Modeling and analysis of parts and devices of genetic regulation

THÈSE N^{O} 8081 (2018)

PRÉSENTÉE LE 16 FÉVRIER 2018
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE BIOTECHNOLOGIE COMPUTATIONNELLE DES SYSTÈMES
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Yves BERSET

acceptée sur proposition du jury:

Prof. B. Deplancke, président du jury
Prof. V. Hatzimanikatis, Prof. J. R. van der Meer, directeurs de thèse
Prof. Y. Schaerli, rapporteuse
Prof. C. Mazza, rapporteur
Prof. B. Fierz, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Acknowledgements

I would like to thank my supervisors, Vassily and Jan. Your expertise, passion and enthousiasm have been the greatest source of inspiration for this work. Thank you!

I would also like to thank all my former and present labmates at the DMF and LCSB. It was amazing to be surrounded everyday by such a diversity of smart and friendly people. Thanks to Davide, Diogo and Vladimir for their patience when performing experiments with me, Tiziano, Julien R., Xavier and Georgios for the discussions and help in modeling. Special thanks to Joana for helping me structure the writing of the thesis, and for your proofreading, comments and motivational speeches.

My past and current office mates of the hottest room in EPFL (in all senses of course!): Keng, Meriç, Tiziano, Julien L., Daniel, Anush, Milenko, Homa, Aarti, Thomas and Liliana. Special thanks to the unflinching rebels to the salad club, Milenko and Tuure for going outside for lunch!

To all my musical heroes I have listened to when coding, plotting, reading and writing for this work.

Merci aux copains du Stamm du mercredi, Evelyne, Pierre, Lucas, Thierry, Guillaume, aux co-pendulaires Julien, Laurence, Alessio et Cécika, et aux copains de l'orchestre Noémie, Julien, Adrien et Adrien. Une pensée pour Julien et Gaël qui sont partis et qui me manquent.

À ma famille pour leur soutien, mes grands-parents, mes frères Nicolas et Xavier. Enfin à mes parents Josiane et Jean-Pierre pour tout ce que vous avez fait pour moi.

Abstract

DNA encrypts the composition of the cellular material that is synthesized through transcription and translation. Nevertheless, gene regulation mechanisms determine the final *amount* of transcribed and translated material. Transcription factors (TF) are a class of proteins that bind to DNA motifs and can either facilitate or prevent the RNA polymerase to transcribe a strand of DNA into an mRNA. These mechanisms allow the cell to sense the environment and adapt to different environmental conditions, e.g. presence of toxic compounds, oxidative stress or absence of nutrients.

TF interactions with DNA are depicted by networks of molecular interactions. Some TFs bind to very specific DNA sites, and others have a broad range of binding sites. Moreover, TFs often interact each other, e.g. through heterodimerization prior to bind to DNA, leading to changes in their binding specificities. However, these effects remain poorly understood. By combining detailed mathematical modeling and high-throughput experimental techniques for quantification of molecular interactions, we built a heterodimer-DNA specificity model with higher predictive power than a one-site model.

Bioreporters are living cells that emit a signal in the presence of a chemical compound. In arsenic bioreporters, a TF triggers the detoxification response in presence of arsenic. To better understand the key mechanisms involved in the response, we built a detailed mechanistic model of the gene regulatory circuits of different bioreporters.

In this study, we used mathematical modeling (ODE, SSA) to create, calibrate and analyze detailed networks of molecular interactions involved in gene regulations. We quantified the cooperative binding of transcription factors forming heterodimers on a DNA library, and optimized bioreporters for the detection of arsenic by modeling feedback, uncoupled and toggle switched-based gene regulatory circuits.

Keywords

DNA binding protein, gene regulation, transcriptional regulation, molecular interactions, cooperativity, heterodimers, bacterial bioreporters, ordinary differential equations, stochastic simulation, bistability.

Résumé

L'ADN crypte la composition du matériel cellulaire, synthétisé par transcription et traduction. Néanmoins, les mécanismes de régulation génique déterminent la quantité finale de matériel transcrit et traduit. Les facteurs de transcription (TF) sont une classe de protéines qui se lient à des motifs d'ADN et peuvent soit faciliter, soit empêcher l'ARN polymérase de transcrire un brin d'ADN en un ARNm. Ces mécanismes permettent à la cellule de détecter l'environnement et de s'adapter à différentes conditions environnementales, par ex. présence de composés toxiques, stress oxydatif ou absence de nutriments.

Les interactions TF avec l'ADN sont représentées par des réseaux d'interactions moléculaires. Certaines TF se lient à des sites d'ADN très spécifiques et d'autres ont un large éventail de sites de liaison. De plus, les TF interagissent souvent les uns avec les autres, par ex. par hétérodimérisation avant de se lier l'ADN, conduisant à des changements dans leurs spécificités de liaison. Cependant, ces effets restent mal compris. En combinant une modélisation mathématique détaillée et des techniques expérimentales à haut débit pour la quantification des interactions moléculaires, nous avons construit un modèle de spécificité hétérodimère-ADN avec un pouvoir prédictif supérieur à celui d'un modèle à un site.

Les biorapporteurs sont des cellules vivantes qui émettent un signal en présence d'un composé chimique. Chez les biorapporteurs d'arsenic, une TF déclenche la réaction de détoxification en présence d'arsenic. Pour mieux comprendre les mécanismes clés impliqués dans la réponse, nous avons construit un modèle mécaniste détaillé des circuits de régulation des gènes de différents biorapporteurs.

Dans cette étude, nous avons utilisé la modélisation mathématique (ODE, SSA) pour créer, calibrer et analyser des réseaux détaillés d'interactions moléculaires impliquées dans la régulation des gènes. Nous avons quantifié la liaison coopérative des facteurs de transcription formant des hétérodimères sur une banque d'ADN et optimisé les bioreporters pour la détection de l'arsenic en modélisant les circuits de régulation des gènes à rétroaction, découplés et à bascule.

Mots-clés

Proténe de liaison à l'ADN, régulation de l'expression des gènes, régulation de la transcription, interactions moléculaires, coopérativité, hétérodimères, biorapporteur bactérien, equation différentielle ordinaire, simulation stochastique, bistabilité.

Contents

Acknowledgements .... **i**

Abstract .... **ii**

Keywords .... **ii**

Résumé .... **iii**

Mots-clés .... **iii**

List of Figures .... **vii**

List of Tables **11**

Abbreviations **12**

Chapter 1 Introduction **13**

- 1.1 Gene regulation 14
- 1.2 Transcription factor function 16
- 1.3 Bioreporters design principles and applications 16
- 1.4 Computational approaches to dissect complex reaction networks 17
- 1.5 Aims of research 18
- 1.6 Thesis structure 18

Chapter 2 Cooperativity in dimer-DNA binding **20**

- 2.1 Introduction 20
  - 2.1.1 Hill Equation and Cooperativity 20
  - 2.1.2 Quantification of Cooperativity in Heterodimer-DNA Binding 23
- 2.2 Methods 25
  - 2.2.1 Experimental Procedures 25
  - 2.2.2 Binding Model 25
  - 2.2.3 Quantification of cooperativity 27
- 2.3 Results 29
- 2.4 Discussion 40
- 2.5 Conclusion 42

Chapter 3 Model of gene regulatory circuits for arsenic bioreporters **53**

- 3.1 Introduction 54

3.2 Results 56
3.2.1 Mechanistic Model for ArsR-P_{ars} Regulation 56
3.2.2 Effect of ArsR Binding Affinity on Circuit Output 60
3.2.3 Including an Arsenite Efflux System Decreases Sensitivity of the Circuit 62
3.2.4 Possible Cross-Binding Effects of Double ArsR Alleles 63
3.2.5 Variability analysis of the Δ *arsRBC* bioreporter strains 67
3.3 Methods 67
3.3.1 Strains, Culturing procedures, Cloning and Molecular Techniques 67
3.3.2 Deletion of the Chromosomal *ars* Genes in *E. coli* 68
3.3.3 Construction of Arsenic Reporter Plasmids 68
3.3.4 Arsenic Bioreporter Assays 69
3.3.5 ArsR Circuit Description and Assumptions 69
3.3.6 Mathematical Model 70
3.3.7 Model Assumptions and Constraints 72
3.3.8 Parameter Estimation 73
3.3.9 Sensitivity Analysis 74
3.3.10 Stochastic formulation Gillespie Algorithm 74
3.4 Conclusion 76
3.5 Supplementary Material 78
Chapter 4 Genetic toggle switch to sharpen arsenic bioreporters 90
4.1 Methods 91
4.1.1 Gardner toggle switch model 91
4.1.2 ArsR-TetR toggle switch model 91
4.2 Results and discussion 92
4.2.1 Analysis of bistability regions 92
Chapter 5 Conclusion and perspectives 95
References 96
Curriculum Vitae 102

