants. *PloS one*. **7**, e49849.

54. Michener, J. K., and Smolke, C. D. (2012) High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab Eng*. **14**, 306-316.
55. Gredell, J. A., Frei, C. S., and Cirino, P. C. (2012) Protein and RNA engineering to customize microbial molecular reporting. *Biotechnol J*. **7**, 477-499.
56. Yang, G., and Withers, S. G. (2009) Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem*. **10**, 2704-2715.
57. Konarzycka-Bessler, M., and Jaeger, K. E. (2006) Select the best: novel biocatalysts for industrial applications. *Trends Biotechnol*. **24**, 248-250.
58. Otten, L. G., Sio, C. F., Vrielink, J., Cool, R. H., and Quax, W. J. (2002) Altering the substrate specificity of cephalosporin acylase by directed evolution of the Beta -subunit. *J Biol Chem*. **277**, 42121-42127.
59. Fernández-Álvaro, E., Snajdrova, R., Jochens, H., Davids, T., Böttcher, D., and Bornscheuer, U. T. (2011) A combination of *in vivo* selection and cell sorting for the identification of enantioselective biocatalysts. *Angew Chem Int Ed Engl*. **50**, 8584-8587.
60. Böttcher, D., Schmidt, M., and Bornscheuer, U. T. (2010) Screens for active and stereoselective hydrolytic enzymes. *Methods Mol Biol*. **668**, 169-176.
61. Willies, S. C., White, J. L., and Turner, N. J. (2012) Development of a high-throughput screening method for racemase activity and its application to the identification of alanine racemase variants with activity towards L-arginine. *Tetrahedron*. **68**, 7564-7567.
62. Ingham, C. J., Sprenkels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J. E. T., and de Vos, W. M. (2007) The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proc Natl Acad Sci U S A*. **104**, 18217-18222.
63. Wagschal, K., and Lee, C. C. (2012) Microplate-based active/inactive 1 degrees screen for biomass degrading enzyme library purification and gene discovery. *J Microbiol Meth*. **89**, 83-85.
64. Kintses, B., Hein, C., Mohamed, M. F., Fischlechner, M., Courtois, F., Lainé, C., and Hollfelder, F. (2012) Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem Biol*. **19**, 1001-1009.
65. Ostafe, R., Prodanovic, R., Commandeur, U., and Fischer, R. (2013) Flow cytometry-based ultra-high-throughput screening assay for cellulase activity. *Anal Biochem*. **435**, 93-98.
66. Yang, G., Rich, J. R., Gilbert, M., Wakarchuk, W. W., Feng, Y., and Withers, S. G. (2010) Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases. *J Am Chem Soc*. **132**, 10570-10577.
67. Prodanovic, R., Ostafe, R., Scacioc, A., and Schwaneberg, U. (2011) Ultrahigh throughput screening system for directed glucose oxidase evolution in yeast cells. *Comb Chem High T Scr*. **14**, 55-60.
68. de Las Heras, A., Carreño, C. A., Martínez-García, E., and de Lorenzo, V. (2010) Engineering input/output nodes in prokaryotic regulatory circuits. *FEMS Microbiol Rev*. **34**, 842-865.
69. Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab Eng*. **14**, 212-222.
70. Kelkar, M., and De, A. (2012) Bioluminescence based *in vivo* screening technologies. *Curr Opin Pharmacol*. **12**, 592-600.
71. Garmendia, J., de las Heras, A., Galvão, T. C., and de Lorenzo, V. (2008) Tracing explosives in soil with transcriptional regulators of *Pseudomonas putida* evolved for responding to nitrotoluenes. *Microb Biotechnol*. **1**, 236-246.
72. de Las Heras, A., and de Lorenzo, V. (2012) Engineering whole-cell biosensors with no antibiotic markers for monitoring aromatic compounds in the environment. *Methods Mol Biol*. **834**, 261-281.
73. Sinha, J., Reyes, S. J., and Gallivan, J. P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol*. **6**, 464-470.
74. Beggah, S., Vogne, C., Zenaro, E., and Van Der Meer, J. R. (2008) Mutant HbpR transcription activator isolation for 2-chlorobiphenyl via green fluorescent protein-based flow cytometry and cell sorting. *Microb Biotechnol*. **1**, 68-78.
75. Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol*. **23**, 88-93.
76. Uchiyama, T., and Watanabe, K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat Protoc*. **3**, 1202-1212.
77. de Las Heras, A., Chavarría, M., and de Lorenzo, V. (2011) Association of *dnt* genes of *Burkholderia* sp. DNT with the substrate-blind regulator DntR draws the evolutionary itinerary of 2,4-dinitrotoluene biodegradation. *Mol Microbiol*. **82**, 287-299.

78. de Lorenzo, V. (2005) Problems with metagenomic screening. *Nat Biotechnol*. **23**, 1045; author reply 1045-1046.
79. Uchiyama, T., and Miyazaki, K. (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl Environ Microbiol*. **76**, 7029-7035.
80. Allert, M., Rizk, S. S., Looger, L. L., and Hellinga, H. W. (2004) Computational design of receptors for an organophosphate surrogate of the nerve agent soman. *Proc Natl Acad Sci U S A*. **101**, 7907-7912.
81. Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature*. **423**, 185-190.
82. Schreier, B., Stumpp, C., Wiesner, S., and Höcker, B. (2009) Computational design of ligand binding is not a solved problem. *Proc Natl Acad Sci U. S. A.* **106**, 18491-18496.
83. Malisi, C., Schumann, M., Toussaint, N. C., Kageyama, J., Kohlbacher, O., and Höcker, B. (2012) Binding pocket optimization by computational protein design. *PloS one*. **7**, e52505.
84. van Sint Fiet, S., van Beilen, J. B., and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci U. S. A.* **103**, 1693-1698.
85. Cebolla, A., Sousa, C., and de Lorenzo, V. (1997) Effector specificity mutants of the transcriptional activator NahR of naphthalene degrading *Pseudomonas* define protein sites involved in binding of aromatic inducers. *J Biol Chem*. **272**, 3986-3992.
86. Mohn, W. W., Garmendia, J., Galvao, T. C., and de Lorenzo, V. (2006) Surveying biotransformations with *à la carte* genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into *lacZ*-based phenotypes. *Environ Microbiol*. **8**, 546-555.
87. Tang, S.-Y., and Cirino, P. C. (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl*. **50**, 1084-1086.
88. Firestine, S. M., Salinas, F., Nixon, A. E., Baker, S. J., and Benkovic, S. J. (2000) Using an AraC-based three-hybrid system to detect biocatalysts *in vivo*. *Nat Biotechnol*. **18**, 544-547.
89. Baker, K., Bleczinski, C., Lin, H., Salazar-Jimenez, G., Sengupta, D., Krane, S., and Cornish, V. W. (2002) Chemical complementation: A reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A*. **99**, 16537-16542.
90. Lin, H., Tao, H. Y., and Cornish, V. W. (2004) Directed evolution of a glycosynthase via chemical complementation. *J Am Chem Soc*. **126**, 15051-15059.
91. Peralta-Yahya, P., Carter, B. T., Lin, H., Tao, H., and Cornish, V. W. (2008) High-throughput selection for cellulase catalysts using chemical complementation. *J Am Chem Soc*. **130**, 17446-17452.
92. Verhoeven, K. D., Altstadt, O. C., and Savinov, S. N. (2012) Intracellular detection and evolution of site-specific proteases using a genetic selection system. *Appl Biochem Biotechnol*. **166**, 1340-1354.
93. Ruigrok, V. J., Levisson, M., Eppink, M. H. M., Smidt, H., and van der Oost, J. (2011) Alternative affinity tools: more attractive than antibodies? *Biochem J*. **436**, 1-13.
94. Ruigrok, V. J., van Duijn, E., Barendregt, A., Dyer, K., Tainer, J. A., Stoltenburg, R., Strehlitz, B., Levisson, M., Smidt, H., and van der Oost, J. (2012) Kinetic and stoichiometric characterisation of streptavidin-binding aptamers. *ChemBioChem*. **13**, 829-836.
95. Desai, S. K., and Gallivan, J. P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc*. **126**, 13247-13254.
96. Kostallas, G., Löfdahl, P. A., and Samuelson, P. (2011) Substrate profiling of tobacco etch virus protease using a novel fluorescence-assisted whole-cell assay. *PloS one*. **6**, e16136.
97. O'Loughlin, T. L., Greene, D. N., and Matsumura, I. (2006) Diversification and specialization of HIV protease function during in vitro evolution. *Mol Biol Evol*. **23**, 764-772.
98. Owen, J. G., Robins, K. J., Parachin, N. S., and Ackerley, D. F. (2012) A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environ Microbiol*. **14**, 1198-1209.
99. Gupta, R. D., and Tawfik, D. S. (2008) Directed enzyme evolution via small and effective neutral drift libraries. *Nat Methods*. **5**, 939-942.
100. Japrung, D., Chusacultanachai, S., Yuvaniyama, J., Wilairat, P., and Yuthavong, Y. (2005) A simple dual selection for functionally active mutants of *Plasmodium falciparum* dihydrofolate reductase with improved solubility. *Protein Eng Des Sel*. **18**, 457-464.
101. García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R. M., Arís, A., Ventura, S., and Villaverde, A. (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb Cell Fact*. **4**, 27.

