



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2020

Changing the ligand-binding specificity of E. coli periplasmic binding protein RbsB by rational design and screening

Tavares Diogo

Tavares Diogo, 2020, Changing the ligand-binding specificity of E. coli periplasmic binding protein RbsB by rational design and screening

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_F6204BD647734

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de Microbiologie Fondamentale

**Changing the ligand-binding specificity of *E. coli*
periplasmic binding protein RbsB by rational design
and screening**

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Diogo TAVARES

Master en Microbiologie de l'Université de Lisbonne

Jury

Prof. Edward Elliston FARMER, Président
Prof. Jan Roelof VAN DER MEER, Directeur de thèse
Prof. Sven PANKE, Expert
Prof. Bruno CORREIA, Expert

Lausanne
2020



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof. Edward Elliston	Farmer
Directeur·trice de thèse	Monsieur	Prof. Jan Roelof	van der Meer
Expert·e·s	Monsieur	Prof. Sven	Panke
	Monsieur	Prof. Bruno	Correia

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Diogo Alexandre Martins Aires Tavares

Master, Université de Lisbonne, Portugal

intitulée

**Changing the ligand-binding specificity
of *E. coli* periplasmic binding protein RbsB
by rational design and screening**

Date de l'examen : 22 janvier 2021

Date d'émission de l'imprimatur : Lausanne, le 17 février 2021

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Table of Content

Summary (English)	i
Résumé (Français)	iii
CHAPTER 1	1
General Introduction	
Aims of the thesis	25
CHAPTER 2	27
Computational redesign of the <i>Escherichia coli</i> ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol	
CHAPTER 3	63
Random and semi-random mutagenesis to improve and understand the binding mechanism of ribose-binding protein mutants towards a non-natural ligand	
CHAPTER 4	99
An <i>in vivo</i> system to follow wild-type and mutant ribose-binding protein expression and localization in <i>Escherichia coli</i>	
CHAPTER 5	139
General Discussion	
CURRICULUM VITAE and Publication List	149
Acknowledgements	153

Summary

Periplasmic binding proteins (PBPs) form a superfamily of bacterial proteins with a conserved bilobal structure, which are involved in substrate scavenging for bacterial cells. A wide variety of natural ligand-binding domains has evolved. PBPs are composed of two domains connected by a hinge region, which form a binding pocket between the two domains. They can be found in two stable conformations; in absence of ligand the PBP adopts an open conformation, where the binding pocket is exposed. In presence of the ligand, the protein changes to the closed conformation where the ligand is buried in the middle of the protein. This project focused on the ribose-binding protein of *Escherichia coli* (RbsB). Ribose binding to RbsB stabilizes the closed state. RbsB-bound ribose is presented to a cytoplasmic transport channels (RbsAC), from where it is imported into the cell, or interacts to membrane receptors (i.e., Trg) and can elicit a chemotactic signal.

Due to their unique ligand-binding characteristics and wide variety of natural binding pockets PBPs have been of interest for the development of biosensors and bioreporter systems. PBP bioreporters were initiated over 20 years ago by a development in the group of Hazelbauer, who fused the C-terminal part of the *E. coli* EnvZ osmoregulation histidine kinase to the N-terminal part of the Trg methyl-accepting chemotaxis receptor protein, creating a hybrid receptor Trz1. Ligand bound galactose-binding protein (GBP) and ribose-bound RbsB interact with Trz1, which eventually leads to phosphorylation of the response regulator OmpR, activating transcription from the *ompC* promoter (and any reporter gene fused to this). In 2003, Hellinga's group proposed that based on crystal structure information of ligand-bound PBPs variants with new ligand recognition specificities could be designed by computational approaches. Notably, they claimed the design of a RbsB-variant with nM affinity for recognition of 2,4,6-trinitrotoluene (TNT). This idea inspired the scientific community, because it could easily extend PBP-binding to a tremendous variety of compounds, including non-natural molecules, and would thus permit a wide variety of biosensor and bioreporter systems based on RbsB/GBP and Trz1. Unfortunately, independent engineering of some of the most promising published mutants failed to reproduce the reported *in vivo* and *in vitro* results. These studies further concluded that the published variants were actually misfolded proteins and/or impaired in stability as result of the introduced ligand-pocket mutations. This fact was largely ignored by Hellinga's publications.

Still inspired by the concept and trying to understand the reason of such limited success, our group raised the hypothesis that changing from ribose to TNT in a single step was likely unfeasible, but – given the wide range of naturally evolved PBP ligand binding pockets, a step by step change of ribose

binding to a non-natural analogue should be possible. To test this, we selected compounds with distinct differences but still chemically similar to ribose: 1,3-cyclohexanediol (13CHD) and cyclohexanol (CH). Mutant ligand binding pockets that might accommodate 13CHD and/or CH were computationally simulated and calculated using Rosetta, from which a list of critical amino acid residues to mutate in RbsB was selected. These were then synthesized and cloned into *E. coli*; a resulting set of 2 million mutants containing one of five possible substitutions at each of 9 selected critical amino acid positions. The library was introduced into an *E. coli* bioreporter strain, which carries the Trz1 hybrid signaling pathway coupled to GFP production when the (new) ligand would bind the (mutant) RbsB. The main goals of this work were to screen and characterize mutants from this first library, and potentially improve mutants for the new ligand binding in further rounds of mutagenesis.

In the first part of this work a precise and user-friendly high-throughput strategy to screen the mutant library was developed. Clones were grown as individual microcolonies in alginate beads, to reduce single cell GFP expression variability, which were screened by fluorescence activated cell sorting (FACS) for gain-of-function GFP expression in presence of 13CHD. Six mutants with modest (1.5-fold) but consistent induction with 1 mM 13CHD were isolated. Moreover, these mutants completely lost the capacity to react to ribose. The RbsB mutants were characterized in terms of periplasmic space abundance, stability, secondary structure and ligand affinity. Isothermal microcalorimetry confirmed 13CHD binding, although only two mutants were sufficiently stable upon purification. Circular dichroism and quantification of periplasmic space abundance suggested the mutants to be prone to misfolding and/or defects in translocation.

In the second part of this work, we used random and semi-random mutagenesis to improve the affinity and/or stability of the six isolated mutants with 13CHD binding capacity. Several mutant libraries were produced and screened with the previous described strategy. Variants displaying higher expression levels of GFP in presence of 13CHD were collected by FACS, and were used as starting point for the next round of evolution. This mutagenesis and rigorous screening strategy allowed us to isolate 7 mutants with improved (3.2-fold) GFP induction in presence of 13CHD and in a concentration-dependent manner. Several variants were observed that displayed open and closed conformations simultaneously, suggesting they were impaired in transition dynamics. Moreover, our screening strategy largely ignores potential variants with improved binding and closed conformation stability, but that are unable to interact with Trz1 receptor (e.i., trigger the signaling cascade).

Finally in the third part of this work, we developed and tested an *in vivo* system to characterize the quality of the translocation process and receptor interactions. Wild-type- and mutant-RbsB

proteins were fused to mCherry reporter protein to study protein abundance and subcellular localization. Whereas RbsB-mCherry proteins clearly localized to the periplasmic space and centered in polar regions depending on chemoreceptor availability, mutant-RbsB-mCherry expression resulted in high proportions of cells devoid of clear foci and low proportions of cells with multiple fluorescent foci, suggesting poorer translocation and mislocalisation. In addition, polar foci of mutants were less fluorescent, suggesting poorer chemoreceptor binding. By spiking further derivative mutant libraries generated by error-prone PCR without or with different proportions of *E. coli* expressing wild-type RbsB-mCherry we could estimate the potential improvement and deterioration of mutants with wild-type-like periplasmic localisation. The *in vivo* translocation system may thus be used to detect mutants with better signal transduction capacity.

In conclusion, we firmly showed that design of PBP receptor proteins with new binding capacities for non-natural compounds is feasible, but still largely a matter of trial and error. The combination of computational simulations, random mutagenesis and rigorous screening allowed us to isolate variants with new recognition for 13CHD and loss of ribose binding. However, our results also showed that most predicted ligand-binding pocket mutations lead to poorly folding and functioning proteins, and it is likely that the dynamic transition needed between open and closed conformations of (here) RbsB is insufficiently understood and currently predictable to allow rational expansion to a wide range of new ligands.

Résumé

Les protéines de liaison péripasmiques (PLP) constituent une superfamille de protéines bactériennes avec une structure bilobée. Elles sont impliquées dans la captation de substrats pour les cellules bactériennes, et montrent grande diversité de domaines de liaison à des composés naturels. Les PLP sont composées de deux domaines connectés par une région charnière, ce qui forme une poche de liaison au substrat entre les deux domaines. Les PLP montrent deux états stables : ouverte en l'absence de ligand, conformation dans laquelle la poche de liaison est exposée, et fermée quand le ligand est séquestré dans la poche de liaison. Ce projet a porté sur l'étude de la PLP RbsB liant le ribose chez *Escherichia coli*. La liaison du ribose stabilise l'état fermé de RbsB et permet l'interaction avec le transporteur cytoplasmique RbsAC et son passage dans le cytoplasme de la cellule, ou son interaction avec des récepteurs membranaires tels que Trg permettant en une réponse chimiotactique.

Étant données leurs caractéristiques uniques de liaison aux ligands et la grande variété de poches de liaison naturellement observée chez les PLP, elles présentent un grand intérêt pour le développement de biosenseurs et de systèmes biorapporteurs. Les premiers biorapporteurs basés sur des PLP ont été développés 20 ans auparavant par le groupe de Hazelbauer. Cette équipe a fusionné la partie C-terminale de la protéine kinase à histidine impliquée dans l'osmorégulation (EnvZ) et l'extrémité N-terminale du récepteur chimiotactique accepteur de groupement méthyle (Trg), pour créer le récepteur hybride Trz1. Les PLP liant le galactose (GBP) et le ribose (RbsB) interagissent avec Trz1, ce qui entraîne la phosphorylation du régulateur réponse OmpR qui lui-même va activer la transcription à partir du promoteur du gène *ompC* (ou n'importe quel système rapporteur placé en aval de ce promoteur). En 2003, le groupe de Hellinga proposait que, sur la base de la structure cristallographique de différents PLP liées à leur ligand, des variants reconnaissant de nouveaux ligands pourraient être générés sur la base d'une approche informatique. En particulier, cette équipe se targue d'avoir généré un variant de RbsB permettant de lier le 2,4,6-trinitrotoluène (TNT) avec une affinité de l'ordre du nanomolaire. Cette idée a inspiré la communauté scientifique car cette approche pourrait s'étendre à une diversité incroyable de composés naturels ou non, ce qui permettrait le développement de biosenseurs et biorapporteurs variés basés sur ce système. Malheureusement, la construction des mutants les plus prometteurs par des équipes indépendantes n'ont pas permis de rapporter de l'activité *in vivo* et/ou *in vitro*. Cela a été ignoré dans les publications du groupe Hellinga.

Inspirés par ce concept et voulant savoir quelles étaient les raisons de ce succès quelque peu limité, notre groupe a émis l'hypothèse que le changement de spécificité de RbsB du ribose au TNT en

une étape était probablement infaisable mais, étant donnée la grande diversité de poches de liaisons naturellement observées chez les LPL, un changement pas à pas du ribose vers un composé analogue non naturel devrait être possible. Pour tester cela, nous avons sélectionné des composés distincts du ribose mais présentant tout de même des similarités : 1,3-cyclohexanediol (13CHD) and cyclohexanol (CH). Des mutants qui pourraient accueillir le 13CHD et/ou CH ont été générés par simulation informatique en utilisant le programme Rosetta, lequel a fourni une liste d'acides aminés critiques à muter. Une librairie de mutant a été synthétisée, celle-ci contenant 2 millions de variants de RbsB avec 1 substitution parmi 5 possibles à 9 positions sélectionnées pour leur aspect critique dans la reconnaissance du substrat. La librairie a été introduite et criblée chez une souche reportrice d'*E. coli* contenant la chaîne de signalisation hybride Trz1 couplée à la production de la protéine fluorescente verte (GFP) lorsque le (nouveau) ligand se liera à la protéine RbsB (sauvage ou mutante). Le but principal de ce travail était de caractériser cette librairie de mutants, et éventuellement d'améliorer la capacité de ces mutants à lier un autre composant par des cycles de mutagénèses additionnels.

Dans la première partie de ce travail, une stratégie simple et efficace pour cribler la librairie de mutant a été développée. Les différents clones/variants ont été cultivés individuellement en microcolonies dans des billes d'alginate afin de réduire la variabilité du signal GFP observé au niveau de la cellule unique. Les billes ont été analysées par trieur de cellules reposant sur la fluorescence (FACS) afin de détecter des mutants présentant une activité GFP accrue en présence de 13CHD. Six mutants ont été isolés pour leur modeste mais significative induction (1,5 fois) en présence de 1 mM de 13CHD. De plus, ces mutants avaient totalement perdu leur capacité à réagir au ribose. Les mutants RbsB ont été caractérisés plus en détails pour leur localisation dans périplasme, leur stabilité, leur abondance et leur affinité pour le ligand. La technique de microcalorimétrie isotherme a confirmé que ces mutants lient le 13CHD, bien que seulement 2 de ces protéines mutantes se soient révélées suffisamment stables après purification. L'analyse par dichroïsme circulaire et la quantification de l'abondance des protéines dans l'espace périplasmique suggèrent que les protéines mutantes sont sujettes à un mauvais repliement et/ou un problème dans la translocation du cytoplasme au périplasme.

Dans une seconde partie, nous avons muté les six mutants isolés précédemment de façon aléatoire ou semi-aléatoire afin d'améliorer leur affinité pour le 13CHD et/ou leur stabilité. Plusieurs librairies de mutants ont été produites et analysées selon la méthode décrite plus tôt. Les variants montrant une plus forte expression du système rapporteur GFP en présence de 13CHD ont été isolés par FACS, et utilisés comme point de départ pour la prochaine étape d'évolution. Cette mutagénèse et

l'analyse rigoureuse des bibliothèques nous ont permis d'isoler 7 mutants avec une augmentation de 3,2 fois du signal GFP en présence de 13CHD, et d'une façon dose-dépendante. Plusieurs variants ont montré qu'ils adoptaient la conformation ouverte et fermées au sein de la population bactérienne. Cette dernière observation suggère que ces mutants sont affectés dans leur capacité à passer d'une conformation à l'autre. De plus, notre stratégie de criblage ne tient pas compte des variants qui montreraient une liaison accrue et une bonne stabilité de la conformation fermée, mais qui seraient incapables d'interagir avec le récepteur Trz1 (et donc de déclencher la cascade de signalisation du rapporteur).

Finalement, dans la troisième partie de ce travail, nous avons développé et testé un système *in vivo* permettant de caractériser la qualité du processus de translocation dans l'espace périplasmique et l'interaction avec les récepteurs. Les protéines RbsB sauvage et mutantes ont été fusionnées à la protéine fluorescente rouge mCherry afin de visualiser l'abondance et la localisation sub-cellulaire des protéines au niveau de la cellule unique en utilisant la microscopie à épifluorescence et le traitement des images obtenues. Alors que la protéine de fusion RbsB sauvage montre une localisation périplasmique centrée au niveau des pôles de la cellule dépendamment de la disponibilité des chimiorécepteurs, les fusions avec les variants de RbsB montraient une forte proportion de cellules dépourvues de foci, et une faible proportion de cellules avec de multiples foci, suggérant une plus faible liaison aux chimiorécepteurs. En analysant plus en détail des bibliothèques de mutants générées par PCR mutagène, en mélangeant ou non avec des cellules contenant la protéine de fusion RbsB sauvage, nous avons pu estimer l'amélioration potentielle ou la détérioration des qualités des mutants RbsB par rapport au sauvage en terme de localisation périplasmique. Ce système de translocation *in vivo* pourrait être utilisé afin de détecter des mutants permettant une meilleure transduction du signal.

En conclusion, nous avons montré que la conception de protéines réceptrices PLP présentant de nouvelles capacités de liaison pour des composés non naturels est bien faisable, mais repose encore sur une stratégie d'essais et erreurs. La combinaison de simulations informatiques, de mutagénèses aléatoires et de criblage rigoureux nous a permis d'isoler des variants de RbsB avec une capacité à reconnaître le 13CHD, tout en ne liant plus le ribose. Néanmoins, nos résultats ont également montré que la plupart des prédictions de mutations au niveau de la poche de liaison ont mené à un mauvais repliement ou fonctionnement des protéines. Il est très probable que la dynamique de transition entre la conformation ouverte et fermée (de RbsB pour cette étude) ne soit pas encore assez bien comprise, et donc actuellement non prédictible pour permettre le test d'une grande variété de nouveaux ligands.

CHAPTER 1

General Introduction

Engineering of sensory proteins with new ligand- binding capacities

Published as: **Diogo Tavares**, Vitali Maffenbeier, and Jan Roelof van der Meer (2019). "Engineering of sensory proteins with new ligand-binding capacities." [*Handbook of Cell Biosensors*](#).

Abstract

Biosensors based on whole cell living bacteria or their isolated components mostly use the capacity of natural protein classes to 'sense', in other words to interact with, chemical ligands. Frequently deployed sensor protein classes consist of transcription regulators, two-component sensory proteins, methyl-accepting chemoreceptors and periplasmic binding proteins. All classes have been linked to some form of transducer and sensor output upon binding of the ligand, such as by de novo synthesis of reporter proteins, through Föster resonance energy transfer, from cell accumulation from chemotaxis, or in dye-encapsulating competitive liposome assays. Despite the partially successful deployment of such biosensors, one of their current limitations is the covered spectrum of chemical ligands, which in most cases reflects the cognate ligand of the used sensory protein. Here we will summarize some of the past and recent efforts to obtain new ligand-binding specificities in these sensory protein classes. Most strategies have followed combinations of random mutagenesis, selection and screening methodologies, and protein-structure guided predictions. They have led to some success of obtaining new specificities, but which not unexpectedly lay relatively close to the original ligand. Computational structure-function predictions have become more important to reduce the mutagenesis efforts to smaller library sizes, but their success has been limited so far to those sensory proteins that do not undergo major conformational changes upon ligand-binding. In particular for the class of periplasmic binding proteins, new ligand-binding specificities have been notoriously difficult to predict, likely as a result of the current limitations in dynamic structure predictions. It is to be expected that future advances in computational algorithms will facilitate this process, which will make it more straightforward to obtain sensory proteins with targeted ligand-binding properties, enabling plug-and-play biosensor design.

Key words:

Bioreporters, Transcription Regulators, Periplasmic binding proteins, Methylaccepting chemotaxis proteins, Protein-structure guided predictions, Random mutagenesis, Selection, Screening

Introduction: What are bioreporters

Bioreporters are engineered living micro-organisms, which contain a synthetic genetic circuit that enables the cell to sense one or more input signals or conditions, execute the instructions imposed by the circuit and produce one or more outputs (Figure 1) ^[1-3]. Traditionally, outputs of bioreporters consist of *de novo* synthesized (reporter) protein(s) formed in response to the input signal, yielding some sort of proportional and quantifiable relation between input and output in the context of the assay ^[4, 5]. For ease of quantification, the choice of deployed reporter output has frequently fallen on light-emitting or color-producing enzymes or autofluorescent reporter proteins, whose (spectral) activity can be easily and/or non-invasively measured ^[4]. Bioreporter assays are relatively simple and consist in essence of an incubation of the bioreporter cells with a liquid sample for a defined duration, after which the reporter output is recorded ^[6, 7]. The ease of designing a bioreporter genetic circuit, the quantifiable aspect and the assay simplicity have attracted both scientific interest and popularity as potential alternative for environmental, medical or food analytics ^[2, 8-11]. Numerous bioreporters have thus been produced and tested with different applications in mind (Table 1). Some have been quite successful in terms of (correctly) quantifying inputs, although eventually very few have been tested sufficiently rigorously for robustness and on real samples ^[2, 8, 12, 13] (Table 1). Engineering and application aspects of bioreporters have been reviewed elsewhere and are not but briefly rehearsed here ^[6]. Instead, the main focus of the underlying review is the question of the sensing elements that are central for the bioreporter circuitry.

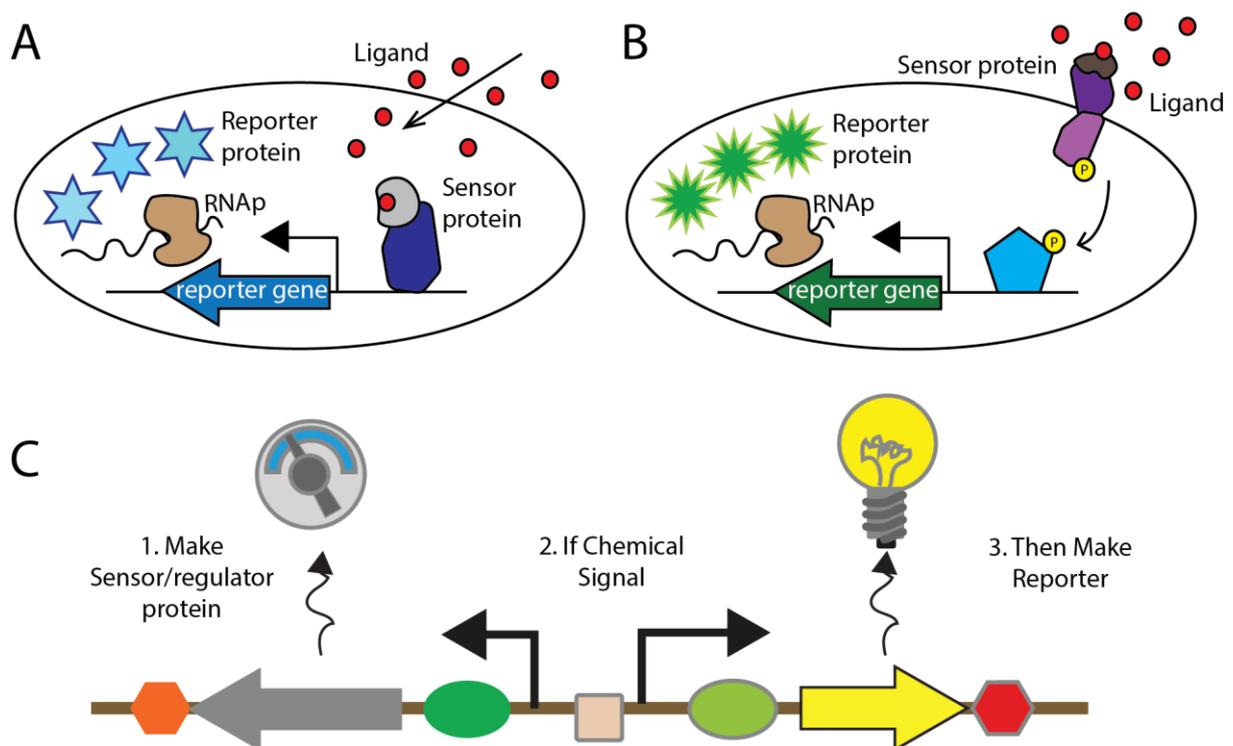


Figure 1 – Conceptual idea of bioreporter cells. Two scenarios are depicted, where the bioreporter senses the ligand inside (A) or outside (B) the cell and, as a consequence, trigger *de novo* expression of a reporter gene. This leads to formation of reporter protein in the cell, which is measured in the bioreporter assay after a defined duration of the incubation with the sample. C) Schematic simplified instructions to be engineered on a synthetic reporter gene circuit. Colored arrows: coding regions of the sensing/regulatory protein and reporter protein; colored ellipses: ribosome binding sites; colored hexagons: transcription terminators; 90°-angle arrows: promoter sequences; colored square: operator site for the interaction of the sensor/regulatory protein with the DNA that leads to transmission of the sensed signal to altered reporter gene expression.

Sensor elements for bioreporters

The sensing elements that enable the bioreporter cell to recognize an input signal are either protein or RNA (e.g., aptamers) ^[14]. Sensor proteins and aptamers form specific three-dimensional domains/configurations for interaction with ligands. Ligand-binding effectuates an intramolecular conformational change, which allosterically may effect a different domain in the protein, for example, one that is involved in DNA binding or yielding a chemical modification of specific amino acid residues on the protein. In other cases, the sensory proteins transmit ligand binding intermolecularly in a chain of phosphorylation events to one or more separate protein(s). In particular transcription regulators (TRs) have been frequently deployed as sensory switches or as synthetic building blocks for hybrid switches in bioreporter circuits. TRs combine sensing and transducing domains in a single protein (although they are active mostly in multimeric forms) ^[15]. Ligand binding leads to altered DNA-binding properties of the TR on an ‘operator’ site, which directly influences transcription efficiency or frequency from the promoters controlled by the TR ^[16]. The coupling of sensing to ‘actuation’ in a single protein leading to the change in occupancy of a promoter has been one of the most useful properties for bioreporter constructions, since the promoter can be used to control expression of the reporter gene. The wide repertoire of regulatory proteins in microorganisms has been exploited intensively for the construction of bioreporters, capitalizing on the native ligands recognized by the individual TRs ^[6].

Various other protein families carry sensing domains useful for bioreporting purposes, but do not directly interact themselves with operator sites in or nearby promoters. Examples include periplasmic binding proteins (PBPs), which have high affinity for specific ligands that they scavenge for the cell, frequently in order to present those to import channels ^[17]. PBPs can also interact with methyl-accepting chemotaxis proteins (MCPs) ^[18], a wide class of proteins that enable bacterial cells to recognize gradients of chemical compounds and either become attracted or repelled ^[19]. Further sensory proteins operate in the so-called bacterial two-component signaling pathways ^[20]. Ligand-

binding to the sensory protein results in autophosphorylation, which is transmitted to a response regulator protein that controls gene expression or specific physiological reactions. Sensory proteins of two-component signaling pathways may span across the cytoplasmic membrane or be found within the cytoplasm itself. Finally, antibodies or specific designed ligand-binding proteins have been deployed as sensory proteins for bioreporter developments ^[14, 21]. Ligand perception by aptamers influences the efficiency of translation of the reporter gene rather than its transcription ^[14].

Strategies for obtaining new sensory proteins

The design of bioreporter bacteria has so far mostly exploited natural compounds and their cognate sensory proteins (TRs, PBPs, MCPs, etc.) ^[8, 9, 22-26]. This has yielded a number of very sensitive and potentially applicable bioreporters, but the current weak part in bioreporter deployment is their limited range of detectable compounds, which may not align with the environmental, industrial or medical interest in the analysis ^[21]. Characterization of promoter activation as a consequence of ligand exposure by global gene expression analysis, may in some cases reveal novel signaling chains or TRs exploitable as sensory elements in bioreporters ^[27, 28]. One of the alternatives to obtain sensory proteins that recognize non-natural ligands is to alter or extend the ligand-binding capacities of existing sensory proteins through mutagenesis. Different strategies have been deployed for this, which usually consist of some form of DNA gene shuffling, or directed mutagenesis procedure on the full gene coding for the sensory protein, followed by rigorous selection and or screening of promising mutants ^[29]. Alternatively or as a complement to this strategy, computational approaches have been proposed to better predict the amino acid residues or protein areas to target and to limit the mutagenesis effort to restricted or site-specific libraries ^[30].

In the following we will briefly expose in general terms a number of current mutagenesis and selection or screening strategies, after which we will focus more specifically on actual obtained sensory mutants in the TR, PBP and MCP families, respectively.

Most sensory domains of TRs, PBPs or MCPs are too large to produce completely randomized mutant gene libraries that would encode all possible variants. And if this were possible, the screening of these variants would not be realistically achievable. Practically speaking, random mutant libraries span sizes from 10^4 (clone libraries) to 10^{12} (e.g., phage or ribosome display) mutants. Mutants are typically produced by error-prone DNA amplification, sometimes in combination with DNA synthesis strategies, that aim to obtain between on average 1–3 mutations per gene^[31]. These DNAs are subsequently cloned into appropriate vectors and introduced into bacterial cells, or expressed on bacteriophages, or kept as *in vitro* transcription-translation systems. The next step consists of an

appropriate library selection or screening strategy to find potential mutants fulfilling the intended new or altered ligand-binding properties. Screening can be based on the ligand-binding reaction itself, in which case one can, for example, present the ligand and capture mutant protein from a library display^[32]. In other cases ligand binding can be detected indirectly through, e.g., ATPase activity of the TR or phosphorylation of the sensory protein^[33]. Selection can also be based on conditional or exclusive mutant growth in the presence of the new ligand, which has the advantage that in principle only positive genotypes survive^[31,34]. The problem in screening new sensing specificities from large libraries and in contrast to, for instance, screening of enzymatic variants using color reactions, is that sensor selection is indirect. The sensory element is coupled to a conditionally lethal output and only if the mutant displays the proper new properties, the cell can grow and multiply^[31]. Several strategies for conditional lethality and counterselection have been developed and have shown some degree of success^[31]. In practice, however, a significant proportion of false positives appears that have overcome the conditionality in a non-intended manner.

As a third variant, sensor mutant libraries may be screened rather than selected for proper signaling in the presence of the new ligand. In this case, the sensory element is coupled to the synthesis of, for instance, a reporter protein^[35]. Cells with outlier reporter protein production are then indicative for mutant protein behaviour, and can be detected and separated. This procedure has the disadvantage that high-throughput screening strategies are required for large mutant library sizes^[36]. On the other hand, there is less chance for selection of off-target false-positive mutations. Particularly autofluorescent protein reporters can be easily measured in high-throughput on individual cells, on *in vitro* transcription/translation systems in water-in-oil emulsion droplets, or on cells or microcolonies immobilized in microbeads using flow cytometry^[35,37]. Potential mutants with outlier fluorescence can be separated using fluorescence activated cell sorting (FACS) and recovered. As the signal increase in positive mutants need not be *a priori* extremely large, the separation of interesting mutants from the rest of the population remains challenging. Furthermore, single cell screening is prone to significant cell-to-cell variability, which adds to the complexity of identifying *bona fide* mutants with outlier fluorescence. To overcome these limitations, new strategies have been proposed recently. Van Rossum and coworkers used a conditional resistance to kanamycin in combination with a fluorescent bioreporter system^[38], to reduce the complexity of a large mutant library to a smaller subset based on growth of surviving clones. Others showed how phenotypic variation among individual cells of different mutants can be reduced by screening libraries embedded and grown to microcolonies in alginate beads^[37].

Mutagenesis and selection of mutant transcription activators

TRs are proteins that influence the transcription of a gene or a set of genes. TRs that carry ligand-sensing and DNA-binding domains can be used as part of a bioreporter system, by fusing a promoterless reporter gene downstream of the promoter that the TR controls (Figure 1C). Both TR activators and repressors, single- or two-component systems, have been deployed within reporter gene circuitry [6]. The affinity of the TR towards its specific ligand(s) and to its operator site in the promoter region determine the sensitivity of the obtained bioreporter system [39, 40]. A number of well-characterized TRs have been used as test cases to attempt to obtain mutants with new ligand-binding properties. Early attempts focused on well-known members of activator and repressor TR families, such as TetR- [41, 42], LuxR-, LysR- [43], NtrC- and AraC-families [44]. Notably NahR and DntR, the LysR-type transcription activators for salicylate, were used for random as well as computation-inferred mutagenesis. Lonneborg and colleagues [45] used crystal structure information to predict the binding pocket of DntR for salicylate and propose specific mutations that would result in binding of 2,4-dinitrotoluene. Interestingly, although a number of designed mutants were obtained, this procedure was at that point not extremely effective, apparently because the ligand-binding site in the mutant protein was slightly distorting the protein elsewhere, which influenced its activation process. The NtrC-type sigma⁵⁴-dependent transcription activators XylR, DmpR and HbpR, for xylenes, phenol and 2-hydroxybiphenyl, respectively, have been used in directed evolution and DNA shuffling experiments to screen for ligand-binding specificities expanding the original spectrum. Indeed, even with limited library sizes of up to 10⁵ clones mutant proteins could be retrieved that displayed activation by non-native ligands, such as dinitrotoluene [46-48], chlorophenols [49] or 2-chlorobiphenyl [50]. More recently, the crystal structure of the related NtrC-type protein MopR with its ligand phenol was solved [33]. This was used to computationally infer amino acid changes in the binding pocket that resulted in binding and activation by methyl- and ethyl-substituted benzenes. Also the structure of LacI with its ligand allolactose was used to predict and build variant libraries for four new sugar ligands, which could subsequently be retrieved through selection and screening approaches [51].

These examples have shown that randomized mutagenesis strategies are more likely to show fruitful results when supported by computational predictions of the ligand-binding cavities in known or closely related crystal structures of the TRs. Although most examples have expanded or shifted the range of native ligands to closely related molecules, it shows that computationally inferred mutant production is a valuable method to obtain new sensory capacity for bioreporters. Protein sensory domains can be further shuffled with different other protein domains as blocks to create artificial TRs in biosensing pathways [52]. It is to be expected that as predictions of structural protein domains improve further, computational-supported designs will gain further in advantage.

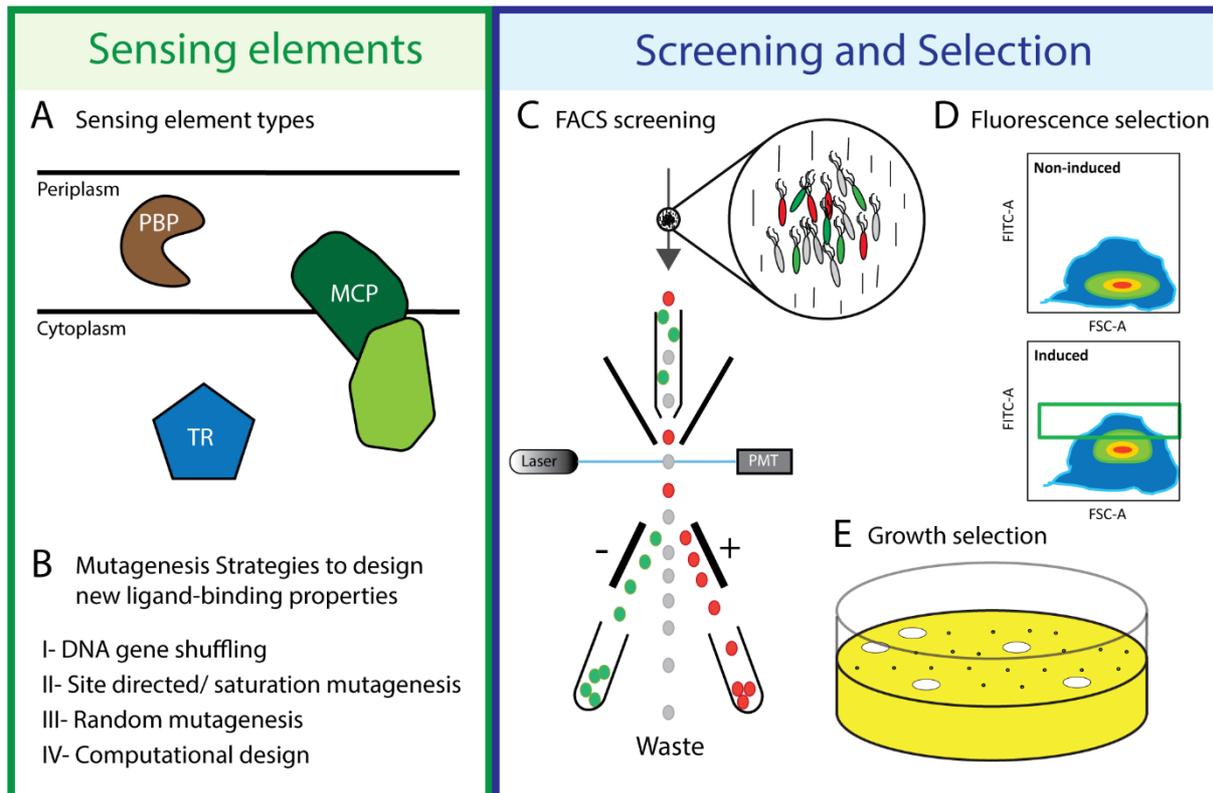


Figure 2 – Strategies for obtaining new ligand-binding properties in sensory proteins. (A) Sensory protein classes: PBP- Periplasmic Binding Proteins; MCP- Methyl-accepting Chemotaxis Proteins; TR- Transcription regulators. (B) Mutagenesis strategies to design new sensing elements. (C), (D) and (E) Screening and selection strategies to find and isolate potential candidate proteins with new or altered ligand-binding properties from mutant libraries. FACS- Fluorescence-activated cell sorting; PMT- Photomultiplier; FSC-A, Forward Scatter Area.

Periplasmic binding proteins and selection

Periplasmic binding proteins have a conserved protein structure, named the bilobal structural fold ^[53], which consists of two domains connected by a hinge region, with the binding pocket formed between the two domains. Based on the connecting hinges, PBPs have been classified in three major groups (I–III). Class I PBPs have three β -sheets connecting the two lobes; which in Class II are two β -sheets only. Class III PBPs have a single α -helix between the two lobes ^[17, 18]. In absence of ligand, PBPs adopt an open conformation, in which the binding site is exposed. Ligand-binding stabilizes the closed form of the protein, with the two lobes approaching each other and burying the ligand within the surrounding protein like a ‘Venus fly-trap’ ^[54, 55]. PBPs facilitate nutrient and trace mineral scavenging for the bacterial cell, by binding the ligand at high affinity and ‘presenting’ the bound ligand to specific transport channels. Binding to the transport channel elicits ATP hydrolysis and compound transport.

Some PBPs are additionally involved in chemotactic sensing and also interact in ligand-bound form with MCPs in the cytoplasmic membrane. Galactose-binding protein (GBP) and ribose-binding protein (RBP) of *E. coli* are two well-known examples involved in chemosensing as well as sugar scavenging. Ligand-bound GBP and RBP interact with the Trg chemoreceptor, which leads to a phosphorylation cascade that biases flagellar movement of the cell [56, 57].

PBPs constitute a widely distributed and evolutionary ancient protein family, members of which bind a large range of natural organic and inorganic ligands. The primary structure homology of PBPs that even bind structurally close ligands is surprisingly rather low. For example, GBP and RBP only share 29% amino acid conservation. Ligand-binding to PBPs has been exploited intensively for the development of biosensors. The configuration change of both protein lobes along the hinge upon ligand-binding can be followed by a variety of biochemical techniques. Most interesting for biosensing purposes is, for example, Förster resonance energy transfer (FRET) [58-60]. For FRET to occur, two different fluorophores are attached to either side of the PBP lobes. The primary fluorophore (the *donor*) is excited by the illuminating light and its emission light excites the second fluorophore, which is measured. Depending on the type of PBP and the placement of the fluorophores, ligand binding can either diminish or increase the FRET signal [61]. FRET on PBPs is commonly engineered by using native Cys-residues or modifying other PBP residues to Cys, which can be coupled in a thiol reaction to the fluorophore on the purified protein. The second fluorophore is frequently an autofluorescent protein (for example, yellow fluorescent protein), which is attached to the N-terminal region of the PBP [60]. Recently, it was shown how fluorophore-bound amino acids, such as L-(7-hydroxycoumarin-4-yl)ethylglycine, can be incorporated directly in the protein using the capacity of *E. coli* engineered amber mutants to incorporate non-natural amino acids [60]. PBP ligand binding can also be detected in competitive surface or liposome-based assays, leading to low sensitivities of detection in the nM- μ M-range [62]. Different groups have demonstrated ligand detection by purified PBPs of, for example, thiamine [63], glucose [64-66], amino acids [60, 67, 68], phosphate [69], phosphonate [70], or maltose [71-73].

Whereas most groups have exploited the natural binding properties of the investigated PBPs, very few have actually attempted to change their binding specificity. PBP binding specificity alteration through protein structure computational guidance was heralded in the early 2000's in a number of conceptual publications [74, 75]. Computational modeling has been used to explain how mutations in the D-glucose/D-galactose-binding protein (GGBP) reduced affinity for glucose by 5000 times, bringing it to physiological human range (mM). The modified PBP may be used as a glucose blood biosensor [65]. Mutations improved the specificity of a leucine-binding protein (LBP) towards L-Leu by reducing cross-reaction to structurally similar amino acids. Further modification increased the affinity for L-Leu by 14 times and led to moderate recognition of L-Met [60].

PBPs have been integrated in whole cell living bioreporters thanks to a discovery in the group of Hazelbauer in 1994 ^[76] that ligand-bound RBP and GBP can activate an *E. coli* hybrid chemoreceptor formed between Trg and EnvZ. This hybrid chemoreceptor (Trz1) consists of the 265 N-terminal amino acids of the Trg chemoreceptor linked to the 230 C-terminal amino acids of the EnvZ histidine kinase of the osmoregulation system. Ligand-bound GBP or RBP interaction to the periplasmic Trg¹-domain of Trz1 triggers histidine kinase activity of the cytoplasmic EnvZ-domain, leading to autophosphorylation and subsequent phosphorylation of the cognate response regulator OmpR ^[77]. Phosphorylated OmpR has increased affinity for the *ompC* promoter, and ligand-binding to RBP or GBP thus causes an increased transcription rate from this promoter. Coupling a reporter gene to the *ompC* promoter yields a bioreporter for galactose and ribose with good sensitivity in the low μM (galactose) to nM-range (ribose, 50 nM detection limit) ^[78].

In addition to proposing computational prediction of PBPs with new ligand-binding capacities on purified protein, several engineered PBP variants were included in the *E. coli* Trz1-OmpR platform. In this case ligand interaction with the engineered mutant RBP would bind Trz1 and trigger reporter gene expression. One of the publications specifically reported the successful design of mutant proteins based on the native *E. coli* ribose binding protein (RbsB) yielding detection of non-natural substrates trinitrotoluene (TNT), lactate or serotonin, down to the nM-mM range. The publication claimed a design of a mutant protein (TNT.R3) that would effectively and sensitively recognize trinitrotoluene (TNT) with a constant of affinity (K_D) of 2 nM ^[75]. Unfortunately, this work was largely questioned and independent studies were unable to reproduce the initial findings ^[78, 79]. Further subsequent studies have shown that RbsB becomes very sensitive to misfolding upon mutations that locate both in the ribose binding pocket as well as elsewhere in the protein ^[80, 81]. The initial idea of computational redesigned ligand-binding based on predictions of the (wild-type) binding pocket, which works successfully in TRs (see above), may have largely underestimated the necessary conformational characteristic changes in PBPs that determine its activity. Computational predictions of amino acid changes that would drastically alter the PBP's ligand, and at the same time maintain correct protein stability and allow its intramolecular conformation changes are currently unreliable ^[82, 83].

Bacterial bioreporters based on methylaccepting chemotaxis proteins

Methyl-accepting chemotaxis proteins (MCPs) are membrane-spanning proteins that provide sensory information to bacteria about chemical gradients in which they live. This permits the cells to swim towards possible attractants such as sugars or amino acids or retreat from toxic repellents. MCPs have four structural domains, which are called the ligand binding (LB), the transmembrane (TM), the histidine kinases, adenylate cyclases, methyl accepting proteins and phosphatases (HAMP), and the signal transmitter (ST) domains ^[84].

The LB domain is responsible for the specificity of chemical recognition, which can consist of one or a group of closely related chemical compounds. LB domains have been clustered into two main groups ^[85] with two subgroups each ^[19, 86], although this classification is not sufficient to predict their cognate ligands. The first cluster is comprised of the four helix bundle (4HB) and single CACHE (sCACHE) classes. *E. coli* MCPs all belong to the 4HB class ^[19]. In contrast to the 4HB class with its four helical bundles, the sCACHE class also displays a β -sheet structure that interacts with the ligand. The second cluster is comprised of the helical bimodular (HBM) and double CACHE (dCACHE) classes. HBM in essence consist of two coupled 4HB domains, one of which is proximal and the other distal to the cytoplasmic membrane. Both 4HB domains in the HBM can have different ligand specificities. The dCACHE class is comprised of two sCACHE subdomains, but in contrast to HBM, only a single one of these binds the ligand ^[19]. LB domains can bind ligands directly, like for the Tar-MCP of *E. coli* ^[87], or interact with a PBP that is associated with the ligand, as is the case for Trg ^[88].

The membrane-spanning TM domain locates the MCPs to the cytoplasmic membrane. The HAMP domain is commonly seen as a linker to the ST domain ^[89]. Some MCP-like proteins have several HAMPs, which are referred to as poly-HAMP arrays ^[90]. The ST domain interacts with the soluble cytoplasmic response regulator (RR) and determines the further downstream reaction of the cell upon perception of the stimulus. Frequently, signal transduction is achieved by protein phosphorylation, either by the ST itself or with help of an associated kinase ^[91]. On the basis of extensive homology searches, 7 different cytoplasmic ST-domains of MCPs have been recognized, relative to the number of helical bundles (named 24H–44H) ^[92].