List of Figures

Figure 1.1 **Comparison of bacterial and eukaryotic genome organization. A:** bacterial gene transcription and translation steps. **B:** After transcription, RNA of eukaryotic cells is processed and transported out of the nucleus before translation. 14

Figure 1.2 **Gene expression of a eukaryotic cell.** Regulation of gene expression can occur at every step from translation to post-translational modification. 15

Figure 1.3 **Example of bioreporter design**. Top: in the absence of the chemical compound of interest, the reporter gene is shut down by a transcription factor, and the RNA polymerase cannot start the transcription. Bottom: The presence of the chemical compound affects the transcription factor and the reporter gene is expressed. 17

Figure 2.1 **One ligand, two sites cooperativity model. A**: Occupancy states of the receptor with two identical ligand-binding sites (0: free binding site, 1: occupied binding site). In absence of cooperativity, the states of the neighboring sites do not influence each other, hence ω = 1. With cooperativity, an occupied site reduces (ω<1) or increases (ω>1) the affinity for a ligand to bind its neighboring site. **B**: Estimation of cooperativity effects through Hill and Adair-Klotz fits. 22

Figure 2.2 **On-chip heterodimer-DNA assembly. A**: Schematic representation of the experimental set-up. (1) PPARγ fused to an eGFP tag is immobilized on the surface of a MITOMI chip with an anti-GFP antibody. (2) RXRα tagged with mCherry and Cy5-labeled DNA baits are introduced into the system and (3) incubated for one hour to allow system equilibration and complex assembly. (4) Newly formed complexes are trapped under a flexible PDMS membrane and unbound molecules as well as molecular complexes are washed away. **B**: Fluorescence-based read-out of PPARγ-GFP, RXRα-mCherry, and Cy5-labeled target DNA from ten MITOMI units. The two upper panels represent PPARγ-GFP and RXRα-mCherry detected in the center of each unit (under the PDMS membrane). The two lower panels represent the variable amounts of Cy5-labeled target DNA molecules detected in the same ten MITOMI units, before (DNA free) and after (DNA bound) trapping. **C**: Corresponding quantitative readout of B where the quantified amounts of both PPARγ and RXRα remain constant but the amount of bound DNA increases with the input DNA concentration until it reaches saturation. The corresponding quantities of proteins and DNA are expressed in relative fluorescent units (RFU). 24

Figure 2.3. **States of the DNA target bound to PPARγ and RxRα proteins**. 0: free, 1: PPARγ, 2: RxRα. States are linked if they are separated by one binding event from each other. Pink nodes represent states involving PPARγ, which will mechanically trap the DNA to be measured as bound. States with 3 numbers mean that the outermost protein (labelled 1 or 2) is only attached to the other protein, not to the DNA site... 27

Figure 2.4 **DNA binding preferences of PPARγ, RXRα, as well as the PPARγ:RXRα heterodimer**. A, linear fits of binding data. Examples of binding curves (arbitrary units) and corresponding linear fits of PPARγ, RXRα and PPARγ:RXRα heterodimer interactions with DNA sequences containing putative nuclear receptor binding sites. B, Relative DNA binding affinities of PPARγ, RXRα, and the PPARγ:RXRα heterodimer to five putative nuclear receptor-binding sites and variants thereof. Each sequence family is defined by the orientation of the canonical hexameric sites (represented by arrows) and the spacing between them.32

Figure 2.5 **The DNA binding behavior of PPARγ** and **RXRα on PPRE, PAL3, or variants thereof. A**: Heterodimer formation in the presence of PPRE and PAL3 DNA at different concentrations. **B**: The DNA binding landscape of RXRα monomer to single nucleotide variants of PPRE. The heatmap represents the mean of ddG values (the difference in Gibbs energy of RXRα binding to a mutant site compared to the energy of RXRα binding to canonical PPRE) derived from two independent MITOMI experiments. The sequence of the canonical PPRE is indicated along the x axis. Two core hexamer repeats, constituting the DR1, are highlighted in red. Below heatmap: nergy normalized sequence logo (enoLOGOS) [54] derived from the matrix of the binding energy contribution for each base at each position of PPRE. **C**: Binding affinities of

PPARγ or RXRα to DR1 and PAL3 sites or truncated variants thereof. **D**: Same as for B, but for PPARγ instead of RXRα. **E**: Binding affinities of PPARγ to variants of DR1 and PAL3 sites. **F**: Visualization of on-chip assembly of putative PPARγ and RXRα dimers. **G:** DNA binding landscape of PPARγ monomer to PAL3 or single nucleotide variants thereof. Each bar represents the mean and standard deviation of ddG derived from two independent MITOMI experiments. Below heatmap : energy normalized sequence logo [59] derived from the matrix of the binding energy contribution for each base at each position in the PAL3 element. . 33

Figure 2.6 **Cooperative TF-DNA interactions. A**: Examples of binding curves representing PPARg:RXRα binding to PPRE or variants thereof. The nucleotide that was substituted in each sampled sequence is highlighted in red. **B**: Binding of the PPARγ:RXRα heterodimer to the DR1 element in function of different DNA and PPARγ concentrations. One example of respectively a strongly (left) and weakly (right) bound sequence is shown. Raw experimental data are represented by black dots and the surface plot represents the regression of the data using Voronoi interpolation. The amount of bound DNA is expressed in arbitrary units (a.u.). : Schematic representation of various scenarios of heterodimer formation. We allow the heterodimer to be formed through either the monomer or dimer scenarios.................................... 36

Figure 2.7 **Significance of cooperative effects in PPARγ:RXRα-DNA binding. A**: The cooperativity map represents log $\omega_{1,2}$ values calculated for each PPRE variant. **B**: DNA affinity change (σ) upon PPARγ heterodimerization with RXRα. 192 sequences were clustered using MAFFT and plotted as a phylotree. The representative sequence of each subtree is denoted outside of the tree circle. The values of occupancy change observed for each sequence are plotted as color plots at the terminal nodes of the phylotree. **C**: Same as B, but for RXRα heterodimerization with PPARγ. 38

Figure 2.8 **Prediction of in vivo binding. A**: An affinity map as well as the corresponding sequence logo (enoLOGOS) [54] of PPARγ:RXRα heterodimer binding to PPRE. The affinity map represents the K_{DoD} values as calculated based on our cooperativity model. **B**: Venn diagram of the number of PPARγ:RXRα binding sites predicted by three different specificity models independently. The PPARγ:RXRα motif occurrence predicted within 200bp genomic regions identified through ChIP-seq at day 6 of 3T3-L1 adipocyte differentiation. 39

Figure 2.9 **Model fits of RXRα-PPRE experimental binding curves**. The mechanistic model is streamlined for monomeric TF-DNA binding in this case and thus restrained to consider RXRα, PPRE, RXRα-PPRE interactions. Residuals, calculated for each sequence, are represented on each plot above the respective sequence fit and are randomly scattered around zero, indicating an accurate model fit. 46

Figure 2.10 Same as Figure 2.9 but for PPARγ. 48

Figure 2.11 **Performance of the mechanistic model solved for equilibrium**. Examples of experimental data corresponding to each tested PPRE variant and corresponding binding curves as predicted by the model when ether accounting for cooperativity (red curves) or not (green curves)..................................... 50

Figure 2.12 (continued) **Performance of the mechanistic model solved for equilibrium**. Examples of experimental data corresponding to each tested PPRE variant and corresponding binding curves as predicted by the model when ether accounting for cooperativity (red curves) or not (green curves). The sum of squared residuals of both cooperativity and “no cooperativity” model fits are indicated for each PPRE mutant (in red and green respectively) and plotted for each sequence as a bar plot at the bottom of the figure for a direct comparison. 51