102. Pourmir, A., and Johannes, T. W. (2012) Directed evolution: selection of the host organism. *CSBJ*. **2**, e201209012.
103. Hu, J. C., Kornacker, M. G., and Hochschild, A. (2000) *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods*. **20**, 80-94.
104. Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A*. **97**, 7382-7387.
105. Marvin, J. S., Schreiter, E. R., Echevarría, I. M., and Looger, L. L. (2011) A genetically encoded, high-signal-to-noise maltose sensor. *Proteins*. **79**, 3025-3036.
106. Xiong, A. S., Peng, R. H., Zhuang, J., Davies, J., Zhang, J., and Yao, Q. H. (2012) Advances in directed molecular evolution of reporter genes. *Crit Rev Biotechnol*. **32**, 133-142.
107. Cho, J. H., Lee, D. Y., Lim, W. K., and Shin, H. J. (2014) A recombinant *Escherichia coli* biosensor for detecting polycyclic aromatic hydrocarbons in gas and aqueous phases. *Prep Biochem Biotechnol*. **44**, 849-860.
108. Reed, B., Blazeck, J., and Alper, H. (2012) Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *J Biotechnol*. **158**, 75-79.
109. Webster, D. P., TerAvest, M. A., Doud, D. F. R., Chakravorty, A., Holmes, E. C., Radens, C. M., Sureka, S., Gralnick, J. A., and Angenent, L. T. (2014) An arsenic-specific biosensor with genetically engineered *Shewanella oneidensis* in a bioelectrochemical system. *Biosens Bioelectron*. **62**, 320-324.
110. Choi, S. L., Rha, E., Lee, S. J., Kim, H., Kwon, K., Jeong, Y. S., Rhee, Y. H., Song, J. J., Kim, H. S., and Lee, S. G. (2014) Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. *ACS Synth Biol*. **3**, 163-171.
111. Jha, R. K., Kern, T. L., Fox, D. T., and Strauss, C. E. M. (2014) Engineering an *Acinetobacter* regulon for biosensing and high-throughput enzyme screening in *E. coli* via flow cytometry. *Nucleic Acids Res*. **42**, 8150-8160.
112. Siedler, S., Schendzielorz, G., Binder, S., Eggeling, L., Bringer, S., and Bott, M. (2014) SoxR as a single-cell biosensor for NADPH-consuming enzymes in *Escherichia coli*. *ACS Synth Biol*. **3**, 41-47.
113. Mahr, R., and Frunzke, J. (2013) Transkriptionsregulatoren im Dienste der Biotechnologie. *BIOspektrum*. **19**, 739-741.
114. Schendzielorz, G., Binder, S., and Marienhagen, J. (2014) Biosensoren für die mikrobielle Stammentwicklung im Hochdurchsatzformat. *BIOspektrum*. **20**, 228-230.
115. Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014) Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metab Eng*. **21**, 2-8.
116. Hynninen, A., and Virta, M. (2010) Whole-cell bioreporters for the detection of bioavailable metals. In *Whole Cell Sensing System II* Belkin, S. and Gu, M. B., (eds). pp. 31-63, Springer Berlin Heidelberg
117. Merulla, D., Hatzimanikatis, V., and van der Meer, J. R. (2013) Tunable reporter signal production in feedback-uncoupled arsenic bioreporters. *Microb Biotechnol*. **6**, 503-514.
118. van der Meer, J. R., and Belkin, S. (2010) Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat Rev Microbiol*. **8**, 511-522.
119. Ganesh, I., Ravikumar, S., Lee, S. H., Park, S. J., and Hong, S. H. (2013) Engineered fumarate sensing *Escherichia coli* based on novel chimeric two-component system. *J Biotechnol*. **168**, 560-566.
120. Jeong, Y. S., Choi, S. L., Kyeong, H. H., Kim, J. H., Kim, E. J., Pan, J. G., Rha, E., Song, J. J., Lee, S. G., and Kim, H. S. (2012) High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. *Protein Eng Des Sel*. **25**, 725-731.
121. Schallmey, M., Frunzke, J., Eggeling, L., and Marienhagen, J. (2014) Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr Opin Biotech*. **26**, 148-154.
122. Binder, S., Schendzielorz, G., Stäbler, N., Krumbach, K., Hoffmann, K., Bott, M., and Eggeling, L. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol*. **13**, R40.
123. Gupta, S., Saxena, M., Saini, N., Mahmooduzzafar, Kumar, R., and Kumar, A. (2012) An effective strategy for a whole-cell biosensor based on putative effector interaction site of the regulatory DmpR protein. *PloS one*. **7**, e43527.
124. van der Meer, J. R., Tropel, D., and Jaspers, M. (2004) Illuminating the detection chain of bacterial bioreporters. *Environ Microbiol*. **6**, 1005-1020.
125. van Rossum, T., Kengen, S. W. M., and van der Oost, J. (2013) Reporter-based screening and selection of enzymes. *FEBS J*. **280**, 2979-2996.
126. Checa, S. K., Zurbriggen, M. D., and Soncini, F. C. (2012) Bacterial signaling systems as platforms for rational design of new generations of biosensors. *Curr Opin Biotech*. **23**, 766-772.

127. Park, M., Tsai, S.-L., and Chen, W. (2013) Microbial biosensors: engineered microorganisms as the sensing machinery. *Sensors*. **13**, 5777-5795.
128. Tang, S.-Y., Fazelinia, H., and Cirino, P. C. (2008) AraC regulatory protein mutants with altered effector specificity. *J Am Chem Soc*. **130**, 5267-5271.
129. Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C. S., Gredell, J. A., and Cirino, P. C. (2013) Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. *J Am Chem Soc*. **135**, 10099-10103.
130. Schleif, R. (2010) AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol Rev*. **34**, 779-796.
131. Stringer, A. M., Currenti, S., Bonocora, R. P., Baranowski, C., Petrone, B. L., Palumbo, M. J., Reilly, A. A., Zhang, Z., Erill, I., and Wade, J. T. (2014) Genome-scale analyses of *Escherichia coli* and *Salmonella enterica* AraC reveal noncanonical targets and an expanded core regulon. *J Bacteriol*. **196**, 660-671.
132. Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem*. **62**, 749-797.
133. Wright, G. D., and Thompson, P. R. (1999) Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Front Biosci*. **4**, D9-D21.
134. Somers, J., Amzallag, A., and Middleton, R. (1973) Genetic fine structure of the leucine operon of *Escherichia coli* K-12. *J Bacteriol*. **113**, 1268-1272.
135. Close, D. M., Ripp, S., and Sayler, G. S. (2009) Reporter proteins in whole-cell optical bioreporter detection systems, biosensor integrations, and biosensing applications. *Sensors*. **9**, 9147-9174.
136. Hakkila, K., Maksimow, M., Karp, M., and Virta, M. (2002) Reporter genes *lucFF*, *luxCDABE*, *gfp*, and *dsred* have different characteristics in whole-cell bacterial sensors. *Anal Biochem*. **301**, 235-242.
137. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U. S. A.* **97**, 6640-6645.
138. Grenier, F., Matteau, D., Baby, V., and Rodrigue, S. (2014) Complete genome sequence of *Escherichia coli* BW25113. *Genome Announc*. **2**, e01038-01014.
139. Morgan-Kiss, R. M., Wadler, C., and Cronan, J. E. (2002) Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity. *Proc Natl Acad Sci U. S. A.* **99**, 7373-7377.
140. Albert, H., Dale, E. C., Lee, E., and Ow, D. W. (1995) Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *Plant J*. **7**, 649-659.
141. Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*. **25**, 1203-1210.
142. Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T. A., and Keasling, J. (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol*. **182**, 7029-7034.
143. Kumar, S., and Venkatesh, K. (2010) Application of methylene blue dye reduction test (MBRT) to determine growth and death rates of microorganisms. *Afr J Microbiol Res*. **4**, 061-070.
144. Uhlin, B. E., and Nordström, K. (1977) R plasmid gene dosage effects in *Escherichia coli* K-12: Copy mutants of the R plasmid R1*drd*-19. *Plasmid*. **1**, 1-7.
145. Beverin, S., Sheppard, D. E., and Park, S. S. (1971) D-Fucose as a gratuitous inducer of the L-arabinose operon in strains of *Escherichia coli* B/r mutant in gene *araC*. *J Bacteriol*. **107**, 79-86.
146. Shetty, R. S., Ramanathan, S., Badr, I. H., Wolford, J. L., and Daunert, S. (1999) Green fluorescent protein in the design of a living biosensing system for L-arabinose. *Anal Chem*. **71**, 763-768.
147. Lis, J. T., and Schleif, R. (1973) Different cyclic AMP requirements for induction of the arabinose and lactose operons of *Escherichia coli*. *J Mol Biol*. **79**, 149-162.
148. Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M., and Frunzke, J. (2012) The development and application of a single-cell biosensor for the detection of L-methionine and branched-chain amino acids. *Metab Eng*. **14**, 449-457.
149. Choi, O., Lee, Y., Han, I., Kim, H., Goo, E., Kim, J., and Hwang, I. (2013) A simple and sensitive biosensor strain for detecting toxoflavin using β-galactosidase activity. *Biosens Bioelectron*. **50**, 256-261.
150. Xu, Z., Li, S., Feng, X., Liang, J., and Xu, H. (2014) L-Arabinose isomerase and its use for biotechnological production of rare sugars. *Appl Microbiol Biotechnol*. **98**, 8869-8878.
151. Englesberg, E. (1961) Enzymatic characterization of 17 L-arabinose negative mutants of *Escherichia coli*. *J Bacteriol*. **81**, 996-1006.
152. Tewari, Y. B., and Goldberg, R. N. (1985) An Investigation of the Equilibria between Aqueous Ribose, Ribulose, and Arabinose. *Biophys Chem*. **22**, 197-204.