Chemotaxis mediated by MCPs is one of the best described two-component signaling pathways. *E. coli* has four MCPs involved in chemotaxis and one in aerotaxis, which are called Trg, Tar, Tsr and Tap, and Aer ^[93]. Trg provides taxis towards ribose and galactose, Tar towards aspartate and maltose, Tsr towards serine and Tap towards dipeptides^[93]. The RR for MCPs in *E. coli* is CheY, which is phosphorylated by the ST-associated kinase CheA in presence of an MCP-activating ligand. This complex is stabilized by CheW^[94]. Phosphorylated CheY has increased affinity for FliM, a component of the flagellum. CheY~P–FliM binding will bias the rotation of the flagellum from anticlockwise into

clockwise direction, leading to more frequent tumbling ^[95, 96]. When the cells perceive a gradient of attractant, the MCPs in the ligand-bound form reduce phosphorylation. This will lower the phosphorylation rate of CheY, which as a result of constant dephosphorylation by CheZ, decreases the intracellular CheY~P levels^[97]. The result is less frequent tumbling and more frequent long directional swims. This on average leads the motile cells to move in the direction of the higher attractant concentration. Repellents lead to the opposite reaction ^[96]. MCP-chemotaxis responses are transient and, to some extent relative (or insensitive to the actual chemical concentration itself). Longer exposure to higher attractant concentrations leads to a reset of the system through MCP methylation at the ST-domain by CheR ^[98]. Demethylation is regulated by CheB which is activated by phosphorylation by CheA as CheY ^[97]. The transient reaction of MCPs upon instant change of chemical concentration can be measured by FRET between CheY~P and CheZ, and yields a concentration-dependent signal which can be interpreted as a biosensor measurement ^[99]. Also the accumulation of motile cells as a result of chemotaxis in a constant chemical gradient can be quantified as a biosensor response ^[100].

MCPs share the HAMP domain with a large range of other two-component signal transduction systems ^[89, 101]. This actually allowed the construction of the first functional chimeric transmembrane receptor Trz1, mentioned above ^[76], by fusing the LB domain of the chemoreceptor Trg to the ST-kinase domain of the osmosensor EnvZ. The structure of EnvZ is slightly different from MCPs as it has no LB domain but a cytoplasmic sensory domain sensitive to osmotic stress. Upon osmotic stress the ST of EnvZ phosphorylates OmpR, which then activates the *ompC* promoter and expression of the porin OmpC to compensate the turgor pressure ^[102-104]. In the Trz1 hybrid system, OmpR phosphorylation is triggered upon binding of the ribose-bound-RBP to the Trz1-LB domain. As mentioned, by fusing a reporter gene like *gfp* to the *ompC* promoter, a very sensitive and quantitative bioreporter for ribose can be obtained ^[78].

Understanding the mode of action of two-component systems was greatly improved by a resolved almost complete structure of the NarQ nitrate sensory histidine kinase ^[105]. The NarQ N-terminus is located inside the cytosol, after which the protein spans the cytoplasmic membrane in a first TM helix. In the periplasm, the NarQ-LB domain is then formed by four antiparallel helices, which connect to the membrane by a small loop between the second and third helix. A next TM helix connects the periplasmic LB domain to the cytosol via the HAMP domain. The HAMP domain consists of a helix-loop-helix motif with a parallel helix bundle and four rotations ^[105]. NarQ forms a dimer whereas MCPs form trimers of dimers, which assemble themselves into MCP clusters located at the cell poles ^[94, 106]. Comparison of the NarQ-structure with and without its ligand nitrate, indicated that binding of nitrate

induces a displacement of the LB-helices of up to 1 Å, resulting into a ‘piston-like’ shift in the TM helices. This shift translates into a ‘scissor-like’ motion of the HAMP-domain through the interaction with the N-terminus of the first TM helix that pushes the second helix of the HAMP further into the cytoplasm. This motion displaces the end of the HAMP-domain 7 Å between the bound and unbound state, and is assumed to lead to activation of the ST domain ^[105].

The exploitation of MCPs has a good potential for future classes of biosensors or bioreporter cells. The natural variability of the LB-domains of MCPs deduced from genome information is large, although their cognate ligands are currently not well described and in most cases not predictable from protein homology ^[95]. Some MCPs have been reported to react to environmentally relevant compounds, such as 2,4-dichlorophenoxyacetic acid (2,4-D; a previously widely applied herbicide) ^[107], naphthalene ^[108] or toluene ^[109]. The ligand-properties of MCPs can be changed to some extent by site-directed or random mutagenesis. For example, Tar mutants have been described that permit a chemotactic response towards cysteic acid, phenylalanine or glutamate ^[110]. Bi and coworkers used structural prediction of the Tar LB-domain to produce binding to eight non-cognate ligands, including a variety of aspartate derivatives and phthalic acid ^[111].

MCP LB-domains may be connected in a plug-and-play manner to other protein domains, in order to obtain quantitative standardized output. This principle was shown in a recent impressive work where a variety of MCP LB-domains were fused to the Tar cytoplasmic parts at sites close to the HAMP linker ^[99]. Activation of the MCP-Tar hybrid by ligand-interaction was then quantified through FRET between CheY~P and CheZ. In this case a yellow fluorescent protein (YFP) was fused to CheY and a cyan fluorescent protein (CFP) was fused to CheZ. CheY~P binding to FliM triggers the FRET signal in seconds, and is influenced by ligand interactions at the MCP-LB domains ^[99, 112]. Since the response is ephemeral, the cells have to be exposed to instantly changing chemical concentrations in order to observe maximal FRET output, after which the signal within a few minutes gradually returns to baseline ^[99].

MCP-LB domains can also be coupled to the cytoplasmic signaling domains of related two-component systems, in which case ligand-interaction with the receptor can be transmitted into *de novo* gene expression of a reporter gene. This was the concept behind the Trz1 fusion as mentioned earlier ^[78]. Apart from the Trz1 Trg-EnvZ hybrid, also Tar-EnvZ hybrids ^[113] DcuS-EnvZ ^[114] and NarX-EnvZ ^[115] have been constructed. Activation of the EnvZ autokinase activity would then in all cases lead to reporter gene production from the *ompC* or *ompF* promoter ^[115]. In contrast to the MCP-LB-domain-Tar fusions measured by FRET, very little response was observed from MCP-EnvZ hybrids ^[115], except for Trz1. This suggests that obtaining correctly functioning hybrid proteins fused by HAMP linkers is still not trivial, and the rules for hybrid MCP functioning are ill understood.

Finally, MCP sensory input for biosensors can be exploited from chemotaxis itself, by measuring the accumulation of motile chemotactic cells in chemical gradients. This exploitation has somewhat been hindered by the inherent difficulty to produce stable gradients, but microfluidic devices have been instrumental in this respect ^[116-118]. Recent work showed the quantitative and temporal response of *E. coli* motile and chemotactic cells to serine, aspartate and methylaspartate in a microfluidic system ^[100].

Conclusions

Sensory proteins are pivotal for the development of whole cell based bacterial sensors. Past work has successfully exploited the wide variety of sensory protein classes, either directly, or in different plug-and-play manners. But the difficulty for future sensor development is to achieve a more straightforward manner to predict and mutate existing (or *de novo*) sensory proteins for non-cognate ligands. For the case of PBPs, computational predictions of new ligand properties have not been successful (yet), and some of the past claims have proven irreproducible. For the case of TRs, computational-guided predictions have been more successful in obtaining new ligand-binding properties, perhaps because these proteins undergo less drastic conformational changes upon ligand interaction. It can be expected that the accuracy of the systems in free energy scoring function calculations, molecular dynamics involved on protein-ligand interactions and prediction of conformational changes will improve in the near future by better algorithms.

Table 1 – Selected examples of periplasmic binding proteins, transcription activators and methylaccepting chemotaxis proteins used in bioreporter systems.

Sensor protein class	Sensor Protein	Host chassis	Reporter system	Ligand molecules	Detection sensitivity	Reference
TR	XylR of <i>Pseudomonas putida</i>	<i>Escherichia coli</i>	<i>lucFF</i>	Benzene, Toluene and xylene	40 μ M	[119]
TR	HbpR of <i>Pseudomonas nitroreducens</i>	<i>E. coli</i>	<i>luxAB</i>	Hydroxylated biphenyls	0.4 μ M	[120, 121]
TR	AlkS of <i>Pseudomonas oleovorans</i>	<i>E. coli</i>	<i>luxAB</i>	C ₆ -C ₁₀ alkanes	10 nM	[122]
TR	ArsR of <i>E. coli</i>	<i>E. coli</i>	<i>luxAB</i>	Arsenite	5 nM	[13, 123]
TR	MerR of <i>E. coli</i>	<i>E. coli</i>	<i>luxCDABE</i>	Hg ²⁺	1 nM	[124]
TR	TetR of <i>E. coli</i>	<i>E. coli</i>	<i>luxCDABE</i>	Tetracyclines	45 nM	[125]
TR	YqjF of <i>E. coli</i>	<i>E. coli</i>	<i>luxCDABE</i>	2,4-dinitrotoluene and 2,4,6-trinitrotoluene		[28, 126]
PBP	RbsB of <i>E. coli</i>	<i>E. coli</i>	<i>gfp</i>	D-Ribose	50 nM	[78]
PBP	LivK (LBP) of <i>E. coli</i>	<i>E. coli</i>	FRET	L-Leu	400 nM	[60]
PBP	MalE of <i>E. coli</i>	<i>E. coli</i>	FRET	Maltose	4 mM	[71]
PBP	TBP of <i>E. coli</i>	Cell free	liposomes	Thiamine	0.5 nM	[62]
MCP	Tsr, Tar of <i>E. coli</i>	<i>E. coli</i>	Cell accumulation	Serine, aspartate	10 μ M	[127]
MCP	Hybrids to Tar	<i>E. coli</i>	FRET	Nitrite, nitrate, L-malate	~0.3 μ M	[99]

TR, transcription regulator; PBP, periplasmic binding protein; MCP, methylaccepting chemotaxis protein.

References

1. van der Meer, J.R. and S. Belkin, *Where microbiology meets microengineering: design and applications of reporter bacteria*. Nat. Rev. Microbiol., 2010. **8**(7): p. 511-522.
2. Xu, T., et al., *Genetically modified whole-cell bioreporters for environmental assessment*. Ecol Indic, 2013. **28**: p. 125-141.
3. Way, J.C., et al., *Integrating biological redesign: where synthetic biology came from and where it needs to go*. Cell, 2014. **157**(1): p. 151-161.
4. Daunert, S., et al., *Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes*. Chem. Rev., 2000. **100**(7): p. 2705-2738.
5. van der Meer, J.R., D. Tropel, and M.C.M. Jaspers, *Illuminating the detection chain of bacterial bioreporters*. Environ. Microbiol., 2004. **6**: p. 1005-1020.
6. van der Meer, J.R., *Bacterial Sensors: Synthetic Design and Application Principles*. Synthesis Lectures on Synthetic Biology, ed. M. Amos. Vol. Lecture #2. 2010: Morgan & Claypool.
7. van der Meer, J.R., *Analytics with engineered bacterial bioreporter strains and systems*. Curr Opin Biotechnol, 2006. **17**(1): p. 1-3.
8. Siegfried, K., et al., *Field Testing of Arsenic in Groundwater Samples of Bangladesh Using a Test Kit Based on Lyophilized Bioreporter Bacteria*. Environmental Science & Technology, 2012. **46**(6): p. 3281-3287.
9. Paton, G.I., B.J. Reid, and K.T. Semple, *Application of a luminescence-based biosensor for assessing naphthalene biodegradation in soils from a manufactured gas plant*. Environmental Pollution, 2009. **157**(5): p. 1643-1648.
10. Merulla, D., et al., *Bioreporters and biosensors for arsenic detection. Biotechnological solutions for a world-wide pollution problem*. Curr. Opin. Biotechnol., 2013. **24**(3): p. 534-41.
11. Courbet, A., et al., *Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates*. Sci. Transl. Med., 2015. **7**(289): p. 289ra83.
12. Harms, H., M.C. Wells, and J.R. van der Meer, *Whole-cell living biosensors--are they ready for environmental application?* Appl. Microbiol. Biotechnol., 2006. **70**(3): p. 273-80.
13. Trang, P.T., et al., *Bacterial bioassay for rapid and accurate analysis of arsenic in highly variable groundwater samples*. Environ. Sci. Technol., 2005. **39**(19): p. 7625-30.
14. Bazin, I., et al., *New biorecognition molecules in biosensors for the detection of toxins*. Biosens Bioelectron, 2017. **87**: p. 285-298.
15. Tropel, D. and J.R. van der Meer, *Bacterial transcriptional regulators for degradation pathways of aromatic compounds*. Microbiol. Mol. Biol. Rev., 2004. **68**(3): p. 474-500.
16. Browning, D.F. and S. Busby, *The regulation of bacterial transcription initiation*. Nat. Rev. Microbiol., 2004. **2**: p. 1-9.

17. Berntsson, R.P.A., et al., *A structural classification of substrate-binding proteins*. FEBS Letters, 2010. **584**(12): p. 2606-2617.
18. Quijoch, F.A. and P.S. Ledvina, *Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes*. Molecular Microbiology, 1996. **20**(1): p. 17-25.
19. Matilla, M.A. and T. Krell, *Chemoreceptor-based signal sensing*. Curr Opin Biotechnol, 2017. **45**: p. 8-14.
20. Stock, A.M., V.L. Robinson, and P.N. Goudreau, *Two-component signal transduction*. Annu Rev Biochem, 2000. **69**: p. 183-215.
21. Chang, H.J., et al., *A Modular Receptor Platform To Expand the Sensing Repertoire of Bacteria*. ACS Synth Biol, 2018. **7**(1): p. 166-175.
22. Tecon, R., et al., *Development of a multistrain bacterial bioreporter platform for the monitoring of hydrocarbon contaminants in marine environments*. Environ. Sci. Technol., 2010. **144**: p. 1049-1055.
23. Yoon, Y., et al., *Use of Tunable Whole-Cell Bioreporters to Assess Bioavailable Cadmium and Remediation Performance in Soils*. PLoS ONE, 2016. **11**(5): p. e0154506.
24. Xu, T., et al., *Detection of Organic Compounds with Whole-Cell Bioluminescent Bioassays*. Advances in biochemical engineering/biotechnology, 2014. **144**: p. 111-151.
25. Hynninen, A. and M. Virta, *Whole-Cell Bioreporters for the Detection of Bioavailable Metals*, in *Whole Cell Sensing System II: Applications*, S. Belkin and M.B. Gu, Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 31-63.
26. Ivask, A., T. Rõlova, and A. Kahru, *A suite of recombinant luminescent bacterial strains for the quantification of bioavailable heavy metals and toxicity testing*. BMC Biotechnology, 2009. **9**(1): p. 41.
27. Shemer, B., et al., *Aerobic Transformation of 2,4-Dinitrotoluene by Escherichia coli and Its Implications for the Detection of Trace Explosives*. Appl Environ Microbiol, 2018. **84**(4).
28. Belkin, S., et al., *Remote detection of buried landmines using a bacterial sensor*. Nat Biotechnol, 2017. **35**(4): p. 308-310.
29. Vogne, C., S. Beggah, and J.R. van der Meer, *Random and Site-Directed Mutagenesis of Transcriptional Regulator Proteins Implicated in Hydrocarbon Degradation*, in *Handbook of Hydrocarbon and Lipid Microbiology*, K.N. Timmis, Editor. 2010, Springer: Berlin Heidelberg. p. 4429-4446.
30. Tinberg, C.E., et al., *Computational design of ligand-binding proteins with high affinity and selectivity*. Nature, 2013. **501**: p. 212-216.
31. Galvao, T.C. and V. de Lorenzo, *Transcriptional regulators a la carte: engineering new effector specificities in bacterial regulatory proteins*. Curr. Opin. Biotechnol., 2006. **17**(1): p. 34-42.
32. Cirino, P.C. and S.A. Qian, *Protein Engineering as an Enabling Tool for Synthetic Biology*. Synthetic Biology: Tools and Applications, 2013: p. 23-42.

33. Ray, S., et al., *Structural Basis of Selective Aromatic Pollutant Sensing by the Effector Binding Domain of MopR, an NtrC Family Transcriptional Regulator*. ACS Chem Biol, 2016. **11**(8): p. 2357-65.
34. Choi, S.-L., et al., *Toward a Generalized and High-throughput Enzyme Screening System Based on Artificial Genetic Circuits*. ACS Synthetic Biology, 2014. **3**(3): p. 163-171.
35. Beggah, S., et al., *Mutant transcription activator isolation via green fluorescent protein based flow cytometry and cell sorting*. Microb. Biotechnol., 2008. **1**: p. 68-78.
36. Schallmeyer, M., et al., *Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors*. Current Opinion in Biotechnology, 2014. **26**: p. 148-154.
37. Duarte, J.M., I. Barbier, and Y. Schaerli, *Bacterial Microcolonies in Gel Beads for High-Throughput Screening of Libraries in Synthetic Biology*. ACS Synthetic Biology, 2017.
38. van Rossum, T., et al., *A growth- and bioluminescence-based bioreporter for the in vivo detection of novel biocatalysts*. Microbial Biotechnology, 2017. **10**(3): p. 625-641.
39. Berset, Y., et al., *Mechanistic Modeling of Genetic Circuits for ArsR Arsenic Regulation*. ACS Synth Biol, 2017.
40. Nielsen, A.A., et al., *Genetic circuit design automation*. Science, 2016. **352**(6281): p. aac7341.
41. Scholz, O., et al., *Teaching TetR to recognize a new inducer*. J Mol Biol, 2003. **329**(2): p. 217-27.
42. de los Santos, E.L., et al., *Engineering Transcriptional Regulator Effector Specificity Using Computational Design and In Vitro Rapid Prototyping: Developing a Vanillin Sensor*. ACS Synth Biol, 2016. **5**(4): p. 287-95.
43. Jha, R.K., et al., *Rosetta comparative modeling for library design: Engineering alternative inducer specificity in a transcription factor*. Proteins, 2015. **83**(7): p. 1327-40.
44. Tang, S.Y. and P.C. Cirino, *Design and application of a mevalonate-responsive regulatory protein*. Angew Chem Int Ed Engl, 2011. **50**(5): p. 1084-6.
45. Lonneborg, R., et al., *In vivo and in vitro investigation of transcriptional regulation by DntR*. J. Mol. Biol., 2007. **372**(3): p. 571-82.
46. Garmendia, J., et al., *A la carte transcriptional regulators: unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors*. Mol. Microbiol., 2001. **42**(1): p. 47-59.
47. Garmendia, J., et al., *Tracing explosives in soil with transcriptional regulators of Pseudomonas putida evolved for responding to nitrotoluenes*. Microb Biotechnol, 2008. **1**: p. 236-246.
48. Galvao, T.C., M. Mencia, and V. de Lorenzo, *Emergence of novel functions in transcriptional regulators by regression to stem protein types*. Mol. Microbiol., 2007. **65**(4): p. 907-19.
49. Wise, A.A. and C.R. Kuske, *Generation of Novel Bacterial Regulatory Proteins That Detect Priority Pollutant Phenols*. Applied and Environmental Microbiology, 2000. **66**(1): p. 163-169.

50. Beggah, S., et al., *Mutant HbpR transcription activator isolation for 2-chlorobiphenyl via green fluorescent protein-based flow cytometry and cell sorting*. *Microbial Biotechnology*, 2008. **1**(1): p. 68-78.
51. Taylor, N.D., et al., *Engineering an allosteric transcription factor to respond to new ligands*. *Nat Methods*, 2016. **13**(2): p. 177-83.
52. Feng, J., et al., *A general strategy to construct small molecule biosensors in eukaryotes*. *Elife*, 2015. **4**.
53. Chu Byron, C.H. and J. Vogel Hans, *A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport*, in *Biological Chemistry*. 2011. p. 39.
54. Li, H.Y., et al., *Analysis of Conformational Motions and Residue Fluctuations for Escherichia coli Ribose-Binding Protein Revealed with Elastic Network Models*. *International Journal of Molecular Sciences*, 2013. **14**(5): p. 10552-10569.
55. Björkman, A.J. and S.L. Mowbray, *Multiple open forms of ribose-binding protein trace the path of its conformational change*¹¹*Edited by R. Huber*. *Journal of Molecular Biology*, 1998. **279**(3): p. 651-664.
56. Binnie, R.A., et al., *Functional mapping of the surface of Escherichia coli ribose-binding protein: mutations that affect chemotaxis and transport*. *Protein Science : A Publication of the Protein Society*, 1992. **1**(12): p. 1642-1651.
57. Eym, Y., Y. Park, and C. Park, *Genetically probing the regions of ribose-binding protein involved in permease interaction*. *Molecular Microbiology*, 1996. **21**(4): p. 695-702.
58. Wu, B., et al., *Catalyst-Free and Site-Specific One-Pot Dual Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids*. *Chembiochem*, 2012. **13**(10): p. 1405-1408.
59. Ye, K. and J.S. Schultz, *Genetic Engineering of an Allosterically Based Glucose Indicator Protein for Continuous Glucose Monitoring by Fluorescence Resonance Energy Transfer*. *Analytical Chemistry*, 2003. **75**(14): p. 3451-3459.
60. Ko, W., S. Kim, and H.S. Lee, *Engineering a periplasmic binding protein for amino acid sensors with improved binding properties*. *Organic & Biomolecular Chemistry*, 2017. **15**(41): p. 8761-8769.
61. Medintz, I.L. and J.R. Deschamps, *Maltose-binding protein: a versatile platform for prototyping biosensing*. *Curr Opin Biotechnol*, 2006. **17**(1): p. 17-27.
62. Edwards, K.A., et al., *High-Throughput Detection of Thiamine Using Periplasmic Binding Protein-Based Biorecognition*. *Anal Chem*, 2016. **88**(16): p. 8248-56.
63. Hanes, J.W., et al., *Construction of a thiamin sensor from the periplasmic thiamin binding protein*. *Chemical communications (Cambridge, England)*, 2011. **47**(8): p. 2273-2275.
64. Li, D., et al., *Optical surface plasmon resonance sensor modified by mutant glucose/galactose-binding protein for affinity detection of glucose molecules*. *Biomedical Optics Express*, 2017. **8**(11): p. 5206-5217.

65. Amiss, T.J., et al., *Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor*. Protein Science : A Publication of the Protein Society, 2007. **16**(11): p. 2350-2359.
66. Helassa, N., et al., *A novel fluorescent sensor protein for detecting changes in airway surface liquid glucose concentration*. The Biochemical journal, 2014. **464**(2): p. 213-220.
67. Paul, S., S. Banerjee, and H.J. Vogel, *Ligand binding specificity of the Escherichia coli periplasmic histidine binding protein, HisJ*. Protein Science : A Publication of the Protein Society, 2017. **26**(2): p. 268-279.
68. Wada, A., et al., *Design and Construction of Glutamine Binding Proteins with a Self-Adhering Capability to Unmodified Hydrophobic Surfaces as Reagentless Fluorescence Sensing Devices*. Journal of the American Chemical Society, 2003. **125**(52): p. 16228-16234.
69. Solscheid, C., et al., *Development of a Reagentless Biosensor for Inorganic Phosphate, Applicable over a Wide Concentration Range*. Biochemistry, 2015. **54**(32): p. 5054-5062.
70. Rizk, S.S., M.J. Cuneo, and H.W. Hellinga, *Identification of cognate ligands for the Escherichia coli phnD protein product and engineering of a reagentless fluorescent biosensor for phosphonates*. Protein Science : A Publication of the Protein Society, 2006. **15**(7): p. 1745-1751.
71. Iijima, I. and T. Hohsaka, *Position-Specific Incorporation of Fluorescent Non-natural Amino Acids into Maltose-Binding Protein for Detection of Ligand Binding by FRET and Fluorescence Quenching*. ChemBioChem, 2009. **10**(6): p. 999-1006.
72. Park, K., et al., *Detection of conformationally changed MBP using intramolecular FRET*. Biochemical and Biophysical Research Communications, 2009. **388**(3): p. 560-564.
73. Fehr, M., W.B. Frommer, and S. Lalonde, *Visualization of maltose uptake in living yeast cells by fluorescent nanosensors*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(15): p. 9846-9851.
74. Dwyer, M.A. and H.W. Hellinga, *Periplasmic binding proteins: a versatile superfamily for protein engineering*. Current Opinion in Structural Biology, 2004. **14**(4): p. 495-504.
75. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions*. Nature, 2003. **423**(6936): p. 185-190.
76. Baumgartner, J.W., et al., *Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ*. Journal of Bacteriology, 1994. **176**(4): p. 1157-1163.
77. Srividhya, K.V. and S. Krishnaswamy, *A simulation model of Escherichia coli osmoregulatory switch using E-CELL system*. BMC Microbiology, 2004. **4**: p. 44-44.
78. Reimer, A., et al., *Escherichia coli ribose binding protein based bioreporters revisited*. Scientific Reports, 2014. **4**: p. 5626.
79. Schreier, B., et al., *Computational design of ligand binding is not a solved problem*. Proceedings of the National Academy of Sciences, 2009. **106**(44): p. 18491-18496.

80. Antunes, M.S., et al., *Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway*. PLoS ONE, 2011. **6**(1): p. e16292.
81. Reimer, A., et al., *Complete alanine scanning of the Escherichia coli RbsB ribose binding protein reveals residues important for chemoreceptor signaling and periplasmic abundance*. Scientific Reports, 2017. **7**: p. 8245.
82. Feldmeier, K. and B. Höcker, *Computational protein design of ligand binding and catalysis*. Current Opinion in Chemical Biology, 2013. **17**(6): p. 929-933.
83. Yang, W. and L. Lai, *Computational design of ligand-binding proteins*. Current Opinion in Structural Biology, 2017. **45**(Supplement C): p. 67-73.
84. Szurmant, H. and G.W. Ordal, *Diversity in chemotaxis mechanisms among the bacteria and archaea*. Microbiol Mol Biol Rev, 2004. **68**(2): p. 301-19.
85. Lacal, J., et al., *Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions*. Environ Microbiol, 2010. **12**(11): p. 2873-84.
86. Upadhyay, A.A., et al., *Cache Domains That are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of Extracellular Sensors in Prokaryotes*. PLoS Comput Biol, 2016. **12**(4): p. e1004862.
87. Tajima, H., et al., *Ligand specificity determined by differentially arranged common ligand-binding residues in bacterial amino acid chemoreceptors Tsr and Tar*. J Biol Chem, 2011. **286**(49): p. 42200-10.
88. Vyas, N.K., M.N. Vyas, and F.A. Quioco, *Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and Functional Similarity*. J Biol Chem, 1991. **266**(8): p. 5226-37.
89. Mondejar, L.G., et al., *HAMP domain-mediated signal transduction probed with a mycobacterial adenylyl cyclase as a reporter*. J Biol Chem, 2012. **287**(2): p. 1022-31.
90. Natarajan, J., et al., *Biochemical characterization of the tandem HAMP domain from Natronomonas pharaonis as an intraprotein signal transducer*. FEBS J, 2014. **281**(14): p. 3218-27.
91. Bilwes, A.M., et al., *Structure of CheA, a signal-transducing histidine kinase*. Cell, 1999. **96**(1): p. 131-41.
92. Alexander, R.P. and I.B. Zhulin, *Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors*. Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2885-90.
93. Grebe, T.W. and J. Stock, *Bacterial chemotaxis: the five sensors of a bacterium*. Curr Biol, 1998. **8**: p. 154-157.
94. Weis, R.M., et al., *Electron Microscopic Analysis of Membrane Assemblies Formed by the Bacterial Chemotaxis Receptor Tsr*. Journal of Bacteriology, 2003. **185**(12): p. 3636-3643.
95. Sampedro, I., et al., *Pseudomonas chemotaxis*. FEMS Microbiol Rev, 2015. **39**(1): p. 17-46.

96. Vladimirov, N. and V. Sourjik, *Chemotaxis: how bacteria use memory*. Biol Chem, 2009. **390**(11): p. 1097-104.
97. Wadhams, G.H. and J.P. Armitage, *Making sense of it all: bacterial chemotaxis*. Nat Rev Mol Cell Biol, 2004. **5**(12): p. 1024-37.
98. Antommattei, F.M., J.B. Munzner, and R.M. Weis, *Ligand-specific activation of Escherichia coli chemoreceptor transmethylation*. J Bacteriol, 2004. **186**(22): p. 7556-63.
99. Bi, S., et al., *Engineering hybrid chemotaxis receptors in bacteria*. ACS Synth Biol, 2016. **5**(9): p. 989-1001.
100. Roggo, C. and J.R. van der Meer, *Miniaturized and integrated whole cell living bacterial sensors in field applicable autonomous devices*. Curr Opin Biotechnol, 2017. **45**: p. 24-33.
101. Aravind, L. and C.P. Ponting, *The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins*. FEMS Microbiol Lett, 1999. **176**(1): p. 111-6.
102. Aiba, H. and T. Mizuno, *Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, stimulates the transcription of the ompF and ompC genes in Escherichia coli*. FEBS Lett, 1990. **261**(1): p. 19-22.
103. Forst, S., J. Delgado, and M. Inouye, *Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in Escherichia coli*. Proc Natl Acad Sci U S A, 1989. **86**(16): p. 6052-6.
104. Forst, S., et al., *Regulation of ompC and ompF expression in Escherichia coli in the absence of envZ*. J Bacteriol, 1988. **170**(11): p. 5080-5.
105. Gushchin, I., et al., *Mechanism of transmembrane signaling by sensor histidine kinases*. Science, 2017. **356**(6342).
106. Eismann, S. and R.G. Endres, *Protein Connectivity in Chemotaxis Receptor Complexes*. PLoS Comput Biol, 2015. **11**(12): p. e1004650.
107. Hawkins, A.C. and C.S. Harwood, *Chemotaxis of Ralstonia eutropha JMP134(pJP4) to the Herbicide 2,4-Dichlorophenoxyacetate*. Applied and Environmental Microbiology, 2002. **68**(2): p. 968-972.
108. Grimm, A.C. and C.S. Harwood, *NahY, a catabolic plasmid-encoded receptor required for chemotaxis of Pseudomonas putida to the aromatic hydrocarbon naphthalene*. J Bacteriol, 1999. **181**: p. 3310-3316.
109. Laca, J., et al., *Bacterial chemotaxis towards aromatic hydrocarbons in Pseudomonas*. Environ Microbiol, 2011. **13**(7): p. 1733-44.
110. Derr, P., E. Boder, and M. Goulian, *Changing the specificity of a bacterial chemoreceptor*. J Mol Biol, 2006. **355**(5): p. 923-32.
111. Bi, S., et al., *Discovery of novel chemoeffectors and rational design of Escherichia coli chemoreceptor specificity*. Proc Natl Acad Sci U S A, 2013. **110**(42): p. 16814-9.

112. Sourjik, V. and H.C. Berg, *Binding of the Escherichia coli response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12669-74.
113. Zhu, Y. and M. Inouye, *Analysis of the role of the EnvZ linker region in signal transduction using a chimeric Tar/EnvZ receptor protein, Tez1*. J Biol Chem, 2003. **278**(25): p. 22812-9.
114. Ganesh, I., et al., *Engineered fumarate sensing Escherichia coli based on novel chimeric two-component system*. J Biotechnol, 2013. **168**(4): p. 560-6.
115. Lehning, C.E., et al., *A Modular High-Throughput In Vivo Screening Platform Based on Chimeric Bacterial Receptors*. ACS Synth Biol, 2017. **6**(7): p. 1315-1326.
116. Si, G., et al., *A parallel diffusion-based microfluidic device for bacterial chemotaxis analysis*. Lab Chip, 2012. **12**(7): p. 1389-94.
117. Wang, X., J. Atencia, and R.M. Ford, *Quantitative analysis of chemotaxis towards toluene by Pseudomonas putida in a convection-free microfluidic device*. Biotechnol Bioeng, 2015. **112**(5): p. 896-904.
118. Wang, X., T. Long, and R.M. Ford, *Bacterial chemotaxis toward a NAPL source within a pore-scale microfluidic chamber*. Biotechnol Bioeng, 2012. **109**(7): p. 1622-8.
119. Willardson, B.M., et al., *Development and testing of a bacterial biosensor for toluene-based environmental contaminants*. Appl. Environ. Microbiol., 1998. **64**(3): p. 1006-12.
120. Turner, K., et al., *Hydroxylated polychlorinated biphenyl detection based on a genetically engineered bioluminescent whole-cell sensing system*. Anal. Chem., 2007. **79**(15): p. 5740-5.
121. Lewis, C., et al., *Novel use of a whole cell E. coli bioreporter as a urinary exposure biomarker*. Environ. Sci. Technol., 2009. **43**(2): p. 423-428.
122. Sticher, P., et al., *Development and characterization of a whole cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples*. Appl. Environ. Microbiol., 1997. **63**(10): p. 4053-4060.
123. Baumann, B. and J.R. van der Meer, *Analysis of bioavailable arsenic in rice with whole cell living bioreporter bacteria*. J. Agric. Food Chem., 2007. **55**(6): p. 2115-20.
124. Selifonova, O., R. Burlage, and T. Barkay, *Bioluminescent sensors for detection of bioavailable Hg(II) in the environment*. Appl Environ Microbiol, 1993. **59**(9): p. 3083-90.
125. Korpela, M.T., et al., *A recombinant Escherichia coli sensor strain for the detection of tetracyclines*. Anal Chem, 1998. **70**(21): p. 4457-62.
126. Yagur-Kroll, S., et al., *Escherichia coli bioreporters for the detection of 2,4-dinitrotoluene and 2,4,6-trinitrotoluene*. Appl. Microbiol. Biotechnol., 2014. **98**(2): p. 885-95.
127. Roggo, C., et al., *Quantitative chemical biosensing by bacterial chemotaxis in microfluidic chips*. Environ Microbiol, 2018. **20**(1): p. 241-258.

Aims of the thesis

The overall aim of this thesis was to develop new sensory proteins by computational design and mutagenesis that could be used for biosensing purposes in whole cell bacterial bioreporters. Our main focus was on the ribose-binding protein RbsB of *E. coli* that had been proposed as a flexible platform for mutagenesis and ligand-range expansion. Our main strategy consisted of demonstrating to change RbsB ligand specificity from ribose to non-natural compounds. We expected that amino acid substitutions would not only change binding specificity, but also help us to better understand the molecular binding, folding and signaling mechanisms of this protein.

The first aim was to develop a precise and user-friendly high-throughput strategy to screen a library of 2 million mutants, based on a bioreporter assay measured by fluorescence activated cell sorting (FACS) and flow cytometer. Our goal was to isolate RbsB variants with changed specificity from ribose to 1,3-cyclohexanediol (13CHD) and/or cyclohexanol (CH). This is described in Chapter 2.

The second aim was to improve the binding affinity and/or solved stability/unfolding issues of the isolated mutants (Chapter 2) with 13CHD binding capacity. These mutants were used as a scaffold to produce several new mutant libraries based on random mutagenesis, site saturation or DNA shuffling approaches. Here we tried not only to improve the protein function but also learn more about the stability and ligand binding mechanism of RbsB. This is described in Chapter 3.

Finally, our third aim was to develop an *in vivo* system to detect expression and subcellular localisation differences between wild-type RbsB and variants. Our goal was to develop a screening system to select variants with potentially improved folding, translocation and receptor interactions. This is described in Chapter 4.

CHAPTER 2

Computational redesign of the *Escherichia coli* ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol

Published as: **Diogo Tavares**, Artur Reimer, Shantanu Roy, Aurélie Joubin, Vladimir Sentchilo, and Jan Roelof van der Meer (2019). "Computational redesign of the *Escherichia coli* ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol." [Sci Rep 9, 16940](#)

Author contributions

D.T., A.R., A. J. and V.S. performed experiments. S.R. and A.R. conducted computational designs. D. T., A. R. and J.R.M. prepared Figures. D.T., A.R. and V.S. contributed strains. D.T., A.R. and J.R.M. wrote the main manuscript. All authors reviewed the final manuscript.

Abstract

Bacterial periplasmic-binding proteins have been acclaimed as general biosensing platform, but their range of natural ligands is too limited for optimal development of chemical compound detection. Computational redesign of the ligand-binding pocket of periplasmic-binding proteins may yield variants with new properties, but, despite earlier claims, genuine changes of specificity to non-natural ligands have so far not been achieved. In order to better understand the reasons of such limited success, we revisited here the *Escherichia coli* RbsB ribose-binding protein, aiming to achieve perceptible transition from ribose to structurally related chemical ligands 1,3-cyclohexanediol and cyclohexanol. Combinations of mutations were computationally predicted for nine residues in the RbsB binding pocket, then synthesized and tested in an *E. coli* reporter chassis. Two million variants were screened in a microcolony-in-bead fluorescence-assisted sorting procedure, which yielded six mutants no longer responsive to ribose but with 1.2-1.5 times induction in presence of 1 mM 1,3-cyclohexanediol, one of which responded to cyclohexanol as well. Isothermal microcalorimetry confirmed 1,3-cyclohexanediol binding, although only two mutant proteins were sufficiently stable upon purification. Circular dichroism spectroscopy indicated discernable structural differences between these two mutant proteins and wild-type RbsB. This and further quantification of periplasmic-space abundance suggested most mutants to be prone to misfolding and/or with defects in translocation compared to wild-type. Our results thus affirm that computational design and library screening can yield RbsB mutants with recognition of non-natural but structurally similar ligands. The inherent arisal of protein instability or misfolding concomitant with designed altered ligand-binding pockets should be overcome by new experimental strategies or by improved future protein design algorithms.

Introduction

Periplasmic binding proteins (PBPs) form a versatile superfamily of proteins with a conserved protein structure, named the bilobal structural fold ^[1, 2]. PBPs facilitate nutrient and trace mineral scavenging for bacterial cells, by binding the ligand in the periplasmic space at high affinity and delivering the bound-ligand to a specific membrane-spanning transport channel ^[2]. Some PBPs are additionally involved in chemotactic sensing and interact in the ligand-bound state with a membrane-located chemoreceptor ^[3]. The crystal structures of several PBPs have been determined, showing two domains connected by a hinge region, with the binding pocket located between the two domains ^[3]. Both structure and nuclear-magnetic resonance data indicate that PBPs switch between two semi-stable conformations. Without ligand the protein adopts an *open* conformation, in which the binding site is exposed. Suitable ligand molecules become buried within the surrounding protein, and stabilize the *closed* protein form ^[4, 5]. High-quality crystal structures of various PBPs have been determined, and this triggered pioneering ideas more than a decade ago to deploy PBPs as a generalized platform for computational design-based construction of new ligand-binding properties ^[6]. PBPs form an interesting class of proteins for biosensing. Biosensing can be achieved by measuring the intermolecular motion of the purified protein itself upon interaction with the target ligand ^[7]. Alternatively, the PBP protein is expressed in a living bacterial cell and triggers a synthetic signaling cascade upon ligand binding. This principle is embedded in so-called bioreporter cells or bactorsensors ^[8]. By maintaining a single unique signaling cascade and reporter output, but varying the PBP-element with different ligand recognition, one could potentially develop a wide class of applicable bioreporters.

The concept of computational design of PBP variants with novel ligand-binding properties was proposed over a decade ago by the group of Hellinga and coworkers ^[9]. On the examples of five different PBPs in *Escherichia coli* they predicted and constructed mutant variants with binding pockets accommodating the non-natural substrates trinitrotoluene (TNT), lactate or serotonin at reported nM–mM *in vitro* affinities ^[9]. Particularly mutants of the ribose binding protein (RbsB) for TNT were further embedded in an *E. coli* synthetic bioreporter, in which ligand-bound RbsB-mutant contacts the Trz1 hybrid membrane receptor, increasing expression of a reporter gene fused to the *ompC* promoter ^[9]. This Trz1 receptor consists of a fusion of the 230 C-terminal amino acids of the *E. coli* EnvZ osmoregulation histidine kinase to the 265 N-terminal amino acids of the Trg methyl-accepting chemotaxis receptor protein ^[10]. Contact activation of Trz1 by ligand-bound RbsB triggers autophosphorylation of the cytoplasmic EnvZ-domain, leading to subsequent phosphorylation of the cognate response regulator OmpR, which activates the *ompC* promoter ^[11]. Independent engineering

of the most sensitive published RbsB mutant (named TNT.R3), however, failed to reproduce the reported TNT detection at sub- μ M concentrations in the *E. coli* Trz1-OmpR background and also failed to demonstrate TNT binding by a purified TNT.R3 mutant using *in vitro* microcalorimetry ^[12]. Subsequent analysis of effects of alanine-substitutions in wild-type *E. coli* RbsB showed that mutations at 12 positions result in misfolded or poorly translocated proteins, one of which was also targeted in the TNT.R3 variant ^[13]. Purification and biophysical analysis of a further set of published mutant PBPs also failed to reproduce the original measurements, and suggested the cause being their misfolding and unintended oligomerization ^[14]. The initial studies may thus have underestimated to a large extent the propensity of PBPs to become misfolded as a result of binding pocket mutations.

More recently, ligand-specificities have been successfully interchanged between PBPs by using binding-pocket grafting (i.e., exchange of binding pockets between functionally closely related PBPs) ^[15, 16], improved prediction of native ligand binding ^[17] and statistical coupling analysis (i.e., the prediction of mutations based on correlating amino acid residues in sectors of two classes of related proteins) ^[15]. So far, however, there have been no reports of non-cognate altered ligand-binding properties of PBPs. The goal of the underlying work was thus to revisit the concept of computational prediction of altered ligand-binding in RbsB. Because of the apparent difficulties to predict structure-function related side-effects such as protein folding, we hypothesized that predictions of minor changes in ligand-specificity might be more successful than major ones (e.g., from ribose to TNT). We thus chose to target molecules structurally related to ribose, in particular, 1,3-cyclohexanediol (13CHD) and cyclohexanol (CH). The computational protein design was based on exploration of sequence-space and estimations of Free energy of binding using Rosetta ^[18-20], to identify a list of mutated protein sequences with potentially sufficiently low energy of binding with the new target ligands. The DNA encoding for a large set of approximately 2 million mutants was then chemically synthesized and cloned into a vector for screening of inducible GFP expression in the *E. coli* Trz1-OmpR, *ompCp::gfp* signaling reporter background ^[12]. Mutant libraries were screened on bead-encapsulated microcolony-grown cells by flow cytometry and sorted using fluorescence-assisted bead-sorting (FABS) (Fig. 1). Positively-responding mutant strains were recovered, their RbsB mutant proteins were purified and further characterized for *in vitro* ligand binding by isothermal microcalorimetry. Periplasmic abundance of the mutant proteins was quantified by peptide mass-spectrometry in comparison to wild-type RbsB, and their folding was addressed by circular dichroism spectroscopy. We recovered a small number of mutants with modest inducibility but significant change in ligand-binding specificity compared to wild-type RbsB and ribose, indicating that the computational design correctly targeted

the intended new ligand-binding properties. However, despite some of gain of inducibility, our results indicate the mutant proteins to be unstable, and prone to misfolding during synthesis and secretion.

Results

RbsB mutant library design

Exploration of sequence space using Rosetta enzyme design simulations produced a library of targeted RbsB mutants, which were predicted to have improved affinity for the non-cognate ligands 13CHD and/or CH. The used design template was the scaffold of wild-type ribose-binding protein RbsB of *E. coli* in its closed configuration (PDB ID: 2DRI, Fig. 2A). First, prior to the design simulations, interactions between RbsB and ribose, 13CHD or CH were studied using docking and molecular dynamics simulations in CHARMM and Merck Molecular Force Fields (MMFF), to gain intuition on the stability of the binding pocket. The average spatial deviation of the RbsB binding pocket with placed ribose during 2 ns (as the root-mean squared deviation) was less than 0.2 Å, but for docked 13CHD and CH molecules was around 0.3 Å (Fig. S1). The ligands themselves showed varying positions with an average root-mean squared deviation of 0.7 Å for ribose, 1 Å for 13CHD and 1.5 Å for CH (Fig. S1). Calculated free energies of binding ($\Delta G_{\text{binding}}$) from CHARMM and MMFF were lowest for ribose, as expected, with $-38.35 \text{ kcal mol}^{-1}$, but $-19.57 \text{ kcal mol}^{-1}$ for 13CHD and $-14.31 \text{ kcal mol}^{-1}$ for CH. These results indicated unstable interactions of 13CHD and CH in the wild-type RbsB binding pocket. We conducted a per-residue free energy decomposition analysis^[21] using the simulation trajectories. The poorer $\Delta G_{\text{binding}}$ of 13CHD and CH seemed largely contributed by the RbsB residues D89, R90 and D215 (for 13CHD), and D89, D191 and D215 (for CH) (Fig. S1).