Figure 3.1 **Constructs and configurations of the arsenite bioreporter circuits**. **A**: Feedback configuration, with *arsR* and *egfp* transcription under control of ArsR from the *ars*-promoter. Derepression in presence of arsenite (As_{III}). Note the formation of two mRNAs, depending on the occupancy of the ArsR operator (ABS) in between *arsR* and *egfp*. **B**: Uncoupled configuration, with *arsR* transcription controlled by the constitutive P_{AA} promoter and *egfp* under control of the *ars*-promoter, ArsR and arsenite. Note the different *arsR* alleles in pPR-ars-ABS/pAAUN versus pPRK12/pAAK12. N, NheI; H, HindIII; P, PseI; E, EcoRI; Sc, ScaI; X, XhoI; Sa, SalI; B, BamHI. 55

Figure 3.2 **Conceptualization of the ArsR arsenic reporter circuitry**. Cells carry the *egfp* reporter circuitry on a plasmid (as example here the feedback circuit) in presence or absence of an additional chromosomal *arsRBC* operon. Straight lines and open arrows depict DNA and location of open reading frames, respectively, with hooked arrows showing the *ars*-promoters. Sinoid lines represent relevant transcribed mRNAs. Colored circles or ovals represent relevant proteins with their names. Two-facing line arrows indicate binding equilibria between relevant partners and their corresponding affinity constant (*K*, see table 1). Transcription, translation, maturation and degradation rate constants not indicated (see Table 3.2). As, arsenite; ABS_x, chromosomal ArsR operator; ABS_p, plasmid ArsR operator (from the R773 system). Note that, depending on the plasmid construct, ArsR can be $ArsR^{K12}$ or $ArsR^{R773}$, and that the modeling further allows formation of heterodimers (i.e., $ArsR^{K12}$-$ArsR^{R773}$). Further note that the host strain can have a deletion of the full *arsRBC* operon or only of the *arsBC* genes. 57

Figure 3.3 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As_{III} concentrations after 3h induction for the four plasmid bioreporter configurations in an *E. coli* host without chromosomal *arsRBC* operon**. Lines (*sim*) show results of the best 100 fits from the parameter estimation, simulated and plotted in all different configurations. Line darkness increases with the overall fitting score, the overall best fit (*best*) being represented with the darkest, and the average of all parameter sets (*mean*) with a dashed line. Data points show the mean of independent biological triplicates. Error bars are smaller than the used symbol size and are therefore not indicated. 60

Figure 3.4 **Sensitivity analysis of key parameters in the model.** Horizontal bars show the relative deviation in EGFP fluorescence output from the respective circuit (i.e., pAAUN etc.) in *E. coli* MG1655 Δ*arsRBC* at a 5% difference of the mean parameter value (*mean*, as in Figure 3.3). Scenarios calculated for circuit output in absence of arsenite (0 µg As l^{-l}), at 50 µg As l^{-1} and at 150 µg As l^{-1}). Note the relatively large sensitivities to overall transcription, translation and degradation rates of mRNA and protein (which the model implicitely assumes constant), and the large sensitivities to promoter strength, the specific binding affinities to the ABS (K_A) and the binding affinity of ArsR to As (K_C). 61

Figure 3.5 **Sensitivity analysis of the effect of further model parameters on EGFP output from the four reporter circuits in *E. coli* MG1655** (**A-D**, as indicated). g_p, plasmid copy number; $K_{A2,\ K12\text{-}K12}$, DNA binding constant for having both chromosomal and plasmid *arsR*^{K12} allele; $K_{A2,\ K12\text{-}R773}$, DNA binding constant for having chromosomal *arsR*^{K12} but plasmid *arsR*^{773} allele; $K_{M,\ ArsB}$, Michaelis-Menten constant of the ArsB efflux pump; $k_{cat,ArsB}$, catalysis rate of the ArsB efflux pump. Horizontal bars show the relative deviation in EGFP fluorescence output from the respective circuit (i.e., pAAUN etc.) at a 5% difference of the mean parameter value (*mean*, as in Figure 3.3). Scenarios calculated for circuit output in absence of arsenite (0 µg As l^{-l}), at 50 µg As l^{-1} and at 150 µg As l^{-1}. 62

Figure 3.6 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As_{III} concentrations after 3 h induction for the four plasmid bioreporter configurations in *E. coli* MG1655 with the arsenite ArsBC efflux pump.** See further legend to Figure 3.3. 63

Figure 3.7 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As_{III} concentration after 3 h induction for the three series of four plasmid bioreporter configurations in *E. coli* MG1655 with Δ*arsRBC*, wild-type and Δ*arsBC* background**. Data points (round, triangle and square markers) and error bars, when larger than the markers, show the mean and standard deviation of independent biological triplicate measurements, respectively. Bars show the median output of the best 100 fits from the parameter estimation, with lower and upper quartiles (black) and minimum and maximum output (gray). When larger than symbol sizes, error bars indicate the standard deviation. 64

Figure 3.8 **A**: **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-AsIII concentrations.** The results are obtained after 3 hours of induction for four plasmid bioreporter configurations in an *E. coli* host without chromosomal arsRBC operon (left column: uncoupled circuits, right column: feedback circuits; upper row: carrying *arsR*^{R773} allele, lower row: carrying *arsR*^{K12} allele). Crosses and error bars (Data) show average fluorescence and standard deviations from single cell measurements. Solid lines (ODE) were simulated from ordinary differential equations and squares with error bars (SSA) result from stochastic simulations of the model (10^4 realizations). Dots denote the minimum and maximum values across the stochastic simulations. **B**: Coefficient of variation of fluorescence output at different arsenite-AsIII concentrations for the four bioreporter configurations from experimental measurements (Data) and stochastic simulations (Model). **C**: Coefficient of variation of short mRNA strands, *i.e.* encoding only ArsR, measured in the stochastic simulations at different arsenite-AsIII concentrations for the four bioreporter configurations. **D**: Same as **C** but for long mRNA strands encoding for EGFP. In the feedback circuits pPR-ars-ABS and pPRK12, long mRNA strands encode for both EGFP and ArsR proteins. 66

Figure 3.9 **Distribution of starting population (gray) and top100 best individual parameter values after the scatter search parameter estimation algorithm (green).** Mean values and standard deviations are indicated by the red dot and lines, respectively. 83

Figure 3.10 **Sequence alignments of the ArsR^{K12} and the ArsR^{R773} proteins (upper part) and of the corresponding nucleotide sequences, including the promoter regions (lower part).** 85

Figure 3.11 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As$_{III}$ concentrations after 3 h induction for the four plasmid bioreporter configurations in *E. coli* MG1655 Δ*arsBC* (including the chromosomal *arsR* gene)**. Lines (*sim*) show results of the best 100 fits from the parameter estimation, simulated and plotted in all different configurations. Line darkness increases with the overall fitting score, the overall best fit (*best*) being represented with the darkest, and the average of all parameter sets (*mean*) with a dashed line. Data points show the mean of independent biological triplicates. Error bars are smaller than the used symbol size and are therefore not indicated. 86

Figure 3.12 **Graphical overview of the *ars* operons and the constructed deletions**. **A**: Schematic detail of the regulatory elements in the *ars* promoter from plasmid R773 with the ArsR binding site (ArsR b.s.), the -35 and -10 promoter and the transcription (arrow facing to the right) and translation starts (ATG). Full sequence shown in Figure 3.10. **B**: Organization of the R773 *ars* operon. Dotted region corresponds to panel **A**. **C**: Organization of the chromosomal *ars* operon of *E. coli* K12 and the location of the produced deletions (dotted lines, number indicating the length). Regions amplified for the I-SceI based recombination selection system are shown as solid lines flanked by (introduced) restriction sites. The 190 bp insert was a sequence difference found in comparison to the published K12 sequence (Accession number U00096.3). 87

Figure 3.13 **Sequence of the resynthesized *arsR*^{K12} gene with the *ars* promoter from R773, the ArsR binding site (abs, in red) and the downstream linkage to the *egfp* gene as in plasmid pPRK12**. 89