153. Yamanaka, K. (1960) Studies on the Pyruvate and Carbohydrate Metabolisms by Lactic Acid Bacteria .10. Some Properties of Pentose Isomerase by Hetero-Type of Lactic Acid Bacteria. *B Agr Chem Soc Japan*. **24**, 305-310.
154. Copp, J. N., Williams, E. M., Rich, M. H., Patterson, A. V., Smaill, J. B., and Ackerley, D. F. (2014) Toward a high-throughput screening platform for directed evolution of enzymes that activate genotoxic prodrugs. *Protein Eng Des Sel*. **27**, 399-403.
155. Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013) Transcription Factor-Based Screens and Synthetic Selections for Microbial Small-Molecule Biosynthesis. *ACS Synth Biol*. **2**, 47-58.
156. Galvão, T. C., and de Lorenzo, V. (2005) Adaptation of the yeast *URA3* selection system to gram-negative bacteria and generation of a Δ*betCDE Pseudomonas putida* strain. *Appl Environ Microbiol*. **71**, 883-892.
157. Dutra, B. E., Sutera, V. A., and Lovett, S. T. (2007) RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U. S. A.* **104**, 216-221.
158. Taylor, S. V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection. *Angew Chem Int Ed Engl*. **40**, 3310-3335.
159. Galvão, T. C., and de Lorenzo, V. (2006) Transcriptional regulators à la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr Opin Biotech*. **17**, 34-42.
160. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning. Cold spring harbor laboratory press, New York.
161. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. **2**, 2006.
162. Uliczka, F., Pisano, F., Kochut, A., Opitz, W., Herbst, K., Stolz, T., and Dersch, P. (2011) Monitoring of gene expression in bacteria during infections using an adaptable set of bioluminescent, fluorescent and colorigenic fusion vectors. *PloS one*. **6**, e20425.
163. Glascock, C. B., and J. Weickert, M. (1998) Using chromosomal *lacI*^{Q1} to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene*. **223**, 221-231.
164. Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*. **158**, 9-14.
165. Westra, E. R., Pul, Ü., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., Mastop, M., Wagner, E. G. H., Schnetz, K., van der Oost, J*., et al.* (2010) H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol Microbiol*. **77**, 1380-1393.
166. Palmeros, B. z., Wild, J., Szybalski, W., Le Borgne, S., Hernández-Chávez, G., Gosset, G., Valle, F., and Bolivar, F. (2000) A family of removable cassettes designed to obtain antibiotic-resistance-free genomic modifications of *Escherichia coli* and other bacteria. *Gene*. **247**, 255-264.
167. Wild, J., Sektas, M., Hradečná, Z., and Szybalski, W. (1998) Targeting and retrofitting pre-existing libraries of transposon insertions with *FRT* and *oriV* elements for in-vivo generation of large quantities of any genomic fragment. *Gene*. **223**, 55-66.
168. Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol*. **189**, 8746-8749.
169. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389-3402.
170. Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., and Yu, Y. K. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J*. **272**, 5101-5109.
171. Carver, T. E., Hochstrasser, R. A., and Millar, D. P. (1994) Proofreading DNA: recognition of aberrant DNA termini by the Klenow fragment of DNA polymerase I. *Proc Natl Acad Sci U. S. A.* **91**, 10670-10674.
172. Chang, A. C., and Cohen, S. N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol*. **134**, 1141-1156.
173. Bron, S., and Luxen, E. (1985) Segregational instability of pUB110-derived recombinant plasmids in *Bacillus subtilis*. *Plasmid*. **14**, 235-244.
174. Carra, J. H., and Schleif, R. F. (1993) Variation of half-site organization and DNA looping by AraC protein. *EMBO J*. **12**, 35-44.
175. Dunn, T. M., and Schleif, R. (1984) Deletion analysis of the *Escherichia coli ara* P_C and P_{BAD} promoters. *J Mol Biol*. **180**, 201-204.

176. Eggeling, L., Bott, M., and Marienhagen, J. (2015) Novel screening methods - biosensors. *Curr Opin Biotech*. **35**, 30-36.
177. Mahr, R., and Frunzke, J. (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl Microbiol Biot*. **100**, 79-90.
178. Zhang, J., Jensen, M. K., and Keasling, J. D. (2015) Development of biosensors and their application in metabolic engineering. *Curr Opin Chem Biol*. **28**, 1-8.
179. Liu, D., Evans, T., and Zhang, F. Z. (2015) Applications and advances of metabolite biosensors for metabolic engineering. *Metab Eng*. **31**, 35-43.
180. Ndu, U., Mason, R. P., Zhang, H., Lin, S. J., and Visscher, P. T. (2012) Effect of inorganic and organic ligands on the bioavailability of methylmercury as determined by using a *mer-lux* bioreporter. *Appl Environ Microbiol*. **78**, 7276-7282.
181. Santos, C. L., Tavares, F., Thioulouse, J., and Normand, P. (2009) A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiol Rev*. **33**, 411-429.
182. de los Santos, E. L. C., Meyerowitz, J. T., Mayo, S. L., and Murray, R. M. (2016) Engineering transcriptional regulator effector specificity using computational design and *in vitro* rapid prototyping: developing a vanillin sensor. *ACS Synth Biol*. **5**, 287-295.
183. Jha, R. K., Kern, T. L., Kim, Y., Tesar, C., Jedrzejczak, R., Joachimiak, A., and Strauss, C. E. M. (2016) A microbial sensor for organophosphate hydrolysis exploiting an engineered specificity switch in a transcription factor. *Nucleic Acids Res*. **44**, 8490-8500.
184. van Rossum, T., Muras, A., Baur, M. J. J., Creutzburg, S. C. A., van der Oost, J., and Kengen, S. W. M. (2017) A growth- and bioluminescence-based bioreporter for the *in vivo* detection of novel biocatalysts. *Microb Biotechnol*. **10**, 625–641.
185. Lewis, M. (2005) The *lac* repressor. *C. R. Biol.* **328**, 521-548.
186. Wilson, C., Zhan, H., Swint-Kruse, L., and Matthews, K. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci*. **64**, 3-16.
187. Davey, J. A., and Wilson, C. J. (2017) Deconstruction of complex protein signaling switches: a roadmap toward engineering higher-order gene regulators. *Wires Nanomed Nanobi*. **9**, e1461.
188. Fulcrand, G., Chapagain, P., Dunlap, D., and Leng, F. F. (2016) Direct observation of a 91 bp LacI-mediated, negatively supercoiled DNA loop by atomic force microscope. *FEBS Lett*. **590**, 613-618.
189. Fulcrand, G., Dages, S., Zhi, X. D., Chapagain, P., Gerstman, B. S., Dunlap, D., and Leng, F. F. (2016) DNA supercoiling, a critical signal regulating the basal expression of the *lac* operon in *Escherichia coli*. *Sci Rep-Uk*. **6**, 19243.
190. Gatti-Lafranconi, P., Dijkman, W. P., Devenish, S. R., and Hollfelder, F. (2013) A single mutation in the core domain of the *lac* repressor reduces leakiness. *Microb Cell Fact*. **12**, 67.
191. Schofield, D. A., Westwater, C., Dolan, J. W., Norris, J. S., and Schmidt, M. G. (2002) Doc-mediated cell killing in *Shigella flexneri* using a C1/LacI controlled expression system. *FEMS Microbiol Lett*. **215**, 237-242.
192. Poelwijk, F. J., de Vos, M. G. J., and Tans, S. J. (2011) Tradeoffs and optimality in the evolution of gene regulation. *Cell*. **146**, 462-470.
193. Richards, D. H., Meyer, S., and Wilson, C. J. (2017) Fourteen ways to reroute cooperative communication in the lactose repressor: engineering regulatory proteins with alternate repressive functions. *ACS Synth Biol*. **6**, 6-12.
194. Meinhardt, S., Manley, M. W., Becker, N. A., Hessman, J. A., Maher, L. J., and Swint-Kruse, L. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res*. **40**, 11139-11154.
195. Shis, D. L., Hussain, F., Meinhardt, S., Swint-Kruse, L., and Bennett, M. R. (2014) Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras. *ACS Synth Biol*. **3**, 645-651.
196. Taylor, N. D., Garruss, A. S., Moretti, R., Chan, S., Arbing, M. A., Cascio, D., Rogers, J. K., Isaacs, F. J., Kosuri, S., Baker, D., Fields, S., Church, G. M., and Raman, S. (2016) Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods*. **13**, 177-183.
197. Wu, J., Jiang, P., Chen, W., Xiong, D., Huang, L., Jia, J., Chen, Y., Jin, J. M., and Tang, S. Y. (2017) Design and application of a lactulose biosensor. *Sci Rep-Uk*. **7**, 45994.
198. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Method Enzymol*. **185**, 60-89.
199. Lee, S. G., and Liao, J. C. (2008) Control of acetate production rate in *Escherichia coli* by regulating expression of single-copy *pta* using *lacI(Q)* in multicopy plasmid. *J Microbiol Biotech*. **18**, 334-337.