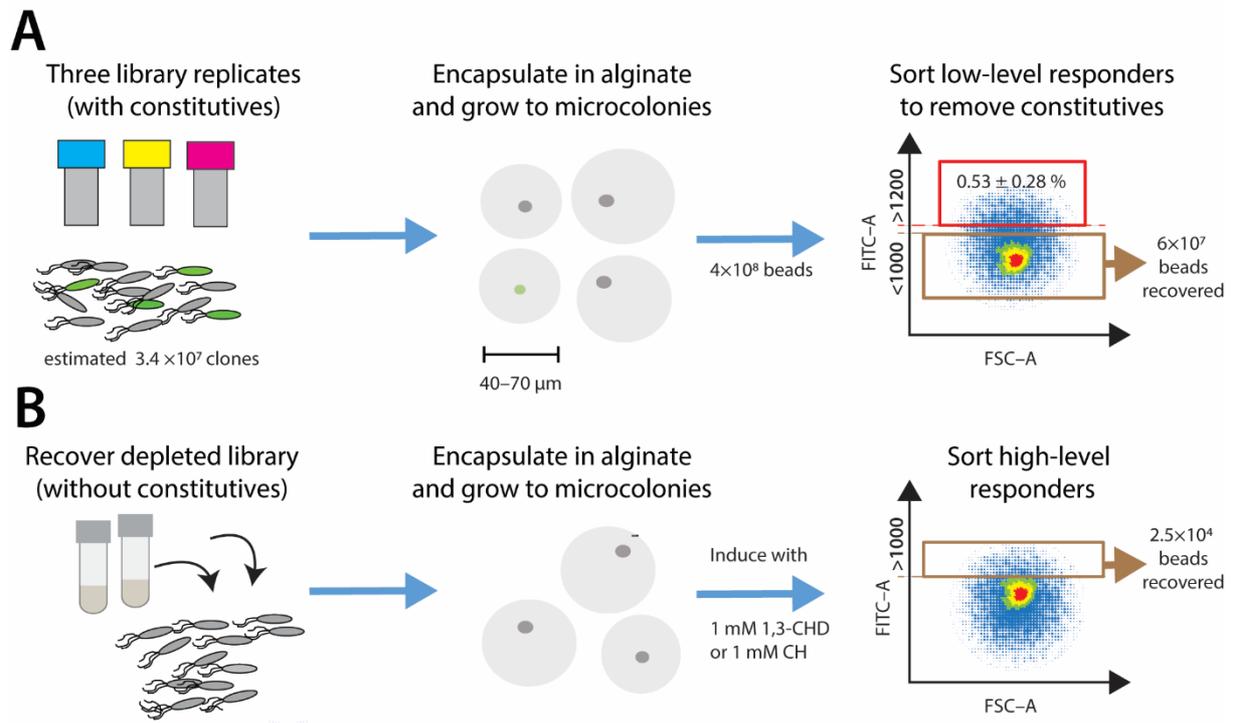
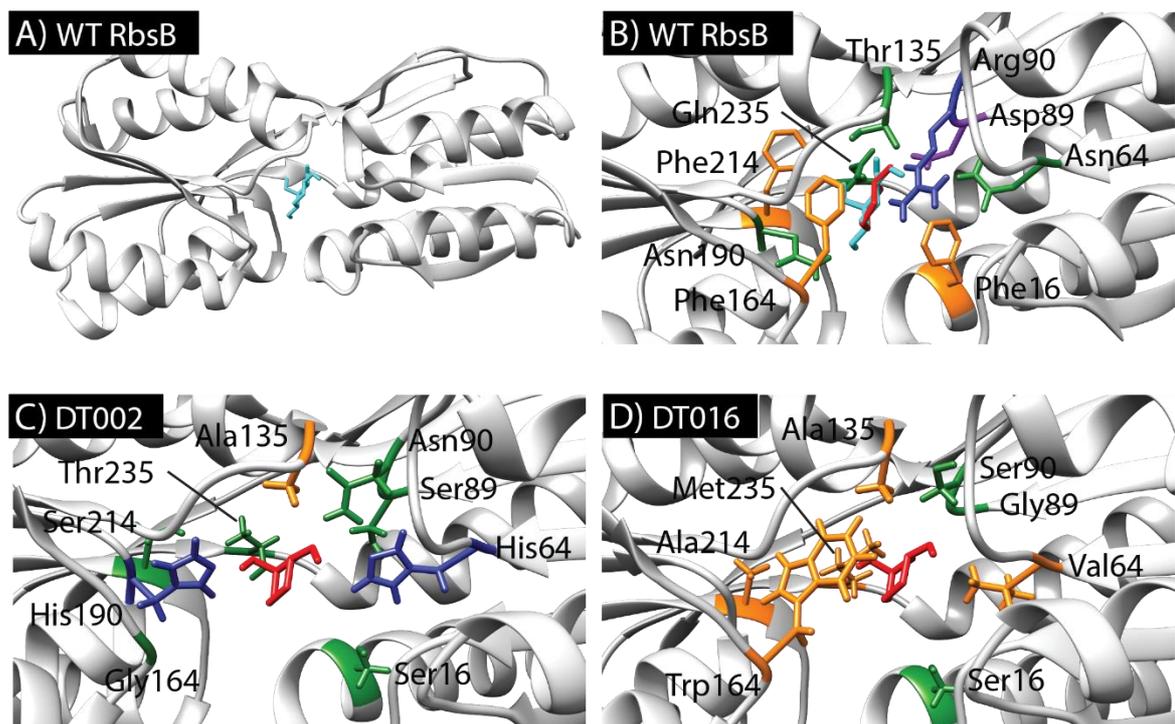


Figure 1- Overview of the mutant library screening strategy. A) Library replicates are encapsulated as on average one cell per agarose bead and grown to microcolony size in fumarate medium without inducer. Beads are passed on flow cytometry and beads with GFP fluorescence intensity below 1000 U are sorted and recovered. B) Cells are recovered from collected beads, again encapsulated and grown to microcolonies, after which they are exposed to the new ligands. Beads with GFP fluorescence intensity higher than 1000 U are recovered and further screened.

Next, we used Rosetta to predict potential beneficial mutations in RbsB for binding of 13CHD and CH. Defined key residues in the RbsB binding pocket (Table 1, Fig. 2B) were computationally replaced by alanine. The 13CHD and CH molecules were then docked 1000 times independently into the “stripped” binding pocket to obtain the positions with the lowest minimal binding energy, which were used as a starting point for the design mode. Design simulations (100 repetitions each producing 100 designs) then explored the combinatorial mutations on the defined 9 positions, from which pool 200 sequences were ranked according to the lowest predicted binding energy, minimal packing energy, hydrogen bond counts, and ligand-solvent exposure. This yielded a list of one of four possible amino acids at the 9 positions in RbsB (Table 1). The DNA encoding these RbsB variants in all their combinations plus the original wild-type residue was produced as a mixed library by DNA synthesis, cloned and introduced into an *E. coli* host enabling GFPmut2 production through the hybrid Trz1-OmpR signaling chain (strain 4172, Table 2, Fig. S2). Independent cloning reactions resulted in three mutant libraries with estimated sizes of 7×10^5 , 24×10^6 and 3.3×10^6 primary transformants.

A total of 4×10^8 alginate beads encapsulating individual cells from the mutant libraries and grown to microcolonies was screened by FABS for GFPmut2 fluorescence, in first instance in absence of inducer (Fig. 1A). An estimated 0.53 ± 0.28 % of the screened beads displayed fluorescence above 1200 units under non-induced conditions, and were considered constitutive–ON mutants. Approximately 60 million beads with fluorescence below 1000 units were sorted and recovered as mixture. Cells were released from the beads, freshly cultured, encapsulated in new alginate beads, regrown to microcolonies and induced with 1 mM 13CHD (Fig. 1B). In this second phase, beads with a fluorescence level higher than 1000 units were sorted and plated to grow individual colonies (a total of 2.3×10^4). After rescreening six mutants displayed consistently between 1.2–1.5-fold higher GFP fluorescence upon incubation with 1 mM 13CHD in comparison to media alone, which was a moderate response but statistically significant (p -values < 0.05). These mutants were no longer inducible and even slightly inhibited with 0.1 mM ribose (Table 3). In contrast, the same *E. coli* host expressing wild-type RbsB was not inducible with 13CHD but is 13-fold inducible with 0.1 mM ribose (Table 3). Only one of the six mutants (DT016) responded to 1 mM CH with a statistically significant increase in GFP fluorescence (Table 3).



E

	Residue number									
Protein	16	64	89	90	135	164	190	214	235	
RbsB	Phe	Asn	Asp	Arg	Thr	Phe	Asn	Phe	Gln	<ul style="list-style-type: none"> Hydrophobic Polar Positively charged Negatively charged
DT001	Phe	Asn	Asp	Leu	Arg	Gly	His	Phe	Thr	
DT002	Ser	His	Ser	Asn	Ala	Gly	His	Ser	Thr	
DT011	Phe	Val	Asp	Thr	Val	Gly	His	His	Thr	
DT013	Ala	Ser	Thr	Ser	Ala	Asp	Ala	Ser	Leu	
DT015	Ala	Asn	Thr	Thr	Ser	His	His	Ala	Thr	
DT016	Ser	Val	Gly	Ser	Ala	Trp	Asn	Ala	Met	

Figure 2 – Structures of wild-type RbsB (WT RbsB) and DT002, DT016 mutants. (A) Global view of closed RbsB (PDB ID: 2DRI) molecular structure with ribose (cyan) bound in its pocket. (B) Details of the RbsB binding pocket with 13CHD (red) and ribose (cyan) molecules. Critical amino acid residues for substrate binding are indicated and color-coded based on amino acid characteristics (nonpolar- orange; positively charged- blue; polar- green; negatively charged- purple). (C) Details of the DT002 binding pocket (threaded on the RbsB structure, PDB ID: 2DRI) with 13CHD (red, placed according to docking with Rosetta) with indication of mutated amino acids. (D) Same as in (C) but for the DT016 mutant. (E) Overview of the targeted residues in the recovered mutants compared to wild-type RbsB.

All six recovered mutants contained different amino acid substitutions, with some, but little overlap (Fig. 2E). Mutant DT001 displayed five mutations and four wild-type residues at the 9 targeted positions, followed by DT011 with 7 mutations, DT015 and DT016 with 8, and clones DT002 and DT013 with all 9 targets substituted (Fig. 2E). In the majority of the 13CHD-responsive mutants, positions D89, R90 and Q235 were replaced by a polar residue, whereas position T135 was replaced by a non-polar residue. Also, in four of six mutants, N190 was substituted by a histidine (Fig. 2E).

Reduced periplasmic space abundance of RbsB mutants responsive to 13CHD

The relative periplasmic space abundance of four RbsB mutants determined by quantitative mass spectrometry was lower compared to wild-type RbsB (Table 4). The DT002 and DT015 proteins displayed the lowest relative abundance, followed by DT011 and DT001. The periplasmic abundance of mutant DT016 was 2 times higher than RbsB. Interestingly, the relative abundance of MglB (galactose-binding protein) was higher in the periplasmic space of *E. coli* expressing mutants DT011 or DT015, in comparison to those expressing wild-type RbsB or the other mutant proteins (Table 4). Also, the summed abundance of all periplasmic binding proteins (excluding RbsB) was higher in all *E. coli* expressing RbsB mutant proteins than wild-type, with up to between 4.5 and 6 times increase in mutants DT015 and DT011 (Table 4). Quantitative mass spectrometry data thus suggest that translocation was affected for most RbsB mutant proteins and that this also influenced the translocation of other periplasmic binding proteins to the periplasm.

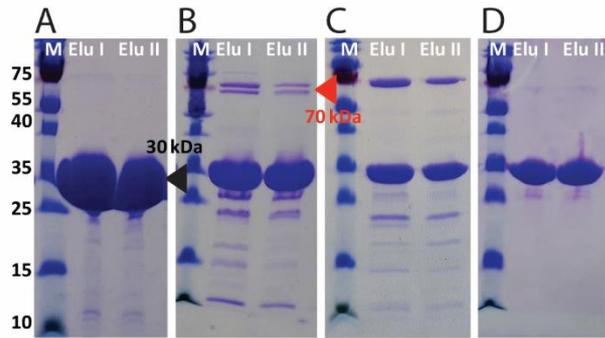


Figure 3- Overexpression and purification of RbsB-His₆ and mutants DT002-His₆ and DT016-His₆. (A) SDS-PAGE gel of purification steps with a HisTrap column of RbsB-His₆. (B) and (C) as for panel (A) but for mutants DT016-His₆ and DT002-His₆, respectively. (D) SDS-PAGE gel of elution steps after a gel filtration column for mutant DT016-His₆. M, Marker; Elu, Elution step. Black triangle indicates the expected position of RbsB-His₆, DT002-His₆ and DT016-His₆ proteins. Red triangle indicates the position of the assumed *E. coli* chaperones. Images in panels (A)-(D) stem from single individual SDS-PAGE gels, as indicated by the white line separator and panel lettering. Individual panel images and lanes were not further combined digitally and show the full protein size range.

In vitro 13CHD binding by mutant RbsB

Cytoplasmic overexpressed His₆-tagged RbsB was readily purified and resulted in protein with >97% purity on SDS-PAGE and a molecular mass of around 30 kDa as expected (Fig. 3A, black triangle). Contrary to wild-type RbsB, contaminating proteins were consistently observed in purified His₆-tagged mutant RbsBs. One or two prominent contaminants with a mass of around 70 kDa were observed after affinity (Fig. 3B and C, red triangle) and gel-filtration columns. These contaminants contributed to an estimated 5-15 % of the total protein quantity. Interestingly, addition of 10 mM ATP to the eluted protein fraction after affinity purification but before gel filtration led to removal of these contaminants (Fig. 3D). Possibly, therefore, they consisted of *E. coli* chaperones such as Hsp70 or DnaK, involved in protein folding and refolding^[22], which remained attached to the mutant RbsBs and detached upon addition of ATP. This suggests that the mutant RbsB proteins produced in the *E. coli* cytoplasm suffer from partial misfolding and are stabilized by chaperones^[23].

Binding of ribose to wild-type purified RbsB-His₆ in isothermal titration calorimetry (ITC) resulted in a clear heat release with an estimated binding affinity constant K_D of 530 nM (Fig. 4A), which is similar to literature values^[14]. Purified RbsB-His₆ showed no significant interaction with 13CHD (Fig. 4D). In contrast, modest but consistent heat release was observed with purified DT002-His₆ and DT016-His₆ in presence of 13CHD in comparison to buffer alone (Fig. 4B, C). Assuming binding of a

single 13CHD ligand per protomer, we found an apparent K_D of 190 μM for DT002-His₆ and 5 μM for DT016-His₆. Kinetic heat release and molar ratios suggested that actually only part of the purified protein fraction engages in binding of the ligand, possibly because another fraction was misfolded and inactive (Fig. 4B, C). No binding of ribose by either of the two mutant proteins was observed (Fig. 4E, F), and none of the other purified mutant proteins (DT001-His₆, DT011-His₆, DT013-His₆ or DT015-His₆) yielded measurable heat release with either 13CHD or ribose as substrates (not shown).

The RbsB protein fraction after affinity purification and gel filtration was stable and displayed consistent binding to ribose in ITC, even upon -80°C freezing and thawing of aliquoted fractions. In contrast, mutant protein fractions purified in the same manner except for addition of ATP before gel filtration were unstable. Purified DT002-His₆ fractions could be kept on ice for at least 4 h and produced similar heat release in ITC upon addition of 13CHD for three consecutive measurements. In contrast, after freezing at -80°C and thawing, the apparent binding affinity was reduced and sometimes even lost. 13CHD-binding to purified DT016-His₆ disappeared within 2 or 3 h after purification, even while maintaining the protein solution on ice. After -80°C freezing and thawing, the DT016-His₆ protein fraction no longer showed any heat-release from added 13CHD in ITC. These observations and the poor molar ratio of 13CHD binding (Fig. 4B, C) suggested that the DT002-His₆ and DT016-His₆ mutant proteins have strongly reduced stability and spontaneously misfold during purification and ITC. Not unlikely, the other four RbsB mutant proteins already completely misfolded during purification, and no sufficiently stable fractions were obtained to measure productive ligand-binding in ITC.

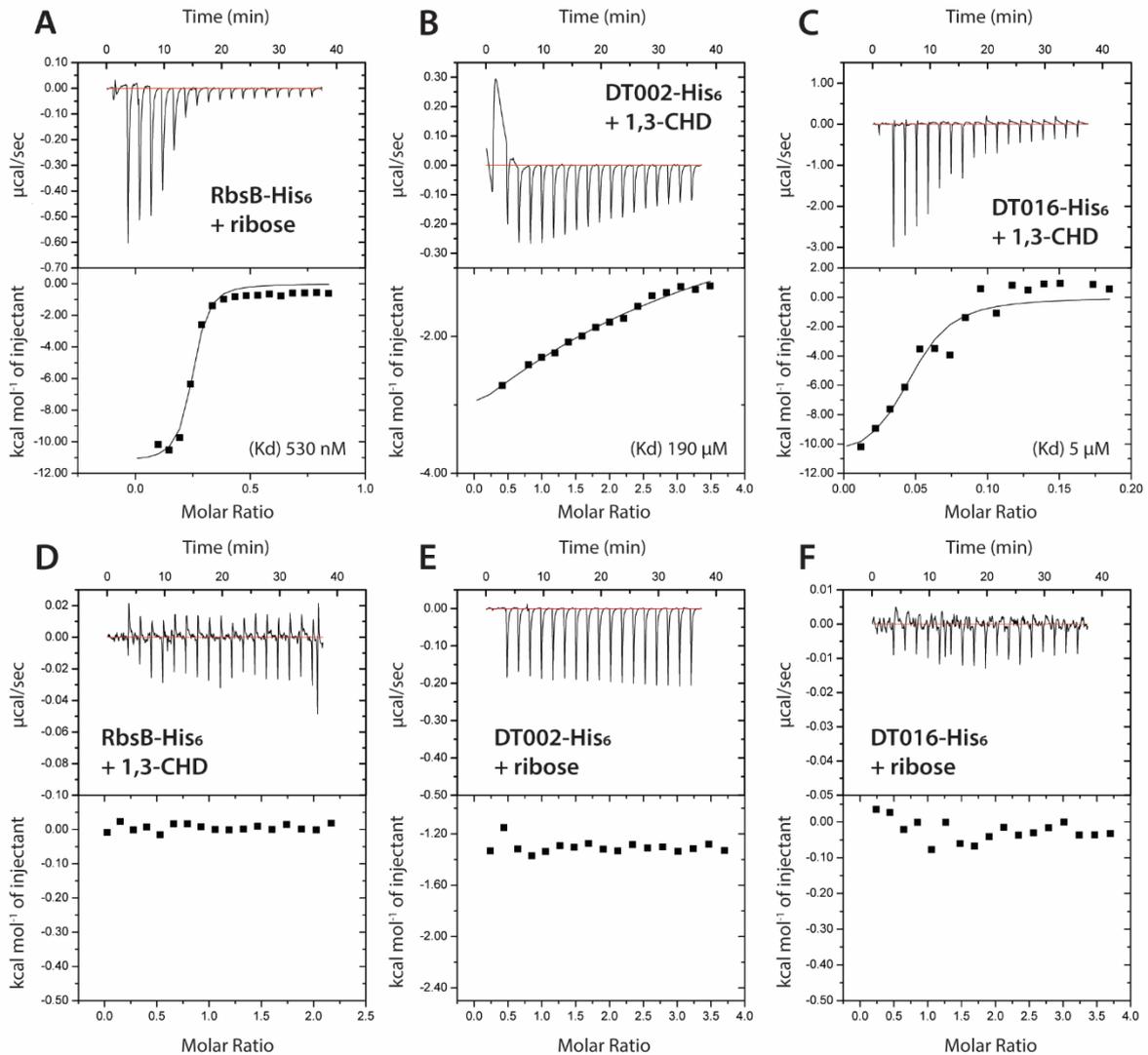


Figure 4- *In vitro* ligand binding measurements using isothermal microcalorimetry (ITC) with purified proteins. (A) Binding affinity of RbsB-His₆ protein with ribose. (B) Binding affinity of DT002-His₆ protein with 13CHD. (C) DT016-His₆ protein with 13CHD. (D) RbsB-His₆ protein with 13CHD. (E) DT002-His₆ protein with ribose. (F) DT016-His₆ protein with ribose. Kd, constant of affinity, assuming a single-ligand per protomer binding model. Graphs display immediate heat release in $\mu\text{cal s}^{-1}$ (upper panels) and calculated heat released per mol of injectant (lower panels).

Secondary structure changes in mutant proteins compared to RbsB

To detect secondary structure differences between wild-type and mutant proteins and observe ligand-induced changes, we analyzed purified protein fractions by circular dichroism spectroscopy in absence and presence of ligand (ribose or 13CHD, Fig. 5). All three His-tagged proteins (RbsB, DT002 and DT016) had similar circular dichroism spectra but with different $\Delta\epsilon$ intensities, which slightly (RbsB and DT002) or more importantly (DT016) increased upon addition of their ligands (Fig.

5A). Secondary structure protein-fold predictions from circular dichroism spectra using recently published tools ^[24] on repeated independently purified protein batches indicated DT002 and DT016 to carry smaller proportions of helices but increased proportions of anti-parallel/parallel and ‘turn’-folds compared to RbsB (Fig. 5B). This suggests notable distortions in the RbsB-folds as a result of the introduced mutations (Fig. 2). Addition of ribose to RbsB resulted in a notable predicted reduction of the antiparallel-2 relaxed fold (for definition, see Ref ^[24]) and an increase of ‘other’ folds and turns (Fig. 5B). This might correspond to the closed configuration of the protein (see, e.g., Fig. 3B in Reimer *et al.* ^[13]). This decrease of the proportion of antiparallel-2 relaxed fold was also observed in one preparation of the purified DT016-protein after addition of 13CHD (see asterisk within Fig. 5B), but not with addition of ribose or CH. Addition of ligands to DT002 protein preparations did not cause any consistent or pronounced changes in the predicted secondary structure fold composition (Fig. 5B).

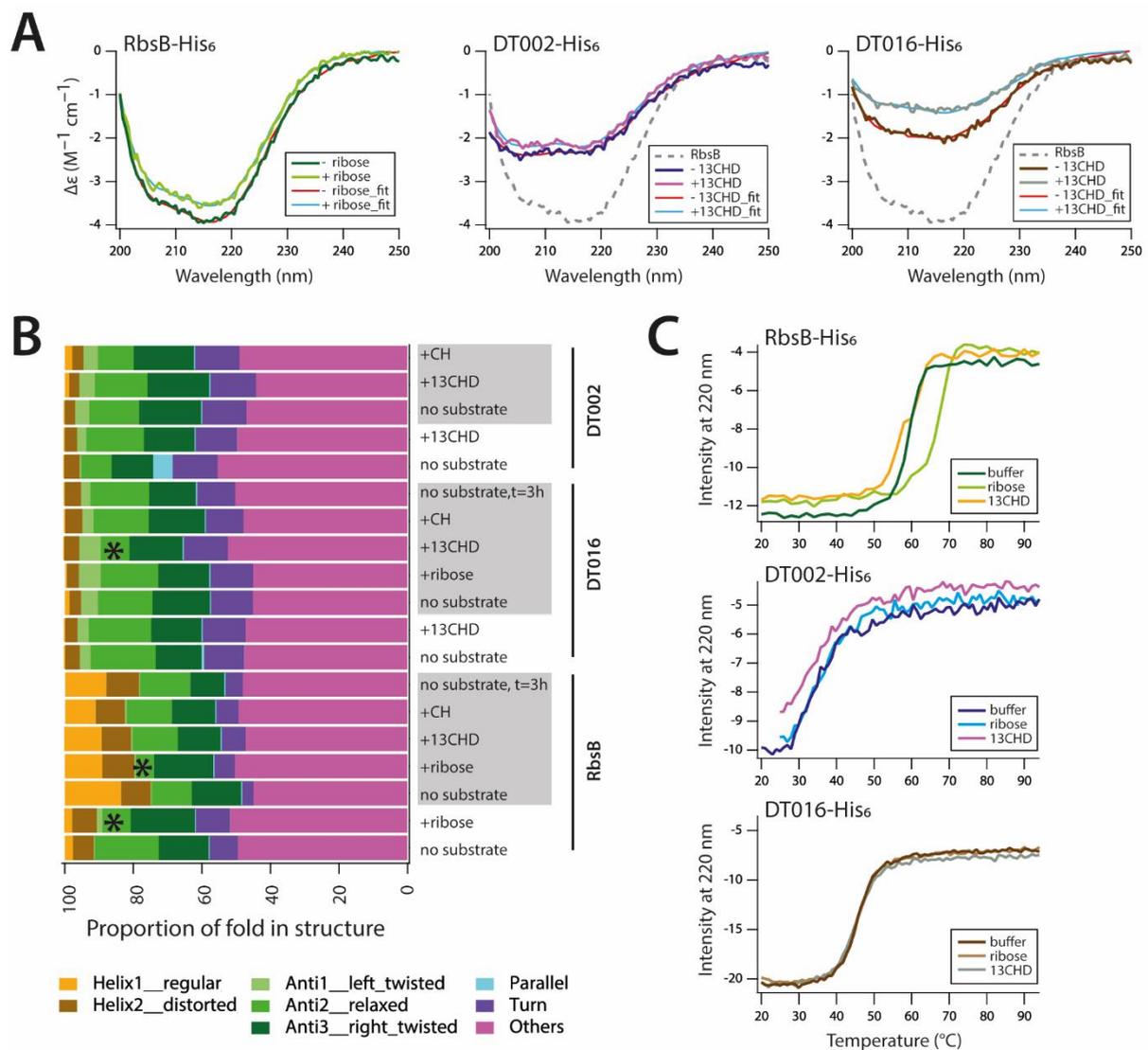


Figure 5- Secondary structure analysis and temperature melting curves of purified wild-type RbsB-His₆, DT002-His₆ and DT016-His₆. (A) Circular dichroism spectra of purified proteins in buffer A (without imidazole) at a protein concentration of 0.1 mg ml⁻¹, in absence or presence of inducer (ribose, 0.1 mM, CH or 13CHD, 1 mM). Spectra were fitted according to reference^[24]. (B) Inferred secondary structure fold composition of the three purified proteins in two independent purified batches (alternating white or grey background), in presence or absence of ligands. Asterisks note the significant changes in antiparallel-2 relaxed protein fold upon productive ligand-addition. Protein fold terminology as in reference^[24]. (C) Melting curves of purified proteins at a protein concentration of 0.3 mg ml⁻¹, in buffer or in presence of ribose (0.1 mM) or 13CHD (1 mM).

Melting curves of wild-type, DT002 and DT016 purified protein fractions indicated important differences in thermal stability (Fig. 5C). Whereas wild-type RbsB showed a melting temperature (T_m) of 58.9±0.1°C (Sigmoidal curve fitting), that of DT016 was only 46.1±0.1°C and that of DT002 not more than 34.7±0.7°C (Fig. 5C). A robust shift in T_m of ~8°C was observed for RbsB upon addition of 0.1 mM ribose (Fig. 5C). This result is in accordance with previously reported data^[14, 25]. In presence of 1 mM 13CHD, however, the T_m of RbsB was slightly reduced to 58.5±0.2°C (Fig. 5C). In contrast, the melting temperatures of mutants DT016 and DT002 were not measurably affected by addition of ribose or 13CHD (Fig. 5C). This indicated that both DT002 and DT016 mutant proteins were indeed much less stable than wild-type RbsB and that interaction with 13CHD did not further stabilize the proteins.

Effect of secondary mutations

In order to potentially improve the stability and/or translocation to the periplasm of the two most promising mutants (DT002 and DT016), further targeted mutations were introduced on these proteins. By site-directed mutagenesis H190 was reverted to wild-type N190 in mutant DT002 (Fig. 2E). Previous studies indicated the importance of N190 on RbsB stability and/or translocation^[13, 25]. In the mutant DT016 the residue 235, which had also been suggested to be implicated in RbsB stability^[25], was changed from M235 to V235. Position W164 in the DT016 protein, which is very close to the binding pocket and might block ligand access, was randomized (Fig. 2D). Back mutation of H190N in DT002 led to complete loss of inducibility by 13CHD (Table 3). Also the M235V mutation in mutant DT002 led to complete loss of 13CHD inducibility. Of the randomized positions at W164 in mutant DT016, only a replacement to Gly maintained 13CHD and CH induction (Table 3). The periplasmic abundance of DT002_{H190N} improved compared to DT002, whereas that of DT016_{W164G} remained the same as that of DT016 (Table 4).

Discussion

PBPs have attracted wide interest because of their potential for biosensing and as universal scaffold for engineering ligand-binding properties *à la carte* [6]. However, despite detailed structure information on a number of PBPs [2], and their biochemical, biophysical and genetic characterization, this *à la carte* design has remained largely elusive [14, 26]. Structure-guided computational predictions to change ligand-binding specificities have been to some extent successful for other sensory-type proteins, such as transcription factors [26-28], but for PBPs have remained limited to small modifications of binding properties in existing ligands [29]. Using the well-characterized RbsB protein from *E. coli* as a model, we showed here that computational predictions of altered amino acid residues in the RbsB binding pocket can indeed lead to a change of functional binding of the cognate substrate (ribose) to foreign but chemically related ligands (13CHD and CH). We acknowledge that although the loss of ribose-binding by the derived mutant RbsB proteins is very clear, the gain of new functionality is detectable but small. The very modest functional gain is not surprising and has been more frequently observed in similar library-screening efforts for altered PBP ligand-binding pocket designs [15]. Our data suggest that the main reason for the limited functional gain is the apparent propensity of RbsB to become misfolded upon mutational redesign of the binding pocket. So far, we have not been able to improve these mutants further by secondary mutations.

Several lines of evidence support our conclusion that we obtained a true change of cognate ligand-binding specificity of RbsB to a non-natural ligand, starting from Rosetta simulations and predictions of improved 13CHD- and CH-binding by changes in 9 amino acids. First of all, six different mutant proteins were isolated from the synthesized clone library, with up to 1.5-fold times induction with 13CHD in the *E. coli trz1-ompCp-gfpmut2* reporter strain. These mutants had lost completely the capacity to become induced *in vivo* by ribose, and RbsB as well as the majority of other mutants in the library showed no induction with 13CHD (Table 3). Further mutation of a number of altered residues in the mutants DT002 and DT016 resulted in loss of 13CHD induction (Table 3). Secondly, ITC measurements confirmed binding of 13CHD by the mutants DT002 and DT016, with estimated K_D of 190 μM and 5 μM , respectively (Fig. 4B, C). This is indicative for poor binding, but a K_D of 5 μM is in the range of measured affinity (1.6 μM) of a grafted L-glutamine-binding domain on the *Salmonella typhimurium* LAO periplasmic binding protein [15]. Finally, circular dichroism spectroscopy and secondary structure fold-decomposition using a recent new approach [24] indicated structural changes to occur in DT016 upon 13CHD addition (Fig. 5B), although this was not consistently observed in independent protein preparations.

Our results corroborated previous observations that computationally designed PBP variants suffer from misfolding and instability ^[14, 15]. Despite showing some inducibility in the *E. coli* signaling reporter chain, four out of six proteins failed to show 13CHD binding in ITC, likely because they unfolded during purification. The two most stable mutant proteins DT002 and DT016, quickly lost functional activity in ITC upon purification and were dramatically less stable than wild-type in thermal denaturation (Fig. 5C). All mutant proteins showed evidence for chaperone co-purification after affinity chromatography, indicative for misfolding. This obviously hampers the screening of mutant libraries, given that the procedure relies on functional gain-of-GFP fluorescence (Fig. 1). If we assume that the observed induction is a combination of the 'true' affinity of the mutant protein for its new ligand and the ratio of correctly versus misfolded protein, the actual gain of 13CHD- and CH-binding may be higher than an induction factor of 1.5 suggests.

Detecting and separating mutants from the initial library with such small improvement of inducibility by 13CHD required optimization of the screening method. We initially screened the library for gain-of-fluorescence upon induction on individual cells, but found that single cell variation was too high and resulted in many false-positive signals ^[30]. Instead, therefore, we switched to growing reporter cells inside alginate beads to microcolonies, which improved the reproducibility and screening efficiency, and reduced the number of false positives. Others have recently compared such procedures and have come to the same conclusion ^[31]. Surprisingly, around 0.5 % of the clones in our library were constitutively 'ON', showing high GFP fluorescence even in absence of inducer. Constitutives may be the result of combinations of mutations stabilizing the closed configuration without inducer being present. Once in the closed conformation the RbsB 'ON'-mutant may bind to the membrane receptor Trz1 and trigger the bioreporter system. The addition of a step to first sort mutants with lower fluorescence levels in absence of inducer was essential to remove the constitutive 'ON'-mutants and improve the screening efficiency (Fig. 1A).

What do the recovered mutants tell us about potential ligand-binding in their designed pockets? The ribose-binding pocket of RbsB has been investigated in detail in previous studies. Vercillo *et al.* reported 13 amino acid positions (S9, N13, F15, F16, N64, D89, S103, I132, F164, N190, F214, D215 and Q235) to play an important role in ribose binding ^[25]. Molecular dynamics simulations suggested several of those to be limiting 13CHD- or CH-binding by RbsB (Fig. S2). In the final computational strategy we stripped the presumed RbsB binding pocket at nine positions (changing virtually to Ala-residues), sampled the positions for 13CHD and CH with the lowest ΔG , and predicted the sets of amino acid residues to contribute with improved 13CHD and CH binding. Although this seemed the best strategy at the time, a more recent and complete Ala-substitution screening of RbsB

by our group found four crucial residues for ribose recognition (D89, N190, D215, R141) ^[13], two of which (D215, R141) were not included in the library predictions. In contrast, that study found several residues critical for RbsB folding and/or translocation, notably D89 and N190, which were targeted here. Indeed, in four out of the six isolated 13CHD-responsive mutants the D89 residue was substituted by a polar amino acid. The N190 residue was substituted in 5 out of the 6 mutants, of which four times by a histidine, a positively charged amino acid (Fig. 2E). This may thus have indirectly contributed to mutant proteins with poorer stability. Perhaps not surprisingly at this stage, several different new binding pocket configurations appeared to confer measurable gain-of-function of 13CHD binding and loss of ribose binding (Fig. 2E). These converge to some extent in the character of the newly positioned amino acid residue, but not in their exact type. Given that the modeled mutant protein binding pockets may in reality deviate more than is suggested in Figure 2C and D, it is too speculative to infer how the introduced new amino acid residues might be contributing to the binding of 13CHD.

Mass spectra analysis revealed that all 13CHD-responsive mutant proteins, except DT016 and its derivative DT016_{W164G}, were less abundant in the periplasmic space than wild-type RbsB (Table 4). This is indirect evidence that the mutant proteins may have additional difficulties in translocation to the periplasm. Cells expressing DT011 and DT015 displayed higher periplasmic levels of MglB and other proteins (Table 4), which might be due to their increased flux through the Sec-translocation channel in absence of lesser abundant mutant RbsB. In case of cells expressing DT016 or DT016_{W164G}, periplasmic space abundance of the mutant protein was higher but that of other PBPs was lower, perhaps because of competition with RbsB through the Sec-channel (Table 4). In case of DT002, both its own periplasmic space abundance, as well as that of MglB and other PBPs, were lower. This may be the result of a partial blocking of the translocation system by the DT002 protein. Four mutant proteins (DT001, DT002, DT011 and DT015) with lower abundance in the periplasm than wild-type RbsB carried an amino acid substitution at the N190 position. Ala-substitutions at this position resulted in loss of ribose-induction, potential misfolding and/or poor translocation into the periplasm ^[13, 25]. Back mutation of the H190 residue in DT002 to an Asn, indeed increased its periplasmic space abundance (Table 4), but also resulted in loss of induction by 13CHD (Table 3). Mutant DT016, on the other hand, still retained asparagine at position 190 and its periplasmic space abundance was twice as high as wild-type RbsB protein (Table 4). Instead, we suspected that the bulky Trp-residue at position 164 in DT016 would limit protein flexibility in the entry and hinge regions (Fig. 2D), and perhaps be responsible for the high observed fluorescence background in absence of inducer (Table 3). Indeed, replacing the Trp by a Gly (DT016_{W164G}) resulted in a much lower background, similar periplasmic space abundance (Table 4) and retainment of 13CHD and CH induction potential (Table 3).

Our results underscore that design of new ligand properties in highly flexible proteins such as PBPs is very challenging. Scoring functions developed for evaluation of protein-ligand binding free energies are not accurate enough ^[26, 32], and it is extremely difficult to predict the intrinsic dynamics and conformational changes at the binding pocket caused by the interaction with the ligand ^[17, 33]. Further important advances have been made by grafting binding pockets between related PBPs ^[15, 16], although this has so far not expanded the spectrum to non-natural ligands. Therefore, we believe that our work is a crucial step forward and shows unmistakably non-natural ligand-binding properties in RbsB mutants. Future studies on PBPs should focus on either improving experimental methods to select for better folders while maintaining designed or grafted new ligand-binding pockets, or on improving computational predictions of stability and translocation of designed mutants.

Materials and Methods

Computational design

Affinities of 13CHD and CH for the wild-type RbsB binding pocket were estimated by molecular dynamics simulations. First, the energetically best docking position of ribose, 13CHD and CH on RbsB in its closed configuration removed of ribose (PDB: 2RDI) were simulated from 15,000 binding modes using SwissDock^[34]. The bound and unbound states of protein, ligand (best docked position) and solvent were then simulated *in silico* in an all-atom description. Solvation equations were simplified by restricting water molecules to a 20 Å-sphere around the RbsB binding pocket (this includes approximately 5000 water molecules). The sphere was divided into an inner reaction region (15 Å ϕ) and an outer buffer region, forming the stochastic boundary. Atom interactions were then simulated during 2 ns with 1 fs time-steps using CHARMM and MMFF forcefields^[35, 36] implemented in Stochastic Boundary Molecular Dynamics (SBMD^[37]), from where the root mean square deviation of the ligand and the protein was calculated.

Free energy of binding was estimated by using the Molecular Mechanics with Generalized Born and Surface Area approach (MM-GBSA), which takes bonded and non-bonded energy, electrostatic and non-polar parts of desolvation energy into account^[37]. The free energy was averaged from 250 frames of the 2 ns SBMD simulations.

For *in silico* design of the mutant library the Rosetta protein design software package was used. In particular, the ligand docking and the enzyme design modules within the Rosetta framework^[18] were used to predict amino acid changes in RbsB to potentially allow binding of 13CHD and CH. The protein design in Rosetta was carried out as a probabilistic simulated annealing algorithm for exploring the sequence space via rotamer replacement and optimization. The scheme has the following components: I- To parametrize and optimize the interaction via force-field terms; II- To determine the target residue positions to set the design and the one to repack; III- To iterate the cycles of sequence design and minimization; IV- To optimize structures using fixed rotamers without constraints. This defined the key residues in the RbsB binding pocket to be targeted (Table 1, Fig. 2) and computationally to be replaced by alanine. Subsequently, the 13CHD and CH molecules were docked 1000 times independently into the "stripped" (Ala-substituted) RbsB binding pocket. The docked conformations were selected according to their energy values, and the ones with the lowest predicted energy values were used as a starting point for remodeling amino acid substitutions at the nine targeted positions. The final list of mutant positions was filtered according to the lowest minimal energy values, packing, hydrogen bond counts and ligand-solvent exposure.

Mutant library and plasmid construction

Based on the *in silico* computational predictions, all the possible combinations of four alternate residues plus wild-type at nine positions (5^9 combinations, Table 1) were produced by DNA synthesis as a mixture of linear DNA fragments (DNA2.0, USA). Delivered fragments were amplified by PCR with primers carrying tails that incorporated Sall and NdeI restriction sites. After restriction enzyme digestion and purification, the fragments were ligated with plasmid pSTV28P_{AA}^[12] digested with Sall and NdeI, which brings expression of the *rbsB*- or its mutant gene under control of the P_{AA} promoter^[38] (Fig. S2). Multiple ligation reactions were independently transformed into batches of *E. coli* DH5 α or MegaX-DH10B™ T1R Electrocomp™ cells (ThermoFisher Scientific). Transformants were cultured *en masse* on Luria-Bertani (LB) medium with chloramphenicol (Cm), from which the pool of pSTV28P_{AA}-library plasmids was isolated and purified. Batches of 200 ng library-plasmid DNA were subsequently transformed into competent cells of the reporter strain *E. coli* BW25113 $\Delta rbsB$ containing plasmid pSYK1 (containing *trz1* under control of the LacI-repressed *tac*-promoter, and the *ompCp-gfpmut2* reporter; strain 4172)^[12] (Fig. S2, Table 2). Small proportions of these transformed batches were plated to estimate the number of viable clones in the libraries. The remaining pooled library cultures were grown for 16 h in 10 mL of low phosphate minimal medium (MM LP) (Table S1) containing 20 mM fumarate as sole carbon and energy source, and supplemented with Ampicillin (Ap) at 100 $\mu\text{g ml}^{-1}$ and Cm at 30 $\mu\text{g ml}^{-1}$ to select for both plasmids. Batches of 1.5 mL were aliquoted and stored in 15% (v/v) glycerol at -80°C.

Individual mutant clones selected from FABS screening (see below) were grown on LB plus Cm and Ap, and both plasmids (the pSTV28P_{AA}-*rbsB*-mutant and pSYK1) were purified using NucleoSpin Plasmid columns (Machery-Nagel, Germany). Mutant *rbsB* genes were recovered on a fragment obtained by digestion with Sall and BstXI or XcmI, which was ligated into vector pET3d cut with the same enzymes^[39]. This places the *rbsB* (mutant) gene with the hexahistidine tag at the C-terminal end under control of the T₇ promoter, and removes the *rbsB* signal sequence for protein translocation to the periplasmic space. Ligations were transformed into *E. coli* BL21 (DE3) containing pLysS, for RbsB overexpression and purification (see below).

Two further RbsB mutant derivatives were produced individually by site-directed mutagenesis (DT002_{H190N} and DT016_{M235V}) and one by site-saturation mutagenesis (DT016_{W164G}), as follows. pSTV28P_{AA}-derivative plasmids containing the respective mutant *rbsB* gene were amplified by PCR using overlapping but reverse complementary primers with point mutations at the desired positions. PCR products were digested with DpnI to remove template DNA^[40]. After enzyme inactivation the PCR products were transformed into *E. coli* DH5 α cells. Transformant colonies were selected on LB with

Cm, plasmids were purified and the mutations in *rbsB* were verified by sequencing, after which they were transformed into the *E. coli* signaling reporter strain 4172 (see above) ^[12].

RbsB-based bioreporter assays

The capacity of RbsB or its mutants to induce the Trz1-OmpR *ompCp-gfpmut2* signaling chain in *E. coli* strain 4172 (Fig. S2) was assessed by flow cytometry, either on individual clones, uninduced or induced with the appropriate ligand in 96-well plates, or on mutant libraries with encapsulated cells grown to microcolonies in alginate microbeads, incubated as mixtures with or without inducer.

E. coli library aliquots of 50 μl (containing approximately 10^8 cells) were inoculated in 10 ml MM LP medium (Table S1) containing 20 mM fumarate, and supplemented with 100 $\mu\text{g ml}^{-1}$ Ap and 30 $\mu\text{g ml}^{-1}$ Cm. Library batches were grown overnight at 37°C and with 180 rpm shaking. The next morning, cultures were diluted with MM LP to a turbidity of 0.03 and mixed in a 10:1 v/v ratio with 1% (w/v) alginate (PRONOVA UP LVG, FMC, Norway) in MM LP solution, to encapsulate the cells at approximately one starting cell per bead. Alginate beads were then formed using a VAR J30 bead machine (Nisco, Switzerland) at nozzle size of 150 μm and pressure set to 4-5 bar, and sprayed into 100 mM CaCl_2 solution under constant stirring to solidify the alginate. This produces beads with an average diameter of 50 μm . After 1 h hardening in solution, cell-loaded beads were filtered sequentially through 40 and 70 μm mesh size nylon strainers (Corning Inc.) and washed with MM LP (Table S1). Recovered beads in the 40–70 μm –diameter range were then incubated for 16 h at 37°C in 5 ml MM LP containing 1 mM fumarate, Ap and Cm, in a rotating wheel (TC-7, New Brunswick/Eppendorf, Belgium). The next day, the cells had grown to microcolonies and individual beads were screened for fluorescence in a FACS Aria flow cytometer particle sorter (BD FACSAria Cell Sorter, Becton Dickinson, USA), equipped with a 100- μm nozzle at a flow rate of 2-5 $\mu\text{l s}^{-1}$ and a density of between 100-1000 particles μl^{-1} . Sensitivities for the FSC and FITC channels were set to 291 V and 435 V, respectively.

Microcolony-in-bead suspensions were screened first without induction and beads with fluorescence less than 1000 units were sorted to deplete the library of constitutive 'ON'-mutants (Fig. 1A). Sorted beads were collected in LB medium supplemented with Ap and Cm, and incubated overnight at 37°C and with 180 rpm rotary shaking to dissolve the alginate beads and grow the *E. coli* cells. Multiple library batches were sorted sequentially to cover the entire mutant library.