Figure 4.1 **Gardner model of toggle switch**. Bottom left: Model in coupled equations, with species A and B being mutual repressors. Top: depending on the parameters a and b, nullclines intersect to give bistable states (two stable steady-states and one unstable steady-state, left) or monostable states (one stable steady-state, right). Bottom right: region of bistability on the (a,b) plane. The “wing” shape delimits the region of bistability, while the outside of the shape corresponds to monostable regions. 91

Figure 4.2 **Model of genetic toggle switch for detection of arsenic.** Left: ArsR and TetR are mutual repressors, weakened in presence of Arsenic (As) and anhydrotetracycline (aTc), respectively. Right: Formulation of the model incorporating the effectors. 92

Figure 4.3. **Bistability regions in the ArsR$_{max}$,TetR$_{max}$ and K$_{TetR\text{-}atc}$ parameter space as functions of As and aTc.** The reference parameters provide a barely visible bistability region (top right panel, red line). The bistable regions are shown as function of the unrepressed levels of ArsR and TetR, and on each panel with the influence of a reduced TetR-aTc interaction (K$_C$ in Table 4.1) 93

Figure 4.4. **Switch behavior of the ArsR and TetR concentrations by induction of arsenic. A:** Relative levels of ArsR and TetR at steady state as function of increasing and decreasing arsenic concentrations. **B**: Same as **A** with 25 µg/L aTc. **C**: Distributions of ArsR and TetR expressions based on 1000 stochastic simulations. **D**: same as **C** with 25 µg/L aTc. 94

List of Tables

Table 2.1 **Model input parameters, K_{DoD} and cooperativity parameters of PPARγ:RxRα heterodimer binding to PPRE as estimated by the cooperativity model** 43

Table 2.2 **Position specific scoring matrices of PPARγ:RxRα towards PPRE.** 45

Table 3.1 **Model species, synthesis, and degradation rates**. 58

Table 3.2 **Binding affinities and reactions.** 59

Table 3.3 **Figures of merit for the different arsr reporter circuits and genetic backgrounds as a function of arsenite exposure.** 65

Table 3.4 **Stochastic model reactions** 76

Table 3.5 **Bacterial strains used in the study** 78

Table 3.6 **Primers used in this study**. 79

Table 3.7 **Constraints on fluxes, plasmid and chromosomal types for construction of bioreporter models.** 80

Table 3.8 **Reaction fluxes used in the models.** 81

Table 3.9 **Stoichiometric matrix used for the construction of the models**. 82

Table 4.1 **Binding affinities used** 92

Abbreviations

ABS: ArsR binding site

ChIP: Chromatin Immunoprecipitation

DNA: Deoxyribonucleic acid

GTF: General transcription factor

MITOMI: Mechanically induced trapping of molecular interactions

mRNA: Messenger RNA

ODE: Ordinary differential equation

PABP: Poly(A) binding protein

PSSM : Position specific scoring matrix

PTM: Post-translational modification

RBPs : RNA-binding proteins

RNA: Ribonucleic acid

SSA: Stochastic simulation algorithm

TF: Transcription factor

TOR: Target of rapamycin

Chapter 1 Introduction

Cells are the basic building blocks of life. They convert nutrients into energy and other molecules that are needed to sustain the growth, survival and adaptation of an organism to its environment. These biological processes consist of a complex network of molecular interactions whose instructions are encoded in the deoxyribonucleic acid (DNA), which is then passed from generation to generation. The central dogma of molecular biology explains how the genetic information is processed in cells – it is an irreversible process where a DNA strand encoding a gene is transcribed to a messenger ribonucleic acid (mRNA) strand (transcription) that is then translated to the corresponding protein sequence (translation) [1]. These proteins subsequently fold into their native shapes, which are required for them to perform tasks involved in cellular maintenance.

Cells are classified into two fundamental types, prokaryotes and eukaryotes. Prokaryotic cells are encompassed by a cytoplasmic double phospholipid membrane either with a thick layer of peptidoglycane, or with a thin layer and a second outer (double) membrane. The prokaryotic cell mostly has no further compartmentalization of 'organelles'. Their genetic material is usually composed of one copy of a single circular chromosomal DNA, carefully folded and wrapped inside the cell. Smaller and multiple copies of additional circular DNA, called plasmids, can complement the genetic material of prokaryotes. These frequently confer more specific functions, such as antibiotic resistance. Another particularity of prokaryotic cells is that their genes are usually grouped based on them being controlled by the same operator, which are called operons. In contrast to prokaryotic cells, eukaryotic cells organize their cellular contents within organelles. In particular, the DNA is stored in the nucleus. As a consequence, the transcription occurs in the nucleus, and the mRNA is transported out of the nucleus before the translation takes place (Figure 1.1). Eukaryotes have linear DNA chromosomes, usually longer than prokaryotes. To fit into the nucleus, the DNA is folded around histone proteins to allow for tighter packing.

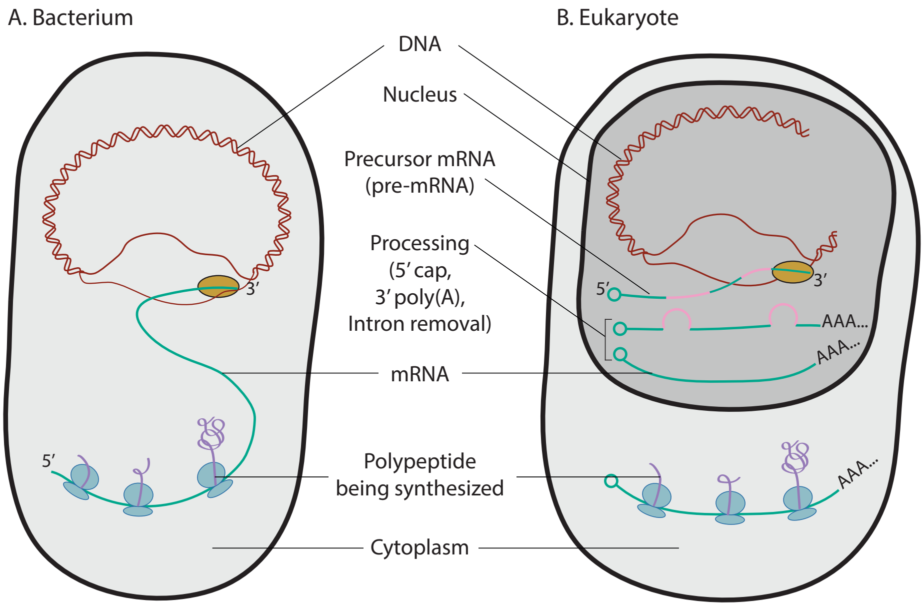


Figure 1.1 **Prokaryotic and Eukaryotic gene transcription and translation and their subcellular location.** A. Both gene transcription and translation steps occur in the cytosol for Prokaryotes. B. In Eukaryotes, transcription occurs in the nucleus while translation occurs in the cytosol. After transcription, the mRNA is duly transported out of the nucleus and processed before translation.

1.1 Gene regulation

Several mechanisms influence the *amount* of genetic material that is transcribed and translated. The ensemble of these mechanisms is called regulation of gene expression, or in short, gene regulation. Each step taking place from transcription to post-translational modifications can be subject to different types of regulation. The gene expression at different stages of an organism life cycle leads to different phenotypes. In particular, for eukaryotic cells, gene expression is controlled by very intricate circuitry (i.e. molecular networks) that can perturb the system at different steps of transcription and translation. Deciphering the complexity of these networks is a fundamental challenge in the field.