200. Atsumi, S., and Little, J. W. (2004) Regulatory circuit design and evolution using phage lambda. *Gene Dev*. **18**, 2086-2094.
201. Hasche, A. A., and Voss, C. (2005) Immobilisation of a repressor protein for binding of plasmid DNA. *J Chromatogr A*. **1080**, 76-82.
202. Manoil, C., and Bailey, J. (1997) A simple screen for permissive sites in proteins: analysis of *Escherichia coli lac* permease. *J Mol Biol*. **267**, 250-263.
203. Marbach, A., and Bettenbrock, K. (2012) *lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol*. **157**, 82-88.
204. Daber, R., Sochor, M. A., and Lewis, M. (2011) Thermodynamic analysis of mutant *lac* repressors. *J Mol Biol*. **409**, 76-87.
205. Stenberg, K. A. E., and Vihinen, M. (2009) Crystal structure of a 1.6-hexanediol bound tetrameric form of *Escherichia coli* lac-repressor refined to 2.1 angstrom resolution. *Proteins*. **75**, 748-759.
206. Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in synthesis of proteins. *J Mol Biol*. **3**, 318-356.
207. Müller-Hill, B., Rickenberg, H. V., and Wallenfels, K. (1964) Specificity of induction of enzymes of lac operon in *Escherichia coli*. *J Mol Biol*. **10**, 303-318.
208. Barkley, M. D., Riggs, A. D., Jobe, A., and Bourgeois, S. (1975) Interaction of effecting ligands with *lac* repressor and repressor-operator complex. *Biochemistry*. **14**, 1700-1712.
209. Wilson, C. J., Zhan, H., Swint-Kruse, L., and Matthews, K. S. (2007) Ligand interactions with lactose repressor protein and the repressor-operator complex: The effects of ionization and oligomerization on binding. *Biophys Chem*. **126**, 94-105.
210. Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007) Structural analysis of *lac* repressor bound to allosteric effectors. *J Mol Biol*. **370**, 609-619.
211. Barry, J. K., and Matthews, K. S. (1999) Substitutions at histidine 74 and aspartate 278 alter ligand binding and allostery in lactose repressor protein. *Biochemistry*. **38**, 3579-3590.
212. Rodgers, M. E., and Schleif, R. (2012) Heterodimers reveal that two arabinose molecules are required for the normal arabinose response of AraC. *Biochemistry*. **51**, 8085-8091.
213. Wilcox, G. (1974) The interaction of L-arabinose and D-fucose with AraC protein. *J Biol Chem*. **249**, 6892-6894.
214. Quax, T. E., Wolf, Y. I., Koehorst, J. J., Wurtzel, O., van der Oost, R., Ran, W., Blombach, F., Makarova, K. S., Brouns, S. J., and Forster, A. C. (2013) Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep*. **4**, 938-944.
215. Yagur-Kroll, S., Lalush, C., Rosen, R., Bachar, N., Moskovitz, Y., and Belkin, S. (2014) *Escherichia coli* bioreporters for the detection of 2,4-dinitrotoluene and 2,4,6-trinitrotoluene. *Appl Microbiol Biot*. **98**, 885-895.
216. Mahr, R., von Boeselager, R. F., Wiechert, J., and Frunzke, J. (2016) Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Appl Microbiol Biot*. **100**, 6739-6753.
217. Rogers, J. K., and Church, G. M. (2016) Genetically encoded sensors enable real-time observation of metabolite production. *Proc Natl Acad Sci U. S. A.* **113**, 2388-2393.
218. Liu, Y., Zhuang, Y., Ding, D., Xu, Y., Sun, J., and Zhang, D. (2017) Biosensor-based evolution and elucidation of a biosynthetic pathway in *Escherichia coli*. *ACS Synth Biol*. **6**, 837-848.
219. Xiong, D., Lu, S., Wu, J., Liang, C., Wang, W., Wang, W., Jin, J. M., and Tang, S. Y. (2017) Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. *Metab Eng*. **40**, 115-123.
220. De Paepe, B., Peters, G., Coussement, P., Maertens, J., and De Mey, M. (2017) Tailor-made transcriptional biosensors for optimizing microbial cell factories. *J Ind Microbiol Biot*. **44**, 623-645.
221. Libis, V., Delépine, B., and Faulon, J. L. (2016) Sensing new chemicals with bacterial transcription factors. *Curr Opin Microbiol*. **33**, 105-112.
222. Shi, S., Choi, Y. W., Zhao, H., Tan, M. H., and Ang, E. L. (2017) Discovery and engineering of a 1-butanol biosensor in *Saccharomyces cerevisiae*. *Bioresource Technol*
223. Li, H., Liang, C., Chen, W., Jin, J. M., Tang, S. Y., and Tao, Y. (2017) Monitoring *in vivo* metabolic flux with a designed whole-cell metabolite biosensor of shikimic acid. *Biosens Bioelectron*. **98**, 457-465.
224. Harton, M. D., Wingler, L. M., and Cornish, V. W. (2013) Transcriptional regulation improves the throughput of three-hybrid counter selections in *Saccharomyces cerevisiae*. *Biotechnol J*. **8**, 1485-1491.
225. Chou, H. H., and Keasling, J. D. (2013) Programming adaptive control to evolve increased metabolite production. *Nat Commun*. **4**, 2595.
226. Ganesh, I., Ravikumar, S., Yoo, I. K., and Hong, S. H. (2015) Construction of malate-sensing *Escherichia coli* by introduction of a novel chimeric two-component system. *Bioproc Biosyst Eng*. **38**, 797-804.

227. Wong, T. S., Tee, K. L., Hauer, B., and Schwaneberg, U. (2004) Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res*. **32**, e26.
228. Zhao, J., Kardashliev, T., Ruff, A. J., Bocola, M., and Schwaneberg, U. (2014) Lessons From Diversity of Directed Evolution Experiments by an Analysis of 3,000 Mutations. *Biotechnol Bioeng*. **111**, 2380-2389.
229. Scheib, U., Shanmugaratnam, S., Farías-Rico, J. A., and Höcker, B. (2014) Change in protein-ligand specificity through binding pocket grafting. *J Struct Biol*. **185**, 186-192.
230. Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins. *Nat Rev Genet*. **16**, 379-394.
231. Rosenfeld, L., Heyne, M., Shifman, J. M., and Papo, N. (2016) Protein engineering by combined computational and *in vitro* evolution approaches. *Trends Biochem Sci*. **41**, 421-433.
232. Stoddard, B. L., ed. (2016) Computational design of ligand binding proteins. Humana Press
233. Verma, D., Grigoryan, G., and Bailey-Kellogg, C. (2015) Structure-based design of combinatorial mutagenesis libraries. *Protein Sci*. **24**, 895-908.
234. Bendl, J., Stourac, J., Sebestova, E., Vavra, O., Musil, M., Brezovsky, J., and Damborsky, J. (2016) HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Res*. **44**, W479-W487.
235. Rockah-Shmuel, L., Tawfik, D. S., and Goldsmith, M. (2014) Generating targeted libraries by the combinatorial incorporation of synthetic oligonucleotides during gene shuffling (ISOR). *Methods Mol Biol*. **1179**, 129-137.
236. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., and Bocola, M. (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J Am Chem Soc*. **132**, 9144-9152.
237. Currin, A., Swainston, N., Day, P. J., and Kell, D. B. (2014) SpeedyGenes: an improved gene synthesis method for the efficient production of error-corrected, synthetic protein libraries for directed evolution. *Protein Eng Des Sel*. **27**, 273-280.
238. Acevedo-Rocha, C. G., and Reetz, M. T. (2014) Assembly of designed oligonucleotides: a useful tool in synthetic biology for creating high-quality combinatorial DNA libraries. *Methods Mol Biol*. **1179**, 189-206.
239. Pines, G., Pines, A., Garst, A. D., Zeitoun, R. I., Lynch, S. A., and Gill, R. T. (2015) Codon compression algorithms for saturation mutagenesis. *ACS Synth Biol*. **4**, 604-614.
240. Parra, L. P., Agudo, R., and Reetz, M. T. (2013) Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBioChem*. **14**, 2301-2309.
241. Lowe, M., Gullotti, D., Damjanovic, A., Cheng, A., Dirla, S., and Schleif, R. (2014) Computational and experimental investigation of constitutive behavior in AraC. *Proteins*. **82**, 3385-3396.
242. Malaga, F., Mayberry, O., Park, D. J., Rodgers, M. E., Toptygin, D., and Schleif, R. F. (2016) A genetic and physical study of the interdomain linker of *E. coli* AraC protein-a *trans*-subunit communication pathway. *Proteins*. **84**, 448-460.
243. Damjanovic, A., Miller, B. T., and Schleif, R. (2013) Understanding the basis of a class of paradoxical mutations in AraC through simulations. *Proteins*. **81**, 490-498.
244. Lee, S. K., Chou, H. H., Pfleger, B. F., Newman, J. D., Yoshikuni, Y., and Keasling, J. D. (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl Environ Microbiol*. **73**, 5711-5715.
245. Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science*. **276**, 421-425.
246. Chen, W., Zhang, S., Jiang, P. X., Yao, J., He, Y. Z., Chen, L. C., Gui, X. W., Dong, Z. Y., and Tang, S. Y. (2015) Design of an ectoine-responsive AraC mutant and its application in metabolic engineering of ectoine biosynthesis. *Metab Eng*. **30**, 149-155.
247. Schellenberg, G. D., Sarthy, A., Larson, A. E., Backer, M. P., Crabb, J. W., Lidstrom, M., Hall, B. D., and Furlong, C. E. (1984) Xylose isomerase from *Escherichia coli* - characterization of the protein and the structural gene. *J Biol Chem*. **259**, 6826-6832.
248. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985) Positive selection procedure for entrapment of insertion-sequence elements in gram-negative bacteria. *J Bacteriol*. **164**, 918-921.
249. Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997) The 1.6 angstrom crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *J Mol Biol*. **273**, 226-237.
250. Dirla, S., Chien, J. Y. H., and Schleif, R. (2009) Constitutive mutations in the *Escherichia coli* AraC protein. *J Bacteriol*. **191**, 2668-2674.