The library depleted of constitutives was grown in multiple batches on MM with fumarate, cells were encapsulated and grown to microcolonies as described above, followed by 2.5 h induction with 1 mM 13CHD. Microcolony-in-bead fluorescence was again screened by flow cytometry, and beads displaying fluorescence levels higher than 1000 units were sorted in pools by FABS into tubes

containing LB plus Ap and Cm medium (Fig. 1B). The collected beads were dissolved, regrown, and stored in 15% glycerol (v/v) at -80°C.

The resulting sub-libraries containing candidate 13CHD-responsive mutants were streaked on LB plates with Ap and Cm, and grown colonies were replica plated on MM agar with 20 mM fumarate, Ap and Cm (MM-FUM-ApCm) in presence or absence of 1 mM 13CHD. After 48 h incubation at 37°C the colonies were photographed under blue-light (Safe Imager Transilluminator, ThermoFisher) and their fluorescence intensities were compared by image analysis using the open source software: http://www.cheminfo.org/Image/Biology/Counting_plates/index.html?viewURL=https://couch.cheminfo.org/cheminfo-public/b616aba5eda653bf97ce9b776976aa4d/view.json?rev=37-e0a7762615ad2a544a1e5149ed1a2f21#). Colonies showing at least 1.25-fold increase in fluorescence in presence compared to absence of 13CHD were restreaked on LB-Ap-Cm agar plates and purified. Individual colonies were then inoculated in eightfold replicates in 96-well plates, containing per well 200 µl of MM-FUM-ApCm. 96-well plates were incubated overnight at 37°C and 700 rpm in a THERMOstar shaker (BMG LABTECH, Germany). The next morning, 5 µl culture of each well was transferred into a new well in a 96-well plate with 195 µl of fresh MM-FUM-ApCm. After 2 h incubation at 37°C, 100 µl from each well was transferred to the corresponding position of a new 96-well plate and immediately analyzed by flow cytometry to measure the uninduced fluorescence levels. To the remainder, 95 µl of fresh MM-FUM-ApCm, and 5 µl of inducer solution were added. This plate was incubated for another 2 h at 37°C, after which each well was again sampled for cellular fluorescence. As inducers we tested 0.1 mM ribose, 1 mM 13CHD and 1 mM CH (final concentrations in the assay). Cellular fluorescence was measured in 20 µl-aliquots, autosampled from each well by a Novocyte flow cytometer (ACEA Biosciences, USA), at an aspiration rate of 14 µl min⁻¹ and culture density between 1000-5000 cells s⁻¹. GFP fluorescence was recorded in the FITC-channel, which was set at a sensitivity of 441 V, and is reported as the average of the mean in each of the 8 replicates ± calculated standard deviations. Note that the fluorescence units of the FACS Aria (in Fig. 1) are not the same as the ones from the Novocyte (as in Table 4). Statistical significance was tested in pair-wise t-tests (one-sided, assuming increased response of the mutant).

Expression and periplasmic space abundance analysis of RbsB wild-type or mutant proteins

The abundance of RbsB wild-type and mutants in the *E. coli* periplasmic space was analyzed using direct peptide mass identification, as described previously^[12]. Periplasmic fraction was prepared by EDTA-ice treatment^[12] from *E. coli* BW25113 $\Delta rbsB$ carrying pSYK1 and the pSTVP_{AA}-*rbsB* derivatives (Table 2). Periplasmic protein fractions were separated by SDS-PAGE and proteins in the

size range between 28 and 36 kDa were excised from the gel. Proteins were analyzed by the UNIL Proteome Facility (<https://www.unil.ch/paf/en/home.html>). In short: samples were digested with trypsin and peptides were separated on an Ultimate 3000 Nano LC System (Dionex), followed by detection in a Thermo Scientific LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Mass spectra were analyzed using Scaffold Viewer 4, using thresholds of 99.9%. The minimum number of peptides for identification was 2.

RbsB-His₆ overexpression and purification

For purification of wild-type or mutant RbsB His₆-tagged protein, 250 ml *E. coli* BL21 (pLysS) cultures with the corresponding pET3d-derivative plasmids grown in LB-Ap-Cm medium at 37°C until a culture turbidity at 600 nm of 0.3, were induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration). Cultures were incubated further for 16 h at 20°C, after which the cells were harvested by centrifugation at 3,600 × *g* for 5 min at 4°C. Cell pellets were stored at –80°C until purification.

Thawed cell pellets were resuspended in 15 ml of buffer A (500 mM NaCl, 50 mM NaH₂PO₄, pH 8.0), containing 20 mM imidazole. The cell suspension was transferred to a metallic chamber containing a single metallic bead (50 ml chamber, Retsch, Germany) and frozen in liquid nitrogen for 1 minute. The cold chamber was transferred to a bead-beater machine (Oscillating Mill MM400, Retsch, Germany), and cells were crushed by constant vigorous shaking for 3 min at 30 s⁻¹. The extract was transferred to a 50 ml centrifuge tube, which was centrifuged at 16,000 × *g* at 4°C for 30 min, after which the lysate supernatant was transferred to a clean tube.

The clean lysate was next loaded onto a HisTrap HP column (HisTrap FF crude 1 ml, GE Healthcare) at 4°C and flow rate of 0.5 ml min⁻¹, followed by washes of, consecutively, 10 column volumes (cv, equal to 1 ml) of buffer A with 20 mM imidazole, 1.5 cv of buffer A with 40 mM imidazole and 1.5 cv of buffer A with 80 mM imidazole. Proteins were eluted with buffer A containing 250 mM imidazole in a total volume of 4 ml. The HisTrap eluate was subsequently loaded on a Superdex 200 10/300 GL 24 ml gel filtration column (GE Healthcare), and eluted with buffer A plus 250 mM imidazole at a flow rate of 0.75 ml min⁻¹. Protein eluates were collected in aliquots of 150 μl, which were immediately frozen in liquid nitrogen and stored at –80°C, or used immediately for ITC assays. In case of RbsB mutant proteins we tested the effect of adding 10 mM ATP to the eluted protein solution directly after the HisTrap column, in order to disassociate and remove contaminating *E. coli* chaperones before loading onto the Superdex 200 10/30 GL column.

Protein concentrations were determined by NanoDrop spectrophotometry (Thermo Scientific, USA), using absorbance at 280 nm. The theoretical molar extinction coefficient and molecular weight were used as parameters. Subsamples of 20 μl were analyzed by SDS-PAGE to examine protein purity.

Analysis of ligand binding using isothermal microcalorimetry (ITC)

A volume of 280 μl of purified protein extract (mostly the Superdex gel filtration eluate; between 5 and 10 mg protein ml^{-1}) was pipetted into the measurement cell of a MicroCal ITC200 isothermal titration calorimetry instrument (GE Healthcare Life Sciences, USA). To avoid potential further unfolding of mutant protein we directly analysed them in imidazole-containing buffer A without previous dialysis, but maintained exactly the same volume of buffer A with 250 mM imidazole in the reference cell. An appropriate concentration of the test ligand (ribose or 13CHD; either at 0.1 or at 1 mM in buffer A with 250 mM imidazole) was filled into the injection syringe. Heat release was measured at 25°C with a reference power of 11 $\mu\text{cal s}^{-1}$ and a stirring velocity of 1000 rpm. Raw data were recorded as changes in $\mu\text{cal s}^{-1}$, and regression curves were fitted based on a one-binding site model using the Microcal Origin software (GE Healthcare).

Circular dichroism and variable temperature measurements

Purified wild-type RbsB-His₆, and DT002-His₆ and DT016-His₆-mutant proteins were analyzed by circular dichroism spectroscopy and variable temperature measurements using a J810 spectropolarimeter (Jasco, Japan). A volume of 100 μl of purified protein immediately after Superdex gel filtration or from thawed protein fraction stored at -80°C , was loaded on a PD minitrapp G-25 column (GE Healthcare) and eluted with 500 μl of buffer A to remove the imidazole at 4°C. Protein in buffer A was then kept on ice and analyzed within 2 h for its circular dichroism spectrum.

Circular dichroism and thermal melting curves were determined in a quartz cuvette with a 0.1 cm path length (L). Spectra (θ , mdeg) were measured at room temperature between 200 and 260 nm at a scanning speed of 10 nm min^{-1} and a protein concentration of 0.1 mg ml^{-1} . Buffer A alone was used as negative control and its circular dichroism spectrum was subtracted from that of the protein fractions. Data were further normalized for $\Delta\epsilon$ ($\text{M}^{-1} \text{cm}^{-1}$) using the effective protein concentration (c , mg ml^{-1}) and the mean residue weight of RbsB (MRW, 109 Da), as follows:

$$\Delta\epsilon = (\theta \times 0.1 \times \text{MRW}) / (3298 \times c \times L)$$

Circular dichroism spectra were further analyzed on the BeStSel webserver for secondary structure fold composition, as per instructions in Ref ^[24].

Variable temperature measurements were conducted with a protein concentration of 0.3 mg ml^{-1} in buffer A (without imidazole) and the following parameters: Start temperature 20°C;

CHAPTER 2

temperature increment of 1°C; target temperature 95°C; temperature ramp rate of 2°C min⁻¹. Measurements were performed in buffer alone or in presence of ribose at 0.1 mM or 13CHD at 1 mM final concentration.

Table 1 – Mutations introduced into *E. coli* wild-type RbsB protein

Amino acid	Wild-type residue	Tested mutations
16	F	H, T, S, A
64	N	V, I, H, A
89	D	G, T, S, V
90	R	S, N, T, L
135	T	S, A, R, V
164	F	D, G, H, W
190	N	H, S, T, A
214	F	H, A, S, N
235	Q	T, D, L, M

Table 2 – List of strains used in this study

Strain	Host	Plasmid(s)	Relevant characteristics	Reference or source
97	<i>E. coli</i> BL21 (DE3)	pLysS	Host strain for overexpression from the T7 promoter	[39]
3725	<i>E. coli</i> BL21 (DE3)	pLysS, pAR1	Cytoplasmic overexpression of RbsB-His ₆	This work
4172	<i>E. coli</i> BW25113 Δ rbsB	pSYK1	Host strain containing the P _{tac} -trz1, P _{ompC} -gfpmut2 bioreporter system	[12]
4175	<i>E. coli</i> BW25113 Δ rbsB	pAR3, pSYK1	Expression of RbsB with signal sequence for periplasmic translocation	[12]
5913	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT001, pSYK1	As 4175, but for DT001 mutant protein of RbsB	This work
5903	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT002, pSYK1	As 4175, but for DT002 mutant protein of RbsB	This Work
5904	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT011, pSYK1	As 4175, but for DT01 mutant protein of RbsB	This Work
5905	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT013, pSYK1	As 4175, but for DT013 mutant protein of RbsB	This Work
5906	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT015, pSYK1	As 4175, but for DT015 mutant protein of RbsB	This Work
5907	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT016, pSYK1	As 4175, but for DT016 mutant protein of RbsB	This Work
5999	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT016 (W164G), pSYK1	As 4175, but for DT016 _{W164G} mutant protein of RbsB	This Work
6054	<i>E. coli</i> BW25113 Δ rbsB	pSTV-DT016 (M235V), pSYK1	As 4175, but for DT016 _{M235V} mutant protein of RbsB	This Work
5927	<i>E. coli</i> BL21 (DE3)	pET3d-DT001, pLysS	Cytoplasmic overexpression of DT001-His ₆	This Work
5908	<i>E. coli</i> BL21 (DE3)	pET3d-DT002, pLysS	Cytoplasmic overexpression of DT002-His ₆	This Work
5909	<i>E. coli</i> BL21 (DE3)	pET3d-DT011, pLysS	Cytoplasmic overexpression of DT011-His ₆	This Work
5910	<i>E. coli</i> BL21 (DE3)	pET3d-DT013, pLysS	Cytoplasmic overexpression of DT013-His ₆	This Work
5911	<i>E. coli</i> BL21 (DE3)	pET3d-DT015, pLysS	Cytoplasmic overexpression of DT015-His ₆	This Work
5912	<i>E. coli</i> BL21 (DE3)	pET3d-DT016, pLysS	Cytoplasmic overexpression of DT016-His ₆	This Work
6016	<i>E. coli</i> BL21 (DE3)	pET3d-DT016(W164G), pLysS	Cytoplasmic overexpression of DT016 _{W164G} -His ₆	This Work

Table 3 – GFPmut2 induction in *Escherichia coli* expressing either wild-type or mutant RbsB.

Protein	GFPmut2 uninduced fluorescence	Fold induction ^a		
		Ribose ^b	1,3-Cyclohexanediol ^b	Cyclohexanol ^b
Wild-type	6180 ± 1580 ^c	13 ± 2.2^d	0.8 ± 0.03	0.8 ± 0.05
DT001	6846 ± 1658	0.9 ± 0.05	1.3 ± 0.16	0.9 ± 0.13
DT002	6019 ± 886	1.0 ± 0.05	1.4 ± 0.13	0.9 ± 0.08
DT011	8462 ± 440	0.9 ± 0.09	1.2 ± 0.04	0.8 ± 0.05
DT013	6154 ± 1319	0.9 ± 0.09	1.3 ± 0.6	0.8 ± 0.08
DT015	4481 ± 694	0.9 ± 0.16	1.3 ± 0.12	0.8 ± 0.11
DT016	24430 ± 3460	1.0 ± 0.08	1.5 ± 0.01	1.5 ± 0.05
DT002_{H190N}	5054 ± 805	0.9 ± 0.07	0.9 ± 0.15	0.7 ± 0.03
DT016_{W164G}	11627 ± 2028	0.8 ± 0.12	1.3 ± 0.04	1.3 ± 0.01
DT016_{M235V}	5960 ± 1300	0.8 ± 0.09	1.2 ± 0.08	1.2 ± 0.06

^a Mean GFPmut2 fluorescence in the assay with inducer divided by that of the assay with buffer only. Assay incubation time is 2h30 at 37°C. Values from eight replicates, ± calculated SD. ^b Final concentration of inducers in the assay: ribose, 0.1 mM; 13CHD, 1 mM; CH, 1 mM. ^c Mean values from eight replicates, ± calculated SD. ^d Values in bold indicate statistically significant induction (pair-wise T-test; p < 0.05, one-sided, equal variance).

Table 4 – Periplasmic abundance of wild-type or mutant RbsB proteins in *Escherichia coli*.

Expressed protein	Periplasmic abundance					
	Exclusive peptide count ^a			Normalized count ^b		
	(mutant) RbsB	MgIB ^c	Periplasmic binding proteins ^d	(mutant) RbsB	MgIB ^c	Periplasmic binding proteins ^d
RbsB	115	46	91	72	29	57
DT001	56	38	235	46	31	193
DT002	13	7	55	24	13	101
DT002 _{H190N}	59	38	301	53	34	268
DT011	39	161	348	42	175	379
DT015	24	202	407	16	133	268
DT016	143	19	141	181	24	179
DT016 _{W164G}	148	20	183	159	21	197

^a Absolute number of peptides exclusive to the indicated respective protein(s) detected by mass spectrometry. ^b Exclusive peptide count normalized to the total number of identified peptide fragments per sample. ^c MgIB, galactose-binding protein. ^d Number of all counted peptide fragments belonging to known other *Escherichia coli* periplasmic binding proteins (i.e., excluding RbsB and MgIB).

References

1. Chu Byron, C.H. and J. Vogel Hans, *A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport*, in *Biological Chemistry*. 2011. p. 39.
2. Berntsson, R.P.A., et al., *A structural classification of substrate-binding proteins*. *FEBS Letters*, 2010. **584**(12): p. 2606-2617.
3. Quioco, F.A. and P.S. Ledvina, *Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes*. *Molecular Microbiology*, 1996. **20**(1): p. 17-25.
4. Li, H.Y., et al., *Analysis of Conformational Motions and Residue Fluctuations for Escherichia coli Ribose-Binding Protein Revealed with Elastic Network Models*. *International Journal of Molecular Sciences*, 2013. **14**(5): p. 10552-10569.
5. Björkman, A.J. and S.L. Mowbray, *Multiple open forms of ribose-binding protein trace the path of its conformational change*¹¹*Edited by R. Huber*. *Journal of Molecular Biology*, 1998. **279**(3): p. 651-664.
6. Dwyer, M.A. and H.W. Hellinga, *Periplasmic binding proteins: a versatile superfamily for protein engineering*. *Current Opinion in Structural Biology*, 2004. **14**(4): p. 495-504.
7. Medintz, I.L. and J.R. Deschamps, *Maltose-binding protein: a versatile platform for prototyping biosensing*. *Curr Opin Biotechnol*, 2006. **17**(1): p. 17-27.
8. van der Meer, J.R. and S. Belkin, *Where microbiology meets microengineering: design and applications of reporter bacteria*. *Nat. Rev. Microbiol.*, 2010. **8**(7): p. 511-522.
9. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions*. *Nature*, 2003. **423**(6936): p. 185-190.
10. Baumgartner, J.W., et al., *Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ*. *Journal of Bacteriology*, 1994. **176**(4): p. 1157-1163.
11. Srividhya, K.V. and S. Krishnaswamy, *A simulation model of Escherichia coli osmoregulatory switch using E-CELL system*. *BMC Microbiology*, 2004. **4**: p. 44-44.
12. Reimer, A., et al., *Escherichia coli ribose binding protein based bioreporters revisited*. *Sci Rep*, 2014. **4**: p. 5626.
13. Reimer, A., et al., *Complete alanine scanning of the Escherichia coli RbsB ribose binding protein reveals residues important for chemoreceptor signaling and periplasmic abundance*. *Scientific Reports*, 2017. **7**: p. 8245.
14. Schreier, B., et al., *Computational design of ligand binding is not a solved problem*. *Proceedings of the National Academy of Sciences*, 2009. **106**(44): p. 18491-18496.

15. Banda-Vázquez, J., et al., *Redesign of LAOBP to bind novel l-amino acid ligands*. Protein Sci, 2018. **27**(5): p. 957-968.
16. Scheib, U., et al., *Change in protein-ligand specificity through binding pocket grafting*. J Struct Biol, 2014. **185**(2): p. 186-92.
17. Boas, F.E. and P.B. Harbury, *Design of protein-ligand binding based on the molecular-mechanics energy model*. J Mol Biol, 2008. **380**(2): p. 415-24.
18. Leaver-Fay, A., et al., *Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules*, in *Methods in Enzymology*, M.L. Johnson and L. Brand, Editors. 2011, Academic Press. p. 545-574.
19. Fleishman, S.J., et al., *Computational design of proteins targeting the conserved stem region of influenza hemagglutinin*. Science (New York, N.Y.), 2011. **332**(6031): p. 816-821.
20. Chevalier, A., et al., *Massively parallel de novo protein design for targeted therapeutics*. Nature, 2017. **550**(7674): p. 74-79.
21. Zoete, V., et al., *SwissParam: A fast force field generation tool for small organic molecules*. Journal of Computational Chemistry, 2011. **32**(11): p. 2359-2368.
22. Bertelsen, E.B., et al., *Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(21): p. 8471-8476.
23. Tokuriki, N. and D.S. Tawfik, *Stability effects of mutations and protein evolvability*. Curr Opin Struct Biol, 2009. **19**(5): p. 596-604.
24. Micsonai, A., et al., *BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra*. Nucl Acids Res, 2018. **46**(W1): p. W315-W322.
25. Vercillo, N.C., et al., *Analysis of ligand binding to a ribose biosensor using site-directed mutagenesis and fluorescence spectroscopy*. Protein Science : A Publication of the Protein Society, 2007. **16**(3): p. 362-368.
26. Yang, W. and L. Lai, *Computational design of ligand-binding proteins*. Current Opinion in Structural Biology, 2017. **45**(Supplement C): p. 67-73.
27. Taylor, N.D., et al., *Engineering an allosteric transcription factor to respond to new ligands*. Nat Methods, 2016. **13**(2): p. 177-83.
28. Ray, S., et al., *Structural basis of selective aromatic pollutant sensing by the effector binding domain of MopR, an NtrC family transcriptional regulator*. ACS Chem Biol, 2016. **11**(8): p. 2357-65.
29. Ko, W., S. Kim, and H.S. Lee, *Engineering a periplasmic binding protein for amino acid sensors with improved binding properties*. Organic & Biomolecular Chemistry, 2017. **15**(41): p. 8761-8769.

30. Reimer, A., *Development of synthetic signaling pathways based on periplasmic binding proteins and hybrid chemoreceptors*, in *Department of Fundamental Microbiology*. 2017, University of Lausanne: Lausanne. p. 205.
31. Duarte, J.M., I. Barbier, and Y. Schaerli, *Bacterial Microcolonies in Gel Beads for High-Throughput Screening of Libraries in Synthetic Biology*. ACS Synthetic Biology, 2017.
32. Feldmeier, K. and B. Höcker, *Computational protein design of ligand binding and catalysis*. Current Opinion in Chemical Biology, 2013. **17**(6): p. 929-933.
33. Stank, A., et al., *Protein Binding Pocket Dynamics*. Accounts of Chemical Research, 2016. **49**(5): p. 809-815.
34. Grosdidier, A., V. Zoete, and O. Michielin, *SwissDock, a protein-small molecule docking web service based on EADock DSS*. Nucleic Acids Res, 2011. **39**(Web Server issue): p. W270-7.
35. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
36. Neria, E., S. Fischer, and M. Karplus, *Simulation of activation free energies in molecular systems*. Journal of Chemical Physics, 1996. **105**(5): p. 1902-1921.
37. Zoete, V., M.B. Irving, and O. Michielin, *MM-GBSA binding free energy decomposition and T cell receptor engineering*. J Mol Recognit, 2010. **23**(2): p. 142-52.
38. Alper, H., et al., *Tuning genetic control through promoter engineering*. Proc. Natl. Acad. Sci. U. S. A., 2005. **102**(36): p. 12678-83.
39. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes*, in *Methods in Enzymology*, D.V. Goeddel, Editor. 1992, Academic Press, Inc.: San Diego. p. 60-89.
40. Vogne, C., S. Beggah, and J.R. van der Meer, *Random and site-directed mutagenesis of transcriptional regulator proteins implicated in hydrocarbon degradation pathways*, in *Handbook of Hydrocarbon and Lipid Microbiology*, K.N. Timmis, Editor. 2010, Springer: Heidelberg. p. 4429-4444.

Acknowledgements

The authors thank Luc Patigny and Julien Dénéréaz for their help in image analysis, and Jonas Canton for his help in initial ligand binding simulations. The Gruber lab at the Department of Fundamental Microbiology is thanked for their help and advice in protein purification. This work was supported by grants 244405 (Biomonar) and OCEAN-2013-614010 (BRAAVOO) from the European Seventh Framework Program (FP7). Patrice Waridel and Manfredo Quadroni from the University of Lausanne Protein Analysis Facility are thanked for their help in peptide analysis.

Supplementary Information

Table S1- Composition of mineral medium (MM) and low phosphate mineral medium (MM LP) used in this study.

Component	Mineral medium	Low phosphate mineral medium
Na ₂ HPO ₄	60 g	0.36 g
KH ₂ PO ₄	30 g	0.33g
NaCl		5 g
NH ₄ Cl		10 g

Recipe for 1L. pH set to 7.4

Supplemented with 1 ml l⁻¹ of Hutner's trace mineral solution as per reference ^[1] for 21C medium.

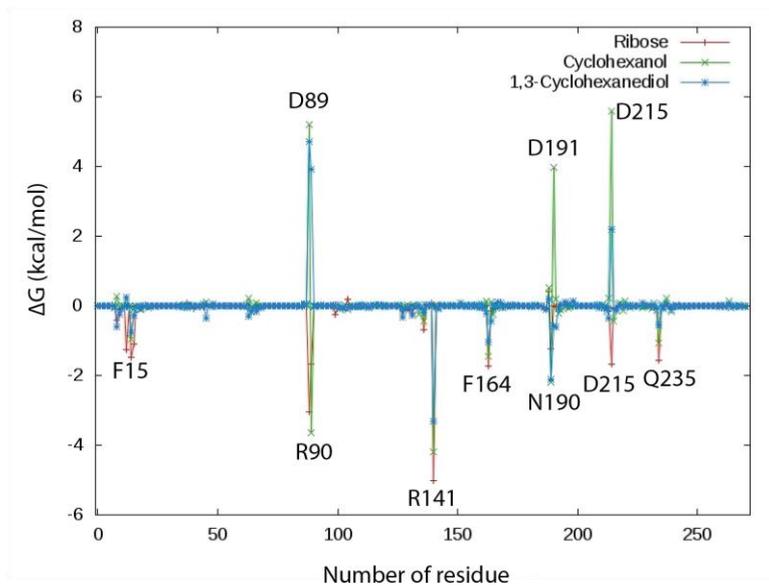
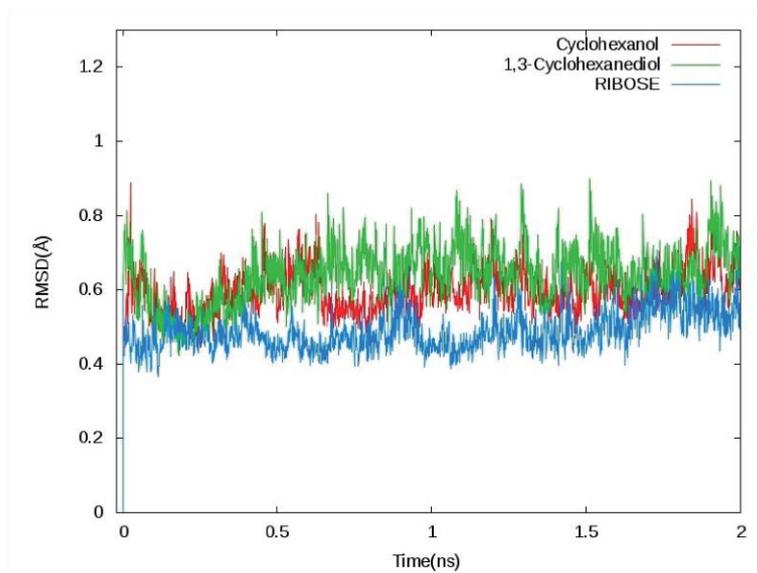
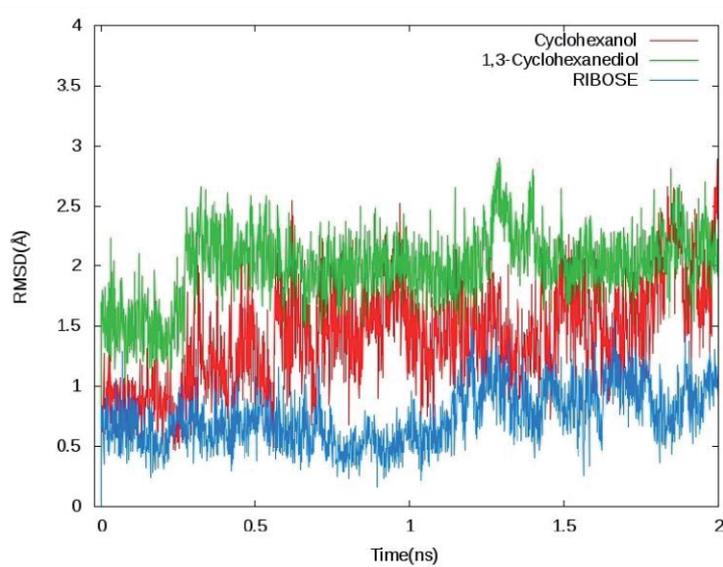
A**B****C**

Figure S1- Interactions between RbsB wild-type and the ligands ribose, 13CHD or CH using docking and molecular dynamics simulations. A) Contribution of each residue of RbsB to the change of Gibbs free energy ΔG (kcal/mol) during binding of the indicated ligand molecules using per-residue binding free energy decomposition based on Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method. B) Average spatial deviation of the RbsB binding pocket with placed ligands (ribose, 13CHD or CH) during 2 ns as root-mean squared deviation. C) As B, but taken from the ligand positions.

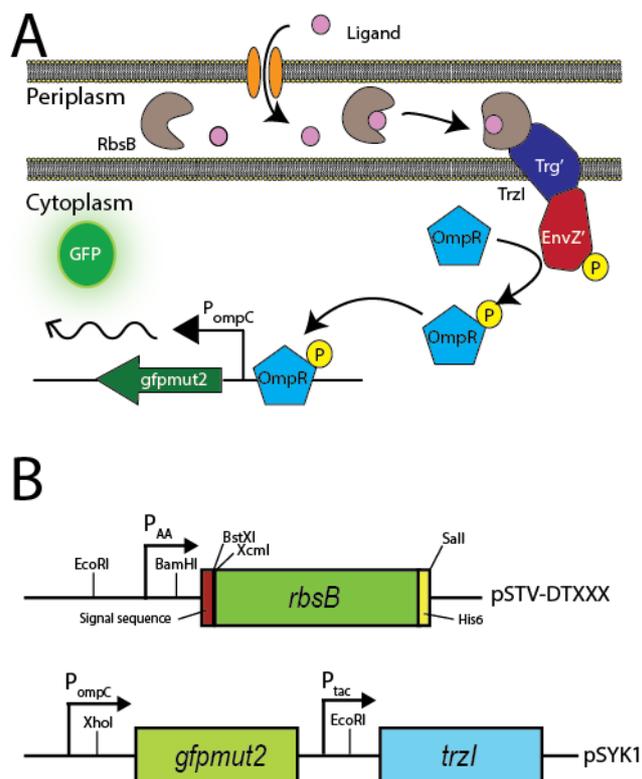


Figure S2 - The hybrid RbsB-OmpR signaling chain of the *E. coli* Trz1-OmpR bioreporter strain. (A) Ribose (ligand) is bound by the ribose-binding protein (RbsB), which docks to the Trz1 hybrid receptor (fusion between the periplasmic part of Trg and the cytoplasmic part of EnvZ). The binding starts a phosphorylation cascade, where OmpR is phosphorylated and increases transcription of *gfp* from the *ompC* promoter. (B) Scheme of relevant plasmids used in this work. Plasmid pSTV-DTXXX expresses the *rbsB* or mutant *rbsB* gene with its translocation signal sequence and hexahistidine tag (His₆) under control of the weak constitutive P_{AA} promoter [2]. Plasmid pSYK1 contains the *gfpmut2* gene under the *ompC* promoter control and the *trzI* gene under control of P_{tac}. Relevant restriction sites are indicated.

Supplementary References

1. Gerhardt, P., et al., eds. *Manual of methods for general bacteriology*. 1981, American Society for Microbiology: Washington, D.C.
2. Alper, H., et al., *Tuning genetic control through promoter engineering*. Proc. Natl. Acad. Sci. U. S. A., 2005. **102**(36): p. 12678-83.

CHAPTER 3

**Random and semi-random mutagenesis to improve
and understand the binding mechanism of ribose-
binding protein mutants towards a non-natural ligand**

Abstract

Periplasmic binding proteins (PBPs) have been proposed as a general scaffold to design new receptor proteins. Designed PBPs can be integrated in bioreporter systems to detect the presence of molecules of interest in the cell's environment. For the past years many groups have explored this exciting concept. However, very limited success was achieved. Most studies working on PBPs have therefore explored their natural properties and studied ligand binding through binding pocket grafting from existing other PBPs or have tried to improve native ligand binding capacity by mutagenesis.

So far, only one successful attempt to completely change the ligand specificity from a natural ligand to a non-natural compound has been reported. This was accomplished on the ribose binding protein (RbsB) from *E. coli*, which was redesigned based on computational simulations to bind 1,3-cyclohexanediol (13CHD). Six proteins with modest but consistent 13CHD binding were isolated. *In vitro* characterization of these RbsB variants confirmed they were no longer responsive to ribose and had gained binding to 13CHD, but were impaired in their stability and/or folding capacity.

The goal of this work was to improve binding capacity and/or solve stability issues of the six 13CHD-binding mutants by random mutagenesis, site saturation and DNA shuffling approaches. Several mutant libraries were constructed, cloned in an appropriate *E. coli* bioreporter system and screened for improved induction of the GFPmut2 reporter fluorescence in presence of 13CHD. Screening was carried out with encapsulated cells in alginate beads grown to microcolonies, induced or not in presence of 13CHD, thresholded at different fluorescence settings. Collected clones by FACS were then used as starting point for the next round of evolution. After three rounds of mutagenesis and screening we isolated 7 mutants with 13CHD concentration-dependent GFPmut2 induction up to 3.2-fold and a lower detection limit of 0.25 mM. Random mutagenesis was the only strategy yielding improved mutants. All observed mutations except one were located outside the direct ligand-binding pocket, suggesting they were compensatory and helping protein folding or functional behaviour other than interaction with the ligand.

Introduction

Proteins are important “players” with multiple applications in health, research and industry ^[1, 2]. Protein engineering techniques have been used to design new proteins, including enzymes and antibodies, or improve natural properties of existing proteins ^[3-6]. The most common approaches used to engineer a protein of interest are rational design, directed evolution and *de novo* design ^[2].

Rational design has successfully improved the functions of proteins ^[7-9]. However, this approach requires an extensive and frequently unreachable knowledge of the relationship between structure and function of proteins ^[10]. This strategy involves site directed mutagenesis (SDM), where defined protein residues are targeted by mutagenesis, replacing the residue by a different one. SDM has the advantage to produce small mutant libraries, reducing significantly the screening time ^[2].

To overcome the situations where limited knowledge about the protein of interest is available, a technique named “directed evolution” has emerged. Over the past years directed evolution has generated powerful and interesting results ^[11-13]. The power of this technique has been internationally recognized by awarding its pioneers with the Nobel Prize of Chemistry in 2018. Directed evolution involves four key steps: (i) selecting a starting gene sequence, (ii) creating a library of variants, (iii) selecting variants by high-throughput screening with improved function, (iv) repeating the process until the improvement or function is achieved ^[14]. Error-prone PCR (ep-PCR), site saturation mutagenesis (SSM) and DNA shuffling are the common mutagenesis methods creating the variants underlying directed evolution ^[15].

Protocols of ep-PCR consist in small modifications of the standard PCR methods in order to enhance the natural error rate of the DNA polymerase. The ep-PCR protocols usually contain high concentrations of MgCl₂, in order to stabilize non-complementary base pairs. Other ways to increase the mutation rate include variation of the nucleotide ratios, inclusion of multiple pairing nucleotides or the addition of MnCl₂ ^[16]. Due to its simplicity and versatility, ep-PCR has emerged as the most common mutagenesis technique and can result in mutation frequencies as high as 2% per nucleotide position ^[17]. SSM enables to create libraries of mutants containing all possible mutations at one or more pre-determined target positions in a target gene sequence ^[15]. This is achieved by introducing all possible base triplets at a given codon, thereby resulting in the insertion of all 20 amino acids at this position of the protein. This method, however, restricts random mutations to predefined sites in the protein of interest, creating therefore what is called focused libraries. SSM requires structural information in order for the appropriate residues sites to be mutagenized ^[18]. Finally, DNA shuffling randomly recombines several related and previously selected gene templates of interest. This *in vitro* recombination reassembles the whole gene (by PCR) from a pool of short overlapping DNA sequences,

generated by random fragmentation of the different parental genes ^[17]. This thus results in assembly of genes that include different combinatorial parts of the parental genes. This technique is extremely useful to reveal synergistic effects between mutations. Although mostly effective for enzymes, random mutagenesis was also successfully applied to alter transcription factor binding specificity ^[19-21].

Periplasmic binding proteins (PBPs) form a family of proteins with a conserved bilobal structure ^[22, 23]. PBPs scavenge molecules (ligands) for the cell, which upon binding are presented to transport channels and/or to membrane receptors involved in chemotaxis ^[24]. Crystal structures and nuclear-magnetic resonance data have shown that PBPs can adopt two semi-stable conformations. In absence of the ligand the majority of the protein adopts an open conformation, in which the binding site is exposed. In presence of the ligand, the molecule is buried in the binding pocket and the PBP adopts a closed conformation ^[25, 26]. Due to this particular characteristics PBPs are an attractive protein class for biosensing purposes. They can be integrated in an *in vivo* hybrid signaling chain leading to expression of a reporter protein, which can be easily measured ^[27, 28]. In the context of this work, we focused on the ribose binding protein (RbsB) of *E. coli*, which in presence of the natural ligand ribose changes from open to close conformational. RbsB ribose binding has been deployed as a bioreporter system through use of a hybrid membrane receptor named Trz1 ^[29]. This hybrid receptor is formed by fusion between the C-terminal part of the *E. coli* natural EnvZ osmoregulation histidine kinase and the N-terminal part of the natural Trg chemotaxis receptor ^[30]. Ribose-bound RbsB interacts with Trz1, triggering its autophosphorylation cascade and yielding increased expression of the *ompC* promoter fused reporter gene.

PBPs have been proposed as a flexible platform to design new receptor proteins based on protein engineering approaches ^[31], with, however, very limited and controversial successes. Several studies attempted to engineer PBPs based on rational design and computational approaches ^[32-34]. Most studies exploited the natural properties of the PBPs to reduce or increase binding specificity ^[35, 36] or to graft binding-pockets between closely related PBPs ^[32, 33]. In the case of RbsB, two studies used alanine replacement mutagenesis to describe and understand the role of each residue in the overall function of the protein ^[37, 38]. The goal of this work was to improve by directed evolution the only so far reported RbsB designed mutants for binding of the non-natural ligand 1,3-cyclohexanediol (13CHD) ^[39].

Previously, we showed how six RbsB mutants with moderate but significant response to 13CHD were obtained through a combination of computational prediction of binding pocket mutations in RbsB and screening for gain of GFPmut2 fluorescence output in an *E. coli* bioreporter system^[39]. Despite *de novo* 13CHD binding, all six mutant proteins showed signs of poor stability, mis- or unfolding and potentially translocation problems compared to wild-type RbsB, indicating that the poor

inducibility by 13CHD may partly be due to protein instability caused by the introduced binding pocket substitutions. Several studies have shown that mutations both in the ribose binding pocket and protein periphery can destabilize RbsB [38-40]. Our hypothesis was thus that it should be possible to increase binding capacity and stability by directed evolution, by using the isolated mutants as starting template. In addition, we expected that random mutagenesis may also lead to mutations in other parts of the RbsB variants, which may help to understand the effects of those positions on stability and overall functioning of the protein.

We used three different random and semi-random mutagenesis approaches to create allelic variants in six variants of RbsB with primitive 13CHD binding. The libraries were produced in the *E. coli* GFPmut2 bioreporter strain and extensively screened for improved GFPmut2 induction in presence of 13CHD, using different thresholding methods. Recovered mutant clones by fluorescence assisted cell sorting were then used in fresh rounds of random mutagenesis of either single or mixed templates, for up to three rounds of directed evolution. We describe 7 mutants of the final screening round, showing two with up to 3.2-fold induction by 13CHD and a lower detection limit of 0.25 mM 13CHD. Mutations were mapped and threaded on inferred protein structure, mostly located outside the direct binding pocket. Our results thus showed that it is possible to improve the overall function of RbsB mutants by random mutagenesis, likely due to better conformational stability.

Results

Random mutagenesis of RbsB protein variants with primitive 13CHD affinity

We previously obtained six mutants in RbsB (DT001, DT002, DT011, DT013, DT015 and DT016) that had lost binding to ribose and instead gained primitive affinity to 13CHD as new ligand ^[39]. Purified mutant proteins, however, displayed severe problems of folding, stability and/or translocation into the periplasmic space ^[39], which we hypothesized could be partly responsible for their relatively poor ligand binding. We set out in this first part, therefore, to randomly create further mutations in each of those DT mutants, under the assumption that this might improve their stability and lead to a higher induction signal with 13CHD. The respective *rbsB-DT* mutant genes were first used as individual templates for random mutagenesis library (RML) preparation, using error prone PCR (ep-PCR). Ep-PCR in presence of different MnCl₂ concentrations (0.025-0.06 mM) resulted in between 1 to 3 mutations per 1000 bp. RMLs produced from each starting RbsB-variant were transformed into the *E. coli* bioreporter strain carrying the *Trz1-ompR-ompC':gfpmut2* signaling chain ^[29]. Library clones were tested individually encapsulated in alginate beads and grown to microcolonies, which were then induced with 1 mM 13CHD as previously described ^[39].

Some 10 million beads, covering three times the estimated sizes of the RML002 and RML016 (derivates of DT002 and DT016, respectively) were screened by fluorescence-activated cell sorting (FACS), in first instance separating beads with an 13CHD-induced fluorescence above the 98th percentile threshold observed in a separate screen of the same amount of beads in presence of 13CHD. 10⁵ beads were recovered, from which plasmid DNA was isolated and used as template for a new round of random mutagenesis. The new libraries (with estimated sizes of 5.5 x 10⁶ and 8.5 x 10⁶ clones) were again encapsulated, induced and screened, but now restricting recovery to the top 0.1 % of GFPmut2 fluorescence (from all induced beads). Some 6 x 10³ beads were collected, purified to individual clones, and screened in 8-fold replicates in 96-well plates for 13CHD induction. This resulted in finding three mutants (named: DT020, DT021 and DT022) with consistent and up to 2.1-fold 13CHD induction, a significant increase and/ or reduction of fluorescence background when compared with parental DT002 and DT016 (Table 1). Sequencing revealed a single different amino acid substitution in each of the three mutants (Table 1). Mutant DT020 had the exact same 1.5-fold induction as its parent DT016, but showed a 30 % reduced fluorescence background intensity ($p = 2.25 \times 10^{-5}$, $n = 12$ replicates, Table1). Mutants DT021 and DT022 displayed a small increase in fold induction to 1.66 ± 0.09 ($n = 13$ replicates) and 2.09 ± 0.16 ($n = 14$ replicates), respectively (Table 1).

Separate RMLs produced from the initial variants DT001, DT011, DT013 and DT015 (named RML001, RML011, RML013 and RML015, and with library sizes of 3.4 x 10⁶, 2.2 x 10⁶, 1 x 10⁶ and 4.1 x

10^6 clones, respectively), were encapsulated, induced with 13CHD and similarly screened on an estimated three times library coverage for higher GFPmut2 fluorescence compared to non-induced conditions. In total, 151 beads were recovered that showed GFPmut2 fluorescence higher than any bead observed in non-induced conditions, which were purified and individually tested for 13CHD inducibility in eight-fold replicates. Unfortunately, all mutants also showed significant increase in fluorescence in absence of 13CHD and none had induction levels above 1.5. Sequencing of some of these mutants showed gene deletions resulting in truncated RbsB mutant proteins. Pooled DNA from those 151 mutants used as template for a new library (estimated size of 1.5×10^6 variants) did not yield further improvements. In contrast, almost all tested clones showed deletions of the *rbsB* variant open reading frames resulting in truncated proteins. These RMLs were therefore abandoned for further investigation.

Table 1- GFPmut2 fluorescence in *E. coli* expressing wild-type- and mutant-RbsB proteins under uninduced and 13CHD-induced conditions.

Wild-type or Variant	Parental Protein	GFPmut2 uninduced fluorescence ^a	Fold induction ^b 1 mM 1,3- cyclohexanediol	Additional mutation(s) ^c
RbsB	-	25226 ± 4066 ^d	0.91 ± 0.05	-
DT016	RbsB	127887 ± 12650	1.51 ± 0.02 ^[39]	^e
DT020	DT016	87877 ± 21152	1.48 ± 0.07	<u>V10I</u> ^f
DT021	DT016	151009 ± 22735	1.66 ± 0.09	K206R
DT022	DT016	132129 ± 25700	2.09 ± 0.16	G89V
DT032	DT022	95492 ± 34689	2.97 ± 0.37	L170S
DT033	DT022	100256 ± 50431	2.63 ± 0.59	L201P S207P K250R
DT035	DT022	95073 ± 23566	2.60 ± 0.50	<u>K5N</u>
DT038	DT022	49432 ± 13382	3.19 ± 0.48	L201V

^a Mean (\pm one *SD*) GFPmut2 fluorescence values of uninduced cultures ($n = 8-14$ replicates). ^b Mean fold induction \pm one *SD*, as the ratio of mean GFPmut2 fluorescence of induced cultures with 1 mM 1,3-cyclohexanediol, by that of uninduced cultures. ^c Additional mutations in comparison to the parental protein. ^d Values in bold indicate statistically significant fluorescence background and/or fold induction compared to DT016 ($p < 0.05$, t-test, equal variance). ^e DT016 carries eight mutations compared to wild-type RbsB: F16S, N64V, D89V, R90S, T134A, F164W, F214A and Q235M. ^f Underlined substitutions are located on the 25 amino acids long signal peptide.