At the transcriptional level, the binding of a DNA binding protein, such as a transcription factor (TF), to the DNA plays a very important role in the regulation of gene expression (Figure 1.2). For example, the *lac* operon, required to metabolize lactose in *Escherichia coli* (*E. coli*) and many other enteric bacteria, is regulated at the transcription level [2]. This regulation allows these organisms to switch their carbon source utilization to lactose when their primary source, glucose, is depleted. In the absence of lactose, a TF (LacI) binds to DNA and turns off the transcription of the operon. However, the presence of lactose impairs the binding of LacI and the transcription can occur. As a consequence, the cell only expresses genes responsible for transport and metabolism of lactose when required. In eukaryotes, transcription can only occur if an assembly of TFs, called general transcription factors (GTF), has bound to the DNA. Transcription is then regulated by competitive binding of other TFs that either facilitate or prevent the transcription initiation [3]. Other processes influence transcription, such as interaction of chromatin remodelers [4] or noncoding [5] RNA with regulatory DNA, as well as defects in the organization of chromatin. Gene regulation also occurs at the post-transcriptional level and involves transcript modification and editing mediated by RNA-binding proteins (RBPs). The transcript can be processed and modified through alternative splicing, a process that removes noncoding regions (introns), and capping, the addition of a protective 7-methylguanylate cap (m^7G) in the 5’ end. Additionally, polyadenylation – addition of a poly(A) tail at the 3’ end of the transcript can influence the efficiency of nuclear export and translation. The poly(A) tail also impacts the stability of the mRNA strand and hence regulates its degradation.

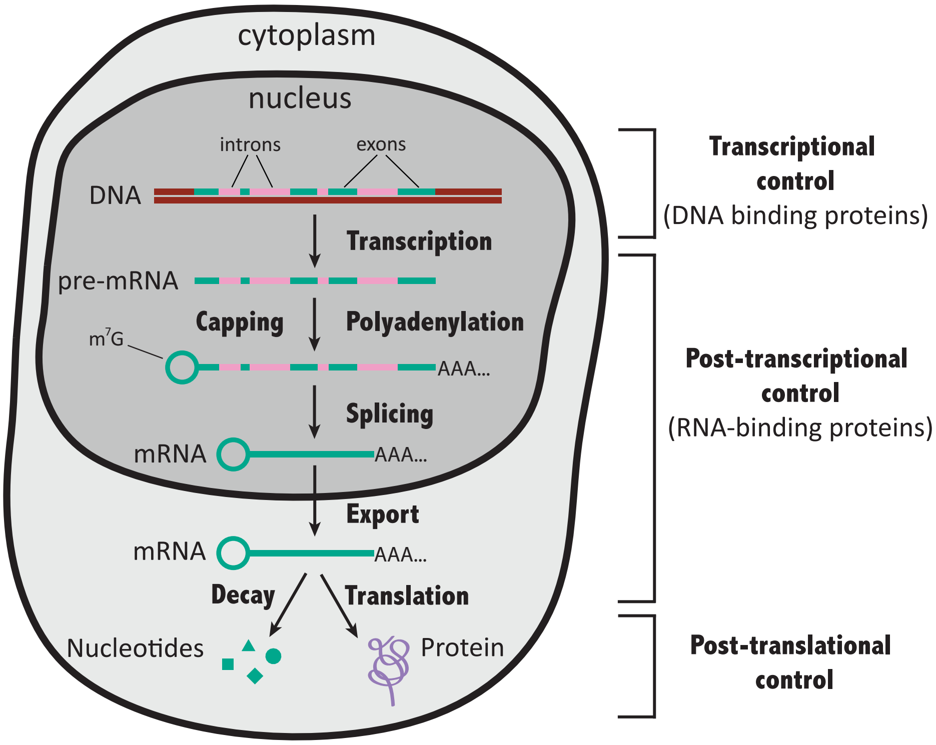


Figure 1.2 **Regulation steps during gene expression in a eukaryotic cell.** By binding to the DNA, TF regulate the transcription. After the transcription, precursor mRNA (pre-mRNA) is processed through capping (m^7G addition), polyadenylation and splicing. The resulting mRNA is transported out of the nucleus to be translated or recycled into nucleotides. Proteins can be further subjected to post-translational modifications.

At the translation level, the regulation occurs mainly at the initiation step through the phosphorylation and binding of translation initiation factors or the formation of mRNA secondary structure that control the ribosome recruitment [6]. Signaling pathways can also influence the translation activity. For instance, the target of rapamycin (TOR) signaling pathway affects translation by altering the amount of ribosomal RNA and ribosomal proteins, as well as by phosphorylating translation initiation factors [7]. Furthermore, post-translational modifications (PTMs) consisting in the addition of a functional group to a protein such as phosphorylation, ubiquitination, acetylation and palmitoylation can modulate gene expression [8].

Perturbations at the different stages of gene expression and regulation have been associated with many diseases. Impairment of transcription regulation has been linked to cancer, autoimmunity, neurological disorders and diabetes, among others [9]. Similarly, translational misregulation is involved in several diseases such as cancer [10], tissue hypertrophy and neurodegeneration [11]. Understanding the mechanisms behind these processes helps to identify potential disease triggers and formulate new therapies.

1.2 Transcription factor function

Transcription is the process that assembles an mRNA transcript from a DNA strand. This process is carried out by an enzyme called RNA polymerase, which reads the DNA strand. In order to start the transcription, the RNA polymerase binds to specific DNA regions called promoters. TFs are proteins that are also able to bind to DNA and either help or prevent the RNA polymerase to attach to the promoter and initiate transcription. Because transcription is the first step in gene expression, its regulation is the most important step of the overall process.

Many biological processes depend on the regulation of transcription, such as cell cycle control, maintenance of physiological balance, cellular differentiation, development, and response to environmental conditions [12]. In recent years, an increasing number of transcription factors and their target have been identified in the human genome [13] and in *E. coli* [14] that are being progressively integrated in databases of gene regulation. These advances are mainly due to improved methods of chromatin immunoprecipitation (ChIP) [15], which associate TFs to genomic regions. However, despite the mappings of TF to the DNA regions with which they interact, it still lacks quantitative information on their binding affinities to better characterize their effect. To this aim, new high-throughput methods have been developed to quantify the absolute binding affinities, such as the mechanically induced trapping of molecular interactions (MITOMI) [16]. MITOMI consists of a microfluidic platform where TFs, held by antibodies, are mixed with fluorescently labeled DNA. After equilibration, a mechanically induced trapping removes the unbound material, leaving only the bound TF-DNA complexes. The binding affinity can be subsequently quantified by measuring the remaining fluorescence [17]. MITOMI was also used to measure the simultaneous binding of two TFs on DNA, after measuring their individual binding affinity [18]. This approach allowed the quantification of cooperative effects of the TFs, presented in Chapter 2.

1.3 Bioreporters design principles and applications

The combined increased knowledge on gene regulatory networks and genetic engineering methodologies has led to the development of bioreporters. These bioreporters are living cells that have been genetically engineered to produce an easily measurable reporter signal in the presence of a specific chemical compound. A bioreporter requires a reporter gene that encodes for the reporter protein, and a promoter-operator region that activates the gene in the presence of the target molecule (Figure 1.3).

Bioreporters can be designed to monitor the presence of toxic compounds via activation of specific gene responses, such as for heavy metals, antibiotics or toluene, or via more general cellular responses, such as oxidative stress or heat shock responses [19]. Bioreporters could be a substitute to chemical analysis for the monitoring of chemicals in the environment [20] and their bioavailability [21], because they are cheap to cultivate, able to rapidly generate a fluorescent signal and can be integrated in microfluidic devices.

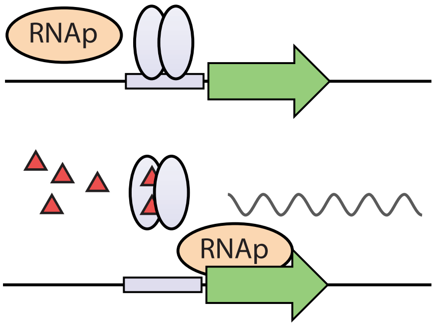


Figure 1.3 **Example of bioreporter design**. Top: the reporter gene is shut down by a transcription factor (TF) in the absence of the chemical compound of interest, and thus the RNA polymerase cannot start the transcription. Bottom: the presence of the chemical compound hinders the binding of the TF and the reporter gene is expressed.

Arsenic bioreporters come in two main types, depending on the gene configuration of their reporter plasmid. The first *E. coli* based bioreporter uses its natural feedback loop for detoxification and reports arsenic concentrations in a linear manner on the 0-80 µg/L concentration range [22]. By removing the feedback loop and placing the repressor gene under constitutive promoter, bioreporters with tunable responses were achieved [23]. To better understand how these circuits work and how we could improve the detection at low arsenic concentration, we created a detailed model that we used for parameter estimation and for prediction of an improved bioreporter variant.