251. Bosley, A. D., and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol Eng*. **22**, 57-61.
252. Henderson, P. J. F., Giddens, R. A., and Jones-Mortimer, M. C. (1977) Transport of galactose, glucose and their molecular analogs by *Escherichia-coli*-K12. *Biochemical Journal*. **162**, 309-320.
253. van Staalduinen, L. M., Park, C. S., Yeom, S. J., Adams-Cioaba, M. A., Oh, D. K., and Jia, Z. (2010) Structure-based annotation of a novel sugar isomerase from the pathogenic *E. coli* O157:H7. *J Mol Biol*. **401**, 866-881.
254. LeBlanc, D. J., and Mortlock, R. P. (1971) Metabolism of D-arabinose - a new pathway in *Escherichia coli*. *J Bacteriol*. **106**, 90-96.
255. Tate, C. G., Muiry, J. A. R., and Henderson, P. J. F. (1992) Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. *J Biol Chem*. **267**, 6923-6932.
256. Wilson, D. M., and Wilson, T. H. (1987) Cation specificity for sugar substrates of the melibiose carrier in *Escherichia coli*. *Biochim Biophys Acta*. **904**, 191-200.
257. Collins, C. H., Leadbetter, J. R., and Arnold, F. H. (2006) Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat Biotechnol*. **24**, 708-712.
258. Ross, J. J., Gryczynski, U., and Schleif, R. (2003) Mutational analysis of residue roles in AraC function. *J Mol Biol*. **328**, 85-93.
259. Tang, S. Y., and Cirino, P. C. (2010) Elucidating residue roles in engineered variants of AraC regulatory protein. *Protein Science*. **19**, 291-298.
260. Lee, N., Wilcox, G., Gielow, W., Arnold, J., Cleary, P., and Englesberg, E. (1974) *In vitro* activation of transcription of *ara*BAD operon by *ara*C activator. *Proc Natl Acad Sci U. S. A.* **71**, 634-638.
261. Ferreira Amaral, M. M., Frigotto, L., and Hine, A. V. (2017) Beyond the Natural Proteome. *Method Enzymol*. **585**, 111-133.
262. Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014) Evolution-guided optimization of biosynthetic pathways. *Proc Natl Acad Sci U. S. A.* **111**, 17803-17808.
263. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., *et al.* (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*. **390**, 249-256.
264. Blomfield, I. C., Vaughn, V., Rest, R. F., and Eisenstein, B. I. (1991) Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol*. **5**, 1447-1457.
265. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. **33**, 103-119.
266. Chang, D. E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J., Anderson, A. B., Grissom, J. E., Laux, D. C., Cohen, P. S., and Conway, T. (2004) Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci U S A*. **101**, 7427-7432.
267. Conway, T., and Cohen, P. S. (2015) Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiology spectrum*. **3**, MBP-0006-2014.
268. Baev, M. V., Baev, D., Radek, A. J., and Campbell, J. W. (2006) Growth of *Escherichia coli* MG1655 on LB medium: determining metabolic strategy with transcriptional microarrays. *Appl Microbiol Biotechnol*. **71**, 323-328.
269. Baev, M. V., Baev, D., Radek, A. J., and Campbell, J. W. (2006) Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols, and organic acids with transcriptional microarrays. *Appl Microbiol Biotechnol*. **71**, 310-316.
270. Baev, M. V., Baev, D., Radek, A. J., and Campbell, J. W. (2006) Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of amino acids, peptides, and nucleotides with transcriptional microarrays. *Appl Microbiol Biotechnol*. **71**, 317-322.
271. Raman, B., Nandakumar, M. P., Muthuvijayan, V., and Marten, M. R. (2005) Proteome analysis to assess physiological changes in *Escherichia coli* grown under glucose-limited fed-batch conditions. *Biotechnol Bioeng*. **92**, 384-392.
272. Aidelberg, G., Towbin, B. D., Rothschild, D., Dekel, E., Bren, A., and Alon, U. (2014) Hierarchy of non-glucose sugars in *Escherichia coli*. *BMC systems biology*. **8**, 133.
273. Desai, T. A., and Rao, C. V. (2010) Regulation of arabinose and xylose metabolism in *Escherichia coli*. *Appl Environ Microbiol*. **76**, 1524-1532.
274. Koirala, S., Wang, X., and Rao, C. V. (2015) Reciprocal regulation of L-arabinose and D-xylose metabolism in *Escherichia coli*. *J Bacteriol*. **198**, 386-393.

275. Fritz, G., Megerle, J. A., Westermayer, S. A., Brick, D., Heermann, R., Jung, K., Rädler, J. O., and Gerland, U. (2014) Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. coli*. *PloS one*. **9**, e89532.
276. Reeder, T., and Schleif, R. (1991) Mapping, sequence, and apparent lack of function of araJ, a gene of the *Escherichia coli* arabinose regulon. *J Bacteriol*. **173**, 7765-7771.
277. Horler, R. S., Müller, A., Williamson, D. C., Potts, J. R., Wilson, K. S., and Thomas, G. H. (2009) Furanose-specific sugar transport: characterization of a bacterial galactofuranose-binding protein. *J Biol Chem*. **284**, 31156-31163.
278. Al Mamun, A. A. (2007) Elevated expression of DNA polymerase II increases spontaneous mutagenesis in *Escherichia coli*. *Mutation research*. **625**, 29-39.
279. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. **5**, 172.
280. Gross, J., and Englesberg, E. (1959) Determination of the order of mutational sites governing L-arabinose utilization in *Escherichia coli* B/r bv transduction with phage Plbt. *Virology*. **9**, 314-331.
281. Gao, S., Yao, S., Hart, D. J., and An, Y. F. (2017) Signal peptide-dependent protein translocation pathway is crucial for the sucrose sensitivity of SacB-expressing *Escherichia coli*. *Biochem Eng J*. **122**, 71-74.
282. Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E., and Hoch, J. A. (1983) Cloning structural gene sacB, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. *J Bacteriol*. **153**, 1424-1431.
283. Englesberg, E., Anderson, R. L., Weinberg, R., Lee, N., Hoffee, P., Huttenhauer, G., and Boyer, H. (1962) L-arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J Bacteriol*. **84**, 137-146.
284. Ackerman, R. S., Cozzarelli, N. R., and Epstein, W. (1974) Accumulation of toxic concentrations of methylglyoxal by wild-type *Escherichia coli* K-12. *J Bacteriol*. **119**, 357-362.
285. Koita, K., and Rao, C. V. (2012) Identification and analysis of the putative pentose sugar efflux transporters in *Escherichia coli*. *PloS one*. **7**, e43700.
286. Koch, A. L. (1984) Shrinkage of growing *Escherichia coli* cells by osmotic challenge. *J Bacteriol*. **159**, 919-924.
287. Purvis, J. E., Yomano, L. P., and Ingram, L. O. (2005) Enhanced trehalose production improves growth of *Escherichia coli* under osmotic stress. *Appl Environ Microbiol*. **71**, 3761-3769.
288. Leahy, J. J. (1968) Diauxie used as a microbiological assay for L-isoleucine or L-valine in *Escherichia coli*. *J Bacteriol*. **95**, 1964-1965.
289. Mostovenko, E., Deelder, A. M., and Palmblad, M. (2011) Protein expression dynamics during *Escherichia coli* glucose-lactose diauxie. *BMC microbiology*. **11**, 126.
290. Horazdovsky, B. F., and Hogg, R. W. (1989) Genetic reconstitution of the high-affinity L-arabinose transport system. *J Bacteriol*. **171**, 3053-3059.
291. Meijnen, J. P., de Winde, J. H., and Ruijssenaars, H. J. (2008) Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose. *Appl Environ Microbiol*. **74**, 5031-5037.
292. Sapunova, L. I., Lobanok, A. G., Kazakevich, I. O., Shliakhotko, E. A., and Evtushenkov, A. N. (2006) Biosynthetic features and properties of xylose isomerases from *Arthrobacter nicotianae*, *Escherichia coli*, and *Erwinia carotovota* subsp. *atroseptica*. *Prikladnaia biokhimiia i mikrobiologiia*. **42**, 279-284.
293. Afroz, T., Biliouris, K., Boykin, K. E., Kaznessis, Y., and Beisel, C. L. (2015) Trade-offs in engineering sugar utilization pathways for titratable control. *ACS Synth Biol*. **4**, 141-149.
294. Vuoristo, K. S., Mars, A. E., Sangra, J. V., Springer, J., Eggink, G., Sanders, J. P., and Weusthuis, R. A. (2015) Metabolic engineering of itaconate production in *Escherichia coli*. *Appl Microbiol Biotechnol*. **99**, 221-228.
295. Younger, A. K., Dalvie, N. C., Rottinghaus, A. G., and Leonard, J. N. (2017) Engineering modular biosensors to confer metabolite-responsive regulation of transcription. *ACS Synth Biol*. **6**, 311-325.
296. Libis, V., Delépine, B., and Faulon, J. L. (2016) Expanding biosensing abilities through computer-aided design of metabolic pathways. *ACS Synth Biol*. **5**, 1076-1085.
297. Domin, G., Findeiss, S., Wachsmuth, M., Will, S., Stadler, P. F., and Mörl, M. (2017) Applicability of a computational design approach for synthetic riboswitches. *Nucleic Acids Res*. **45**, 4108-4119.
298. Etzel, M., and Mörl, M. (2017) Synthetic riboswitches: from plug and pray toward plug and play. *Biochemistry*. **56**, 1181-1198.
299. Autour, A., and Ryckelynck, M. (2017) Ultrahigh-throughput improvement and discovery of enzymes using droplet-based microfluidic screening. *Micromachines*. **8**, 128.