Random mutagenesis of 2nd generation mutant *RbsB* proteins with 13CHD affinity

Because of the accumulation of truncated gene variants in the libraries we decided to create three new RMLs based on the newly isolated (slightly) improved DT variants (DT020, DT021 and DT022, Table 1). These libraries (RML020, RML021 and RML022, with estimated sizes of 1.5×10^6 , 3×10^6 and 2.5×10^6 clones, respectively) were again encapsulated to individual cells, grown to microcolonies and screened both under uninduced and 13CHD-induced conditions. In this screening, only beads with a fluorescence signal higher than the maximum observed signal under non-induced conditions for the same number of screened beads, were collected. These mutants were purified and screened in 96-well plates in 13CHD-induced conditions. This resulted in four mutants with consistent and significant increase of 13CHD-dependent induction of GFPmut2 fluorescence and/or reduced background in absence of 13CHD (Table 1, $n = 8-14$ replicates). Three of those mutants displayed a single amino acid substitution, and one (DT033) showed three substitutions (Table 1). In one case (DT035) the substitution affected an amino acid in the signal peptide. Mutants DT033 and DT035 showed a similar fold induction, around 2.6 times, in presence of 13CHD ($n = 9 - 12$ replicates, Table 1). Mutants DT032 and DT038 were the most promising, with fold-inductions of 2.97 ± 0.37 and 3.19 ± 0.48 times ($n = 13 - 14$ replicates, Table 1). All four mutants displayed a reduction in GFPmut2 fluorescence in absence of inducer in comparison to parental DT016, except DT033. The highest reduction was observed with DT038, with a background reduction of approximately 2 times (Table 1).

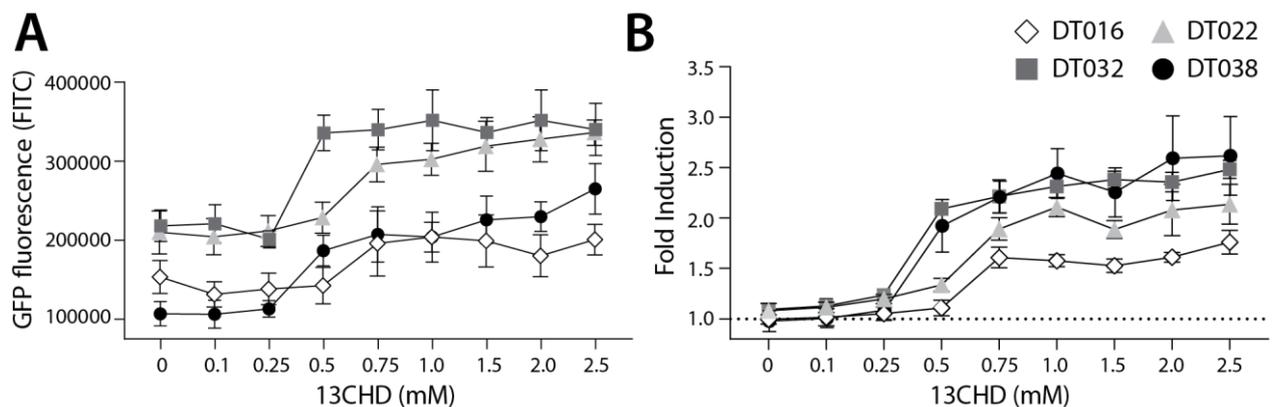


Figure 1- GFPmut2 fluorescence of *E. coli* BW25113 $\Delta rbsB$ expressing DT016 (white diamonds), DT022 (light grey triangles) DT032 (dark grey squares) or DT038 (black circles) in presence of different 13CHD concentrations. A) Average GFPmut2 fluorescence in flow cytometry after 2 h incubation with a range of 13CHD concentrations. B) As (A), but as fold induction compared to uninduced conditions. Each point shows the mean of 8 biological replicates (each from a 100 000 cells' measurement). Error-bars

indicate calculated *SD* (not visible when inside the symbol). Dashed line represents the non-induction level.

GFPmut2 fluorescence in *E. coli* cells expressing DT016, DT022 DT032 and DT038 displayed a typical dose-dependency at different 13CHD concentrations (Fig. 1). For the four mutants, the GFPmut2 fluorescence signal after 2h induction was saturated at 0.5–0.75 mM 13CHD with 1.5 – 3 times fold induction (Fig. 1B). Higher concentrations of 13CHD, up to 2.5 mM, did not lead to further increase of fluorescence (Fig. 1A, $p = 0.06 - 0.99$, t-test equal variance, $n = 8$ replicates). The lowest concentrations of 13CHD that yielded significant induction compared to medium without inducer after 2h incubation were 0.25 mM for DT022 and DT032 and 0.5 mM 13CHD for DT032 and DT038 (Fig. 1B).

The four 3^d generation (e.i., DT032, DT033, DT035 and DT038) mutants were subsequently used to create four new random mutagenesis libraries, which were screened as before, but this did not lead to isolation of mutant proteins with improved induction with 13CHD (i.e., more than 3 times GFPmut3 fluorescence increase upon induction compared to uninduced levels). Interestingly, however, we noted that populations of several mutants displayed double fluorescence levels simultaneously, almost irrespectively of 13CHD presence (Fig. 2). These subpopulations corresponded to completely uninduced and fully induced fluorescence levels seen from wild-type RbsB with ribose (Fig. 2A, Low_Pop and High_pop). For example, mutant 1F6 displayed one subpopulation with a mean fluorescence value of 15,000 and a second of 220,000 (Fig. 2B). The proportion of cells within the low and high subpopulation was approximately 36 % and 62 %, respectively. Upon 2h incubation with 1 mM 13CHD the proportion of cells within either subpopulation change to 25 % and 74 %, respectively (Fig. 2B). Similar results were obtained with mutant 2C10, showing a reduction of 8 % in the proportion of cells within the low fluorescence subpopulation upon 13CHD induction, and an increase of 9% in the high subpopulation (Fig. 2C). Mutant 1F8 displayed a different behaviour, with an almost equal proportion of cells distributed between the low and high subpopulations under uninduced conditions; but an increase up to 87% within the subpopulation with lowest fluorescence in presence of 13CHD (Fig. 2D).

Across multiple tests and replicates, the proportions of cells within those subpopulations differed substantially, making it hard to judge whether this was consistent behaviour one would expect from an inducible protein. This suggested, therefore, that these mutants had become hypersensitive and spontaneously switched between open (i.e., uninduced signal) and closed (i.e., induced) state at the level of an individual cell.

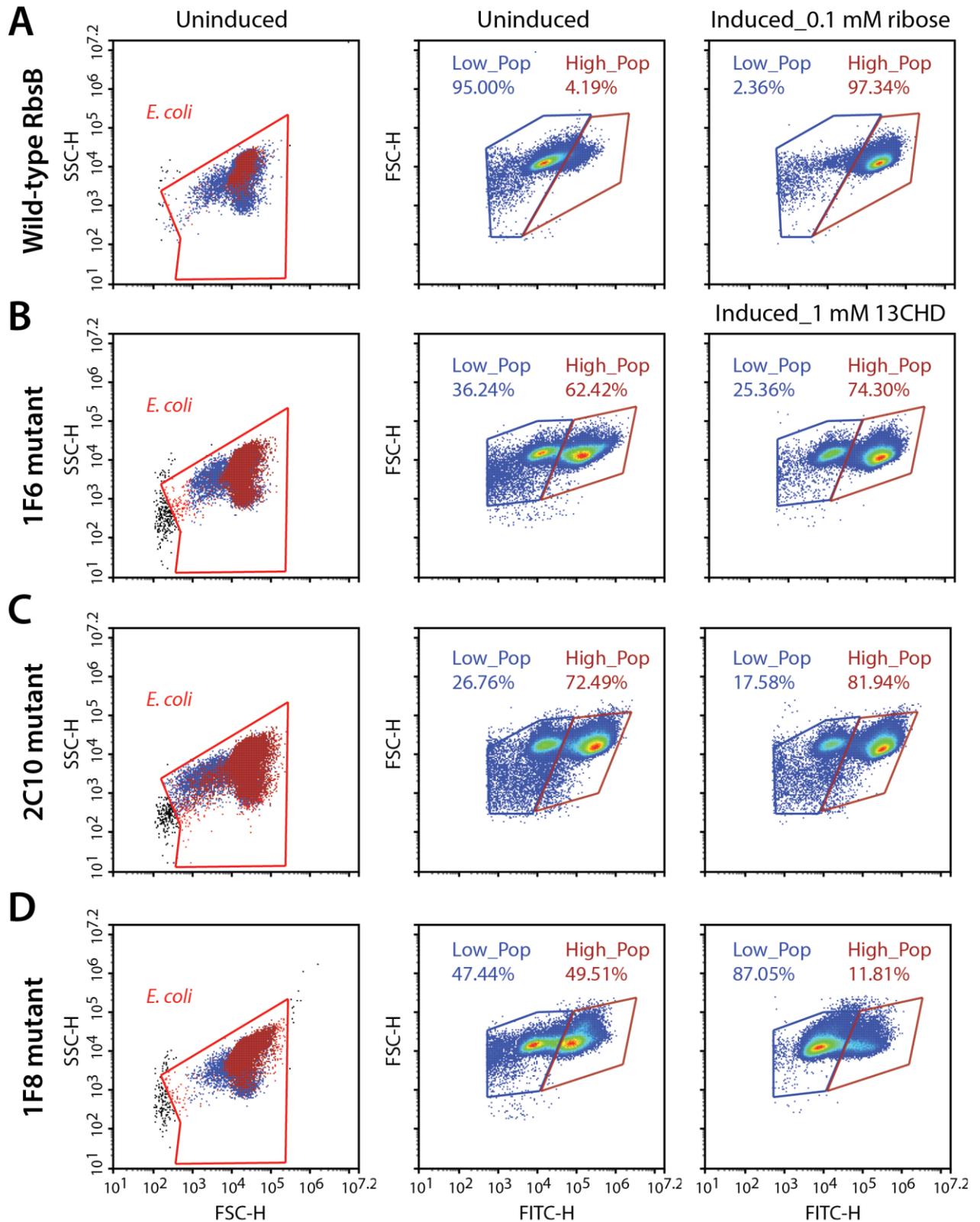


Figure 2- DT038-derivative mutants displaying subpopulations with two fluorescence states. Low_Pop and High_Pop represents cells with low and high fluorescence levels, respectively. Proportions of cells within both gates are indicated. A) Wild-type RbsB uninduced and induced with 0.1 mM. B)–D) As (A), but for mutant 1F6, 2C10 and 1F8 induced with 1 mM 13CHD. Each density plot shows 100 000 cells.

Positions of 2nd and 3rd generation mutations with improved induction with 13CHD

The positions of the amino acid substitutions observed in the various new mutants (Table 1) were threaded on the closed structure of wild-type RbsB (PDB ID: 2DRI). Six out of the seven isolated mutants displayed a single amino acid substitution. Exception was DT033 that displayed three amino acid changes. Mutant DT020 displayed a conservative substitution within the signal peptide (V10I) (Table 1). DT021 and DT022 displayed conservative amino acid substitutions K206R and G89V, respectively (Fig. 3A, Table 1). The V89 residue of DT022 is located within the binding pocket, 1.9 Å from the 13CHD molecule (Fig. 3A). This suggests that the G89V substitution is directly responsible for the 50 % increase of GFPmut2 fluorescence upon 13CHD induction, when compared to its parent DT016 [39]. Previous studies demonstrated the importance of residue 89 for ligand binding [37-39], suggesting that V89 improves the capacity to bind 13CHD in comparison with G89. Finally, mutation K206R found in DT021 is located in a peripheral turn of the structure (Fig. 3A).

Two out of the four 3rd generation mutants displayed non-conservative mutations, notably, DT032 had a leucine at position 170 instead of serine, whereas DT033 displayed a S207P substitution. DT033 showed two conservative mutations (L201P and K250R; Fig. 3A, Table 1). Mutant DT035 again displayed a mutation in the signal peptide (K5N). This and the V10I substitution of DT020 in the signal peptide may have improved the translocation and/or stability of the mutant protein. Of the other five substitutions observed in this 3rd round of evolution, three (L170S, L201V and K250R) were located outside the binding pocket in three different α -helices of the protein (Fig. 3A). The two others (K206R and S207P) were localized in a peripheral turn of the protein structure (Fig. 3A). All of them led to an increase of the fold-induction with 13CHD (Table 1). However, their peripheral position suggested they play a role in protein stability and not directly in ligand binding. Interestingly, five out of the seven substitutions found in 2nd and 3rd generation mutants localized in the same region of the protein (Fig. 3A), suggesting that changes in that area improve protein function (e.g., stability or intramolecular hinge movement). Interestingly, leucine at position 201 was substituted twice independently by two different amino acids (i.e., proline and valine), underscoring its critical role. None of the positions recovered in these DT variants for 13CHD binding had been previously described as critical for the various roles of RbsB in a near-complete Ala-substitution scanning [38].

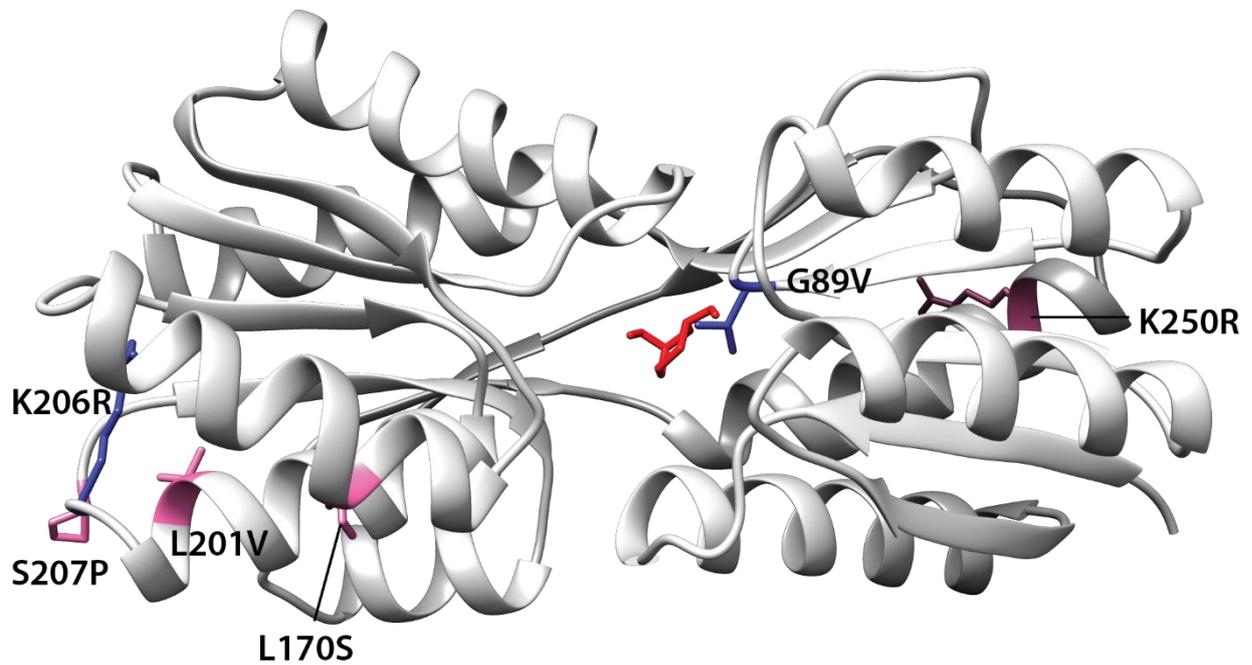
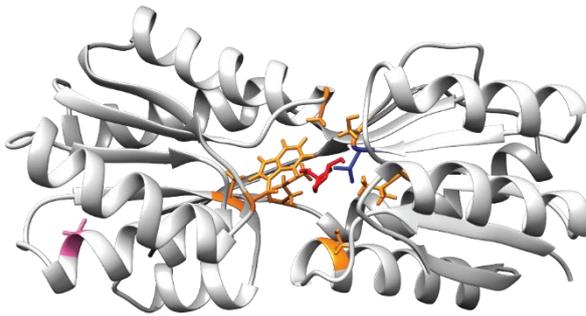
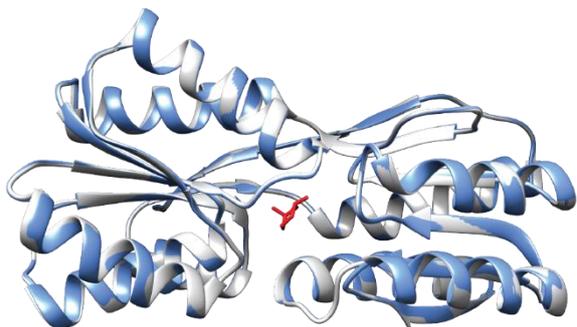
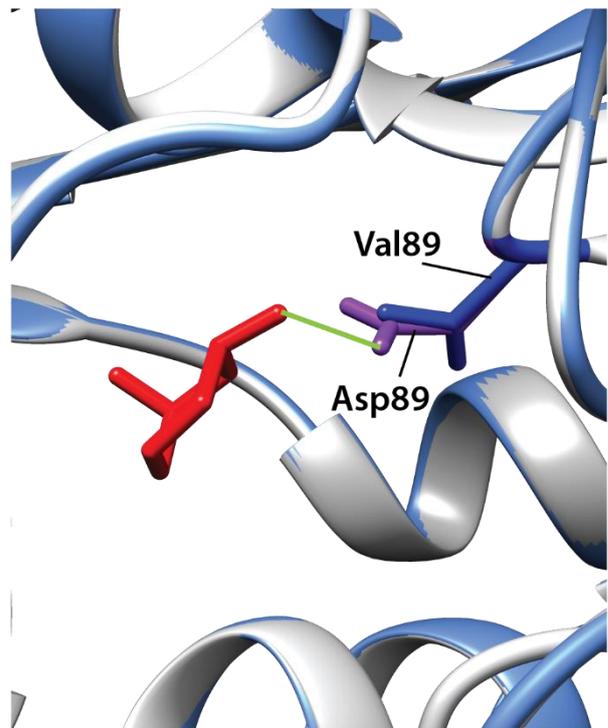
A Overlay of mutations in 2nd and 3rd generations variants**B** DT038 protein**C** RbsB and DT038 superposition**D** Details of 89 residue interaction with 13CHD molecule

Figure 3- Positions of amino acid substitutions observed in various 13CHD mutants, visualized on the closed structure of wild-type RbsB (PDB ID: 2DRI). A) Mutated residues in 2nd (blue) and 3rd generation mutants (pink). B) Acquired mutations in DT038, seen as a superposition of 1st (orange, in DT016; the original Rosetta designed binding pocket mutations), 2nd (blue, in DT022) and 3rd generation mutations

(pink). C) Superposition of wild-type RbsB (PDB ID: 2DRI, blue), and the threaded DT038 structure (light grey). D) Details view from C) panel, showing the interaction between ligand and residues located at 89 position. Asp89 of RbsB and Val89 of DT038 are colored in purple and blue, respectively. Light green line indicates a H-bond. Molecular structures of 13CHD (red) are placed in the binding pocket. Signal peptide mutations are not shown on the structure as this is cleaved off.

To further infer potential structural changes of observed mutants in comparison to wild-type RbsB, we used Swiss-Model^[41-45], Phyre2^[46] and Missense3D^[47]. Swiss-Model, Phyre2 did not predict any structural differences of the new mutations compared to the closed structure of wild-type RbsB (Fig. 3C, shown for DT038). Analysis of each of the eight amino acid substitutions in DT016 by Missense3D indicated expansion of the binding cavity by F16S and R90S, and H-bond breakage by D89V and T135A. This was expected, since these were designed and engineered ligand binding pocket mutations to accommodate 13CHD. However, none of other four mutations in DT016 were predicted by Missense3D to cause any (individual) structural difference compared to RbsB. Interestingly, the G89V mutation in DT022 compared to DT016 (or D89V compared to wild-type) was predicted to cause further expansion of the ligand binding pocket and to H-bond breakage (Fig. 3D). This is probably the consequence of replacing a buried amino acid (Gly) by an exposed one (Val). Other individual amino acid substitutions, found in other isolated mutants, were not predicted to cause any structural difference compared to wild-type RbsB, but we acknowledge that Missense3D only tests single substitutions at the time.

DNA shuffling and site saturation mutagenesis

Rescreening of the 2nd round RML002 and RML016 libraries, led not only to the isolation of the second generation mutants (e.i., DT020, DT021 and DT022) but to 8 more variants as well (Table 2). Individual retesting of those eight variants showed no change in the mean fold-induction of GFPmut2 in presence of 1 mM 13CHD compared to DT016 itself (Table 2, $p = 0.245 - 0.89$, $n = 6$ replicates). On the other hand, five mutants (named here: 2H2, 7B2, 7B9, 7C5 and 7G4) displayed a lower background fluorescence in uninduced conditions, when compared to DT016 (Table 2, $p = 0.01 - 0.00001$, $n = 6$ replicates). The highest background reduction was two-fold, observed in 7G4 mutant (Table 2). The lower fluorescence background suggests a better equilibrium between open and closed conformation. The DNA of the 8 mutants was then shuffled in the hope to create synergetic effects, but no mutant with improved induction by 13CHD was isolated from this screening.

Table 2- List of isolated mutants from RML002 and RML016 2nd round used for DNA shuffling.

RbsB Protein	GFPmut2 uninduced fluorescence ^a	Fold induction ^b 1 mM 1,3- cyclohexanediol	New mutation(s) ^c
Wild-type	25226 ± 4066 ^d	0.91 ± 0.05	-
DT016	160622 ± 33495	1.87 ± 0.26	^e
2H12	112793 ± 16657	2.01 ± 0.13	Q80R
3B11	157805 ± 24521	1.88 ± 0.13	V17E T58A
5H2	128273 ± 13989	1.97 ± 0.11	K29R A214T
6H9	132217 ± 12670	1.92 ± 0.1	I132T
7B2	95166 ± 8064	1.97 ± 0.11	T93M
7B9	91342 ± 14436	1.75 ± 0.06	T10A
7C5	85631 ± 15553	1.89 ± 0.14	N175S K228Q T232D
7G4	78653 ± 8090	1.81 ± 0.16	N73S

^a Mean ± one *SD* GFPmut2 fluorescence values of uninduced cultures ($n = 6 - 8$). ^b Mean fold induction ± one *SD*, as the ratio of the mean GFPmut2 fluorescence of induced cultures with 1 mM 13CHD and the mean fluorescence of uninduced cultures. ^c Newly observed mutations in comparison to DT016. ^d Values in bold are statistically significantly different from those of DT016 ($p < 0.05$, t-test, equal variance). ^e DT016 carries eight mutations compared to wild-type RbsB: F16S, N64V, D89V, R90S, T134A, F164W, F214A and Q235M.

Computational simulations had previously suggested nine amino acids as being critical for changing the specificity of RbsB protein to 13CHD^[39]. Two residues were later found by ala-substitution scanning to be important for ribose binding (R141 and D215)^[38]. We therefore tested whether site-saturation mutagenesis of these residues could further improve DT002 and DT016 variants for 13CHD induction (Fig. 4). Replacement of R141 and D215 by each of the other 20 possible amino acids was confirmed by sequencing and 250 colonies of each site saturation library were tested individually by flow cytometry for gain of 13CHD induction. None of the tested mutants from DT002_{R141X}, DT002_{D215X}, or DT016_{R141X} and DT016_{D215X} libraries showed improved 13CHD induction, compared to parental strains, whereas several were worse. However, this showed that R141 in mutant DT016 can be replaced by a serine without impairing inducibility by 13CHD.

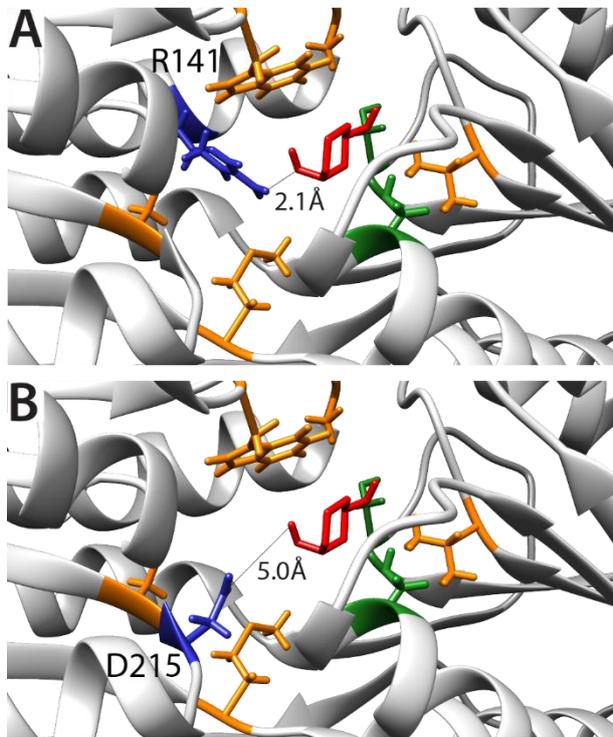


Figure 4- Inferred binding pocket of DT016 in presence of 13CHD (in red) and with the substituted residues for ligand binding color-coded based on characteristics (nonpolar- orange; polar- green). A) Arginine at position 141, selected for site saturation mutagenesis, is highlighted in dark blue. Distance between 13CHD molecule and R141 is indicated. (B) As for (A) but for Aspartic acid at position 215.

Mutagenesis of neighboring residues in DT016

Finally, we tested whether substitutions in the direct neighborhood of the previously engineered ligand binding pocket mutations would affect induction by 13CHD, through synergistic or compensatory effects on the overall protein function or behaviour. For this we focused again on DT016, the most promising mutant with newly obtained specificity to 13CHD [39]. Next, we designed a strategy to mutate the two amino acids flanking (i.e., those before and after) each of the eight ligand binding pocket mutations of DT016 (Fig. S2). We reconstituted the *dt016* open reading frame in 12 overlapping PCR fragments (Fig. 5A). PCR primers covered the regions of the eight introduced mutations in DT016, with flanking amino acids of those being replaced by all other 20 possible amino acids (Table 3). The disadvantage of this strategy was that stop codons could not be avoid in primer design. A library with an estimated size of 1×10^6 clones (RML-DT016_{AA}) was screened as before by agarose encapsulation and flow cytometry. As expected, a large fraction of clones carried truncated proteins (75% from 25 randomly picked colonies from the library on plates). None of the clones displayed higher fold induction than DT016 itself. We concluded that this strategy was not worth further pursuing.

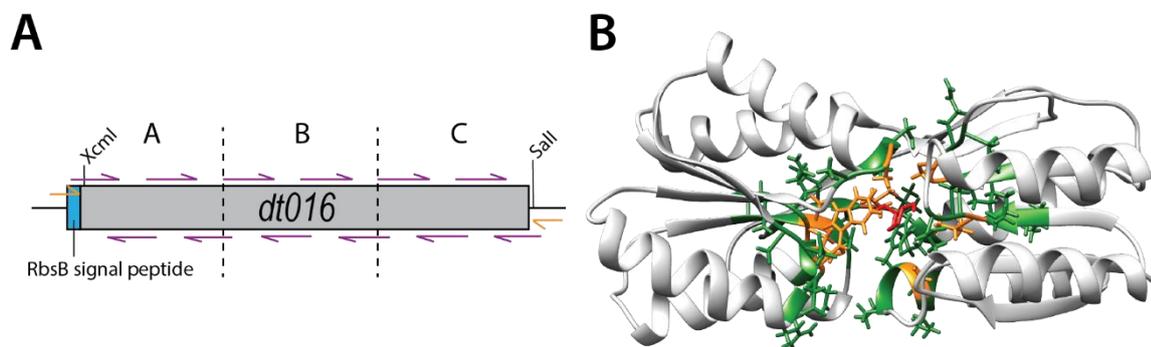


Figure 5- Flanking residue mutagenesis. A) Scheme of the design strategy to reconstitute *dt016* with 12 overlapping primers (purple), positioned at the sites of the previous ligand binding pocket mutations, and each covering the two neighboring codons on each side. The assembly was carried out in 3 steps (A, B and C) and followed by a final extension reaction with external primers (orange). Gene not drawn to scale. B) Inferred structure of DT016 based on RbsB structure (PDB ID: 2DRI). Molecular structure with 13CHD (red) bound in its pocket. The eight designed substitutions in DT016 are highlighted in orange and flanking residues mutated with this strategy are highlighted in green.

Discussion

Periplasmic binding proteins (PBPs) have been used as a starting point to design new receptor proteins^[31]. Despite the vast knowledge on PBPs structures and their natural ligands^[23], the successful design of new ligand binding domains has been very limited so far^[48, 49]. Introducing amino acid substitutions in a protein is challenging, since they can easily lead to an abnormal function or behaviour of the mutated protein^[38]. In a previous study nine residues were identified and substituted in the binding pocket of RbsB, with the goal to change the binding specificity from the natural ligand ribose to the non-natural compound 13CHD^[39]. Despite the modest achieved success (up to 1.5-fold induction with 13CHD), six mutant proteins without ribose recognition but with 13CHD binding were isolated. However, the introduced substitutions were assumed to lead to unstable proteins with translocation and misfolding issues^[39].

The goal here was to produce and select complementary mutations by random or semi-random approaches, which might either have a stabilizing effect or further improve 13CHD ligand binding, or both. We focused on the previously isolated mutants, which we used as scaffolds for mutagenesis. Several rounds of random mutagenesis by ep-PCR and increasing selectivity of sorting of bead-grown microcolonies induced with 13CHD, led to recovery of a few mutants with consistently higher induction of GFPmut2 fluorescence than their parental strains (up to 3.2-fold at 1 mM 13CHD). As these mutants carried substitutions outside the direct ligand binding pocket, we assume that they are compensatory mutations that improve functions other than ligand binding itself, for example, L170S in DT032 or L201V in DT038.

In order to maximize our chances to isolate a variant with improve 13CHD binding capacity and/ or stability, we used different mutagenesis approaches to create genetic variability. Our semi-random approaches did not produce the expected results, since no improved variant was isolated from created libraries. Site saturation mutagenesis of R141 and D215 residues on DT002 and DT016 resulted in decrease of the capacity for induction by 13CHD, except for a R141S substitution in DT016 that did not affect inducibility. This indicated that we could not improve the 13CHD induction by replacing R141 and D215 residues; in contrast, it showed that their presence is essential for 13CHD binding and signaling. The importance of both R141 and D215 residues in RbsB for ribose induction and signaling^[38] and for ligand binding (D215) had been previously demonstrated^[37]. Also DNA shuffling did not lead to isolation of mutants with potential synergistic improvements, although background reduction in absence of inducer was observed (Table 2). Random mutation of the 32 residues flanking the 9 substitutions engineered for 13CHD ligand binding did not yield improved variants either, possibly because of the high percentage of variants with a truncated protein. Some variants displayed a

comparable induction level to their parent DT016 (around 1.5 times). However, given the high number of substitutions in these variants (up to 32 amino acid substitutions) the interpretation of their effect was impossible. These results indicated that changing several parts of the RbsB-mutant proteins at the same time may not be the best way to find variants with improved function. Introducing multiple mutations increases the probability to find proteins with improved capacities, but at the same time increases the chances to introduce mutations that may impair the protein function. This creates an important trade-off, and has to be considered each time when designing and implementing a mutagenesis strategy.

In contrast, random mutagenesis across the complete gene variants led to the isolation of seven mutants with significantly improved 13CHD inducibility (up to 3.2 times, Table 1). As these consisted relatively large libraries, we conceived two selection and sorting strategies that were more or less restrictive in their recovery. In the less restrictive sorting, all beads above the 98th fluorescence percentile of the 13CHD induced library were collected. The DNA of the sorted variants was then used as template for a new random library, from which we recovered the top 0.1 % fluorescence beads. This strategy led to isolation of 3 variants with up to 2.1 times induction with 13CHD. In the more restrictive sorting, only beads with a fluorescence higher than any bead under uninduced conditions were recovered. Here we isolated 4 mutants with up to 3.2-fold induction by 13CHD. Since both strategies allowed us to isolate mutants with improved 13CHD binding capacity, we conclude that the restrictive sorting is a better strategy. Being less restrictive is more time consuming and required two rounds of cloning and FACS screening, plus leading to more downstream screening of individual clones. More restrictive sorting thus significantly reduces the time to screen potential mutants. A disadvantage of the restrictive strategy is that mutants are missed, which have low fluorescence background under uninduced conditions and intermediate fluorescence upon induction (i.e., a fluorescence signal less than the maximum observed in the uninduced library). Alternatively, one could try to 'bin' mutants in different fluorescence categories in the hope of finding some with lower fluorescence backgrounds and still some induction. The difficulty is that *a priori* the evolutionary path of a variant highly inducible by 13CHD is not known and may pass through intermediates with high uninduced levels to regain background, or through those with low uninduced levels and gain specificity [14, 15, 50]. Selectively focusing on the high fluorescence makes for an easier screening.

Multiple rounds of mutagenesis thus allowed to improve 13CHD binding capacity in a step-by-step manner. This suggests that further rounds of random mutagenesis could eventually lead to the isolation of a variant with similar binding capacity to 13CHD as wild-type RbsB towards ribose (13 fold) [39], although we could not achieve that here. Some studies show that multiple rounds of evolution are needed to improve a specific protein ability without impairing the protein [51].

What can we conclude from the obtained DT variants in terms of amino acid substitution effects? Two mutants (DT020 and DT035) displayed an amino acid substitution in the signal peptide (V10I and K5N). The improved 13CHD induction might have been due to higher periplasmic protein levels, being the result of a positive effect on peptide recognition by SecB chaperone, responsible for presenting RbsB to the translocation channel, and or improved stability. Only one variant (DT022) carried a substitution (G89V) in the binding pocket (Fig. 3A and 3D). This residue is less than 2 Å away from inferred position of 13CHD and previous studies demonstrated the importance of residue 89 for ligand binding^[37-39]. An exposed valine residue at this position thus seems to improve 13CHD binding, yielding a 50 % higher fold induction when compared with parent DT016 (Table 1). All other mutations were found outside the binding cavity, and we assume that they must have improved other aspects of protein functionality than ligand binding itself. This could affect, for example, protein stability or improved hinge flexibility, or binding to the chemoreceptor Trz1. Five out the seven mutations were located in the same peripheral region of the protein (Fig. 3A), but none concerned positions previously implicated in RbsB functioning by Ala-substitution scanning^[38]. Leucine at position 201 was replaced by two other non-polar residues in two different isolated mutants and neighboring residues K206 and S207 were replaced by arginine and proline, respectively. The concentration of observed mutations in this region suggests that previous introduced mutations, possibly, disturbed this region of the protein and compensatory mutations were needed. This specific region of the protein, therefore, could be a promising target for future rounds of mutagenesis, aiming to find variants with better overall function. Importantly, the new variants were not only a higher induced by 13CHD, but also displayed reduced fluorescence background, especially DT032, DT035 and DT038. This is further evidence that these mutations play an important role in the stabilization of the protein.

The two most promising mutants DT032 and DT038 displayed similar threefold induction in presence of 13CHD, mostly as a result of background reduction in absence of inducer (Table 1). The variants react in a dose-dependent manner, with the lower detection limit around 0.25 mM 13CHD, which gives further credence that these DT variants are genuine new ligand-binders. Moreover, we could observe the 3rd generation mutation (e.i., DT032 and DT038) displayed a higher fold induction than 1st and 2nd generations mutants (e.i., DT016 and DT022, respectively) (Fig. 1B). Confirming that introduced mutations improve the binding capacity and/or the overall function of these variants. It would be interesting to confirm this by protein purification and determining *in vitro* stability and affinity towards 13CHD.

Creation of new ligand-binding cavities in PBPs had been heralded more than a decade ago as one of the key advance areas for computational protein design ^[52], but more recent *de novo* design of protein (and peptide) structure design have focused more on small-molecule-binding proteins ^[53], switchable/ allosteric capacity ^[54], protein folding ^[55] and epitopes scaffolds design ^[56]. Much of the initial claimed successes of PBP ligand pocket engineering has not hold the scrutiny of independent repetitions ^[29, 48]. Different advances have been reported, such as by grafting of existing ligand pockets and *de novo* design of 13CHD binding ^[32, 33, 39]. It might thus well be that, in contrast to the original assumption of a wide protein family with known crystal structures of open and closed configurations, PBPs are actually particularly difficult to engineer. The reasons may be that PBPs need an inherent intramolecular protein movement between open and closed configuration and have manifold functional constraints, such as ligand binding, binding to the receptor, or translocation. Current ligand pocket predictions do not take the other constraints into consideration, which make complete rational computational design challenging.

For example, in the RbsB-based bioreporter configuration wild-type- and mutant-RbsB proteins have to be expressed and translocated to the periplasmic space. Once in the periplasmic they recognize and bind their ligand, leading to a conformational change of the protein ^[34, 57]. Subsequently, the closed form of the protein binds the hybrid Trz1 receptor, starting a phosphorylation cascade that in the end leads to induction of GFPmut2 expression ^[29]. It is important to understand that if an introduced mutation affects a single of these steps the final outcome (e.i., GFPmut2 signal) is affected. The transition between open and closed conformation is extremely important for PBPs with bilobal structure such as RbsB. It is assumed that PBPs in absence of ligand can be found in a dynamic equilibrium of open and close state ^[48, 58], which is important for their function. Similar characteristics are observed in other PBPs, for example in the closely related galactose-binding protein of *E. coli* ^[59]. In presence of the proper ligand, the closed form is stabilized ^[48] and, like in case of RbsB can present the ligand molecule (ribose) to either the chemoreceptor (i.e., Trg and Trz1) or to ribose transport channels ^[24, 60, 61]. We observed that introduced mutations can block RbsB variants in either of the two states, and consequently, disable its function to bind the ligand and trigger the bioreporter system, or trigger the receptor signaling cascade without binding the ligand. We also observed RbsB-DT variants that in the *E. coli* bioreporter strain caused 'stable' double populations with different GFPmut2 fluorescence intensities both in absence or in presence of inducer. This is in contrast to wild-type RbsB behaviour, which (despite reported open-closed form dynamics in absence of ligand) in absence of ribose results in coherent low reporter output and in its presence in coherent high fluorescence. This suggests that the time-scale of the dynamics is affected by the introduced mutations, blocking the DT variants in either open or closed form long enough to trigger (or not) the bioreporter signaling cascade

leading to GFP expression. This is supported by the fact that these subpopulations corresponded to completely uninduced and fully induced fluorescence levels seen from wild-type RbsB with ribose (Fig. 2). A small percentage of the low fluorescence population shifts to high fluorescence upon induction, indicating that ligand-binding is still affecting the transition states, but is insufficiently discriminating between the two (Fig. 2B and 2C). Further biochemical characterization of such protein variants may be interesting from the point of view of observing the transition states.

In conclusion, the results obtained in this study showed that it is possible to improve the ligand binding capacity and stability of previously designed RbsB mutants with *de novo* ligand binding pockets using random mutagenesis. Unfortunately, despite extensive library screening, we could not improve the 13CHD bioreporter system to more than 3-fold induction and 0.25 mM lower detection limit (in 2 h assay), which is not yet comparable to that of RbsB and ribose (13-fold induction and 50 nM lower detection limit) [29, 39]. This suggests that further randomized screening from existing variants by bioreporter output itself may not be on the proper 'evolutionary path' [15, 50, 62]. Of note that the screening approach we used here would ignore potentially interesting other variants, for example, that have better translocation, binding of 13CHD and stabilizing the closed conformation, but that are unable to interact and bind Trz1. For this reason, follow-up studies should focus on other functional constraints in the bioreporter cascade, such as protein translocation, folding, ligand binding, conformational change, stability and receptor binding. The complete set of mutants thus obtained may help to derive and improve further computational approaches that challenge dynamic protein behaviour.

Table 3- Strain list used in this study.

Strain	<i>E. coli</i> host	Plasmids	Relevant characteristics	Reference or source
3044	DH5 α λPir		Host for plasmid propagation	[63]
3671	DH5 α	pSTVP _{AA} _mcs	pSTVP _{AA} to clone <i>rbsB</i> and its derivatives	[29]
4172	BW25113 $\Delta rbsB$	pSYK1	Host strain containing the P _{tac} - <i>trzl</i> , P _{ompC} - <i>gfpmut2</i> bioreporter system	[29]
4175	BW25113 $\Delta rbsB$	pSTVP _{AA} _rbsB, pSYK1	RbsB expression with signal peptide for periplasmic translocation	[29]
5913	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT001, pSYK1	As 4175, but for DT001 mutant protein of RbsB	[39]
5903	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT002, pSYK1	As 4175, but for DT002 mutant protein of RbsB	[39]
5904	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT011, pSYK1	As 4175, but for DT011 mutant protein of RbsB	[39]
5905	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT013, pSYK1	As 4175, but for DT013 mutant protein of RbsB	[39]
5906	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT015, pSYK1	As 4175, but for DT015 mutant protein of RbsB	[39]
5907	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT016, pSYK1	As 4175, but for DT016 mutant protein of RbsB	[39]
7241	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT020, pSYK1	As 4175, but for DT020 mutant protein of RbsB	This work
7242	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT021, pSYK1	As 4175, but for DT021 mutant protein of RbsB	This work
7243	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT022, pSYK1	As 4175, but for DT022 mutant protein of RbsB	This work
7244	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT032, pSYK1	As 4175, but for DT032 mutant protein of RbsB	This work
7245	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT033, pSYK1	As 4175, but for DT033 mutant protein of RbsB	This work
7246	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT035, pSYK1	As 4175, but for DT035 mutant protein of RbsB	This work
7247	BW25113 $\Delta rbsB$	pSTVP _{AA} _DT038, pSYK1	As 4175, but for DT038 mutant protein of RbsB	This work
2H12	BW25113 $\Delta rbsB$	pSTVP _{AA} _2H12, pSYK1	As 4175, but for 2H12 mutant protein of RbsB	This work
3B11	BW25113 $\Delta rbsB$	pSTVP _{AA} _3B11, pSYK1	As 4175, but for 3B11 mutant protein of RbsB	This work
5H2	BW25113 $\Delta rbsB$	pSTVP _{AA} _5H2, pSYK1	As 4175, but for 5H2 mutant protein of RbsB	This work
6H9	BW25113 $\Delta rbsB$	pSTVP _{AA} _6H9, pSYK1	As 4175, but for 6H9 mutant protein of RbsB	This work
7B2	BW25113 $\Delta rbsB$	pSTVP _{AA} _7B2, pSYK1	As 4175, but for 7B2 mutant protein of RbsB	This work
7B9	BW25113 $\Delta rbsB$	pSTVP _{AA} _7B9, pSYK1	As 4175, but for 7B9 mutant protein of RbsB	This work
7C5	BW25113 $\Delta rbsB$	pSTVP _{AA} _7C5, pSYK1	As 4175, but for 7C5 mutant protein of RbsB	This work
7G4	BW25113 $\Delta rbsB$	pSTVP _{AA} _7G4, pSYK1	As 4175, but for 7G4 mutant protein of RbsB	This work

Table 4- List of primers used to introduce new mutations in RbsB-derivate proteins. NNN indicates mutated positions for production of the libraries.

Mutagenesis Technique	Target	Primers	DNA sequence (5' to 3')
DNA shuffling	<i>rbsB</i> (outer primers)	190101 F	CAGCTGGCGAAAGGGGGATGTG
		130401 R	CTGAGCACATCAGCAGGAC
	<i>rbsB</i> (inner primers)	160401 F	CACGACGTTGTAACGACGGCC
		190102 R	CTGGCTACCCTGGTTTCCGCTG
Site saturation mutagenesis	pSTVP _{AA} -DT002 _{R141X}	180901 F	GAAGCCTTCGCCNNNTTCACGGGCTGCGGACGCACC
		180702 R	GCAGCCCGTGAANNNGGCGAAGGCTTCCAGCAGGCC
	pSTVP _{AA} -DT002 _{D215x}	180705 F	ATCCGGTGTACCNNNAGCTCCGACGACCATCACATC
		180706 R	GTCGTCGGATCGNNNGGTACACCGGATGGCGAAAAA
	pSTVP _{AA} -DT016 _{R141X}	180901 F	GAAGCCTTCGCCNNNTTCACGGGCTGCGGACGCACC
		180702 R	GCAGCCCGTGAANNNGGCGAAGGCTTCCAGCAGGCC
	pSTVP _{AA} -DT016 _{D215x}	180901 F	ATCCGGTGTACCNNNCGCTCCGACGACCATCACATC
		180708 R	GTCGTCGGAGCGNNNGGTACACCGGATGGCGAAAAA

Material and Methods

Bacterial strains and culture conditions

Expression of the RbsB-Trz1-ompCp-gfpmut2 signaling chain (or the RbsB variants) was tested in *E. coli* BW25113 $\Delta rbsB$ as host. In this case, cells were cultured in minimal medium (MM) supplemented with 20 mM fumarate. For selection of mutants by fluorescence activated cell sorting, cells were first grown within alginate beads in low phosphate minimal medium (MM LP) supplemented with 20 mM fumarate and appropriate antibiotics to produce microcolonies, as described previously [39]. The cells-in-beads were then induced with 0.1 mM ribose or 1 mM 1,3-cyclohexanediol (13CHD) for 2 h, as described previously (Chapter 1). *E. coli* DH5 α cells were used for cloning and plasmid propagation. Random libraries were transformed into ElectroMAX DH10B T1 Phage-Resistant competent cells (ThermoFisher).

All strains used in this study are listed in Table 3.