1.4 Computational approaches to dissect complex reaction networks

Synthetic biology combines approaches and methods from electrical and genetic engineering, such as biological engineering, molecular biology, computational biology and computer science, among others, to create biological devices or systems by assembling biological components. The technological advances made in recent years allow precise experimental measurements of gene regulation at all the above-mentioned levels (see Section 1.1). Computational tools and methods have also been developed to model and analyze these networks and gather insights on the causes of variations in gene regulation.

Depending on the level of complexity of the system and on the aim of the research, a wide range of mathematical models of gene regulation can be applied. Bayesian networks combine probability theory and graph theory. These networks depict the conditional dependencies within a set of variables, and can provide the most probable cause of an observation, e.g. what is the most probable gene network structure underlying the observed gene expression data [24]. Gene regulatory networks can be described by Boolean models where genes are represented as being either active or inactive and their regulation is determined by logic functions applied on the states of their regulators. Similarly to Bayesian models, Boolean models can be used to infer regulatory networks from gene expression data [25]. Ordinary differential equation (ODE) models use continuous variables and are convenient for time-series or steady-state simulations. These models can be used to explore quantitative data on parameters (binding affinities, degradation) and variables (concentrations) to simulate the biological system. They can also be used for condition-dependent parameter estimation. However, at the molecular level, molecular concentrations are no longer continuous, which is not captured by ODE models. In such cases, stochastic models such as Gillespie's stochastic simulation algorithm (SSA) provide a method to simulate time-series of a chemical system at low concentrations [26], which at the high concentration limit converges to ODE models.

Understanding how human TFs work to find their DNA target is poorly understood. Even a simple system of two TFs binding to two DNA sites requires an extensive mathematical formulation. With the advent of new methods for experimental measurements of absolute binding affinities, such detailed models of molecular interactions can be developed, but remain uncommon.

Development and optimization of bioreporters rely mainly on experimental trial and error, high-throughput screenings or logic gate modeling and testing. Models of gene regulatory networks usually simplify the gene responses to step, logistic, or Hill functions. A mechanistic description is missing to account quantitatively for the molecular reactions of the regulatory network of a bioreporter, such as dimerization, transport or protein maturation.

1.5 Aims of research

In this thesis, we make an effort to go towards quantitative representations of molecular interactions in biological systems. With the help of high-throughput experimental procedures allowing the absolute measurements of binding affinities, a detailed mechanistic model for a system of two TFs interacting with DNA becomes useful, in particular to estimate the cooperative effects of the TFs upon their binding on DNA.

Gene regulatory systems involved in bioreporters are small enough to be comprehensively described by a mathematical model. The model can be used for parameter estimation to better understand the mechanisms involved in the different bioreporter circuits and hint towards the improvement of arsenic detection at low concentration. With the obtained parameters, we could explore the possibility of development of a bioreporter based on genetic toggle switch, a bistable gene regulatory network, for a switch-like response to arsenic detection.

1.6 Thesis structure

In Chapter 2, I present a collaborative work in which my contribution was to develop a computational model for a system of TFs and their binding motifs to a DNA target. The model encompassed all the molecular interactions and was used to quantify the cooperativity between two TFs upon their binding to DNA.

In Chapter 3, I present a model of the gene circuits assembled in arsenic bioreporter that takes into account all known molecular interactions around the regulatory species. The model was calibrated with experimental data and used to improve the detection of arsenic. Finally, the variability of the regulatory mRNA was analyzed with stochastic simulations.

In Chapter 4, the molecular interactions estimated by the arsenic bioreporter model in Chapter 3 were implemented in a model of bioreporter based on a genetic toggle switch. The model was used to study what parameters should be changed in order to adjust the bistability regions for an efficient detection of arsenic at low concentration.

Chapter 2 Cooperativity in dimer-DNA binding

The contents of this chapter have been previously published with the following details:

Isakova, Alina, Yves Berset, Vassily Hatzimanikatis, and Bart Deplancke. “Quantification of Cooperativity in Heterodimer-DNA Binding Improves the Accuracy of Binding Specificity Models.” *Journal of Biological Chemistry* 291, no. 19 (May 6, 2016): 10293–306. doi:10.1074/jbc.M115.691154.

Author contributions: A.I. and B.D. designed the study and wrote the paper, A.I. performed the in vitro experiments, A.I., Y.B. and V.H. performed mechanistic modeling.

2.1 Introduction

Molecular interactions are at the basis of all cellular function and participate in very different biological processes – e.g. substrate transformation by the binding of an enzyme, or regulation of DNA transcription by the binding of a TF. A TF is a ligand that binds to specific locations. Some TFs target very specific DNA sites, whereas others interact with a broad range of DNA sites. Many of them can form heterodimers and as a result modify their binding affinities towards DNA. For instance, when a ligand (such as a TF) binds a receptor with multiple binding sites (DNA), it often follows that the affinity of the adjacent sites to bind another TF is increased. These mechanisms are referred to as cooperative binding.

2.1.1 Hill Equation and Cooperativity

Mapping the interactions between transcription factors (TFs) and their DNA target sites is essential for elucidating the structural properties of gene regulatory networks [27], [28]. Data on TF-DNA binding specificities have so far revealed that individual TFs can bind to a broad set of target sites that cover a wide affinity range [29]–[32]. In addition, it is now well appreciated that the binding of many TFs is not autonomous but is in fact influenced by a multitude of factors, including chromatin state, post-translational modifications, and interactions with other proteins. One specific form of protein interaction involves two TFs forming one heterodimeric DNA binding complex. Such heterodimers are highly abundant across organisms and exert essential molecular functions [28], [33], [34]. Consequently, a lot of effort has been invested to determine their DNA binding specificities using various *in vitro* and *in vivo* approaches [33], [35]–[41]. Several studies demonstrated the ability of two TFs to cooperate on DNA elements and thus provide an alternative mode of DNA recognition [42], [43]. For example, Hox proteins gain novel specificities when bound to DNA together with the dimeric cofactor Exd [44]. Sox-Oct partners, as well as certain nuclear receptor dimers, have different cooperativity constants when bound to DNA sites separated by spacers of variable length [43], [45], [46].

Historically, experimental measurement of ligand binding was performed without prior knowledge on the number of binding sites present [47]. The fraction of bound receptors can be expressed by the Hill equation as a function of the concentration of ligand ([L]):

$$\theta = \frac{[\text{Bound receptor}]}{[\text{Total receptor}]} = \frac{K_{\mathrm{H}}[L]^{n_{\mathrm{H}}}}{1+K_{\mathrm{H}}[L]^{n_{\mathrm{H}}}}. \tag{2.1}$$

The parameter K_H represents the concentration of ligand at which half of the receptors are saturated, and n_H is the Hill coefficient [48], which relates to the degree of cooperativity at which the ligand binds to the receptor. The Hill coefficient was usually used to report the effect of cooperativity, and is a simple way to do it. Unfortunately, n_H does not relate to anything mechanistically. Originally, Hill introduced his equation in the context of discrepancies in the measurement of hemoglobin molecules binding to oxygen. He suspected that the disagreements in the saturation curves measured by different scientists were due to aggregation of hemoglobin molecules in solutions of different salinities. Because the actual number of hemoglobin molecules in aggregates were not known, he tried to see if the formula would fit all the curves [48]. Hill was aware that he was neglecting the different binding steps, but the calculations for more than two parameters were too tedious. This formula remained famous because it could successfully fit a broad range of saturation curves and as a result n_H remained associated with the number of binding pockets, or the degree of cooperativity.

Only years later the number of binding sites for oxygen on hemoglobin was discovered by Adair [49], who derived an expression for the fractional occupancy of the receptors. Later on, Klotz extended the expression [50] to provide the Adair-Klotz equation

$$\theta = \frac{1}{n}\frac{K_1[L]+2K_1K_2[L]^2+\ldots+nK_1\ldots K_n[L]^n}{1+K_1K_2[L]^2+\ldots+K_1\ldots K_n[L]^n}, \tag{2.2}$$

where n is the number of binding sites (n=4 for hemoglobin) and K_i is the apparent binding constant of the i^{th} binding ligand. From this formula, we remark that the Hill equation (2.1) can accommodate the shape of the more general Adair-Klotz equation (2.2) when cooperativity effects are strong, *i.e.* when the term of power n dominates, or by adjusting the Hill coefficient, which can take any continuous value.