300. Kwon, K. K., Lee, D. H., Kim, S. J., Choi, S. L., Rha, E., Yeom, S. J., Subhadra, B., Lee, J., Jeong, K. J., and Lee, S. G. (2018) Evolution of enzymes with new specificity by high-throughput screening using DmpR-based genetic circuits and multiple flow cytometry rounds. *Sci Rep*. **8**, 2659.
301. Liu, Y., Li, Q., Zheng, P., Zhang, Z., Liu, Y., Sun, C., Cao, G., Zhou, W., Wang, X., Zhang, D., Zhang, T., Sun, J., and Ma, Y. (2015) Developing a high-throughput screening method for threonine overproduction based on an artificial promoter. *Microb Cell Fact*. **14**, 121.
302. Bunzel, H. A., Garrabou, X., Pott, M., and Hilvert, D. (2018) Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. *Curr Opin Struct Biol*. **48**, 149-156.
303. Liu, S. D., Wu, Y. N., Wang, T. M., Zhang, C., and Xing, X. H. (2017) Maltose utilization as a novel selection strategy for continuous evolution of microbes with enhanced metabolite production. *ACS Synth Biol*. **6**, 2326-2338.
304. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M., and Sommer, M. O. A. (2018) Synthetic addiction extends the productive life time of engineered *Escherichia coli* populations. *Proc Natl Acad Sci U S A*. **115**, 2347-2352.
305. Lin, J. L., Wagner, J. M., and Alper, H. S. (2017) Enabling tools for high-throughput detection of metabolites: Metabolic engineering and directed evolution applications. *Biotechnol Adv*. **35**, 950-970.
306. Cheng, F., Tang, X. L., and Kardashliev, T. (2018) Transcription factor-based biosensors in high-throughput screening: advances and applications. *Biotechnol J*, e1700648.
307. Wong, Q. N., Ng, V. C., Lin, M. C., Kung, H. F., Chan, D., and Huang, J. D. (2005) Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic Acids Res*. **33**, e59.
308. DeVito, J. A. (2008) Recombineering with *tolC* as a selectable/counter-selectable marker: remodeling the rRNA operons of *Escherichia coli*. *Nucleic Acids Res*. **36**, e4.
309. Yagi, K. (2007) Applications of whole-cell bacterial sensors in biotechnology and environmental science. *Appl Microbiol Biotechnol*. **73**, 1251-1258.
310. Vollmer, A. C., Belkin, S., Smulski, D. R., Van Dyk, T. K., and LaRossa, R. A. (1997) Detection of DNA damage by use of *Escherichia coli* carrying *recA'*::*lux*, *uvrA'*::*lux*, or *alkA'*::*lux* reporter plasmids. *Appl Environ Microbiol*. **63**, 2566-2571.
311. Vollmer, A. C., and Van Dyk, T. K. (2004) Stress responsive bacteria: biosensors as environmental monitors. *Adv Microb Physiol*. **49**, 131-174.
312. Koga, K., Harada, T., Shimizu, H., and Tanaka, K. (2005) Bacterial luciferase activity and the intracellular redox pool in *Escherichia coli*. *Mol Genet Genomics*. **274**, 180-188.
313. Gibson, D. G., Smith, H. O., Hutchison, C. A., 3rd, Venter, J. C., and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat Methods*. **7**, 901-903.
314. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. **6**, 343-345.
315. Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N.*, et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*. **329**, 52-56.
316. Finnigan, G. C., and Thorner, J. (2015) Complex *in vivo* ligation using homologous recombination and high-efficiency plasmid rescue from *Saccharomyces cerevisiae*. *Bio Protoc*. **5**, e1521.
317. Kuijpers, N. G., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J. T., Daran, J. M., and Daran-Lapujade, P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb Cell Fact*. **12**, 47.
318. Song, C. W., and Lee, S. Y. (2013) Rapid one-step inactivation of single or multiple genes in *Escherichia coli*. *Biotechnol J*. **8**, 776-784.
319. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol*. **81**, 2506-2514.
320. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990) Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng*. **35**, 668-681.
321. Bervoets, I., and Charlier, D. (2018) A novel and versatile dual fluorescent reporter tool for the study of gene expression and regulation in multi- and single copy number. *Gene*. **642**, 474-482.
322. Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res*. **42**, 11383-11392.

323. Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., and Koffas, M. A. G. (2017) Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth Biol*. **6**, 710-720.
324. Kasey, C. M., Zerrad, M., Li, Y., Cropp, T. A., and Williams, G. J. (2018) Development of transcription factor-based designer macrolide biosensors for metabolic engineering and synthetic biology. *ACS Synth Biol*. **7**, 227-239.
325. Wang, B., Barahona, M., and Buck, M. (2015) Amplification of small molecule-inducible gene expression via tuning of intracellular receptor densities. *Nucleic Acids Res*. **43**, 1955-1964.
326. Chen, X. F., Xia, X. X., Lee, S. Y., and Qian, Z. G. (2018) Engineering tunable biosensors for monitoring putrescine in *Escherichia coli*. *Biotechnol Bioeng*. **115**, 1014-1027.
327. Chong, H., and Ching, C. B. (2016) Development of colorimetric-based whole-cell biosensor for organophosphorus cCompounds by engineering transcription regulator DmpR. *ACS Synth Biol*. **5**, 1290-1298.
328. Borkowski, O., Ceroni, F., Stan, G. B., and Ellis, T. (2016) Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr Opin Microbiol*. **33**, 123-130.
329. Mannan, A. A., Liu, D., Zhang, F., and Oyarzún, D. A. (2017) Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synth Biol*. **6**, 1851-1859.
330. Jha, R. K., Bingen, J. M., Johnson, C. W., Kern, T. L., Khanna, P., Trettel, D. S., Strauss, C. E. M., Beckham, G. T., and Dale, T. (2018) A protocatechuate biosensor for *Pseudomonas putida* KT2440 via promoter and protein evolution. *Metab Eng Commun*. **6**, 33-38.
331. Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., Siedler, S., Borodina, I., Jensen, M. K., and Keasling, J. D. (2016) Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat Chem Biol*. **12**, 951-958.
332. Snoek, T., Romero-Suarez, D., Zhang, J., Ambri, F., Skjoedt, M. L., Sudarsan, S., Jensen, M. K., and Keasling, J. D. (2018) An orthogonal and pH-tunable sensor-selector for muconic acid biosynthesis in yeast. *ACS Synth Biol*. **7**, 995-1003.
333. Ambri, F., Snoek, T., Skjoedt, M. L., Jensen, M. K., and Keasling, J. D. (2018) Design, engineering, and characterization of prokaryotic ligand-binding transcriptional activators as biosensors in yeast. *Methods Mol Biol*. **1671**, 269-290.

About the author

Teunke van Rossum was born on August 30th 1986 in Leiden, The Netherlands. In 2005, she started her studies Life Science & Technology, a joined program between Leiden University and Delft University of Technology, from which she obtained her BSc (2008) and MSc (2011) degrees cum laude. She performed her BSc and major MSc research projects in the group of Molecular Biotechnology, Leiden University, under supervision of Prof. Gilles van Wezel, and partly at the Department of Microbial Sciences, University of Surrey, United Kingdom, under supervision of Prof. Colin Smith, Dr. Giselda Bucca and Dr. Emma Laing. During these projects, she aimed to unravel the function of a two-component regulatory system in the filamentous bacterium *Streptomyces coelicolor* and discovered that this system is involved in controlling the development of this bacterium. In her minor MSc research project, she characterized a β-galactosidase at the Genencor division of Danisco (currently DuPont), Denmark, under supervision of Dr. Charlotte Poulsen, Dr. Morten Krog Larsen and Masoud Zargahi. After her studies, she continued her research at the group of Molecular Biotechnology, Leiden University, and discovered that the previously studied two-component system of *S. coelicolor* has resemblance with the dormancy regulatory system of *Mycobacterium tuberculosis*.

In September 2011, Teunke started her PhD project in the Laboratory of Microbiology at Wageningen University & Research, under supervision of Prof. John van der Oost and Dr. Servé Kengen. This project was part of the EU-project Hotzyme. Her work aimed to develop a transcriptional regulator-based bioreporter, as a generic selection method for novel enzymes, to enhance the number of available biocatalysts for the sustainable production of chemicals. To this purpose, she used molecular biology techniques to build the bacterium *Escherichia coli* into such a bioreporter, which she subsequently used to enrich cells with L-arabinose isomerase activity over a million fold, as proof of principle. In collaboration with Prof. Birte Höcker (Universität Bayreuth, Germany) and Dr. Andre Stiel (Helmholtz Zentrum München, Germany), she used a rational design to modify the ligand specificity of the transcriptional regulator.

Since February 2018, she is working as Science manager in Dr. Stan Brouns' lab at Delft University of Technology, where she supports the research and education on the interaction between bacteriophages and microbes.

List of publications

van Rossum, T., van der Oost, J., and Kengen, S.W.M. Inhibitory and stimulatory effects of L-arabinose on growth of *Escherichia coli* BW25113. *Manuscript in preparation*.

Creutzburg, S.C.A., **van Rossum, T.**, S.W.M., and van der Oost, J. *In vivo* selection of riboswitches with an altered specificity. *Manuscript in preparation*.

Rossum, T., Stiel, A.C., Röttjers, L., Makindji, F., Höcker, B., van der Oost, J., and Kengen, S.W.M. Engineering the ligand specificity of the transcriptional regulator AraC and enrichment of desired variants by combined selection and screening. *Manuscript in preparation*.

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., **van Rossum, T.**, Siebers, B., Menzel, P., Kublanov, I.V., Rike, A.G., Skretas, G., Szabo, Z., Peng, X., and Young, M.J. (2018) Discovery of novel hydrolases from hot environments. *Submitted to Biotechnology advances*.

van Rossum, T., Muras, A., Baur, M.J.J., Creutzburg, S.C.A., van der Oost, J., and Kengen, S.W.M. (2017) A growth- and bioluminescence-based bioreporter for the *in vivo* detection of novel biocatalysts. *Microb Biotechnol* 10, 625–641.

Urem, M.*, **van Rossum, T.***, Bucca, B., Moolenaar, G.F., Laing, E., Świątek-Połatyńska, M.A., Willemse, J., Tenconi, E., Rigali, S., Goosen, N., Smith, C.P., and van Wezel, G.P. (2016) OsdR of *Streptomyces coelicolor* and the dormancy regulator DevR of *Mycobacterium tuberculosis* control overlapping regulons. *mSystems* 1: e00014-16.

van Rossum, T., Kengen, S.W., and van der Oost, J. (2013) Reporter-based screening and selection of enzymes. *FEBS J.* 280: 2979-96.