Random mutagenesis libraries and plasmid construction

Mutations in the *rbsB* gene or its *dt* variants were generated by error-prone PCR (ep-PCR). Gene variants were amplified by primers flanking the coding sequence in the plasmids pSTV_{AA}-DTxxx and located up- and downstream of the *Sall* and *NdeI* sites (Fig. S1). Ep-PCR reactions were carried out with 4 ng of DNA template in presence of varying MnCl₂ concentrations (0.025–0.06 mM). Six reactions were prepared simultaneously to average stochastic biases. After an initial denaturation period of 10 min at 94°C, the following steps were repeated for 25 cycles: 1 min at 94°C, 1 min at 70°C and 1.5 min at 72°C, followed by an extension of 10 min at 72°C. Amplicons were then visualized by agarose electrophoresis, and products of around 1 kb were excised, pooled and purified. Purified PCR products and pSTV_{AA} plasmids were digested with *Sall* and *NdeI* at 37°C and 300 rpm for 45 min. Plasmid self-ligation was prevented by treating the digested plasmid with Shrimp Alkaline Phosphatase (rSAP, NEB). The digestion products were visualized by agarose electrophoresis, and the correctly sized digested bands were excised from the gel and purified.

Plasmid and amplicon fragments were ligated with T4 DNA ligase using a ratio of 1:2 vector to insert. The ligation mixture was incubated overnight at room temperature, and aliquots of 100 were electro-transformed into ElectroMAX DH10B cells. Cells were recovered after the electroporation by addition of 1 mL of SOC medium, pooled from five separate reactions and incubated for 1h30 at 37°C, 225 rpm. Small proportions of these mixtures were plated on LB agar (Cm) plates to estimate the

number of CFUs in the libraries. The remainder was cultured *en masse* in 200 ml LB medium (Cm), which was used to isolate and purify a plasmid-library pool. Aliquots of 1.6 mL of the grown culture were stored in 15% (v/v) glycerol at - 80°C. Five aliquots of each 100 ng of purified pSTV_{AA}-mutant plasmid pool were then re-transformed into the bioreporter strain *E. coli* BW25113 $\Delta rbsB$, for testing of ribose- and 13CHD-dependent expression of GFPmut2. Library aliquots were again stored at - 80°C.

DNA shuffling and site saturation mutagenesis

As alternative to ep-PCR we used DNA shuffling to create new *rbsB* variants. For this we used eight *rbsB* variants as template (Table 2), which were amplified by PCR using primers outside the coding regions and beyond the *XcmI* and *Sall* sites (Fig. S1, Table 4). PCR-amplified templates (200 ng each) were mixed and digested with 0.5 U of DNase I for 3 min at 15°C, after which the reaction was inactivated at 80°C for 10 min. Subaliquots of 200 ng of fragmented DNA were then reassembled by PCR in progressive hybridization in presence of 2.5 U of GoTaq polymerase and 200 μ M of each dNTP in the following temperature cycles. After an initial denaturation period of 2 min at 94°C, the following steps were repeated for 35 cycles: 40s at 94°C, 90s from 65°C to 41°C in intervals of 3°C and 90s at 68°C, followed by a final 30 min period at 68°C. The PCR reassembly products were next re-amplified with primers located inside the previous ones (Fig. S1, Table 4), visualized by agarose electrophoresis and 1–kb DNA bands were isolated and purified. These products were digested with *XcmI* and *Sall* and ligated to pSTV_{AA} digested with the same enzymes. Ligation mixture aliquots of 100 ng each were then transformed into the bioreporter strain *E. coli* BW25113 $\Delta rbsB$.

Positions R141 and D215 in the DT002 and DT016 were changed by site-saturation mutagenesis. For this, the entire plasmid(s) pSTV_{AA}-DT002 or -016 (5 ng) was amplified by Q5 High-Fidelity DNA polymerase using overlapping but reverse complementary primers with ambiguous bases at the desired positions (Fig. 4, Table 4). PCR products were digested with *DpnI* to remove template DNA and after enzyme inactivation were directly transformed into the bioreporter strain *E. coli* BW25113 $\Delta rbsB$. Variant genes were confirmed by sequencing.

Mutagenesis of neighboring residues in DT016

In order to reconstitute the *rbsB* gene 12 overlapping primers were designed (Table 5). Assembly of the 12 designed primers was carried out in 2 steps. In the first step, primers were divided in 3 independent annealing groups (Fig. 5A). Reaction for every group was performed with 200 μ M of dNTPs, 50 nM of each primer and 0.02 U/ μ L of Q5 High-Fidelity DNA polymerase. After an initial denaturation period of 30s at 98°C, the following steps were repeated for 10 cycles: 10s at 98°C, 15s

at 45°C and 20s at 72°C, followed by a final 2 min period at 72°C. In the second step, 2 µL of the 3 independent reactions were mixed with 200 µM of dNTPs and 0.02 U/ µL of Q5 High-Fidelity DNA polymerase (same PCR conditions as in step1). PCR products had a final extension reaction with external primers, 1 µL of the 2nd assembly reaction was amplified by PCR and reactions were performed in triplicate to remove any bias. Initial denaturation period of 30s at 98°C, the following steps were repeated for 35 cycles: 10s at 98°C, 30s at 68°C and 30s at 72°C, followed by a final 5 min period at 72°C. Amplified products were visualized by agarose electrophoresis and 1 kb DNA bands were isolated and purified. Final products were digested with *XcmI* and *Sall* and ligated to pSTVP_{AA} digested with same enzymes. After overnight ligation with T4 DNA ligase, D5Hα cells were transformed with 100 ng of DNA for plasmid replication. After rescue a small percentage of cells were plated and were send for sequencing to estimate library size and variability.

RbsB-based bioreporters assays

RbsB- or mutant-RbsB based libraries in *E. coli*, or single purified clones were screened for GFPmut2 expression by FACS and/or flow cytometry, as described in detail previously (Chapter 1, ^[39]). In short, library screening of clones was done of cells encapsulated and grown to microcolonies in alginate beads, before induction. Beads (microcolonies) expressing higher FITC signal than the set thresholds were sorted and collected in tubes containing 1 ml LB medium supplemented with Amp and Cm. Sorted mutants were regrown and re-screened either as alginate-bead mixtures or as pure cultures in 96-well plates (in at least eight individually grown replicates).

Media, incubation conditions, experiments and instruments details are explained in chapter 1 ^[39].

Statistical analysis

Flow cytometry induction of GFPmut2 in *E. coli* cultures was measured in multiple independently grown biological replicates ($n = 6 - 14$). Induction is then expressed as the ratio of the mean GFPmut2 fluorescence of induced cultures by that of their uninduced (split) halves. Differences among the mean GFPmut2 uninduced fluorescence and fold induction were tested using Student's t-test.

Structure threading and analysis

Threaded structures of RbsB variants were determined with Swiss-Model^[41-45] and Phyre2^[46], wild-type RbsB structure (PDB ID: 2DRI) was used as a scaffold. Structural superpositions and distance analysis were calculated with UCSF Chimera^[64].

Acknowledgements

I thank Julien Dénéréaz, Ludivine Jotterand, Thibaut Laurent, Valentina Benigno, Maxime Brunner and Cédric Schneider for helping with libraries preparation, cloning and flow cytometry screening. Without the excellent work of these internship students this project would not be possible.

References

1. Lutz, S. and S.M. Iamurri, *Protein Engineering: Past, Present, and Future*. Methods Mol Biol, 2018. **1685**: p. 1-12.
2. Sinha, R. and P. Shukla, *Current Trends in Protein Engineering: Updates and Progress*. Curr Protein Pept Sci, 2019. **20**(5): p. 398-407.
3. Singh, R.K., et al., *Protein Engineering Approaches in the Post-Genomic Era*. Curr Protein Pept Sci, 2018. **19**(1): p. 5-15.
4. Baweja, M., et al., *Current Technological Improvements in Enzymes toward Their Biotechnological Applications*. Frontiers in microbiology, 2016. **7**: p. 965-965.
5. Kumar, V., et al., *Microbial Enzyme Engineering: Applications and Perspectives*, in *Recent advances in Applied Microbiology*, P. Shukla, Editor. 2017, Springer Singapore: Singapore. p. 259-273.
6. Chiu, M.L. and G.L. Gilliland, *Engineering antibody therapeutics*. Curr Opin Struct Biol, 2016. **38**: p. 163-73.
7. Gaw, S.L., et al., *Rational Design of Amphiphilic Peptides and Its Effect on Antifouling Performance*. Biomacromolecules, 2018. **19**(9): p. 3620-3627.
8. Shen, C., et al., *Protein Engineering on Human Recombinant Follistatin: Enhancing Pharmacokinetic Characteristics for Therapeutic Application*. J Pharmacol Exp Ther, 2018. **366**(2): p. 291-302.
9. Pandelieva, A.T., et al., *Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif*. ACS Chem Biol, 2016. **11**(2): p. 508-17.
10. Kuchner, O. and F.H. Arnold, *Directed evolution of enzyme catalysts*. Trends Biotechnol, 1997. **15**(12): p. 523-30.
11. Brissos, V., et al., *Engineering a Bacterial DyP-Type Peroxidase for Enhanced Oxidation of Lignin-Related Phenolics at Alkaline pH*. ACS Catalysis, 2017. **7**(5): p. 3454-3465.
12. Guntas, G., et al., *Directed evolution of protein switches and their application to the creation of ligand-binding proteins*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(32): p. 11224.
13. Taylor, N.D., et al., *Engineering an allosteric transcription factor to respond to new ligands*. Nature Methods, 2016. **13**(2): p. 177-183.
14. Romero, P.A. and F.H. Arnold, *Exploring protein fitness landscapes by directed evolution*. Nature Reviews Molecular Cell Biology, 2009. **10**(12): p. 866-876.
15. Tracewell, C.A. and F.H. Arnold, *Directed enzyme evolution: climbing fitness peaks one amino acid at a time*. Curr Opin Chem Biol, 2009. **13**(1): p. 3-9.

16. Cirino, P.C., K.M. Mayer, and D. Umeno, *Generating mutant libraries using error-prone PCR*. *Methods Mol Biol*, 2003. **231**: p. 3-9.
17. Brakmann, S., *Discovery of superior enzymes by directed molecular evolution*. *Chembiochem*, 2001. **2**(12): p. 865-71.
18. Reetz, M.T., D. Kahakeaw, and R. Lohmer, *Addressing the numbers problem in directed evolution*. *Chembiochem*, 2008. **9**(11): p. 1797-804.
19. Beggah, S., et al., *Mutant HbpR transcription activator isolation for 2-chlorobiphenyl via green fluorescent protein-based flow cytometry and cell sorting*. *Microbial Biotechnology*, 2008. **1**(1): p. 68-78.
20. Garmendia, J., et al., *À la carte transcriptional regulators: unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors*. *Molecular Microbiology*, 2008. **42**(1): p. 47-59.
21. Garmendia, J., et al., *À la carte transcriptional regulators: unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors*. *Molecular Microbiology*, 2001. **42**(1): p. 47-59.
22. Chu, B.C.H. and H.J. Vogel, *A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport*. *Biological Chemistry*, 2011. **392**(1-2): p. 39.
23. Berntsson, R.P., et al., *A structural classification of substrate-binding proteins*. *FEBS Lett*, 2010. **584**(12): p. 2606-17.
24. Binnie, R.A., et al., *Functional mapping of the surface of Escherichia coli ribose-binding protein: mutations that affect chemotaxis and transport*. *Protein science : a publication of the Protein Society*, 1992. **1**(12): p. 1642-1651.
25. Li, H.Y., et al., *Analysis of Conformational Motions and Residue Fluctuations for Escherichia coli Ribose-Binding Protein Revealed with Elastic Network Models*. *International Journal of Molecular Sciences*, 2013. **14**(5): p. 10552-10569.
26. Björkman, A.J. and S.L. Mowbray, *Multiple open forms of ribose-binding protein trace the path of its conformational change*¹¹*Edited by R. Huber*. *Journal of Molecular Biology*, 1998. **279**(3): p. 651-664.
27. Medintz, I.L. and J.R. Deschamps, *Maltose-binding protein: a versatile platform for prototyping biosensing*. *Current Opinion in Biotechnology*, 2006. **17**(1): p. 17-27.
28. van der Meer, J.R. and S. Belkin, *Where microbiology meets microengineering: design and applications of reporter bacteria*. *Nature Reviews Microbiology*, 2010. **8**: p. 511.
29. Reimer, A., et al., *Escherichia coli ribose binding protein based bioreporters revisited*. *Scientific Reports*, 2014. **4**: p. 5626.
30. Baumgartner, J.W., et al., *Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ*. *Journal of Bacteriology*, 1994. **176**(4): p. 1157-1163.

31. Dwyer, M.A. and H.W. Hellinga, *Periplasmic binding proteins: a versatile superfamily for protein engineering*. Current Opinion in Structural Biology, 2004. **14**(4): p. 495-504.
32. Banda-Vázquez, J., et al., *Redesign of LAOBP to bind novel l-amino acid ligands*. Protein Sci, 2018. **27**(5): p. 957-968.
33. Scheib, U., et al., *Change in protein-ligand specificity through binding pocket grafting*. Journal of Structural Biology, 2014. **185**(2): p. 186-192.
34. Boas, F.E. and P.B. Harbury, *Design of Protein–Ligand Binding Based on the Molecular-Mechanics Energy Model*. Journal of Molecular Biology, 2008. **380**(2): p. 415-424.
35. Amiss, T.J., et al., *Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor*. Protein Science : A Publication of the Protein Society, 2007. **16**(11): p. 2350-2359.
36. Ko, W., S. Kim, and H.S. Lee, *Engineering a periplasmic binding protein for amino acid sensors with improved binding properties*. Organic & Biomolecular Chemistry, 2017. **15**(41): p. 8761-8769.
37. Vercillo, N.C., et al., *Analysis of ligand binding to a ribose biosensor using site-directed mutagenesis and fluorescence spectroscopy*. Protein Science : A Publication of the Protein Society, 2007. **16**(3): p. 362-368.
38. Reimer, A., et al., *Complete alanine scanning of the Escherichia coli RbsB ribose binding protein reveals residues important for chemoreceptor signaling and periplasmic abundance*. Scientific Reports, 2017. **7**: p. 8245.
39. Tavares, D., et al., *Computational redesign of the Escherichia coli ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol*. Scientific Reports, 2019. **9**(1): p. 16940.
40. Antunes, M.S., et al., *Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway*. PLoS ONE, 2011. **6**(1): p. e16292.
41. Bienert, S., et al., *The SWISS-MODEL Repository—new features and functionality*. Nucleic Acids Research, 2016. **45**(D1): p. D313-D319.
42. Guex, N., M.C. Peitsch, and T. Schwede, *Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective*. ELECTROPHORESIS, 2009. **30**(S1): p. S162-S173.
43. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes*. Nucleic Acids Research, 2018. **46**(W1): p. W296-W303.
44. Bertoni, M., et al., *Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology*. Scientific Reports, 2017. **7**(1): p. 10480.
45. Studer, G., et al., *QMEANDisCo—distance constraints applied on model quality estimation*. Bioinformatics, 2019. **36**(6): p. 1765-1771.
46. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nature Protocols, 2015. **10**(6): p. 845-858.

47. Ittisoponpisan, S., et al., *Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated?* Journal of Molecular Biology, 2019. **431**(11): p. 2197-2212.
48. Schreier, B., et al., *Computational design of ligand binding is not a solved problem.* Proceedings of the National Academy of Sciences, 2009. **106**(44): p. 18491-18496.
49. Yang, W. and L. Lai, *Computational design of ligand-binding proteins.* Current Opinion in Structural Biology, 2017. **45**(Supplement C): p. 67-73.
50. Zheng, J., J.L. Payne, and A. Wagner, *Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks.* Science, 2019. **365**(6451): p. 347.
51. Brustad, E.M. and F.H. Arnold, *Optimizing non-natural protein function with directed evolution.* Current Opinion in Chemical Biology, 2011. **15**(2): p. 201-210.
52. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions.* Nature, 2003. **423**(6936): p. 185-190.
53. Polizzi, N.F. and W.F. DeGrado, *A defined structural unit enables de novo design of small-molecule-binding proteins.* Science, 2020. **369**(6508): p. 1227-1233.
54. Langan, R.A., et al., *De novo design of bioactive protein switches.* Nature, 2019. **572**(7768): p. 205-210.
55. Rocklin, G.J., et al., *Global analysis of protein folding using massively parallel design, synthesis, and testing.* Science, 2017. **357**(6347): p. 168-175.
56. Sesterhenn, F., et al., *De novo protein design enables the precise induction of RSV-neutralizing antibodies.* Science, 2020. **368**(6492): p. eaay5051.
57. Stank, A., et al., *Protein Binding Pocket Dynamics.* Accounts of Chemical Research, 2016. **49**(5): p. 809-815.
58. Ravindranathan, K.P., E. Gallicchio, and R.M. Levy, *Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein.* J Mol Biol, 2005. **353**(1): p. 196-210.
59. Unione, L., et al., *Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations.* ACS Chemical Biology, 2016. **11**(8): p. 2149-2157.
60. Riley, M., *Functions of the gene products of Escherichia coli.* Microbiological reviews, 1993. **57**(4): p. 862-952.
61. Stewart, J.B. and M.A. Hermodson, *Topology of RbsC, the membrane component of the Escherichia coli ribose transporter.* Journal of bacteriology, 2003. **185**(17): p. 5234-5239.
62. Zheng, J., N. Guo, and A. Wagner, *Selection enhances protein evolvability by increasing mutational robustness and foldability.* Science, 2020. **370**(6521): p. eabb5962.
63. Platt, R., et al., *Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the Escherichia coli chromosome.* Plasmid, 2000. **43**(1): p. 12-23.

64. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.

Supplementary information

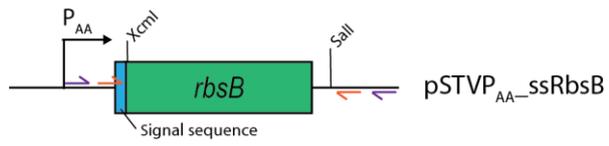


Figure S1- Scheme of used primers for DNA shuffling and important restrictions sites. Outer and inner primers are represented in purple and orange, respectively. Not drawn to scale.

KDTIALVVSTLNN **PFSV**SLKDGAQKEADKLGYNLVVLD SQ
NNPAKELANVQDLTVRGTKILL **LIVPT**DSDAVGNAVKMAN
QANIPVIT **TLGSQ**ATKGEVVS HIASDNVLGGKIAGDYIAKKA
GEGAKVIELQGI **AGASA**ARER GEGFQQAVAAHKFNVLAS
QP **ADWDRI**KGLNVMQNLLTAHPDVQAVF **QND**EMALG
ALRALQTAGKSDVMV **VGADG**TPDGEKAVNDGKLAAT **IA**
MLPDQIGAKGVETADKVLKGEKVQAKYPVDLKL VVKQ

Figure S2 – Amino acid sequence of the DT016 variant. Nine residues mutated based on Rosetta predictions, 1st generation mutations, are highlighted in yellow. Neighboring residues targeted by site saturation mutagenesis are highlighted in blue.

CHAPTER 4

**An *in vivo* system to follow wild-type and mutant
ribose-binding protein expression and localization in
*Escherichia coli***

Abstract

Periplasmic-binding proteins have been proposed as a general scaffold to design sensor proteins with new recognition specificities for non-natural compounds. They can be integrated in bacterial bioreporter systems, which produce a concentration-dependent signal after ligand binding to the sensor. However, computational based design approaches to create new ligand-binding properties ignore more general properties of the protein, such as proper translocation into the periplasm, folding and interaction with membrane receptor. This may be one of the reasons for the poor success in recent periplasmic-binding protein design studies.

To facilitate downstream screening of secondary mutants with potentially improved folding, translocation and receptor interactions of periplasmic-binding proteins we developed and tested an *in vivo* system based on mCherry protein fusions and subcellular localisation measurements. As proof of concept, we studied translocation and periplasmic localisation of the wild-type ribose-binding protein RbsB in *Escherichia coli* and mutant-RbsB with altered ligand-binding properties obtained through protein design and selection. Whereas RbsB-mCherry protein clearly localized to the periplasmic space and centered in polar regions depending on chemoreceptor availability, mutant-RbsB-mCherry expression resulted in high proportions of cells devoid of clear foci and low proportions of cells with multiple fluorescent foci, suggesting poorer translocation and mislocalisation. In addition, polar foci of mutants were less fluorescent, suggesting poorer chemoreceptor binding. By spiking further derivative mutant libraries generated by error-prone PCR without or with different proportions of *E. coli* expressing wild-type RbsB-mCherry we could estimate the potential improvement and deterioration of mutants with wild-type-like periplasmic localisation. Our results show that an *in vivo* system can be used to detect cells expressing mutant proteins with better periplasmic localisation, helping to increase the changes to find candidates with secondary mutations improving their signal transduction capacity.

Introduction

Periplasmic binding proteins (PBPs) are part of a large family of proteins with a typical and conserved bilobal structure ^[1,2]. PBPs are involved in nutrient and mineral scavenging for bacterial cells ^[2], binding specific ligands (e.g., sugars), and presenting the bound molecules to specific transport channels and/ or membrane receptor involved in chemotactic movement ^[3]. Their unique ligand-binding properties have propelled PBPs as an attractive class of proteins to use for biosensing. PBPs can be either purified and their ligand-binding capacity be exploited using in-vitro measurements (e.g., isothermal titration calorimetry), or the PBP ligand-binding activity can be integrated in an *in vivo* hybrid signaling chain leading to fluorescent protein expression ^[4,5]. In that configuration the presence of a ligand to the cell leads to conformational changes of the PBP, which binds to its membrane receptor and triggers the synthetic signaling chain. The synthetic signaling chain that has been perused for this purpose consisted of RbsB (the ribose-binding protein of *E. coli*) and its hybrid receptor Trz1 ^[6], formed by fusion between the C-terminal part of the *E. coli* EnvZ osmoregulation histidine kinase and the N-terminal part of the Trg chemotaxis receptor^[7]. Addition of ribose to cultures carrying the signaling chain starts a phosphorylation cascade that leads to *gfp* expression^[6], which can be easily measured in a variety of ways ^[5].

With the aim to expand detection specificity of the biosensors, PBPs have been proposed as a scaffold to design novel proteins with new binding capacities. Unfortunately, *in vitro* and *in vivo* based mutagenesis approaches to design new receptor PBP-based proteins have had limited success and shown poorly reproducible outcomes ^[8-11]. Several studies have shown ligand-binding reconstruction through grafting of semi-natural binding pockets ^[8,9], but so far there has only been one report of new PBP variants for non-natural compound recognition based on redesign of the RbsB ligand-binding pocket ^[12]. Although the obtained RbsB-variants displayed no more ribose recognition, their affinity for the new ligand 1,3-cyclohexanediol was much lower (mM-range) and the (purified) proteins suffered from poor stability and or misfolding ^[12]. Since ligand-prediction algorithms, mutagenesis and screening methods mostly focused on the ligand-binding properties of the PBP, other key steps of the signal transduction chain by the PBP have been largely ignored. For a protein like RbsB in *E. coli*, this involves (i) proper translocation by SecB into the periplasm through the Sec pathway, (ii) proper folding in the periplasm enabling ribose-binding, (iii) docking of the RbsB-bound ribose to the transport channels (RbsAC) and, partly, to the Trg chemoreceptor ^[13-15]. Ribose transport in *E. coli* involves the cytoplasmic membrane transporter RbsC and the cytosolic ATPase RbsA to provide energy. Ribose is presented to the channel in its bound form by RbsB, and released into the cytoplasm by the ribokinase RsbK, with subsequent phosphorylation ^[15]. The Trg membrane receptor controls chemotaxis of cells

towards ribose upon binding by ribose-bound RbsB^[16]. The primary goal of this work was thus to design an *in vivo* screening system that would enable studying the localisation aspects of the RbsB-signaling chain with the idea that this might help to improve the selection of mutant RbsB with designed different ligand-binding properties for further mutations that favor their proper localisation.

RbsB mutations introduced to improve ligand-binding do not specifically take into account proper translocation into the periplasm after cytosolic expression, and, therefore, such mutants may display misfolding and poor translocation into the periplasmic space, which reduces their ligand and membrane receptor binding capacity and hence, their signaling capacity. The translocation of fusion proteins to the periplasm of Gram-negative bacteria (e.g., *E. coli*) has been widely studied^[17-20]. Due to low costs and high production yields *E. coli* is commonly used as a platform to express and purify target proteins (e.g., therapeutic antibodies)^[21, 22]. Translocation to the extracellular space offers several advantages for protein purification when compared with intracellular expression. Translocated proteins do not require disruption of the cellular membranes, suffer less from inclusion bodies, and have less contamination with cytosolic proteins^[23]. In addition, proteases are less abundant in the periplasmic or extracellular space, which is further a favorable environment for folding^[24].

The two major bacterial translocation mechanisms are i) the twin-arginine translocation (Tat) and ii) the ubiquitous general secretion (Sec) pathways^[25]. The twin-arginine pathway recognizes a twin-arginine pair (RR) in the signal peptide, and translocate its substrate proteins in a complete folded form through the Tat translocase^[26]. After cytosolic expression and folding the RR motif binds to the membrane receptor complex, formed by TatB and TatC. Subsequently, the protein is translocated through a membrane pore formed by multiple homooligomeric complexes of Tata^[27]. A signal peptidase cleaves the signal peptide and the mature protein is released in the periplasm^[25]. The Sec translocation system is the most important route to export proteins from the cytoplasm to periplasmic space^[28]. In the Sec system the substrate proteins are translocated in an unfolded state^[28]. After translation, chaperones (e.g., SecB) protect the substrates from aggregation and maintain them in an unfolded configuration. Subsequently, they present the unfolded protein to SecA, which recognizes the signal peptide and the protein through the translocase pore SecYEG. The signal peptide is removed by peptidase cleavage and the protein is folded and released in the periplasm^[25, 29].

In order to study the translocation and periplasmic localisation of RbsB and a number of previously generated RbsB-mutants with new ligand binding properties^[12], we deployed fluorescent protein fusions and single cell epifluorescence microscopy. We tested *mcherry* gene fusions to either the 5'- or 3'-end of wild-type *rbsB* gene, and used either the natural signal peptide of RbsB (and the corresponding Sec translocation pathway) or introduced a TorA signal sequence for the Tat pathway. PBPs-fluorescent protein variants have been constructed before, but only to improve their purification^[18, 30]. Protein export by the Tat and Sec pathways is overall similar^[31], and differences among their

signal peptides are a characteristic consensus motif (S/T-R-R-X-F-L-K) present in all Tat signal peptides, which is absent in Sec-recognized peptides. The choice of optimal translocation and signal peptide depends on the fused protein of interest and has to be tested for each case individually [18, 30, 32]. RbsB wild-type and mutant-mCherry proteins were expressed in different *E. coli* backgrounds [12], and relative positions of fluorescent foci in individual cells were quantified and compared. Finally, we constructed three further random mutation libraries on the basis of the existing ligand-binding mutants and spiked those with wild-type RbsB/mCherry to estimate how well mutants with potential improved translocation behaviour could be detected using this imaging pipeline. The presented *in vivo* system may be used to highlight signaling deficiencies of PBPs and help to improve mutant selection with better signaling behavior from libraries.

Results

Fusion constructs inform about RbsB expression and translocation

In order to follow wild-type RbsB expression and distinguish the localisation steps in the ribose-RbsB mediated signaling cascade, we produced mCherry fluorescence protein fusions. We first tested functionality and effects on expression levels of C- or N-terminal fusion positions, driven from two different promoters in *E. coli* and in presence or absence of a different translocation signal sequence. RbsB/mCherry fusion constructs introduced on plasmids and expressed from the constitutive P_{AA} promoter with the native *rbsB* signal sequence (nss, Fig. S1) [6], were poorly reproduced in *E. coli*, irrespective of the position of the *mCherry* tag. *E. coli* transformants carrying P_{AA} -nss-*rbsB*-*mCherry* or P_{AA} -nss-*mCherry*-*rbsB* grew poorly, and faster growing colonies harbored plasmids with deletions and mutations (not shown). In contrast, transformants expressing P_{AA} -*rbsB*-*mCherry* without signal sequence were viable, showing bright fluorescence (not shown). This indicated that not expression of RbsB-mCherry itself, but its translocation was deleterious for the cells.

In contrast, plasmid-cloned RbsB/mCherry fusion constructs with either the RbsB native signal sequence (Nss) or the TorA-signal sequence (TorAss) for the TAT translocation pathway, expressed under control of the arabinose-inducible P_{BAD} promoter (Fig. S2) were stably maintained in absence of arabinose in *E. coli* DH5 α . However, after 3 h induction with 0.5% L-arabinose, cells expressing RbsB/mCherry fusions with the TorAss were longer than cells with the same constructs but carrying the wild-type RbsB signal sequence (Fig. 1A–D). Cells expressing TorAss-RbsB/mCherry displayed a size range of $8.46 \pm 4.8 \mu\text{m}$ whereas those expressing Nss-RbsB/mCherry were on average $3.50 \pm 0.7 \mu\text{m}$ ($p = 0.000014$, $n = 18$ replicates, 10^3 cells per replicate, Fig. 1A-D). In case of the TorAss fused to the N- and mCherry to the C-terminal end of RbsB, induction further yielded cells with swollen ends and cytoplasmic accumulation of the fusion protein, but without sign of periplasmic localisation (Fig. 1A). In contrast, fusing both the *torAss* and mCherry to the RbsB N-terminus yielded cells with visible periplasmic fluorescence, and in some cells, a visible fluorescent spot near the cell poles (Fig. 1B). Much clearer periplasmic fluorescence and polar fluorescent spots, however, were obtained for the constructs with wild-type RbsB signal sequence to N- and mCherry fused to the C-terminal end (Fig. 1C). Fusing both wild-type Nss and mCherry to the N-terminus of RbsB again led to very poorly visible periplasmic and polar fluorescence (Fig. 1D).

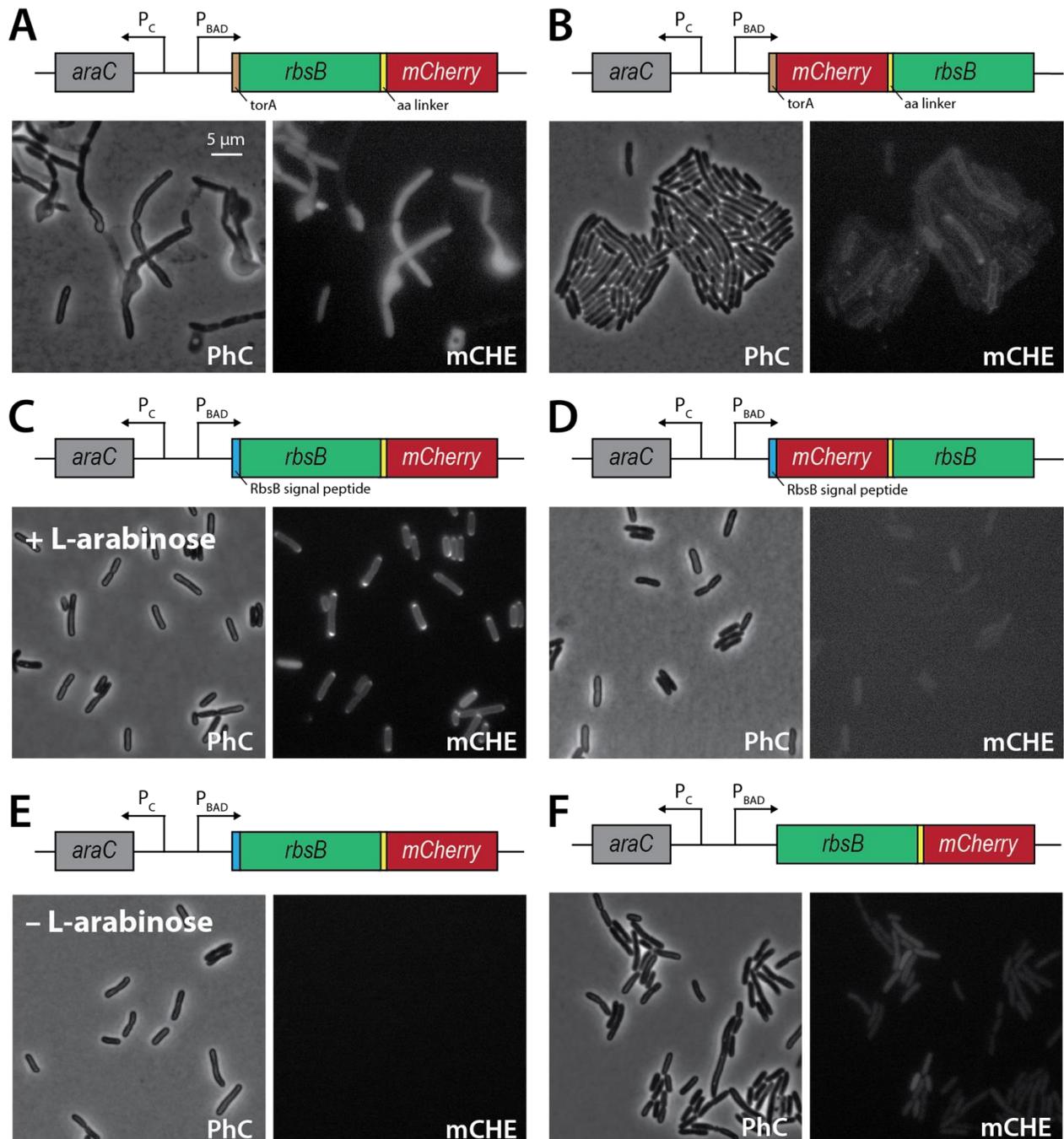


Figure 1- Effect of fluorescent protein fusion positioning and signal sequences on RbsB expression and translocation. A) *E. coli* DH5 α expressing *torA-rbsB-mCherry* from P_{BAD} . B) as (A) but for *torA-mCherry-rbsB*. C) as (A) but for *rbsB-mCherry* with the native RbsB signal sequence. D) as (C) but for *mCherry-rbsB*. E) *E. coli* DH5 α expressing *rbsB-mCherry* with the native RbsB signal sequence without L-arabinose induction. F) Induction of *rbsB-mCherry* from P_{BAD} but without native RbsB signal peptide. Fusion constructs are schematically drawn on top of each micrograph (not to scale). Cells were incubated for 3 h and expression of fusion protein was induced with 0.5% L-arabinose (except in panel E). Arrows indicate promoters and the transcription direction. PhC, phase-contrast. mCHE, mCherry fluorescence. All fluorescence images (mCHE) are scaled to the same intensity (low-high).

In order to test whether the observed periplasmic signal of *E. coli* DH5 α cells expressing *nss-rbsB-mCherry* was specific, we examined cells without induction with arabinose and without Nss signal sequence. Without induction, no fluorescent signal was observed (Fig. 1E). Furthermore, the clear polar fluorescence in arabinose-induced cells, and their weak fluorescence halo of periplasmic localized RbsB-mCherry (Fig. 1C), were lost when the Nss signal sequence was removed (Fig. 1F). As a consequence, a homogenous distribution of RbsB-mCherry fluorescence in the cytoplasm was observed (Fig. 1F), indicating that only in presence of the proper signal sequence, RbsB-mCherry is able to translocate to the periplasm and accumulate at the cell poles. Collectively, these results indicated that the C-terminal end fusion (*rbsB-mcherry*) fused to the native RbsB Nss signal sequence was the optimal configuration for expression and translocation into the periplasm.

Functionality of RbsB-mCherry

Next we studied the functionality of the RbsB-mCherry fusion protein to respond to its native ligand ribose in *E. coli* by measuring induction of GFP formation from the *Trz1-ompR-ompC':gfp* hybrid signaling chain ^[6]. The plasmid carrying the P_{BAD} -*nss-rbsB-mCherry* fusion was hereto cotransformed into *E. coli* $\Delta rbsB$ cells, further containing the ribose-RbsB-dependent *Trz1-ompC':gfp* bioreporter system that was previously developed ^[12] (Fig. 2A, inset). Incubation with 0.1 mM ribose for 2h30 resulted in a 1.7-fold induction of GFP fluorescence in flow cytometry (Fig. 2A), which was statistically significant ($p < 0.001$, ANOVA followed by post-hoc Tukey test), although lower than the 13-fold induction of wild-type *rbsB* in the same *E. coli* background ^[12]. In contrast, none of the other three configurations resulted in any statistically significant change of GFP expression of the *Trz1-ompC':gfp* signaling pathway (Fig. 2A), suggesting these are proteins inactive in signaling or insufficiently present in the periplasm. This confirmed that the C-terminally fused mCherry to RbsB is a functional protein, and thus the best proxy to follow RbsB localization and signaling. Adding ribose caused a slight increase of mCherry fluorescence in all four strains (Fig. 2B), which may have been due to an increased induction from the P_{BAD} promoter. Consequently, we used this construct to further study effects of RbsB mutations on the translocation and signaling process.

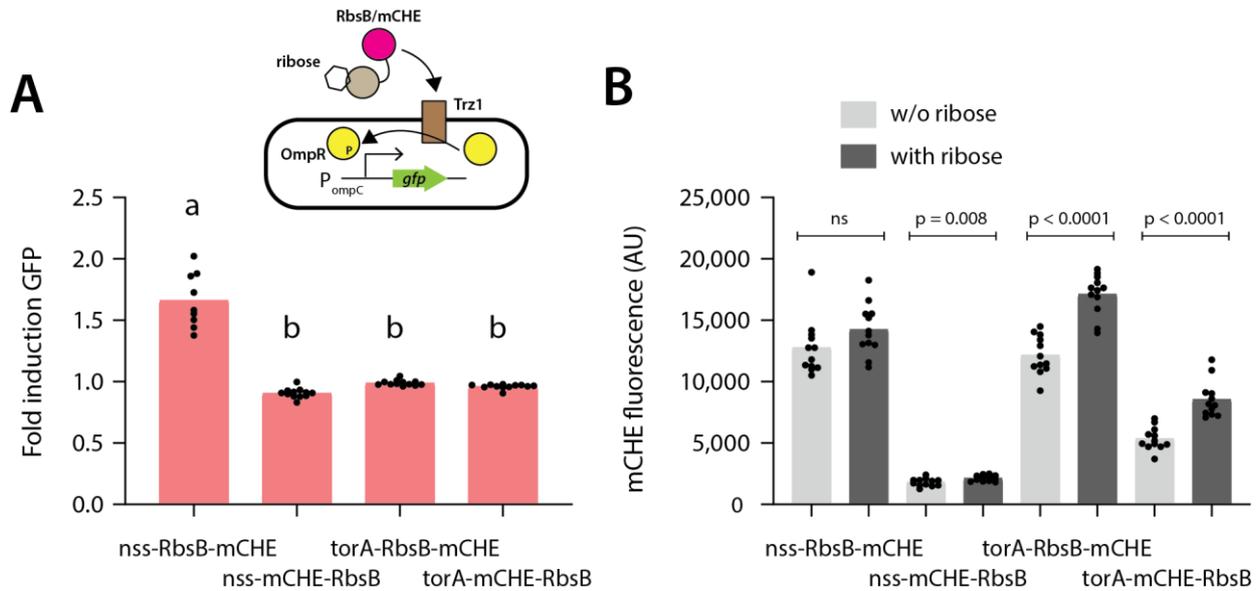


Figure 2- Ribose-dependent induction of the Trz1-*ompC'*::*gfp* signaling chain in *E. coli* $\Delta rbsB$ by different plasmid-expressed RbsB/mCherry fusion constructs. (A) Fold induction of GFP fluorescence in presence of 0.1 mM ribose for 2h30 compared to no ribose. Bars show the mean ratio (bars) from twelve replicates (black dots). Letters indicate significance groups in ANOVA followed by post-hoc Tukey test ($p_{a,b} < 0.001$). Inset shows principle of ribose-dependent RbsB-Trz1 induction of *gfp* from the *ompC* promoter. (B) Effect of ribose addition on RbsB/mCherry signal intensities for four fusion constructs. Bars indicate the mean of population fluorescence means in flow cytometry across twelve replicates (black dots) in absence (light grey) or presence of ribose (dark grey). P-values derived from two-sided t-test ($n = 12$). Ns- not significant p-value.

Wild-type and mutant-RbsB-mCherry proteins localize in the periplasm

In previous work, we isolated a number of RbsB mutants with novel ligand-binding properties [12]. Although these proteins had altered ligand-binding properties, biochemical and cell data showed them to be impaired in stability and potentially, in their translocation behavior [12]. In order to better understand if and where in the signaling pathway those RbsB mutants might be impaired, we constructed similar C-terminal end fusions as above and compared mCherry fluorescence localization to that of the wild-type RbsB. We concentrated on five RbsB mutants named DT002, DT016, DT020, DT021 and DT022, which were no longer responsive to ribose but instead to 1,3-cyclohexanediol [12]. Constructs were in first instance transformed into and tested for expression in *E. coli* DH5 α cells (Table 1), grown in liquid culture until mid-exponential phase in presence of 0.5% (w/v) L-arabinose and imaged after 3 h. All mutants expressed mCherry fluorescence and showed periplasmic and polar

fluorescence, although different in intensity and localization, and with occasional additional non-polar fluorescent spots (Fig. 3, blue arrows). In some cells inclusion bodies were observed (e.g., white arrow in DT002-mCherry, Fig. 3). Side periplasmic foci were in general less intense than polar ones (e.g., blue arrows in DT020- and DT022-mCherry, Fig. 3).

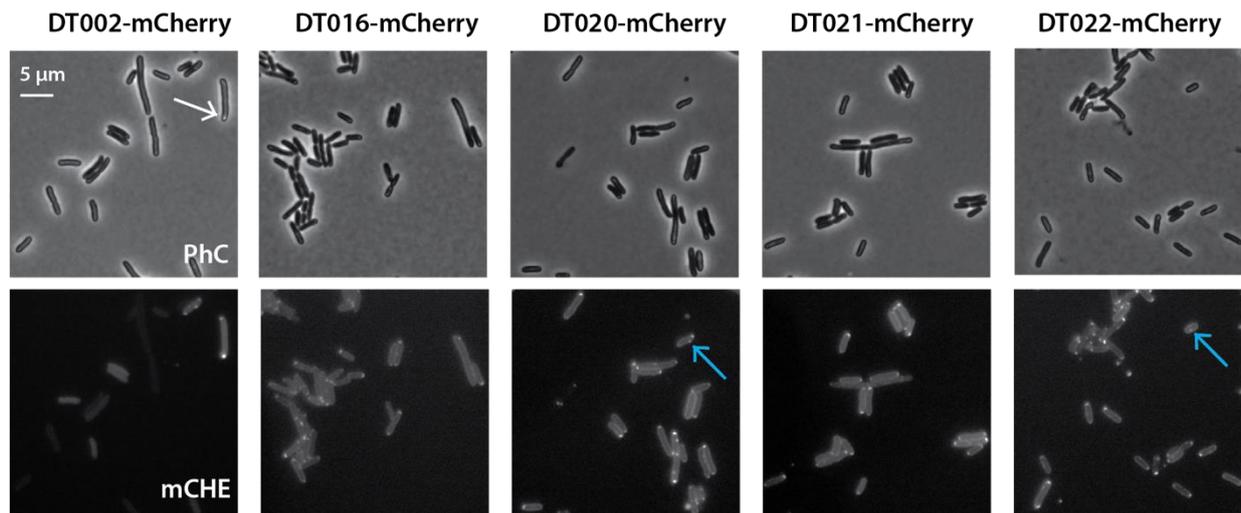


Figure 3- Expression and cellular localization of RbsB-mutant mCherry C-terminal translational fusions in *E. coli* DH5 α cells. Images show phase-contrast (PhC) and mCherry (mCHE) fluorescence of cells induced for 3 h with 0.5% L-arabinose. All mutant-mCherry proteins were expressed from P_{BAD} and with the native RbsB signal sequence. The white arrow in DT002-mCherry points to possible cytoplasmic inclusion bodies. Blue arrows in DT020- and DT022-mCherry indicate non-polar periplasmic fluorescent foci. All fluorescence images (mCHE) were scaled to the same brightness (300-900).

In order to contrast expression and localization differences from RbsB- or mutant-mCherry proteins, and to better understand potential defects in signaling steps, we quantified and compared (relative) foci intensities, foci positions and the number of foci per cell among the various strains. In addition, we tested an *E. coli* host with the native chemoreceptor for RbsB (i.e., Trg) and one in which Trg was deleted, both in absence or presence of a plasmid co-expressing the Trz1 hybrid receptor, to which RbsB is also expected to bind ^[7]. Moreover, we studied expression of the fusion proteins in an *E. coli* host without the chromosomal *rbsB*, and one with *rbsK* deletion (coding for a protein involved in ribose uptake).