Below it is illustrated how the Adair-Klotz equation (2.2) is more adapted to quantify the cooperativity of ligand binding to a receptor. In this example, a receptor has two binding sites (Figure 2.1A). In one case the ligand molecules bind without cooperativity, i.e. the binding affinities of the binding sites are fixed, and in the other case the ligands cooperate positively, i.e. their affinity for a binding site is increased if another ligand is bound to the adjacent site.

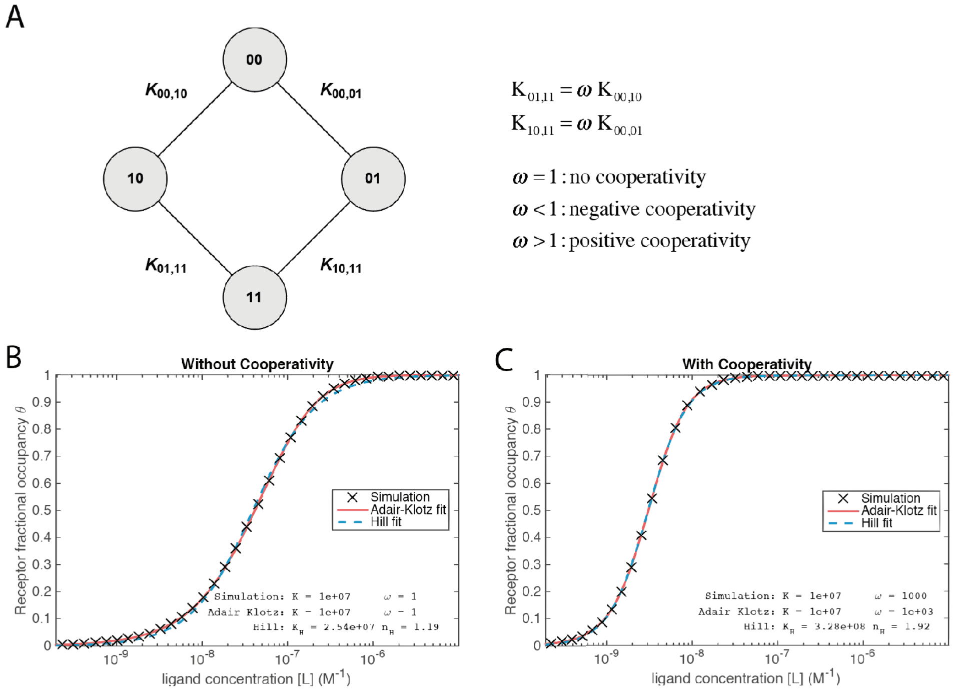


Figure 2.1 **One ligand, two sites cooperativity model. A**: Occupancy states of the receptor with two identical ligand-binding sites (0: free binding site, 1: occupied binding site). In absence of cooperativity, the states of the neighboring sites do not influence each other, hence ω = 1. With cooperativity, an occupied site reduces (ω<1) or increases (ω>1) the affinity for a ligand to bind its neighboring site. **B**: Estimation of cooperativity effects through Hill and Adair-Klotz fits.

We simulated the system (Figure 2.1A) with ODEs for arbitrary values of binding affinity ($K_{00,01} = K_{00,10} = 10^7\ M^{-1}$) and cooperativity ($\omega = 1$, $\omega = 10^3$) and generated the fractional occupancy of the receptors. Then, we fit the data with Hill and Adair-Klotz equations, and reported the parameter obtained (Figure 2.1B). Using Adair-Klotz formula, the fractional occupancy of the receptor simplifies to

$$\theta = \frac{2K[L] + \omega K^2[L]^2}{1 + 2K[L] + \omega K^2[L]^2} \tag{2.3}$$

where K is the binding affinity between the receptor and the ligand. From this formula, it is easy to see that the Hill equation can approximate the Adair-Klotz equation when cooperativity effects are strong (*i.e.* ω>>1), or by adjusting the Hill coefficient.

Even if the Hill equation fits tightly to the data, the reported parameters are not as informative as the fits with Adair-Klotz equation. In the absence of cooperativity, we find a Hill coefficient $n_H > 1$, suggesting that the system has some cooperative effects. With cooperativity, the Hill coefficient increased to indicate a greater cooperativity effect, whereas the Adair-Klotz parameters estimated correctly the cooperative effects in both cases. Using the Hill coefficient makes the silent assumption that the cooperative effects are so strong that we can ignore the intermediate binding states; or equivalently, that n ligands bind the receptor simultaneously. Moreover, the hill exponent n is more a qualitative measure, knowing that above 1 there are cooperativity effects, but does not quantify the strength of a mechanism [51].

2.1.2 Quantification of Cooperativity in Heterodimer-DNA Binding

Despite this clear demonstration of cooperativity phenomena, our ability to integrate its impact in quantitative models of DNA binding, and ultimately gene regulation, remains limited. Consequently, several important questions remain unaddressed. These include whether the perturbation of cooperative TF-DNA binding always involves major rearrangements of interacting molecules such as, for example, the addition or removal of a protein partner or introduction of a different spacer between two binding sites. In addition, it remains unclear whether cooperativity can also be modulated on a much more fine-grained scale such as, for example, at the level of nucleotide variations in target binding sites. More specifically, it has not been comprehensively explored whether the information on the variable “strength” of cooperative effects in dimer binding to sites of different nucleotide composition could be used to refine a quantitative specificity model for the TF pair. Several quantitative models of TF-DNA binding specificity have been developed [29], [37], [52], [53], but none of these include to our knowledge the cooperative determinant of specificity. This knowledge gap reflects in large part the challenging nature of retrieving quantitative DNA binding parameters underlying heterodimer-DNA binding.

In this study, my collaborators addressed this challenge by using a robust microfluidics approach, MITOMI [54], which allowed them to track and characterize the implicated molecular interactions in great quantitative detail. As a model system, they focused on the PPARγ:RXRα heterodimer. PPARγ is well known as one of the major regulators of adipocyte differentiation [55], [56], forming a DNA binding partnership with another nuclear receptor, RXRα, to control the adipogenic gene expression program. Generating a quantitative understanding of the molecular rules underlying the assembly of this heterodimer on DNA is therefore of gene regulatory as well as biomedical relevance. To accommodate the quantitative analysis of PPARγ:RXRα-DNA interactions, they expanded the previously described MITOMI setup by introducing and testing the usage of multiple fluorescent fusions with both heterodimer TFs, aiming to both track individual TFs as well as to monitor homo- and heterodimer formation on DNA (Figure 2.2). Then, they used the MITOMI-derived data to assess the ability of the PPARγ:RXRα heterodimer to change its specificity upon dimerization as well as to support the development of a detailed quantitative binding model, specifically assessing the contribution of cooperativity to the DNA binding process.

In this context, I took a comprehensive mechanistic modeling approach that allowed me to derive binding constants that account for cooperative heterodimer-DNA binding. This allowed me to build a PPARγ:RXRα-DNA binding specificity model of greater predictive power than the one based on a regular one-site equilibrium. As such, the results provide unprecedented insights into the quantitative aspects of PPARγ:RXRα-DNA complex formation, emphasizing the role of binding site composition in influencing the cooperative nature of heterodimeric DNA binding.