*Contributed equally

Overview of completed training activities

Discipline-specific activities

Meetings & conferences

- Microbiology Centennial Symposium (MIB), Wageningen (NL), 2017*
- Dutch Biotechnology Conference (NBV), Wageningen (NL), 2017**
- Fall meeting (KNVM), Delft (NL), 2015*,***, poster prize
- Annual project meeting (Hotzyme), Copenhagen (DK), 2015**
- Conference: Catalytic mechanisms by biological systems (EMBO), Manchester (UK), 2014*
- Annual project meeting (Hotzyme), Exeter (UK), 2014**
- CW Study group meeting (NWO), Veldhoven (NL), 2013*
- Wbox 2 mini-symposium: redox biocatalysis (NBV), Wageningen (NL), 2013
- Annual Molecular Genetics meeting (NWO-ALW), Lunteren (NL), 2013*
- Seminar by David Baker (NVBMB), Utrecht (NL), 2013
- CW Study group meeting (NWO), Veldhoven (NL), 2012*
- Conference: Catalytic mechanisms by biological systems (EMBO), Groningen (NL), 2012*
- Annual project meeting (Hotzyme), Wageningen (NL), 2012**
- Dutch Biotechnology Conference (NBV), Ede (NL), 2012
- Symposium: Microbes for sustainability (SENSE), Wageningen (NL), 2012
- CHAINS (NWO), Maarssen (NL), 2011
- Annual Molecular Genetics meeting (NWO-ALW), Lunteren (NL), 2011

*poster presentation, **oral presentation, ***pitch

Courses

- Autumn school "Biomolecular Structure and Function - Computational Approaches" (UDE BIOME core "Computational Biomedicine", Radboudumc CMBI), Essen (DE) and Nijmegen (NL), 2012
- Food and biorefinery enzymology course (VLAG), Wageningen (NL), 2011

General courses

- Career orientation (WGS), Wageningen (NL), 2015
- Scientific writing (WGS), Wageningen (NL), 2014
- Reviewing a scientific paper (WGS), Wageningen (NL), 2013
- Mini-symposium: How to write a world-class paper (WUR), Wageningen (NL), 2013
- Introduction to qPCR (Philippe Guilla Puylaert), Wageningen (NL), 2013
- Project and time management (WGS), Wageningen (NL), 2013
- Competence assessment (WGS), Wageningen (NL), 2012
- VLAG introduction week (VLAG), Baarlo (NL), 2012
- Introduction to R (Sebastian Tims), Wageningen (NL), 2012

Optionals

- PhD trip (MIB, SSB), California (US), 2015
- Metabolic engineering tools meetings (BacGen, BPE), Wageningen (NL), 2014-2015
- VLAG project mini-symposium (including organization; VLAG, BacGen), Wageningen (NL), 2013
- PhD trip (MIB, SSB), Boston (US) and Toronto (CA) areas, 2013
- Organizing PhD trip (MIB, SSB), Boston (US) and Toronto (CA) areas, 2012-2013
- Seminars at WUR (MIB, SSB, WEES, VLAG, etc.), Wageningen (NL), 2011-2017
- BacGen group meetings (BacGen), Wageningen (NL), 2011-2017
- PhD meetings (MIB), Wageningen (NL), 2011-2015
- Preparation of PhD research proposal, Wageningen (NL), 2011

Acknowledgements

It is finally there: my thesis. The route towards this end goal has not been easy. However, I am still behind my choice to start this project one hundred percent, because I learned a lot both scientifically and personally and I got to know many very interesting and great persons. Although the major task of executing the experiments and writing this thesis was on my shoulders, I could not have done this without the help and support of numerous other persons. Therefore, I would like to take the opportunity here to thank all people involved in realizing my thesis.

John, thank you for giving me the chance to do this project in the great group that you made of BacGen. I very much enjoyed my time working in this inspiring and dynamic group. Thank you also for your everlasting enthusiasm and optimism. You always saw some new opportunities, even in very negative results, and no idea was ever out of reach. Thank you for always being there for me and making time whenever I needed your help. I really appreciate your directness and this helped me many times in deciding which way to go.

Servé, thank you for being there for me during every step of the way, during the hard parts and also the fun parts. I really appreciated that your door was always open whether for a scientific question or just a chat. You gave me much freedom, but you also showed a lot of interest in my progress and my well-being. You often told me I was being too much of a perfectionist, which I really valued, although I was not always listening;) In short, I really enjoyed having you as my supervisor, also when strolling together through Exeter and ending up in an expensive restaurant in Copenhagen.

Many thanks to all my collaborators. Sjoerd, thank you for all your help in my experimental work and for all the discussions. I learned a lot from your detailed knowledge in many areas and it was nice to share ideas on our similar projects. Hotzyme members, it was a great experience to be part of the Hotzyme team. Although the projects of Sjoerd and me were the innovative part of this European project and therefore the collaboration with other partners was not so extensive, discussing my project in such a large consortium gave extra critical mass and helped me making the context of enzyme research more concrete. Thank you all for your input and discussions and of course for all the work on the review. I hope the future will bring more collaborations. Birte and André, thank you very much for all your work on the design of AraC (and LacI). I really appreciate your critical views, your openness and especially your patience. In my view, the contact was always

straightforward and easy going. I enjoyed working with you very much and hope this will not end here.

Prarthana, thank you for being my paranymph. I really appreciate all the support in many different forms that you gave for preparing my defence, but most of all during all these years leading up to this event. With you I could always discuss all the things that were on my mind, either small or big, work related or personal, difficult or fun stuff. I really enjoyed our lunches and afternoon walks, which were a good break in a busy day. It was great hanging out with you on many occasions and sharing an office and it was fun that often just a few words or a look were enough to understand one another. I am positive we will have much more of these moments in the future.

Alex, thank you for being my paranymph. I very much enjoyed sharing these final steps towards our defences together. This really helped keeping motivated, but of course I would like to thank you as well for all the support and fun in the years before this. You were always there for me when needed, either for a chat, some advice, a coffee break or very importantly showing me how to be a bit more flexible and how to put things in perspective, although there are some more steps to take in the latter two;) I very much enjoyed sharing an office, going to TRX together for many years and hanging out during various social activities. I hope we will have many more of these nice activities together.

To all my office mates, thank you for all the discussions, support and most importantly for all the fun. Melvin, thank you for showing me the way when I first arrived at Microbiology, for going through our PhDs together, for your trust in me and for the nice time with you, Sandra, Rene and Robin. John R., thanks for all the fun word jokes, which somehow got lost to the rest of the office and which prepared me for many more jokes over the years. Also thank you and Eline for all the support and dinners/bbqs with you and the kids. Biniam, thank you for the international vibe in our office and for showing me how to cook Ethiopian. Yifan, thank you for being there for me and for the nice statue from China, which is still on my desk, but now in Delft. Daria, Hikmah and Shreyans, thank you for all the nice chats. Nico, thank you for the many many many chats, your great advice, our deal for walking the stairs and looking out for me, even from Potsdam. Hanne, thank you for making the office so lively and for the great chats. Tim, thank you for all your input in my manuscript figures and for your great humor. Lucia, thank you for your optimistic attitude and the nice discussions. Anna, thank you for your great interest and lively chats. Tijs, thank you for being there for me and for the nice discussions. Carrie, thank you for creating a homey atmosphere in the office, the great dinners, afternoon coffee breaks, sporting together, and

for looking out for me. Joep, Belén, Janneke, Ismael, thank you for all the nice chats.

Many thanks to all the BSc and MSc students who helped me in my project. Marco, Aleksandra, Lisa and Fadi, thank you for all your lab work, discussions and other input. Aleksandra, thank you for all the nice emails over the last couple of years.

Thank you I-scream group, a.k.a. Nico, Emmy, Prarthana, Alex, Dorett, Carrie, Wen, Paul, Ruben, Joyshree and Mark, for all the fun pub quizzes, dinners, games and other activities. Dorett, thank you for all the joyful zumba classes and for the true understanding of how I think and feel. Wen, thank you for being so lively and contagiously cheerful. Joyshree, thank you for all the great lunches and chats. Emmy, thank you inspiring me to be more sustainable and for all the support and nice chats.

Mamou, thank you for all your support, but most of all thank you for the Friday afternoons speaking French and Dutch. This was a great diversion of my work and I very much look forward to continue with this.

Many thanks to the great group of BacGen colleagues, the ones mentioned earlier above and Tijn, Elleke, Tom, Richard, Bram, Vincent, Edze, Pierpaolo, Matthijs, Sjon, Tessa, Peter, Daan, Bas, Raymond, Franklin, Becca, Jochem, Patrick, Stan, Jorrit, Lione, Marnix, Ioannis, Jeroen, Reno, Mihris and Costas. Thank you for all the input, discussions but above all thank you for the support and all the fun. Tijn and Richard, thank you for your help with *Geobacillus*.

Willem, thank you for your input in my project and for the great and stimulating working environment that Microbiology is. Anja, Wim, Tom vdW, Carolien, Sjon, Philippe, Merlijn, Ton, Tom S, Monika, Steven, and all other technicians and support staff, thank you for all your technical support. Anja, thank you for all your help in many areas, but especially thank you for all the nice chats and mental support. Wim, thank you for all your help, the nice chats and for all the delicious rhubarb. Tom vdW, thank you and Juliëtte for your support, your help and all the fun.

Many thanks to all the people involved in the PhD trips. Vicente, Floor, Corina, Noora, Javier and Bas, thank you for all the fun and hard work in organizing the PhD trip to the east coast of the US and Canada. For me, it was an unforgettable experience. Floor, thank you for all your support, nice chats and the fun stuff we did together. Also thanks to all the participants who made the trip into two great weeks. Alex, Nico, Ruben, Yue, Anna, Maarten, Jasper and Kees, thank you for organizing a wonderful PhD trip to California and all the participants, thank you for two very nice weeks. Alex, Nico, Tim, Marnix and Yifan, I had a lot of fun with you in the car and hanging out on various beaches and other locations.

Janneke, Detmer, Thomas, Romy, Marcelle, Nam, Tom S, Tom vdB, Nicolas, Gerben and Caroline, thank you for the support and the nice chats over lunch or during other non-work related activities. Also many thanks to all other colleagues whom I met over the years at Microbiology, SSB, BPE and FBR.

Stan, thank you for your input in my project and for seeing my strengths, which resulted in you hiring me as Science manager for your group. I very much enjoy this position. To all my new colleagues in Delft, Franklin, Rita, Anna, Patrick, Sebastian, Cristian, Tobal, Becca, Jochem, Boris, thank you for your warm welcome at BN and for your support during the last stages of my thesis.

Gilles, thank you for your contagious enthusiasm for science and supervising me during my first steps in the lab and many more steps after that. Without the great experience in your lab I would probably not have considered to do a PhD project. I very much hope our paths will cross frequently in the future.

To all my friends, thank you for all your support during my project and for your patience when I did not have much time. It felt good to share my ups and downs with you. Magda, thank you for being an example for me in doing a PhD and for the numerous chats and cooking tips over the years. Nancy and Alexandra, thank you for all the relaxing museum trips and all the lunches/diners which fitted all our tastes so well. Katja, thank you for the warm and homey feeling you always give me. Shokoufeh, thank you for lending me a sympathetic ear whenever needed. Mia, thank you for all our discussions about personal well-being and many other things and of course for all our joined work on OsdR (SCO0204). Esmée, thank you for being there for me almost as long as I remember.