To analyze foci intensities among wild-type RbsB- and mutant-mCherry fusion proteins, we quantified the mean of the 95th percentile fluorescence pixel per cell across all segmented cells in the images ($n = 10^3$ cells per replicate, Fig. 4A). The mean 95th percentile fluorescence of RbsB-mCherry increased tenfold both in presence or absence of Trg upon induction with L-arabinose (Fig. 4A, $p = 0.00018$, two-sided t-test, $n = 3$ biological replicates). In contrast, the mean 95th percentile fluorescence decreased six-fold in *E. coli* DH5 α expressing RbsB-mCherry without the Nss signal sequence (Fig. 4A, $p = 0.0248$, $n = 3$). This indicated that the mean 95th percentile fluorescence is an appropriate measure for foci fluorescence, and showed that appearance of foci is specific for periplasmic translocated RbsB-mCherry.

To study the possible influence of host background in foci formation and intensity, four *E. coli* host cells (i.e., MG1655 wild-type, Δtrg , $\Delta rbsB$ and $\Delta rbsK$) were transformed with wild-type- and mutant-RbsB-mCherry fusion proteins (Table 1). Notably, expression in *E. coli* $\Delta rbsB$ in all cases resulted in a slight increase of the mean 95th percentile fluorescence when compared with wild-type (i.e., MG1655) and Δtrg hosts (Fig. 4B), suggesting there is some competition due to wild-type unlabeled RbsB. The mean 95th percentile fluorescence in *E. coli* $\Delta rbsK$ of both RbsB- and DT016-mCherry, but not DT022-mCherry, was even higher than that in $\Delta rbsB$ background, indicating that more fluorescent protein localized to the poles in absence of ribose transport into the cytoplasm (Fig. 4B). Also the median fluorescence of cells was higher in *E. coli* $\Delta rbsB$ and $\Delta rbsK$ than in MG1655 and Δtrg backgrounds (Fig. S3A), suggesting that these deletions may affect expression levels themselves. In contrast, deletion of the *trg* gene (for the Trg chemoreceptor) did not result in significant differences of the mean 95th percentile fluorescence compared to wild-type (Fig. 4B).

We investigated the importance of Trg and the hybrid Trz1 receptor on RbsB- or mutant-mCherry protein localisation a little further by comparing expression in *E. coli* DH5 α and *E. coli* BW25113 Δtrg , in absence or presence of a plasmid-expressed Trz1 (Fig. 4C). Foci fluorescence was higher for equivalent constructs in *E. coli* BW25113 Δtrg than in DH5 α ($p = 0.009$, Wilcoxon rank-sum test, $n = 15$). Co-expression of the Trz1-hybrid receptor caused an increase of foci fluorescence by 3-4 fold in the Δtrg background ($p = 3.16 \times 10^{-7}$, two-sided t-test, $n = 25$, Fig. 4D), but not in DH5 α ($p = 0.2656$, Fig. 4C). This indicated that specific expression of the Trz1 receptor in absence of the cognate Trg chemoreceptor favors RbsB- or mutant-mCherry binding in the periplasmic polar regions. Globally, expression of RbsB- or mutant-mCherry was also higher in Δtrg background with co-expressed Trz1 than without (Fig. S3B, median fluorescence, $p = 0.0003$, two-sided t-test, $n = 25$).

Visualized across all cells in a biological replicate ($n = 10^3$) in a z-projection of the mean top 20 percentiles per pixel, the higher protein accumulation at the cell poles was visible even in absence of Trg, but not without induction by L-arabinose (Fig. 4E). Comparable results were obtained in *E. coli*

DH5 α cells expressing wild-type- and mutant-RbsB-mCherry proteins (Fig. S4A). This suggested either some inherent property of RbsB-mCherry to accumulate in membrane curved regions or hidden interactions with other *E. coli* chemoreceptors, which are typically located at the poles [33-35]. The localisation of the RbsC-ribose transporter is, to the best of our knowledge, not known. RbsB-mCherry fluorescence intensity in the polar regions further increased upon co-expression of Trz1, indicating positioning of this hybrid chemoreceptor to the polar regions and RbsB-mCherry binding to it. Similar results were obtained in cells expressing DT016- and DT022-mCherry proteins (Fig. 4E), or DT002-, DT020- and DT021-mCherry (Fig. S4B). In summary, this showed that RbsB- or mutant-RbsB-mCherry fluorescence mostly localized at the poles and was dependent to some extent on the presence of chemoreceptor and ribose transport.

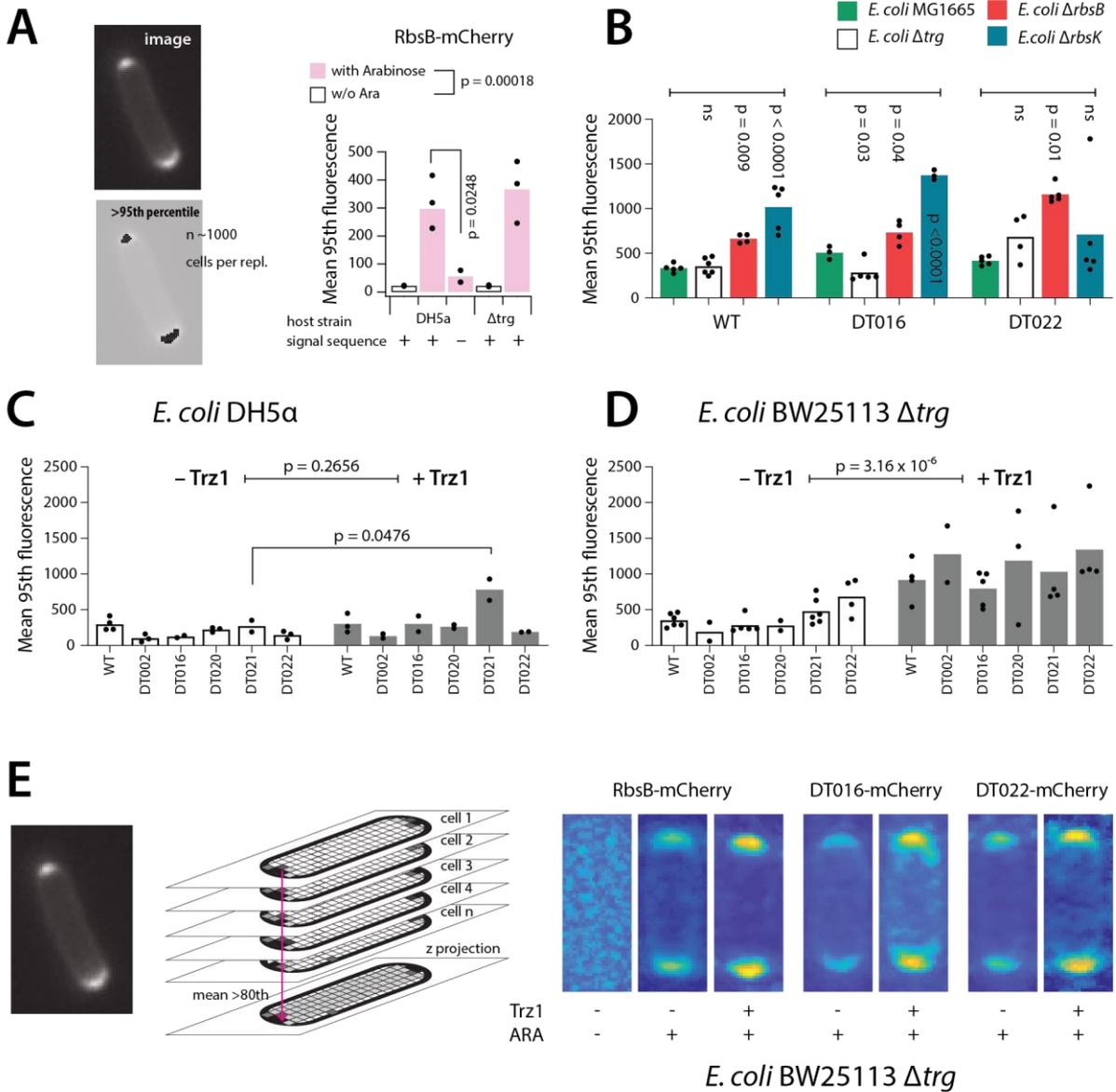


Figure 4- Foci intensity and localisation in *E. coli* cells expressing RbsB- or mutant-mCherry fusion proteins. (A) Mean 95th percentile mCherry fluorescence intensity as measure for foci intensity. Example image and cartoon show the measurement principle. Bars show mean values from biological replicates with black dots representing the individual mean 95th percentiles from around 1000 cells of a single replicate. With arabinose; 0.5 % L-arabinose during 3 h. p-values from two-sided t-tests ($n = 6$ for arabinose difference, $n = 3$ for signal sequence difference). (B) Comparison of mean 95th percentile fluorescence intensity of wild-type-RbsB-, DT016- and DT022-mCherry expressed in four different *E. coli* strains. Bars and dots as in (A). Induction with 0.5% L-arabinose for 3 h. P-values derived from one-way ANOVA, compared to the respective fusion protein expressed in *E. coli* MG1655 background. (C) Effect of host DH5 α background in presence (grey bars) or absence (white bars) of co-expressed Trz1 hybrid signaling receptor. Bars and individual dots as in (A). p-values from two-sided t-test ($n = 15$ for Trz1 difference, $n = 3$ for DT021 difference). (D) As in (C) but for *E. coli* Δ trg background. P-value from two-sided t-test ($n = 25$). (E) z-projection of the mean top 20 percentiles per pixel across all cells in a biological replicate ($n = 10^3$) standardized to the same cell length and width. Heatmap representation of fluorescence intensity in the cells. Blue and orange colors represent low and high intensity, respectively. Scale is not comparable between heat maps.

Mutant RbsB-mCherry proteins tend to localize less at the cell poles

If translocation or periplasmic folding would be affected by ligand-binding mutations (as was suspected for DT002 and DT016^[12]), one would expect to detect differences in the positions and quantities of fluorescent mCherry fusion protein in the cells compared to wild-type. In case of *E. coli* Δ trg expressing RbsB-mCherry (irrespective of co-expressing Trz1), around 33% of imaged cells carried no foci, 50% carried one or two foci, and 17% carried 3 or 4 (Fig. 5A). The proportion of cells without any foci increased for cells expressing mutant-RbsB-mCherry, in particular when Trz1 was co-expressed (Fig. 5B, $p = 0.004$, one-sided t-test, $n = 6$, taking all mutants as a group). Among the cells that showed fluorescent foci, the proportion of those with 1 or 2, or 3 and 4 foci, was not different among expressed RbsB- or mutant-mCherry (Fig. 5C, two-sided t-tests, $n = 6$). Globally, however, the proportion of cells with 3 and 4 foci (among those that showed presence of foci) decreased for cells co-expressing Trz1 (Fig. 5C, $p = 0.037$, one-sided t-test, $n = 20$). This suggested that there are differences in periplasmic translocation between wild-type- and mutant-RbsB mCherry fluorescent fusion proteins, in particular in the increased number of cells without foci.

In order to understand whether the positioning of foci was different between cells expressing wild-type RbsB- or mutant-mCherry, we plotted foci positions across some 800 cells, normalized to the same length and width (Fig. 5D). This showed clearly the strong polar localization for wild-type RbsB-

mCherry, the absence of any localization in uninduced cells, and aberrant localization in e.g., cells expressing mutant DT002-mCherry (Fig. 5D). Note that a threshold of 3 is relatively low and as a consequence, the segmentation program also picks up spurious variations in cytoplasmic fluorescence as in case of uninduced cells (Fig. 5D). In two different backgrounds, however, statistically significant differences occurred among mutants and wild-type RbsB-mCherry in the fraction of polar foci (defined as foci positions within 6 pixels from the detected cell edge on the long axis, Fig. 5E).

Whereas uninduced *E. coli* DH5a cells and those expressing RbsB-mCherry without Nss displayed background proportions of polar foci (15-20%, Fig. 5E), those in presence of arabinose displayed up to 56% of polar foci, irrespective of co-expressed Trz1 (Fig. 5E, $p = 0.86$, Wilcoxon ranked sum test, $n = 9$). Particularly cells expressing mutant DT002-, DT016-, DT020- or DT022-mcherry, showed lower polar foci fractions (between 30–45%, p -values from 3.27×10^{-5} - 0.0057, one-sided t-test, $n = 3-6$, Fig. 5E). In contrast, mutant DT021-mCherry localized indifferently from wild-type ($p = 0.28$, two-sided t-test, $n = 3 - 6$, Fig. 5E). Mutant differences persisted in absence of Trg and/or co-expressed Trz1, except for mutant DT022-mCherry (Fig. 5F). Co-expression of Trz1 in Δtrg background globally decreased the proportion of polar foci for all constructs (Fig. 5F) and leading to a higher proportion of side foci ($p = 0.0131$, Wilcoxon ranked sum test, $n = 14$). These results showed that mutant-mCherry fusions, except for DT022, tend to be poorer localized to the cell poles, which may therefore lead to poorer signal transduction.

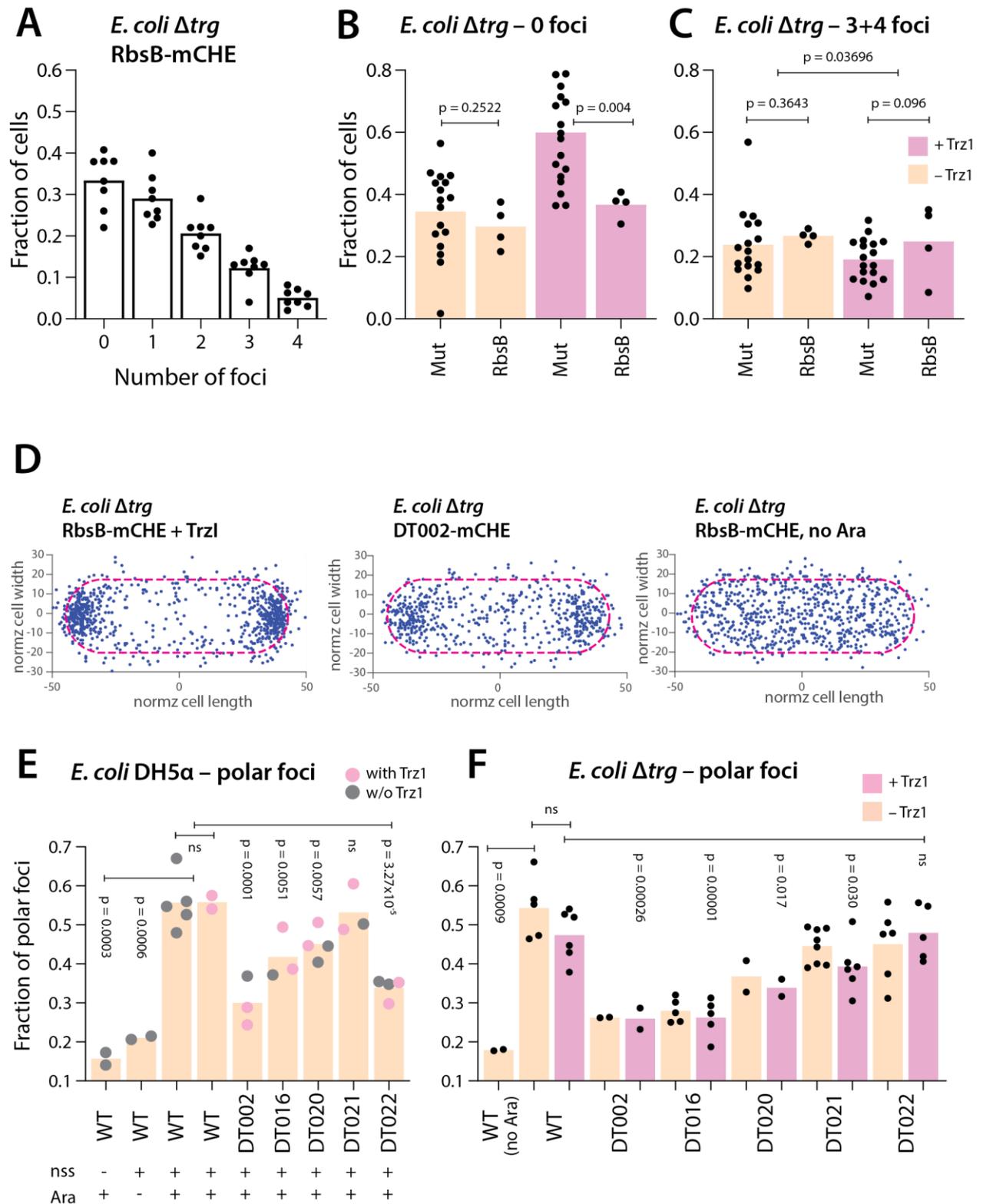


Figure 5- Localization and number of detected mCherry foci in *E. coli* cells. (A) Fraction of cells with detected foci numbers between 0 – 4, for *E. coli* Δtrg expressing wild-type RbsB-mCherry. Bars show means of combined data from 8 biological replicates (black dots), half of which co-expressed Trz1. (B) Proportion of *E. coli* Δtrg cells expressing wild-type RbsB-mCherry (RbsB) or mutant-mCherry (Mut, grouped here for DT002-, DT016-, DT020-, DT021- and DT022-mCherry), in absence (salmon) or

presence (pink) or co-expressed Trz1. Bars show means of combined data from mutants or RbsB-wild-type with dots indicating individual replicate values. P-values derived from paired t-testing. (C) Proportion of cells carrying 3 or 4 foci among those that showed any foci, in absence (salmon) or presence (pink) of co-expressed Trz1. Bars and dots and p-values as in (B). (D) Foci position plots for single replicate experiments of the indicated strains, with normalized cell lengths and widths (dotted pink line), scattering $n = 800$ foci. Length and width in pixel units. Foci threshold = 3. (E) Fraction of foci among all detected (as in panel D) attributed to the cell poles (i.e., within 6 pixels of the defined original cell length) for expressed wild-type (WT) or mutant-mCherry (DT numbers), induced or not with L-arabinose (Ara), in absence or presence of the *nss* signal sequence. P-values calculated from pair-wise t-test compared to combined WT data set (grey dots, in absence of; pink dots, in presence of co-expressed Trz1, all one-sided tests, $n = 3 - 6$). Justified because polar foci fractions for WT in presence or absence of Trz1 are not significantly different ($p = 0.98$). (F) as panel E, but for *E. coli* Δtrg cells, co-expressing or not Trz1. P-values derived from one-sided t-tests compared to WT in absence of Trz1 ($n = 3$), or compared to WT in presence of Trz1 (grouped to $n = 6$, compared to grouped mutant DT in presence of Trz1, $n = 3 - 6$).

Detection of proper signaling behavior in mutant RbsB libraries

In order to test if we could use the differences in periplasmic localisation behaviour to detect RbsB-mutants with different signaling behavior in a library, three mutant libraries generated by error-prone PCR (named epDT016, epDT021 and epDT022; the DT number referring to the parental mutant) were constructed in which the RbsB-variant was C-terminally fused to mCherry. Libraries were induced as before and examined either as such, or after spiking with *E. coli* cells expressing wild-type RbsB-mCherry in proportions of 50 %, 10%, 1% and 0.1% (cell/cell).

The mean 95th fluorescence decreased for most library mixtures at lower proportions of spiked cells expressing RbsB-mCherry, whereas also the proportion of cells with 1-2 foci decreased, as expected (Fig. 6A, B; p-values from $< 0.0001 - 0.0009$, one-way ANOVA, $n = 9 - 17$). This indicated that we would expect to be able to observe and quantify the presence of potentially favorable signaling mutants among a background of improperly folded or translocated mutant protein.

thresholds of (C). Jitter plot indicates the percentage of cells with WT behavior in a technical replicate (individual colored dots) from a single biological replicate ($n= 300-1000$ cells). Black line represents the median and whiskers indicates the 95% confidence intervals for the median. Percentages represent the mean values from 9-15 technical replicates, \pm calculated SD. P-values as in (A). (E) Percentage of cells with 0 and 4 detected foci in same biological replicates as in (D). Bars show means of technical replicates (black dots). P-values as in (A).

Based on cells expressing wild-type RbsB-mCherry we defined gating thresholds to consider mutant cells displaying wild-type-like localisation. Cells with 0 and 4 foci were excluded, and only those with 1–3 foci were retained (as in, Fig. 5A). Secondly, we retained cells with expressing levels within the 5th and 95th percentile of the median fluorescence and above the 25th percentile of the mean 95th fluorescence observed for RbsB-mCherry (Fig. S5). For cells expressing RbsB-mCherry itself, this produced a true positive rate of 65.5 ± 14.1 % (Fig. 6C and D, p-values from $< 0.0001 - 0.002$, one-way ANOVA, $n = 9 - 15$), whereas those with DT002-mCherry showed a true negative rate of 7.1 ± 7.4 % (Fig. 6D, Table S1). Error-prone libraries (epDTxxx) produced from mutants DT022, DT016 and DT021 gave 16.7 ± 9.5 %, 3.0 ± 2.5 % and 11.2 ± 7.0 % of cells passing the thresholds, respectively (Fig. 6D, Fig. S6). Not unexpectedly, this is a drastic decrease compared to the strain expressing parental proteins (Fig. 6D, Fig. S6). Mutant libraries also displayed a higher number of cells with 0 and 4 detected foci, especially epDT021 and epDT022, when compared with wild-type and parental RbsB mutants (Fig. 6E, p-values from $< 0.0001 - 0.0064$, one-way ANOVA, $n = 9 - 15$). This suggested that most library mutants lost capacity to accumulate at the cell poles and/ or interact with membrane receptor. In summary, with this analysis we can clearly detect the signaling capacity differences between a true positive and negative fusion protein (e.i, wild-type RbsB and DT002). Moreover, we could quantify the number of cells with promising signal transduction capacity in three mutant derivative libraries allowing us to understand in each library we have higher chances to isolate a mutant with improved signaling behaviour. Finally, we observed that a high number of derivative mutants have signaling issues as a direct consequence of new introduced mutations.

Discussion

Here we developed an *in vivo* system to follow RbsB or mutant-RbsB expression, their translocation into and localisation within the periplasmic space from mCherry fluorescent protein fusions. Among the different tested genetic configurations only C-terminally fused mCherry, native RbsB signal sequence and inducible expression from P_{BAD} resulted in viable cells with proper periplasmic localized highly fluorescent foci. Translocation of other translational fusion proteins probably blocked the Sec translocation channels, leading to abnormal cells. Despite this, the RbsB-mCherry fusion protein was only partly functional in its capacity to be induced by ribose, but we considered this sufficient to follow its subcellular localisation and to be the best proxy to compare wild-type to mutant behavior to in terms of expression, translocation and localization and signaling capacity [6].

Our results with wild-type RbsB-mCherry indicated clear foci formation at both cell poles in *E. coli* in an L-arabinose-induced dependent manner and depending on translocation signal sequence. From this we conclude that these are not cytoplasmic but periplasmic polar foci. Their localization at the cell poles is most likely the result of RbsB-mCherry binding to the (Trg) chemoreceptors, which preferentially concentrate at the cell pole [33-35]. However, RbsB-mCherry also produced polar foci in an *E. coli* Δtrg background, which does not produce the RbsB-specific Trg chemoreceptor and therefore was expected to show more even fluorescence around the periplasmic space. This would thus indicate that polar accumulation is not only the result of a direct interaction between RbsB and Trg, but either to other chemoreceptors, to the ribose transporter channel, or a result of the intrinsic membrane curvature, as previously reported for other fluorescent proteins [34, 36-41]. On the other hand, there must be a definitive interaction between the periplasmic exposed Trg part and RbsB-mCherry, since expressing the hybrid chemoreceptor Trz1 (a protein fusion of the periplasmic Trg part and cytoplasmic EnvZ part) in Δtrg background increased RbsB-mCherry fluorescence in the polar regions. Expression of Trz1 in *E. coli* DH5 α did not yield higher RbsB-mCherry fluorescence at the cell poles, possibly because expressed Trz1 is in competition for space with the other *E. coli* chemoreceptors in the polar zones. The increased polar fluorescence of RbsB-mCherry in Δtrg background with co-expressed Trz1 also indicates that the majority of Trz1 would be located at the cell pole (despite having EnvZ-cytoplasmic tails). Results with RbsB-mCherry in $\Delta rbsB$ background also showed slight increase of polar foci fluorescence, which we attribute to less competition for chemoreceptor binding sites from native RbsB. Interestingly, polar foci fluorescence increased even more in $\Delta rbsK$ background, which interrupts the ribose transport [42, 43]. This may have two consequences: the *rbsK* deletion interrupts ribose metabolism, yielding more ribose available in the periplasm to bind to (native RbsB) and RbsB-

mCherry, favoring the closed ligand-bound form, which interacts with the chemoreceptor ^[13]. Secondly, there is no turnover of ribose-bound docked RbsB-mCherry to the transport channels in absence of RbsK. If we assume the transporters to be evenly distributed along the cytoplasmic membrane, more ribose-bound RbsB-mCherry would therefore be available to dock to the chemoreceptors, causing the observed increased foci fluorescence. Finally, although the majority of RbsB-mCherry localizes to the cell poles, a distinct diffuse halo is visible in the periplasmic space, suggesting circulating open/closed RbsB-mCherry complexes and/or ephemeral binding to transporter channels. Wild-type RbsB-mCherry rarely produced clear 'side' foci and the majority of cells displayed 1 or 2 polar located foci.

In order to understand if and where mutant-RbsB might differ in their translocation and periplasmic localisation properties we constructed mutant-RbsB-mCherry fusions, and expressed them from identical promoters and using identical translocation signal sequences as in case of RbsB-mCherry. We concentrated on five mutants with changed ligand-binding properties that were designed by computational approaches, and that instead of ribose bind 1,3-cyclohexanediol (but at lower efficiency than wild-type binding to ribose). Two of these have been described previously, and were shown to have intrinsic folding problems, reduced cytoplasmic stability and reduced translocation efficiency ^[12]. We used these mutants, because we wondered whether their poor *in vivo* ligand detection might be partly due to this poorer expression, translocation and possibly, less efficient receptor binding. One of the mutants, (DT002), bound to mCherry was indeed quantifiably less abundant in the periplasmic space and polar regions in comparison to wild-type RbsB-mCherry. This is thus in agreement with previous periplasmic space mass spectra analysis that showed that wild-type RbsB is more abundant than DT002 ^[12]. We speculated that this may be the result of partially blocked Sec translocation channels by misfolded DT002 (or DT002-mCherry). Indeed, our microscopic images also suggested the presence of DT002-mCherry fluorescent foci in the cytoplasm, which might be caused by the protein being stocked in inclusion bodies and was not observed in cells expressing other RbsB-or mutant-RbsB-mCherry fusions. Populations of cells expressing DT002-mCherry also carried a much larger proportion of cells without any polar fluorescence, even in absence of Trg or Trz1, suggesting that there is effectively much less fluorescent protein present in the periplasmic space, which in addition may not dock as efficiently to the chemoreceptors.

The four other expressed mutant-RbsB-mCherry yielded more consistent polar periplasmic foci in cells than DT002-mCherry, although also in those cases overall the proportion of cells without any foci was increased compared to wild-type RbsB-mCherry, their polar foci were less intense and more frequently side foci appeared. Further individual differences were observed in the tested host backgrounds, which suggest specific defects to either of these mutant-RbsB-mCherry proteins in

comparison to wild-type. Of the various mutants, the DT021- and DT022-mCherry proteins behaved most similar as wild-type in their foci distribution pattern, but still tended to be poorer located at the polar regions, which might thus partly explain their observed lower *in vivo* signal transduction capacity through the Trz1 receptor.

Do the periplasmic localisation patterns of RbsB-mCherry permit some kind of testing or screening of further mutant PBP libraries for variants that behave more like wild-type and in that sense may be exemplary for improved secondary mutations beyond just the ligand binding itself? In order to test this, we produced three libraries starting from the *dt016*, *dt021* or *dt022* alleles, introducing further mutations by error-prone PCR at a rate of 1 amino acid substitution per 1000 bp and comparing foci localisation in individual cells with the expected behaviour of wild-type RbsB-mCherry, a true negative like DT002-mCherry and the starting alleles. By mixing libraries in different proportions with cells expressing RbsB-mCherry we could observe expected increase of foci fluorescence and polar localisation the more RbsB-mCherry expressing cells were present. Moreover, the percentage of cells with 1 and 2 foci diminished with decreasing proportions of spiked cells expressing RbsB-mCherry. The three libraries themselves (without spiked wild-type RbsB-mCherry cells) showed pronounced individual variability, with particularly low numbers of cells with potentially 'correctly' localized fluorescent foci in epDT016 (the derivative of the *dt016* allele). On the other hand, their numbers seemed to increase in the epDT021- and epDT022-mCherry libraries. Thresholding cells to within wild-type behaviour on the basis of foci localisation and fluorescence suggested that we could not expect to detect any potentially wild-type-like mutants in the epDT016 library (percent thresholded smaller than in the true negative DT002-mCherry). In contrast, the epDT021- and epDT022-mCherry libraries showed percentages higher than in the negative control (11.2 % and 16.7 %, respectively, compared to 7.1 % in the negative control). This would suggest that the epDT021 and epDT022 libraries have, potentially, a higher number of variants with signaling transduction capacity equaling wild-type RbsB-mCherry behaviour. On the other hand, in comparison to their starting alleles, the epDT021 and epDT022-mCherry libraries has an increased number of cells with apparent deleterious mutations to RbsB-mutant folding, translocation and periplasmic binding. This is not unexpected, given the random nature of the introduced secondary mutations on the alleles by the error-prone PCR. Thresholding for outliers may therefore be important as a criterium to judge appearance of wild-type-like translocation behaviour.

In conclusion, we demonstrate specific localisation patterns mirroring the expected periplasmic targets of RbsB (studied by its proxy of a C-terminal mCherry fusion protein) in *E. coli*, and apparent defects in foci positioning, intensity and proportions of foci numbers per cell across the clonal population, which are evidence for poorly functioning mutant RbsB proteins with new ligand-binding

properties. We propose that visualized expression and subcellular localization of fusion proteins is sensitive enough to detect potential improvements of secondary mutations introduced into RbsB-variants by typical random library screening. Although it is currently technically very challenging to isolate single cells from microscopy studies, knowing population behaviour and observing increased proportions of potentially correctly performing mutants, may help to devise other strategies to isolate such mutants and increasing the chances of finding a protein with improved signal transduction capacity.

Table 1- Strain list used in this study.

Strain	<i>E. coli</i> host	Plasmids	Relevant characteristics	Reference or source
3044	DH5 α Pir		Host for plasmid propagation	[44]
3671	DH5 α	pSTVP _{AA} _mcs	pSTVP _{AA} to clone <i>rbsB</i> and its derivatives	[6]
4076	BW25113 Δ <i>rbsB</i>		Genomic deletion of <i>rbsB</i> gene	[45]
4172	BW25113 Δ <i>rbsB</i>	pSYK1	Host strain containing the P _{tac} - <i>trzI</i> , P _{ompC} - <i>gfpmut2</i> bioreporter system	[6]
4175	BW25113 Δ <i>rbsB</i>	pSTVP _{AA} _rbsB, pSYK1	RbsB expression with <i>nss</i> for periplasmic translocation	[6]
4498	MG1655			E. coli Genetic Center, Yale (CGSC#8237)
4501	BW25113 Δ <i>trg</i>		Genomic deletion of <i>trg</i> gene	[45]
4505	BW25113 Δ <i>rbsK</i>		Genomic deletion of <i>rbsK</i> gene	[45]
6686	BW25113 Δ <i>rbsB</i>	pSTVP _{AA} _nss-rbsB-mCherry, pSYK1	RbsB-mCherry expression with <i>nss</i> for periplasmic translocation	This work
6952	DH5 α Pir	pSTVP _{BAD} _nss-rbsB-mCherry	Inducible rbsB-mCherry expression and translocation (<i>nss</i>)	This work
6953	DH5 α Pir	pSTVP _{BAD} _torAss-rbsB-mCherry	Inducible rbsB-mCherry expression and translocation (<i>torAss</i>)	This work
6955	DH5 α Pir	pSTVP _{BAD} _nss-mCherry-rbsB	Inducible mCherry-rbsB expression and translocation (<i>nss</i>)	This work
6957	DH5 α Pir	pSTVP _{BAD} _torAss-mCherry-rbsB	Inducible mCherry-rbsB expression and translocation (<i>torAss</i>)	This work
6958	DH5 α Pir	pSTVP _{BAD} _nss-DT002-mCherry	As for 6952, but wild-type RbsB was replaced by DT002 mutant	This work
6959	DH5 α Pir	pSTVP _{BAD} _nss-DT016-mCherry	As for 6952, but wild-type RbsB was replaced by DT016 mutant	This work
6960	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-rbsB-mCherry	Same as 6952 in host 4501	This work
6962	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-rbsB-mCherry, pSYK1	Same as 6972 in host 4501	This work
6963	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-DT002-mCherry	Same as 6958 in host 4501	This work
6965	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-DT002-mCherry, pSYK1	Same as 6973 in host 4501	This work
6967	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-DT016-mCherry	Same as 6959 in host 4501	This work
6969	BW25113 Δ <i>trg</i>	pSTVP _{BAD} _nss-DT016-mCherry, pSYK1	Same as 6974 in host 4501	This work
6972	DH5 α Pir	pSTVP _{BAD} _nss-rbsB-mCherry, pSYK1	Inducible rbsB-mCherry expression and translocation (<i>nss</i>), P _{tac} - <i>trzI</i> , P _{ompC} - <i>gfpmut2</i> bioreporter system	This work
6973	DH5 α Pir	pSTVP _{BAD} _nss-DT002-mCherry, pSYK1	As for 6972, but wild-type RbsB was replaced by DT002 mutant	This work
6974	DH5 α Pir	pSTVP _{BAD} _nss-DT016-mCherry, pSYK1	As for 6972, but wild-type RbsB was replaced by DT016 mutant	This work
6975	DH5 α Pir	pSTVP _{BAD} _nss-DT020-mCherry	As for 6952, but wild-type RbsB was replaced by DT020 mutant	This work

CHAPTER 4

6976	DH5 α Pir	pSTVP _{BAD_nss} -DT021-mCherry	As for 6952, but wild-type RbsB was replaced by DT021 mutant	This work
6977	DH5 α Pir	pSTVP _{BAD_nss} -DT022-mCherry	As for 6952, but wild-type RbsB was replaced by DT022 mutant	This work
6991	DH5 α Pir	pSTVP _{BAD_nss} -DT020-mCherry, pSYK1	As for 6972, but wild-type RbsB was replaced by DT20 mutant	This work
6992	DH5 α Pir	pSTVP _{BAD_nss} -DT021-mCherry, pSYK1	As for 6972, but wild-type RbsB was replaced by DT021 mutant	This work
6993	DH5 α Pir	pSTVP _{BAD_nss} -DT022-mCherry, pSYK1	As for 6972, but wild-type RbsB was replaced by DT022 mutant	This work
6996	BW25113 Δ trg	pSTVP _{BAD_nss} -DT020-mCherry	Same as 6975 in host 4501	This work
6997	BW25113 Δ trg	pSTVP _{BAD_nss} -DT021-mCherry	Same as 6976 in host 4501	This work
6998	BW25113 Δ trg	pSTVP _{BAD_nss} -DT022-mCherry	Same as 6977 in host 4501	This work
6999	BW25113 Δ trg	pSTVP _{BAD_nss} -DT020-mCherry, pSYK1	Same as 6991 in host 4501	This work
7000	BW25113 Δ trg	pSTVP _{BAD_nss} -DT021-mCherry, pSYK1	Same as 6992 in host 4501	This work
7001	BW25113 Δ trg	pSTVP _{BAD_nss} -DT022-mCherry, pSYK1	Same as 6993 in host 4501	This work
7007	BW25113 Δ rbsB	pSTV P _{BAD_nss} -RbsB-mCherry, pSYK1	Inducible nss-RbsB-mCherry expression for periplasmic translocation in host 4172	This work
7035	DH5 α Pir	pSTV P _{BAD} -RbsB-mCherry	Inducible rbsB-mCherry expression and without nss for translocation	This work
7146	BW25113 Δ rbsB	pSTV P _{BAD_nss} -mCherry-RbsB, pSYK1	As 7007, but for nss-mCherry-RbsB	This work
7147	BW25113 Δ rbsB	pSTV P _{BAD_torAss} -rbsB-mCherry, pSYK1	As 7007, but for torAss-rbsB-mCherry	This work
7148	BW25113 Δ rbsB	pSTV P _{BAD_torAss} -mCherry-RbsB, pSYK1	As 7007, but for torAss-mCherry-RbsB	This work
7189	BW25113 Δ rbsK	pSTVP _{BAD_nss} -rbsB-mCherry	Same as 6952 in host 4505	This work
7190	BW25113 Δ rbsK	pSTVP _{BAD_nss} -DT016-mCherry	Same as 6959 in host 4505	This work
7191	BW25113 Δ rbsK	pSTVP _{BAD_nss} -DT022-mCherry	Same as 6977 in host 4505	This work
7192	MG1655	pSTVP _{BAD_nss} -rbsB-mCherry	Same as 6952 in host 4498	This work
7193	MG1655	pSTVP _{BAD_nss} -DT016-mCherry	Same as 6959 in host 4498	This work
7194	MG1655	pSTVP _{BAD_nss} -DT022-mCherry	Same as 6977 in host 4498	This work
7195	BW25113 Δ rbsB	pSTVP _{BAD_nss} -rbsB-mCherry	Same as 6952 in host 4076	This work
7196	BW25113 Δ rbsB	pSTVP _{BAD_nss} -DT016-mCherry	Same as 6959 in host 4076	This work
7197	BW25113 Δ rbsB	pSTVP _{BAD_nss} -DT022-mCherry	Same as 6977 in host 4076	This work

rbsB

200111 F

ACTAGTGGTGGCTCTGGCTCTGGCTCGAGAATGGCAAAGACACCATCGCGCT

200107 R

TATGACCATGATTACGAATTTTGCATGCCTGCAGGTCGACTCAGTGGT

Materials and methods

Bacterial strains and culture conditions

Five *E. coli* strains were used as host for expression of fusion constructs (Table 1). Strains 4076, 4501 and 4505 from the Keio Collection carry genomic deletions of the *rbsB*, *trg*, and *rbsK* genes, respectively [45]. As wild-types, we used *E. coli* MG1655, obtained from the *E. coli* Genetic Stock Center (Yale university; CGSC#8237), and DH5 α Pir cells [44]. Cells were grown overnight in LB medium supplemented with appropriate antibiotics. The medium was supplemented with 30 μ g chloramphenicol ml⁻¹ and with 100 μ g ampicillin ml⁻¹ for strains containing pSTV and pSYK1 plasmids, respectively. To maintain the P_{BAD} promoter in the *off* state [46] 0.4 % (w/v) glucose was added to the LB culture medium (LB-Glc). To activate expression from P_{BAD}, overnight cultures were 100 times diluted in fresh LB medium (without glucose) supplemented with 0.5 % L-arabinose (w/v).

All strains and plasmid constructions used in this study are listed in Table 1.

Cloning of mCherry fusion protein constructs

To study localization, we translationally fused *rbsB* and *mcherry* open reading frames in different configurations. Recombinant proteins were fused with the cognate RbsB signal peptide or with the TorA signal peptide. Products amplified with primers listed in Table 2 were cloned into pSTV28P_{AA}mcs [6] digested with *NdeI* and *Sall* using a ClonExpress kit according to manufacturer instructions (Vazyme, China).

In short, the wild-type *rbsB* gene (including or not the signal peptide) was amplified without its stop codon using pSTV28P_{AA}_RbsB6His as template (extracted from strain 4175). The gene coding for *mCherry* was amplified from pARS1003 (extracted from strain 4600). Primers used to amplify *rbsB* and *mCherry* add a unique *SpeI* restriction site and include a coding sequence for an eight-amino acid linker (GGSGSGSR) between the two proteins.

In case of the constructs cloned under control of the inducible P_{BAD} promoter [46] pSTV28P_{AA}mcs was double digested with *EcoRI* (to remove the P_{AA} promoter region). The P_{BAD} promoter region and TorA signal peptide were amplified by PCR from pCRO4 (extracted from strain 4651) [47].

ClonExpress reactions (~ 180 ng DNA) were transformed into DH5 α cells by heat shock. The DNA sequences of the different fusion constructs were confirmed by sequencing on purified plasmids.

Plasmids carrying different fusion constructs were then transformed in five *E. coli* backgrounds (with or without pSYK to co-express *trz1*), as specified in Table 1.

Expression of rbsB-mCherry fusions for epifluorescence microscopy

To study RbsB-mCherry fusion protein expression and localisation, a single *E. coli* colony grown on a selective plate to maintain the respective plasmid(s) was inoculated in 5 mL of LB-Glc medium (supplemented with the appropriate antibiotics) and grown overnight at 37 °C with 180 rpm shaking. The next morning the culture was diluted 100 times in the same medium. Protein expression was induced by addition of 0.5% L-arabinose (*w/v*) to the culture, which was incubated further at 37°C and with 180 rpm shaking until reaching exponential phase (OD₆₀₀ of 0.5).

Samples of 4 µL of cell suspension (at OD=0.5) were pipetted on a 1%-agarose-in-M9 medium coated standard microscopy slide (Menzel-Gläser) and covered with a cover slip. Images were acquired with a Zeiss Axioplan II epifluorescence microscope with a 100 x Plan Apochromat oil objective (Carl Zeiss, Jena, Germany), 10 x ocular and a Sola SE light engine (Lumencor, USA). A SPOT Xplorer slow-can charge coupled device camera (1.4 Megapixels monochrome w/o IR; Diagnostic Instruments) fixed on the microscope was used to capture images. Up to ten images at different positions were acquired using Visiview software (Visitron systems GmbH), with exposures set to 10 ms and 500 ms for phase contrast (PhC) and mCherry, respectively.

Mutant library and spiked library construction

Mutant libraries (epDT016, epDT021 and epDT22) were generated by error-prone PCR. Mutations were introduced in *dt016*, *dt021* and *dt022* genes at a frequency of achieving 1 sense mutation per 1000 bp. Library genetic material (> 10⁶ potential variants) was extracted as a DNA pool. Library mutants were C-terminally fused to mCherry protein as explained above, with the difference that here mutant *rbsB* genes were directly amplified from the plasmid library pool (ClonExpress Kit, Vazyme, China). ClonExpress reactions (~ 180 ng DNA) were transformed into DH5α cells by heat shock. The DNA sequences of 20 fusion constructs were confirmed by sequencing on purified plasmids to confirm correct cloning and variability.

Cells carrying P_{BAD}-*nss-epDTxxx-mCherry* were induced for 3h with 0.5 % L-arabinose before epifluorescence microscopy analysis, as explained above. Cells expressing epDTxxx-mCherry constructs were analyzed alone or after spiking with *E. coli* cells expressing wild-type RbsB-mCherry in proportions of 50 %, 10%, 1% and 0.1% (cell/cell).

Image analysis

Individual cells were segmented on images using SuperSegger^[48] and relevant cell (e.g., length, width, orientation, mean fluorescence) and foci information (e.g., number of foci per cell, foci position, score and intensities) were recorded. Cell and foci lists were combined to extract median fluorescence and mean 95th percentile fluorescence values (MatLab R2019a), which are representative for the mCherry expression in general and foci fluorescence intensity, respectively. The number of foci detected per cell was extracted at a foci intensity threshold of > 3 . Their positions with respect to the cell long and short axis were normalized and plotted to a 'standard' cell length and width, to show the average localization patterns across a cell population. Foci were considered as polar when located less than 6 pixels from the detected cell edge on the long axis.

For visualization of fusion proteins accumulation at the cell poles, we used a custom-written MatLab (R2019a) script to prepare a z-projection of the mean top 20 percentiles per pixel across all imaged cells in a biological replicate ($n = 1000$ cells). Heatmaps were standardized to the same cell length and width.

Statistical analysis

Statistical analysis was done using GraphPad Prism software (Version 8.4.3). Fold induction of cells expressing fusion constructs in presence of ribose (Fig. 2A) were calculated using an ANOVA followed by post-hoc Tukey test. The differences in mCherry fluorescence were tested using Student's t-test on cells with and without ribose induction. Differences in mean 95th percentile fluorescence, median fluorescence of two groups were tested using Student's t-test. To compare multiple groups an ANOVA was used. Numbers of replicates were variable and are indicated in figure legends.

Acknowledgements

I want to thank Jan Roelof van der Meer who developed MatLab scripts for image analysis, and Roxane Moritz for her help with scripting modifications.

References

1. Chu, B.C.H. and H.J. Vogel, *A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport*. Biological Chemistry, 2011. **392**(1-2): p. 39.
2. Berntsson, R.P., et al., *A structural classification of substrate-binding proteins*. FEBS Lett, 2010. **584**(12): p. 2606-17.
3. Quijoch, F.A. and P.S. Ledvina, *Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes*. Mol Microbiol, 1996. **20**(1): p. 17-25.
4. Medintz, I.L. and J.R. Deschamps, *Maltose-binding protein: a versatile platform for prototyping biosensing*. Current Opinion in Biotechnology, 2006. **17**(1): p. 17-27.
5. van der Meer, J.R. and S. Belkin, *Where microbiology meets microengineering: design and applications of reporter bacteria*. Nature Reviews Microbiology, 2010. **8**: p. 511.
6. Reimer, A., et al., *Escherichia coli ribose binding protein based bioreporters revisited*. Scientific Reports, 2014. **4**: p. 5626.
7. Baumgartner, J.W., et al., *Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ*. Journal of Bacteriology, 1994. **176**(4): p. 1157-1163.
8. Banda-Vázquez, J., et al., *Redesign of LAOBP to bind novel l-amino acid ligands*. Protein Sci, 2018. **27**(5): p. 957-968.
9. Scheib, U., et al., *Change in protein-ligand specificity through binding pocket grafting*. Journal of Structural Biology, 2014. **185**(2): p. 186-192.
10. Boas, F.E. and P.B. Harbury, *Design of Protein–Ligand Binding Based on the Molecular-Mechanics Energy Model*. Journal of Molecular Biology, 2008. **380**(2): p. 415-424.
11. Ko, W., S. Kim, and H.S. Lee, *Engineering a periplasmic binding protein for amino acid sensors with improved binding properties*. Organic & Biomolecular Chemistry, 2017. **15**(41): p. 8761-8769.
12. Tavares, D., et al., *Computational redesign of the Escherichia coli ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol*. Scientific Reports, 2019. **9**(1): p. 16940.
13. Binnie, R.A., et al., *Functional mapping of the surface of Escherichia coli ribose-binding protein: mutations that affect chemotaxis and transport*. Protein science : a publication of the Protein Society, 1992. **1**(12): p. 1642-1651.
14. Riley, M., *Functions of the gene products of Escherichia coli*. Microbiological reviews, 1993. **57**(4): p. 862-952.

15. Stewart, J.B. and M.A. Hermodson, *Topology of RbsC, the membrane component of the Escherichia coli ribose transporter*. Journal of bacteriology, 2003. **185**(17): p. 5234-5239.
16. Grebe, T.W. and J. Stock, *Bacterial chemotaxis: the five sensors of a bacterium*. Curr Biol, 1998. **8**(5): p. R154-7.
17. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems*. Applied Microbiology and Biotechnology, 2003. **60**(5): p. 523-533.
18. Cantu-Bustos, J.E., et al., *Expression and purification of recombinant proteins in Escherichia coli tagged with the metal-binding protein CusF*. Protein Expression and Purification, 2016. **121**: p. 61-65.
19. Gopal, G.J. and A. Kumar, *Strategies for the Production of Recombinant Protein in Escherichia coli*. The Protein Journal, 2013. **32**(6): p. 419-425.
20. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms*. Biotechnol Adv, 2009. **27**(3): p. 297-306.
21. Spadiut, O., et al., *Microbials for the production of monoclonal antibodies and antibody fragments*. Trends in biotechnology, 2014. **32**(1): p. 54-60.
22. Robinson, M.-P., et al., *Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria*. Nature Communications, 2015. **6**(1): p. 8072.
23. Costa, S., et al., *Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system*. Frontiers in Microbiology, 2014. **5**(63).
24. Malik, A., *Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of E. coli*. 3 Biotech, 2016. **6**(1): p. 44.
25. Freudl, R., *Signal peptides for recombinant protein secretion in bacterial expression systems*. Microbial Cell Factories, 2018. **17**(1): p. 52.
26. Palmer, T. and B.C. Berks, *The twin-arginine translocation (Tat) protein export pathway*. Nat Rev Microbiol, 2012. **10**(7): p. 483-96.
27. Mori, H. and K. Cline, *A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid Δ pH/Tat translocase*. The Journal of cell biology, 2002. **157**: p. 205-10.
28. Tsirigotaki, A., et al., *Protein export through the bacterial Sec pathway*. Nat Rev Microbiol, 2017. **15**(1): p. 21-36.
29. Rusch, S.L. and D.A. Kendall, *Interactions that drive Sec-dependent bacterial protein transport*. Biochemistry, 2007. **46**(34): p. 9665-9673.
30. Santos, B.D., et al., *Optimizing Periplasmic Expression in Escherichia coli for the Production of Recombinant Proteins Tagged with the Small Metal-Binding Protein SmbP*. Molecular Biotechnology, 2019. **61**(6): p. 451-460.
31. Pradel, N., et al., *Influence of tat mutations on the ribose-binding protein translocation in Escherichia coli*. Biochem Biophys Res Commun, 2003. **306**(3): p. 786-91.

32. Linton, E., et al., *Translocation of green fluorescent protein by comparative analysis with multiple signal peptides*. Biotechnology Journal, 2012. **7**(5): p. 667-676.
33. Thomason, P.A., P.M. Wolanin, and J.B. Stock, *Signal Transduction: Receptor Clusters as Information Processing Arrays*. Current Biology, 2002. **12**(11): p. R399-R401.
34. Pollard, A.M. and V. Sourjik, *Transmembrane region of bacterial chemoreceptor is capable of promoting protein clustering*. The Journal of biological chemistry, 2018. **293**(6): p. 2149-2158.
35. Maddock and L. Shapiro, *Polar location of the chemoreceptor complex in the Escherichia coli cell*. Science, 1993. **259**(5102): p. 1717-1723.
36. Smith, J., et al., *Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli*. Molecular microbiology, 2001. **39**: p. 47-53.
37. Santini, C.L., et al., *Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock*. J Biol Chem, 2001. **276**(11): p. 8159-64.
38. Dammeyer, T. and P. Tinnefeld, *Engineered fluorescent proteins illuminate the bacterial periplasm*. Computational and structural biotechnology journal, 2012. **3**: p. e201210013-e201210013.
39. Govindarajan, S., et al., *The General Phosphotransferase System Proteins Localize to Sites of Strong Negative Curvature in Bacterial Cells*. mBio, 2013. **4**(5): p. e00443-13.
40. Romantsov, T., et al., *Protein Localization in Escherichia coli Cells: Comparison of the Cytoplasmic Membrane Proteins ProP, LacY, ProW, AqpZ, MscS, and MscL*. Journal of Bacteriology, 2010. **192**(4): p. 912.
41. Strahl, H., et al., *Transmembrane protein sorting driven by membrane curvature*. Nature Communications, 2015. **6**(1): p. 8728.
42. Hope, J.N., et al., *Ribokinase from Escherichia coli K12. Nucleotide sequence and overexpression of the rbsK gene and purification of ribokinase*. The Journal of biological chemistry, 1986. **261**(17): p. 7663-7668.
43. Iida, A., et al., *Molecular cloning and characterization of genes required for ribose transport and utilization in Escherichia coli K-12*. Journal of bacteriology, 1984. **158**(2): p. 674-682.
44. Platt, R., et al., *Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the Escherichia coli chromosome*. Plasmid, 2000. **43**(1): p. 12-23.
45. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Molecular Systems Biology, 2006. **2**(1): p. 2006.0008.
46. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter*. Journal of bacteriology, 1995. **177**(14): p. 4121-4130.
47. Roggo, C., N. Carraro, and J.R. van der Meer, *Probing chemotaxis activity in Escherichia coli using fluorescent protein fusions*. Scientific Reports, 2019. **9**(1): p. 3845.

48. Stylianidou, S., et al., *SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells*. Mol Microbiol, 2016. **102**(4): p. 690-700.

Supplementary Information

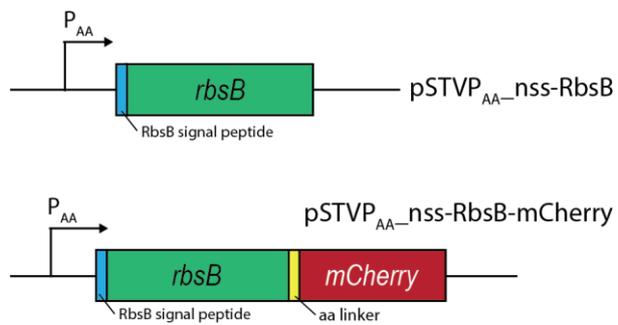


Figure S1- Scheme of the genes for Nss-RbsB and Nss-RbsB-mCherry fusion protein under control of the constitutive P_{AA} promoter. Scheme is not drawn to scale. Arrows indicate promoters and the transcription direction.

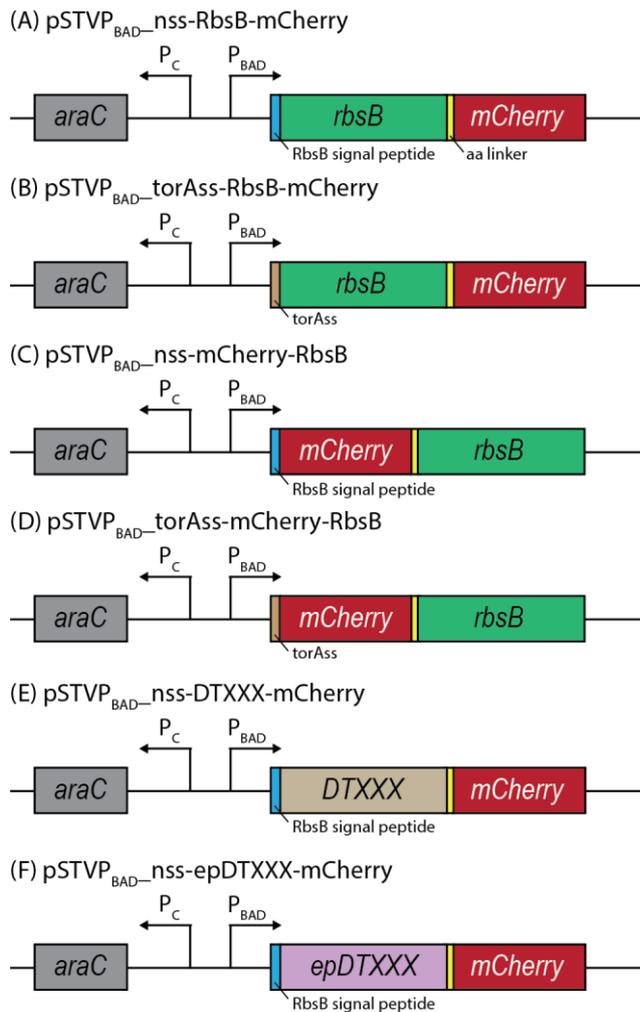


Figure S2- Representative scheme of fusion protein genes under control of inducible P_{BAD} promoter used in this study (not drawn to scale). A) Schematic representation of Nss-RbsB-mCherry construct. B) as (A) but for TorAss-RbsB-mCherry. C) as (A) but for Nss-mCherry-RbsB. D) as (A) but for TorAss-mCherry-RbsB. E) as (A) but for Nss-RbsBmutants-mCherry. F) as (A) but for Nss-epRbsBmutants-mCherry (e.i., error prone PCR library based on mutant-RbsB was C-terminal fused to mCherry). Arrows indicate promoters and the transcription direction.

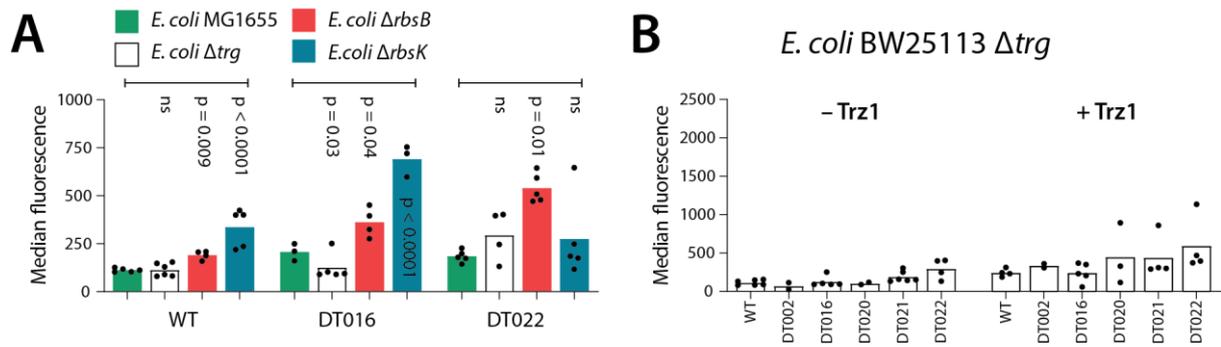


Figure S3- Cellular mCherry fluorescence comparison of four *E. coli* strains expressing wild-type- and mutant-RbsB-mCherry fusion proteins. (A) Bars show means of median cellular fluorescence from biological replicates (black dots, each dot calculated from around 1000 cells). With arabinose, 0.5 % L-arabinose during 3 h. P-values derived from one-way ANOVA, compared to the respective fusion protein expression in *E. coli* MG1655 background. (B) Trz1 co-expression leads to higher median mCherry fusion protein fluorescence. Bars show means of the median cellular fluorescence from biological replicates (black dots). P-value = 0.0003, two-sided t-test, $n = 25$.

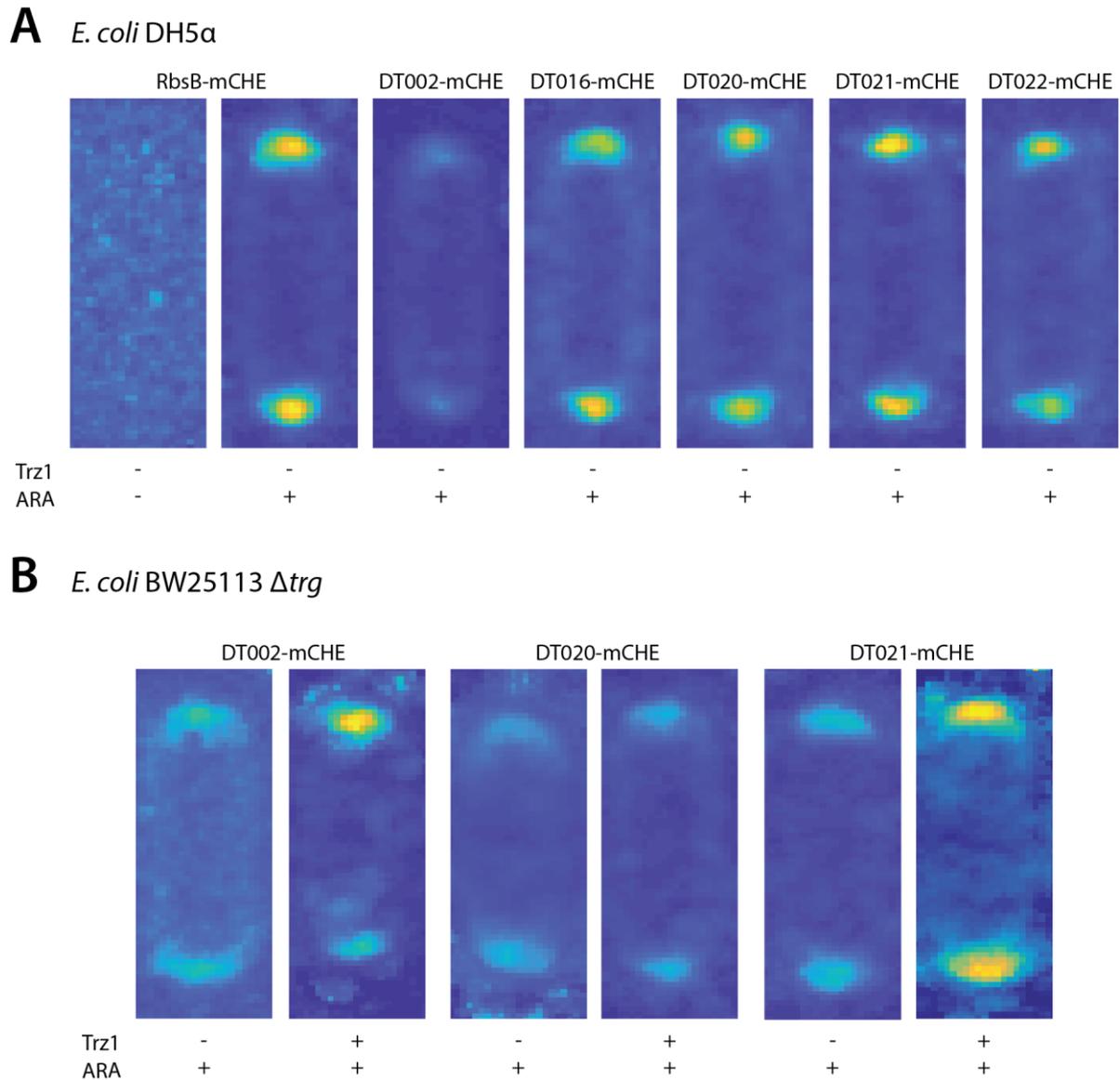


Figure S4- *E. coli* cells expressing wild-type- and mutant-RbsB-mCherry fusions were used to prepare a Z-projection of the mean top 20 percentiles per pixel across all cells in a biological replicate ($n = 10^3$) standardized to the same cell length and width. (A) Heatmap representation of fluorescence intensity in the *E. coli* DH5 α cells after 3h incubation with 0.5% L-arabinose (ARA) and in presence or absence of TrzI chemoreceptor. Blue and orange colors represent low and high intensity, respectively. Scale is not comparable between heat maps. (B) As in (A) but in *E. coli* BW25113 Δ *trg* background.

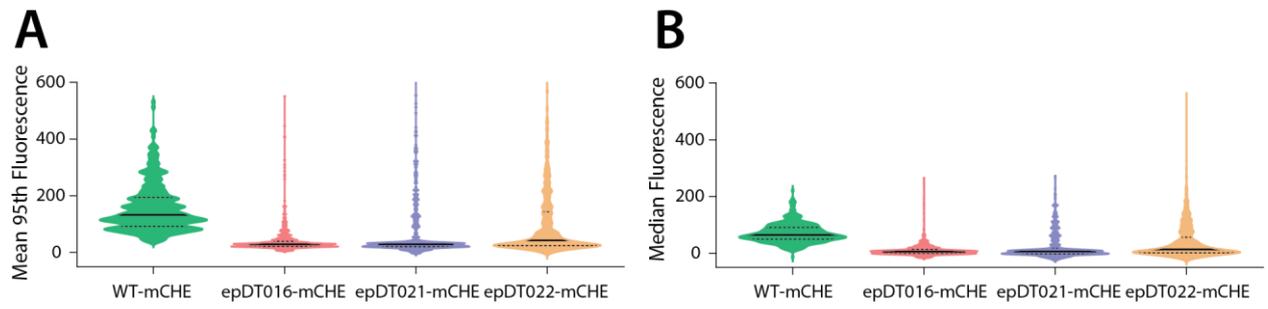


Figure S5- Distribution of mean 95th percentile (A) and median background (B) fluorescence of single cells ($n = 270 - 1000$ cells). Dashed lines in the violin plot show Q1 and Q3 quartiles. Solid line indicates the median fluorescence of the distribution.

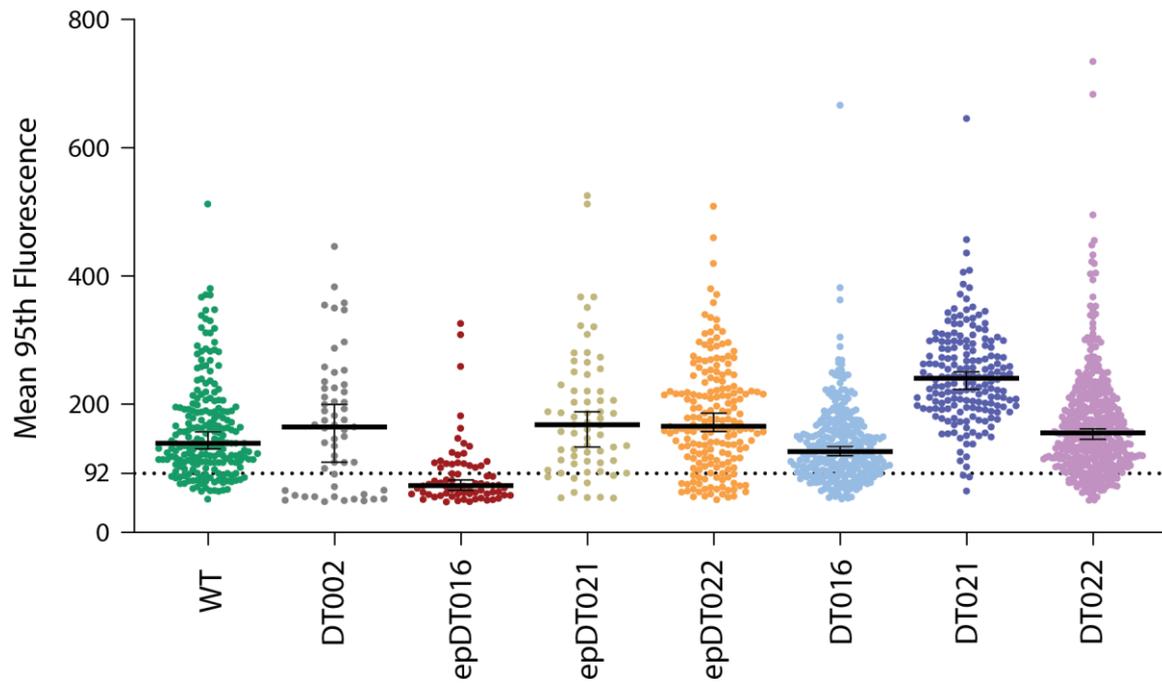


Figure S6- Detection of mutant-RbsB-mCherry fusions with wild-type behavior in *E. coli* cells expressing error-prone libraries (epDTxxx) and RbsB-based mutants. Proportion of cells displaying 1-3 foci and between 5th-95th percentile median fluorescence of wild-type RbsB-mCherry cells. Plots are jittered single cell mean 95th fluorescence values (individual colored dots) in a single replicate ($n = 300-1000$) with black line representing the median and whiskers indicating the 95% confidence intervals for the median. Dashed line indicates cells above the 25th percentile.

Table S1- Indication of the percentage of cells that were removed with application of the different thresholds.

Threshold	% of removed cells							
	WT (n = 267)	epDT016 (n = 798)	epDT021 (n = 475)	epDT022 (n = 988)	DT002 (n = 335)	DT016 (n = 538)	DT021 (n = 595)	DT002 (n = 755)
1 st	14.8	25.5	55.3	65.9	39.7	24.8	31.2	29.2
2 nd	7.5	65.9	29.3	13.8	17.3	7.2	39.2	11.4
3 rd	13.5	6.0	1.9	3.1	4.5	12.5	0.5	7.7

n = total number of cells

1) Remove all cells with 0 and 4 detected foci

2) Remove all cells all cells out of the 5th and 95th percentile of the median fluorescence observed in WT-mChe cells

3) Remove all cells below 25th percentile of the mean 95th fluorescence observed in WT-mChe cells

CHAPTER 5

General Discussion

The aim of this thesis was to develop a high-throughput strategy, based on a bioreporter system, and to isolate variants of the RbsB periplasmic binding protein with new recognition specificities. I performed *in vitro* assays to characterize isolated variants of interest, and executed further rounds of random and semi-random mutagenesis to improve the affinity and overall function of the most promising variants.

Previously in our group the ribose-binding protein of *Escherichia coli* (RbsB) was used as a template to design binding pockets that might accommodate 1,3-cyclohexanediol (13CHD) and cyclohexanol (CH). These two molecules have a similar structure to ribose, but wild-type RbsB shows no binding capacity for either. The Rosetta software was used to simulate and calculate new binding pockets, from which a list of critical amino acid residues to mutate in RbsB was selected. In total 9 positions were selected, with five possible amino acid substitutions each. The library containing 2 million mutants was produced by DNA synthesis and cloned into an *E. coli* bioreporter strain ^[1], which carries the Trz1 hybrid signaling pathway coupled to GFP expression upon (new) ligand binding by the (mutant) RbsB.

My first goal was to develop a strategy to isolate variants with 13CHD and/or CH binding capacity. Clones were grown as individual microcolonies in alginate beads rather than screened as individual cells, which clearly reduced GFP expression variability. Others have compared screening of single cells vs microcolonies and came to the same conclusions ^[2]. Approximately 0.5 % of library variants were capable of interacting with the hybrid receptor Trz1 even in absence of the ligand, which might be the result of some combinations of introduced mutations spontaneously stabilizing the closed conformation of the RbsB mutants, and thus triggering the bioreporter system (i.e., GFP expression). In order to remove such constitutive-ON mutants an additional sorting step under uninduced conditions was added to the screening, and only beads with low level of GFP were collected. These variants were cultured, encapsulated, grown to microcolonies and screened in presence of 13CHD. In the second sorting step, high GFP producers were collected. The decision of screening the library in microcolonies and add a sorting step to deplete library from constitutive-ON mutants was essential to reduce the number of false positives and increase the efficiency of the screening strategy.

Six mutants (DT001, DT002, DT011, DT013, DT015 and DT016) with up to 1.5-fold induction with 13CHD were finally isolated from this procedure. These mutants lost completely the capacity to bind ribose. Isothermal microcalorimetry confirmed *in vitro* 13CHD binding by mutants DT002 and DT016, with estimated K_D of 190 μ M and 5 μ M, respectively. This is indicative of poor binding, especially compared to binding affinity of wild-type RbsB towards ribose, with a K_D of 500 nM.

However, it was the first time of isolating genuine periplasmic binding protein (PBP) variants with affinity towards a non-natural compound. The other four purified mutants did not show any 13CHD binding in isothermal microcalorimetry, likely because they suffered from poor stability and/or unfolded even during purification. These results were in agreement with other studies showing that computationally designed PBP variants yield misfolded and/or instable proteins ^[3, 4]. This notion was further supported by thermal denaturation data showing that DT002 and DT016 were indeed much less stable than wild-type RbsB. Moreover, inclusion of 13CHD did not stabilize the mutant proteins like is the case of wild-type RbsB in presence of ribose. Further evidence for instability and misfolding of DT mutant proteins was the observation of co-purified chaperones. Circular dichroism and mass spectra analysis showed clear secondary structure differences and lower periplasmic abundance of DT mutants in comparison to wild-type RbsB. We concluded that the introduced mutations in the RbsB binding pocket indeed changed the binding specificity from ribose to 13CHD, but severely impaired their secondary structure, stability and translocation process. If we assume that the displayed GFP induction is the result of the complete chain of true ligand-binding affinity, periplasmic abundance and proper folding, the actual gain of 13CHD binding might have been higher than the induction of 1.5 suggested.

Despite the frustration resulting from independent unsuccessful reproduction ^[1, 3] of the initial studies from Hellinga's group suggesting easy and expandable ligand binding pocket design in PBPs ^[5], we could thus demonstrate that *de novo* design of PBPs for a non-natural molecule is feasible. Design of PBP variants with *de novo* binding specificities thus remains very challenging, and, likely, computational simulations are not accurate enough ^[6, 7] to properly predict the intrinsic dynamics and conformational changes caused by the interaction with the ligand ^[8, 9], and primarily by introduced mutations. The resulting designed PBP variants are thus impaired in stability and overall function. As a consequence of this, most recent studies on PBPs have exploited their natural ligand binding properties and used mutagenesis techniques to reduce or increase binding specificity ^[10, 11] or to graft binding-pockets between closely related PBPs ^[4, 12], without new attempts for *de novo* ligand binding design. However, we clearly (and for the first time) demonstrated that a combination of rational design and efficient screening allowed to isolate RbsB variants with recognition of a non-natural compound (i.e., 13CHD) from an initial library with 2 million variants.

In the second part of my thesis the goal was to improve the binding capacity and/or solve stability issues of the six isolated mutants with 13CHD-binding capacity. These mutants were used as a scaffold to produce several new mutant libraries based on random mutagenesis, site saturation or DNA shuffling approaches. Libraries were cloned in an *E. coli* strain carrying the bioreporter system and screened for improved induction of GFPmut2 reporter fluorescence in presence of 13CHD ^[13]. In contrast to our expectations, none of the semi-random approaches, DNA shuffling and random mutation of the 32 residues flanking the 9 substitutions engineered for 13CHD-binding, yielded improved variants. Probably because introducing multiple mutations (in multiple parts of the protein) increased the chance for further functional degeneration, despite a probability to improved binding capacity. This thus creates an important trade-off. I focused specifically on two residues (R141 and D215) that were later shown to be important for ribose induction and signaling ^[14] and for ligand binding (D215) ^[15]. However, also site saturation mutagenesis of both residues in DT002 and DT016 mutants could not improve the GFP mut2 induction of 1.5 times in presence of 13CHD. In fact, all substitutions reduced or abolished 13CHD induction, except for R141S in DT016. This showed that both R141 (or S141) and D215 are currently essential for 13CHD-induction in the DT002 and DT016 proteins.

In contrast, the random mutagenesis approach across the complete gene variants led to isolation of seven mutants with significantly improved 13CHD inducibility (up to 3.2 times). I tried to use two different selection and sorting strategies that were less and more restrictive to separate potential mutant microcolonies in FACS. My results suggested that being more restrictive to the upper outlier fluorescence response is a better strategy, because it involves fewer rounds of mutagenesis and less variants are recovered, reducing significantly the time to screen individual mutants. Although we acknowledge that by selection and sorting of high fluorescence outlier variants we are missing those, which have low fluorescence background under uninduced conditions and intermediate fluorescence upon induction, this is again a trade-off between screening time and chance to find better mutants. It is, however, difficult to predict *a priori* the evolutionary path of a variant to its final aspired success (e.g., high induction, low nM binding affinity) ^[16]. Recent work on protein evolution showed that a strong and restrictive screening is extremely important to improve a desired protein property, and not allow the possibility to accumulate neutral and deleterious mutations ^[17].

Interestingly, all observed mutations in variants recovered from directed evolution, except one, were located outside the direct ligand-binding pocket, suggesting they were compensatory and helping protein folding or functional behaviour other than interaction with 13CHD ligand. Two mutants, DT020 and DT035, displayed a single amino acid substitution in the signal peptide. Our

hypothesis is that these mutations had a positive effect on peptide recognition by the SecB chaperone, resulting in a higher periplasmic presence and/or improved stability of these variants. All other mutations were found outside the binding cavity, and we assume that they must have improved other aspects of protein functionality than ligand binding itself. This could affect, for example, protein stability or improved hinge flexibility, or binding to the chemoreceptor Trz1. Only variant DT022 displayed a mutation (G89V) in the binding pocket. Structure threading indicated this residue to be less than 2 Å away from inferred position of 13CHD and we must assume that the exposed valine is beneficial for 13CHD-binding. Previous studies demonstrated the importance of residue 89 for ligand binding [13-15]. Of note, that these interpretations were based on inferred structures of variants (threaded on the crystal structure of wild-type RbsB) and the real effect of introduced mutations may have been underestimated. Phyre2 and Swiss-Model also did not predict any structural differences caused by introduced mutations in DT002 and DT016. However, *in vitro* characterization showed that these variants have significant secondary structure differences when compared with wild-type RbsB [13].

Some mutants seemed to be blocked in open and closed state simultaneously, possibly impairing their dynamic transition between open and closed conformation. Similar results have been observed in transcription factors with allosteric behaviour [18]. The effect of this was that populations carrying these variants in the bioreporter systems caused 'stable' double populations with different GFPmut2 fluorescence intensities both in absence or in presence of inducer. This suggests that the time-scale of the dynamics is affected by the introduced mutations, blocking the DT variants in either open or closed form long enough to trigger (or not) the bioreporter signaling cascade leading to GFP expression. A small percentage of the low fluorescence population shifts to high fluorescence upon induction, indicating that ligand-binding is still affecting the transition states, but is insufficiently discriminating between the two. This suggested that only a small percentage of the population is still active and able to bind the ligand, changing the conformation and triggering the bioreporter cascade.

The second part of my thesis thus demonstrated that it is possible to improve binding capacity and solve stability issues of variants with *de novo* recognition specificities using a combination of random mutagenesis and restrictive screening. Improved variants displayed a dose-dependent induction in presence of 13CHD, with a maximal induction of 3 times and 0.25 mM of lower detection limit. Despite the clear improvement, the affinity of these variants is not yet comparable to wild-type RbsB towards ribose (i.e., 13-fold induction and 50 nM detection limit) [1, 13]. Of note that our mutagenesis was extremely laborious and time consuming, and did not allow to isolate a variant with similar stability and affinity towards 13CHD as wild-type to ribose. Since our screening relied on higher

GFPmut2 expression as the final step of the signaling cascade, other variants with improved translocation, binding capacity and stabilization of the closed conformation, but unable to interact and bind Trz1 would be ignored.

Finally in the third part of my thesis, I screened the differences in subcellular localisation of RbsB and DT variant proteins, in order to understand their different behaviour. I was also interested to see if subcellular variations could be developed into an *in vivo* screening system for variants with potentially improved folding, translocation and receptor interactions. Wild-type- and mutant-DT proteins were translationally fused to mCherry, as a fluorescent reporter for their abundance and subcellular localization. Despite a small loss of function by carrying the mCherry tag, my results showed that wild-type RbsB-mCherry protein clearly localized to the periplasmic space and centered in polar regions. This may be indicative for its propensity to react to the available chemoreceptors. In contrast to wild-type RbsB, DT-mCherry expression resulted in high proportions of cells devoid of clear foci and low proportions with multiple fluorescent foci, confirming the previously obtained *in vitro* poorer folding, and the observed poorer translocation. In addition, polar foci in cell expressing some DT-mutant-mCherry were less fluorescent, suggesting poorer chemoreceptor binding. This was particularly clear for DT002 and is thus in agreement with previous periplasmic space mass spectra analysis that showed that wild-type RbsB is more abundant than DT002 ^[13]. The four other expressed mutant-RbsB-mCherry yielded more consistent polar periplasmic foci in cells than DT002-mCherry, although also in these cases the overall proportion of cells without foci was increased compared to wild-type RbsB-mCherry, their polar foci were less intense and more frequently side foci appeared. Of the various mutants, the DT021- and DT022-mCherry proteins behaved most similar as wild-type in their foci distribution pattern, but still tended to be poorer located at the polar regions, which might thus partly explain their observed lower *in vivo* signal transduction capacity through the Trz1 receptor. However, this showed that our random mutagenesis approach improved not only the binding capacity, but also the signaling behaviour of DT021 and DT022 (in comparison with their parent DT016).

Microscopy analysis of individual cells expressing random mutagenesis libraries based on DT016, DT021 and DT022 fused to mCherry indeed showed a large proportion of cells without any visible periplasmic mCherry signal nor polar localized foci. This confirmed that a majority of introduced mutations lead to RbsB mistranslocation and/or poorer receptor interaction. Based on cells expressing wild-type RbsB-mCherry I tried to estimate the percentage of cells in the libraries with potentially improved variants having proper expression, translocation and signaling transduction.

The screening approach can thus reveal if PBP variants are impaired in some other part of the signaling cascade than just ligand binding, which previously was impossible to detect with our

bioreporter (ligand binding) system. The *in vivo* system is sensitive enough to detect potential improvements of secondary mutations introduced into RbsB-variants by typical random library screening and may be helpful to find protein variants with improved signaling capacity.

In conclusion, my work showed the importance of combining computational ligand-binding pocket design and further directed evolution approaches. I have no doubt that a successful design of PBP variants for detection of non-natural compounds is dependent on both methods. Completely random and even semi-random approaches, however, are very time consuming and still a matter of “trial and error”. Computational algorithms can drastically reduce the “sampling pool” size, however, they have to improve and take in consideration other aspects of PBP behaviour (not only binding). For example, proper translocation, folding, dynamic transition between open and closed state and receptor interaction are aspects as important as binding capacity in order to design a function receptor protein.

Another challenge encountered during this work was the fact that the structures of isolated mutants with 13CHD binding capacity were not available. It would have been important to better interpret the effect of selected amino acid substitutions on protein functionality, but doing so based on inferred structures is risky and imprecise. However, solving the crystal structure of an unstable protein is not possible, since it is unlikely to crystalize properly in its active form. Moreover, until very recently it was impossible to correctly predict protein structure based on the amino acid sequence. This may soon change, however, with further improvements on *de novo* structure predictions, such as AlphaFold ^[19] (an artificial intelligence-based software), which has shown considerable success in determining protein structures from amino acid sequences alone. AlphaFold achieved a median score of 92.4 GDT (global distance test) in the 14th Critical Assessment of protein Structure Prediction (CASP). Hopefully soon this will become available to the scientific community and the combination of modeling programs and further improvement of machine learning ^[20, 21], will be a tremendous advance in *de novo* design of proteins.

References

1. Reimer, A., et al., *Escherichia coli ribose binding protein based bioreporters revisited*. Scientific Reports, 2014. **4**: p. 5626.
2. Duarte, J.M., I. Barbier, and Y. Schaerli, *Bacterial Microcolonies in Gel Beads for High-Throughput Screening of Libraries in Synthetic Biology*. ACS Synthetic Biology, 2017.
3. Schreier, B., et al., *Computational design of ligand binding is not a solved problem*. Proceedings of the National Academy of Sciences, 2009. **106**(44): p. 18491-18496.
4. Banda-Vázquez, J., et al., *Redesign of LAOBP to bind novel L-amino acid ligands*. Protein Sci, 2018. **27**(5): p. 957-968.
5. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions*. Nature, 2003. **423**(6936): p. 185-190.
6. Yang, W. and L. Lai, *Computational design of ligand-binding proteins*. Current Opinion in Structural Biology, 2017. **45**(Supplement C): p. 67-73.
7. Feldmeier, K. and B. Höcker, *Computational protein design of ligand binding and catalysis*. Current Opinion in Chemical Biology, 2013. **17**(6): p. 929-933.
8. Boas, F.E. and P.B. Harbury, *Design of Protein–Ligand Binding Based on the Molecular-Mechanics Energy Model*. Journal of Molecular Biology, 2008. **380**(2): p. 415-424.
9. Stank, A., et al., *Protein Binding Pocket Dynamics*. Accounts of Chemical Research, 2016. **49**(5): p. 809-815.
10. Amiss, T.J., et al., *Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor*. Protein Science : A Publication of the Protein Society, 2007. **16**(11): p. 2350-2359.
11. Ko, W., S. Kim, and H.S. Lee, *Engineering a periplasmic binding protein for amino acid sensors with improved binding properties*. Organic & Biomolecular Chemistry, 2017. **15**(41): p. 8761-8769.
12. Scheib, U., et al., *Change in protein-ligand specificity through binding pocket grafting*. Journal of Structural Biology, 2014. **185**(2): p. 186-192.
13. Tavares, D., et al., *Computational redesign of the Escherichia coli ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol*. Scientific Reports, 2019. **9**(1): p. 16940.
14. Reimer, A., et al., *Complete alanine scanning of the Escherichia coli RbsB ribose binding protein reveals residues important for chemoreceptor signaling and periplasmic abundance*. Scientific Reports, 2017. **7**: p. 8245.
15. Vercillo, N.C., et al., *Analysis of ligand binding to a ribose biosensor using site-directed mutagenesis and fluorescence spectroscopy*. Protein Science : A Publication of the Protein Society, 2007. **16**(3): p. 362-368.

16. Zheng, J., J.L. Payne, and A. Wagner, *Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks*. *Science*, 2019. **365**(6451): p. 347.
17. Zheng, J., N. Guo, and A. Wagner, *Selection enhances protein evolvability by increasing mutational robustness and foldability*. *Science*, 2020. **370**(6521): p. eabb5962.
18. Taylor, N.D., et al., *Engineering an allosteric transcription factor to respond to new ligands*. *Nature Methods*, 2016. **13**(2): p. 177-183.
19. Senior, A.W., et al., *Improved protein structure prediction using potentials from deep learning*. *Nature*, 2020. **577**(7792): p. 706-710.
20. Wu, Z., et al., *Machine learning-assisted directed protein evolution with combinatorial libraries*. *Proc Natl Acad Sci U S A*, 2019. **116**(18): p. 8852-8858.
21. Yang, K.K., Z. Wu, and F.H. Arnold, *Machine-learning-guided directed evolution for protein engineering*. *Nat Methods*, 2019. **16**(8): p. 687-694.

CURRICULUM VITAE

Name	Diogo
Surname	Tavares
Date of Birth	May 13 th , 1991
Citizenship	Portugal
Residence Permit	Swiss B Permit

University of Lausanne

01.2016 to current – Doctoral Research (PhD candidate)

Environmental and Evolutionary Microbiology lab | Supervisor: Prof. J R van der Meer

Research project:

Developing a strategy to design new receptor proteins for non-natural compounds.

Implement a new screening strategy of mutant libraries combining bioreporter systems and FACS.

06.2015 to 12.2015 – Internship student

Environmental and Evolutionary Microbiology lab | Supervisor: Prof. J R van der Meer

Tasks/ Responsibilities:

- Explored a bacterial protein in a bio-sensor development team.

Faculdade de Ciência de Lisboa, Universidade de Lisboa / Instituto de Tecnologia Química e Biológica (ITQB), Universidade Nova de Lisboa

09.2012 to 12.2014 – Master in Applied Microbiology

Microbial & Enzyme Technology lab | Supervisor: Prof. Lígia O. Martins

Research project:

Engineering new bacterial dye-decolourising peroxidases for lignin degradation

Faculdade de Ciência de Lisboa, Universidade de Lisboa

09.2012 to 12.2014 – Bachelor in Molecular Biology and Genetics

Awards

Best oral presentation award of environmental microbiology session (Swiss Society for Microbiology, 2018)

List of Publications

Vânia Brissos, **Diogo Tavares**, Ana Catarina Sousa, Maria Paula Robalo, and Lúgia O. Martins (2017). "Engineering a Bacterial DyP-Type Peroxidase for Enhanced Oxidation of Lignin-Related Phenolics at Alkaline pH." *ACS Catalysis* 7 (5), 3454-3465.

Artur Reimer, Vitali Maffenbeier, Manupriyam Dubey, Vladimir Sentchilo, **Diogo Tavares**, Manuel Hernandez Gil, Siham Beggah, and Jan Roelof van der Meer (2017). "Complete alanine scanning of the *Escherichia coli* RbsB ribose binding protein reveals residues important for chemoreceptor signaling and periplasmic abundance." *Sci Rep* 7, 8245.

Diogo Tavares, Artur Reimer, Shantanu Roy, Aurélie Joubin, Vladimir Sentchilo, and Jan Roelof van der Meer (2019). "Computational redesign of the *Escherichia coli* ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol." *Sci Rep* 9, 16940

Diogo Tavares, Vitali Maffenbeier, and Jan Roelof van der Meer (2019). "Engineering of sensory proteins with new ligand-binding capacities." *Handbook of Cell Biosensors*.

List of conferences

2016

- 74th Annual Meeting of the Swiss Society of Microbiology (SSM), Bern, Switzerland (Poster presentation)
- BRAAVOO Marine Biosensors Workshop, Villars-sur-Ollon, Switzerland (Oral presentation)

2017

- Batsheva de Rothschild Seminar on New Concepts in Biosensing, Dead Sea, Israel (Oral and poster presentation)
- 75th Annual Meeting of the Swiss Society of Microbiology (SSM), Basel, Switzerland (Poster presentation)

2018

- Annual Meeting of the Swiss Society of Microbiology (SSM), Lausanne, Switzerland (Oral presentation)
- The International Society for Microbial Ecology (ISME17), Leipzig, Germany (Poster presentation)

2019

- International Symposium on Biocatalysis and Biotransformation (BioTrans2019), Groningen, Germany (Poster presentation)

Acknowledgements

My first words go to my thesis director, Jan Roelof van der Meer, who gave me the opportunity to come in Switzerland, work in his lab and obtain a PhD degree. During the past 5 years I learned so much, I am sure I would not finish without your support, guidance and advices. Thank you very much!

To all EEM group (current and former members), without your friendship and multiple scientific discussions this project would be much more difficult to carry on. Special thanks to Vladimir, Clémence, Manu, Marian, Andrea, Nico, Daveri, Roxane, Tania, Gaitan, Senka and Maxime. You are all amazing!!!

I acknowledge all the DMF administrative staff who helped me with multiple situations during my PhD. Thanks Armindo, Nazife, Maria, Lydie, Aline, Nadine and Nassim.

I also want to thank all DMF members, including PIs, Postdocs, PhD students and technicians. DMF is an incredible place to work because is full of extraordinary people. Thank you for all the help, good energy, dinners, hikes and other activities.

I thank Julien Dénéreaz, Ludivine Jotterand, Thibaut Laurent, Valentina Benigno, Maxime Brunner and Cédric Schneider for helping in several parts of the project. Without the excellent work of these internship students this project would not be possible.

I also acknowledge the members of my thesis committee: Bruno Correia, Sven Panke and Edward Elliston Farmer.

Of course I also want to thank my family. My parents, brother, uncles, cousins, grandmother, who were always there to support and encourage me during my all life. Thank you very very very much!

Finally, but not less important, I want to thank my future wife. Joana was always by my side during these difficult years. A PhD project is extremely complicated and a proper support at home is determinate. Thank you for all love and support! Without you this would not be feasible.