Al mijn familie, Vlasbloms, van Rossums, Mulders, Levissons en Leenmannen, heel erg bedankt voor al jullie support en interesse gedurende de vele jaren die mijn PhD heeft geduurd. Pap en mam, bedankt voor al het vertrouwen en er gewoon voor me zijn als dat nodig was. Nelleke, Rick, Anke en Dennis, bedankt voor alle mentale support en de afleiding met leuke dingen. Elly en Maus, bedankt voor alle begrip. Karen, Marc, Jens en Jurre, bedankt voor de aangename afleiding. Cora, bedankt voor al je lieve mails en gesprekken en het meedenken in mijn persoonlijke ontwikkeling en mijn carrière.

Mark, zonder jou had ik dit niet kunnen doen. Heel erg bedankt voor alle hulp, zowel inhoudelijk als mentaal. Het was heel prettig om mijn ideeën met je te kunnen bespreken en het heeft mij heel erg geholpen dat na een drukke dag jij de dingen die in mijn gedachten bleven hangen kon relativeren. Ook heel erg bedankt voor alle geduld met mij in de vele momenten of periodes wanneer ik niet goed in mijn vel zat. Gelukkig heeft je humor me op

veel momenten uit een slechte stemming gehaald. Het is fijn dat je er al die tijd voor me was, bijna vanaf het begin van mijn PhD;) En jou ontmoeten is zeker het leukste dat me is overkomen tijdens deze periode.

About the cover

The cover gives an artistic impression of the bioreporter developed during my thesis. The tree bark motif as background of the cover represents nature as source of novel enzymes for the sustainable production of chemicals. On the backside of the cover, the transcriptional regulator AraC is shown, bound to and looping a DNA helix. In this situation, no target enzyme is present and active and therefore the reporter genes are repressed. On the spine of the book, the Pac-man represents the target enzyme, which converts the substrate to the product. The latter is presented as star. On the front of the cover, three agar plates with bacteria spotted in a raster are shown. From one plate, light originates. This figure represent the selection based on growth and the screening based on bioluminescence in the situation when the target enzyme is present and active. The enzymatic product is bound by AraC, which switches conformation, resulting in transcription of the two reporter genes, one for selection and one for screening. Only when the target enzyme is present and active (the right plate) the bacterial cells can survive the selective pressure and give light, whereas without this enzymatic activity only false positive cells can survive (left bottom plate). The left top plate is a control plate without selective pressure on which all bacterial cells can grow, but without the production of light.

The research described in this thesis was financially supported by the Graduate School VLAG and by the Hotzyme project as part of the 7th framework program for research and technological development (FP7) of the European Union (GA: 265933). Authors BH and ACS (Chapter 5) were supported by DFG grant HO 4022/2-3.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by: Dennis Hendriks, ProefschriftMaken || www.proefschriftmaken.nl
Printed by: ProefschriftMaken || www.proefschriftmaken.nl on recycled, FSC-certified paper

Propositions

1. The power of evolution is a double edged sword.
   (this thesis)

2. Once enzymes can be easily engineered by computational design, transcriptional regulator-based bioreporters will still be useful in obtaining novel biocatalysts.
   (this thesis)

3. In science, serendipity is lost without perceptiveness.

4. Communication is the most important, but also the most challenging element of performing research.

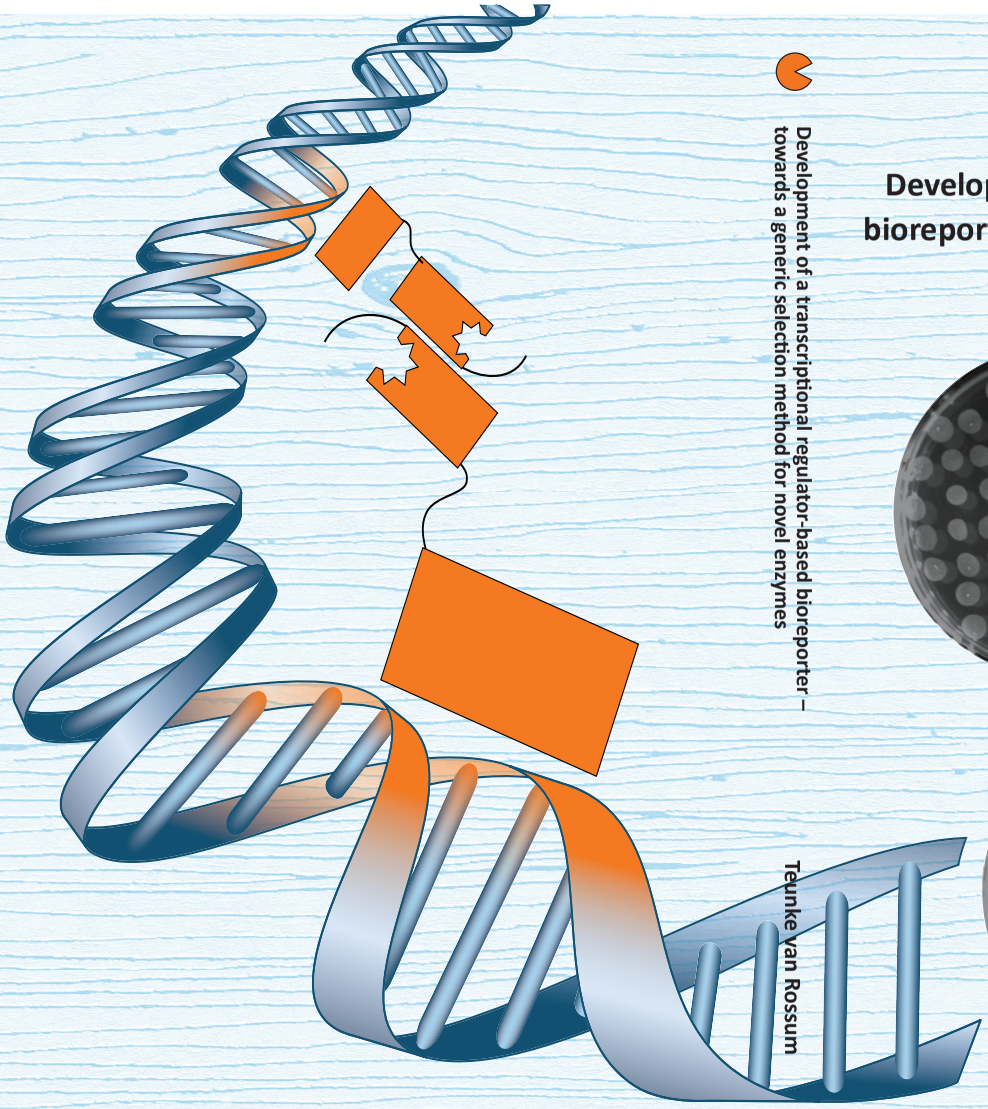
5. Until we completely comprehend the intricate network that makes a prokaryotic cell, the full molecular understanding of more complex life forms is out of reach.

6. Phage therapy is being wrongfully neglected in the Dutch health care system until now.

7. Nature is our best teacher towards a biobased economy.

Propositions belonging to the thesis entitled
'Development of a transcriptional regulator-based
bioreporter - towards a generic selection method for
novel enzymes'

Teunke van Rossum
Wageningen, 14 September 2018

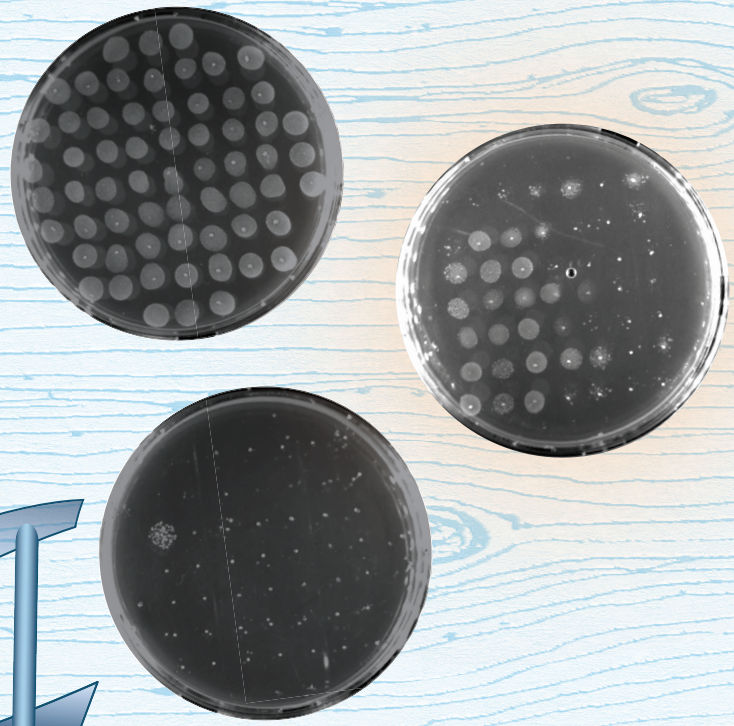


Development of a transcriptional regulator-based bioreporter – towards a generic selection method for novel enzymes

Teunke van Rossum

2018

Development of a transcriptional regulator-based bioreporter - towards a generic selection method for novel enzymes



Teunke van Rossum

Invitation

for the public defence
of my PhD thesis entitled

'Development of a transcriptional regulator-based bioreporter - towards a generic selection method for novel enzymes'

Friday 14th of September 2018
at 11.00 a.m. in the Aula of
Wageningen University & Research
Generaal Foulkesweg 1,
Wageningen

Teunke van Rossum
teunkevanrossum@gmail.com

Paranymphs

Prarthana Mohanraju
prarthana.mohanraju@wur.nl

Aleksander J. Kruis
alex.kruis@wur.nl