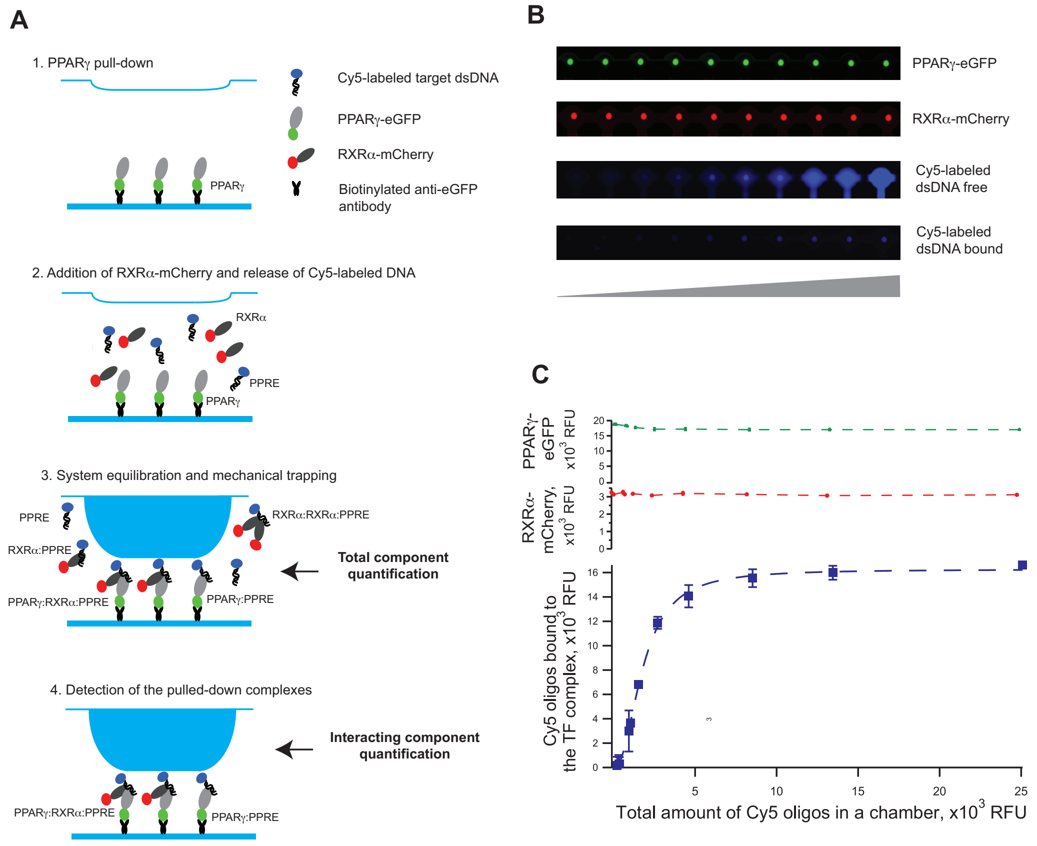

A
1. PPARγ pull-down
Cy5-labeled target dsDNA
PPARγ-eGFP
RXRα-mCherry
Biotinylated anti-eGFP antibody
PPARγ
2. Addition of RXRα-mCherry and release of Cy5-labeled DNA
RXRα
PPRE
PPARγ
3. System equilibration and mechanical trapping
PPRE
RXRα:RXRα:PPRE
RXRα:PPRE
Total component quantification
PPARγ:RXRα:PPRE
PPARγ:PPRE
4. Detection of the pulled-down complexes
Interacting component quantification
PPARγ:RXRα:PPRE
PPARγ:PPRE
B
PPARγ-eGFP
RXRα-mCherry
Cy5-labeled dsDNA free
Cy5-labeled dsDNA bound
C
PPARγ-eGFP x10³ RFU
RXRα-mCherry, x10³ RFU
Cy5 oligos bound to the TF complex, x10³ RFU
Total amount of Cy5 oligos in a chamber, x10³ RFU
0 5 10 15 20
0 1 2 3
0 2 4 6 8 10 12 14 16
5 10 15 20 25

Figure 2.2 **On-chip heterodimer-DNA assembly. A**: Schematic representation of the experimental set-up. (1) PPARγ fused to an eGFP tag is immobilized on the surface of a MITOMI chip with an anti-GFP antibody. (2) RXRα tagged with mCherry and Cy5-labeled DNA baits are introduced into the system and (3) incubated for one hour to allow system equilibration and complex assembly. (4) Newly formed complexes are trapped under a flexible PDMS membrane and unbound molecules as well as molecular complexes are washed away. **B**: Fluorescence-based read-out of PPARγ-GFP, RXRα-mCherry, and Cy5-labeled target DNA from ten MITOMI units. The two upper panels represent PPARγ-GFP and RXRα-mCherry detected in the center of each unit (under the PDMS membrane). The two lower panels represent the variable amounts of Cy5-labeled target DNA molecules detected in the same ten MITOMI units, before (DNA free) and after (DNA bound) trapping. **C**: Corresponding quantitative readout of B where the quantified amounts of both PPARγ and RXRα remain constant but the amount of bound DNA increases with the input DNA concentration until it reaches saturation. The corresponding quantities of proteins and DNA are expressed in relative fluorescent units (RFU).

2.2 Methods

2.2.1 Experimental Procedures

Device fabrication

All the molds for microfluidic devices and devices itself were designed and fabricated as described previously [54], [57].

Synthesis and printing of target DNA libraries

All target DNA fragments were obtained as single-stranded oligonucleotides from Invitrogen. These oligonucleotides were subsequently used to generate fluorescently labeled double-stranded oligonucleotides as described previously [54]. The single base substitution libraries of PPRE, 5'-AAACTAGGTCAAAGGTCA-3', and PAL3, 5'-AAACTAGGTCACCGTGACCT-3', were generated by substituting one nucleotide of the elements at a time to all possible variants. All labeled double-stranded oligonucleotides were spotted onto an epoxy-coated glass slides (CELL Associates) with a SpotBot III microarrayer (ArrayIT) using a 946MP4 pin (European Biotek Network SPRL).

Protein cloning and expression

TFs were expressed *in vitro* using the TnT SP6 High-Yield Wheat Germ protein expression system (Promega). To enable the expression of TFs and their fluorescence-based detection, we generated novel vectors by cutting the pF3A WG (BYDV) Flexi vector (Promega) with *NcoI* and *DraI*, removing the barnase cassette. The *NcoI* site was blunted, and the Gateway reading frame A cassette (Life Technologies, Inc.) was ligated. Subsequenty, the eGFP and the mCherry coding sequence (EUROSCARF) containing a stop codon at the 3'-end were incorporated between the *KpnI* and *SacI* restriction sites using standard cloning techniques. Full-length PPARγ and RXRα ORFs were then subcloned from the Entry clones [58] into the generated vectors by standard Gateway cloning.

MITOMI and Data analysis

The surface chemistry, MITOMI and image acquisition were performed as described previously [54]. We quantified the amount of each mutated sequence bound to the respective TF at the equilibrium state by means of fluorescence in a range of input DNA concentrations. The obtained equilibrium binding curves for each sequence were then fitted with the regression curves generated from the proposed model of cooperative binding.

2.2.2 Binding Model

Monomer-DNA interactions

In case of a single TF-DNA interaction at equilibrium, the reactions are characterized by:

$$[\mathrm{PPAR}\gamma\mathrm{:PPRE}] \rightleftharpoons [\mathrm{PPAR}\gamma] + [\mathrm{PPRE}]$$

$$K_{00,10} = \frac{[\mathrm{PPAR}\gamma\mathrm{:PPRE}]}{[\mathrm{PPAR}\gamma][\mathrm{PPRE}]} \ , \qquad (2.4)$$

and

$$[\mathrm{RxR}\alpha\mathrm{:PPRE}] \rightleftharpoons [\mathrm{RxR}\alpha] + [\mathrm{PPRE}]$$

$$K_{00,20} = \frac{[\mathrm{RxR}\alpha\mathrm{:PPRE}]}{[\mathrm{RxR}\alpha][\mathrm{PPRE}]} \ , \tag{2.5}$$

where $K_{00,10}$ and $K_{00,20}$ are the respective PPARγ- or RXRα-DNA binding constants that are mutation-dependent. For monomer-DNA interactions, the binding curves were fitted with a single-parameter non-linear function. For each sequence, the fit that yielded the lowest χ^2 value was used to compute the function parameter (binding constant). The accuracy of the fitting parameters was assessed via residuals of the fit. The standard deviation (σ) of the binding constant was computed for each sequence (Table 2.1).

Heterodimer-DNA interactions

In the case of heterodimer-DNA interactions, we accounted for the number of all possible molecular species that could be formed between all three components. We formed a system of two different sites and two ligands, similar to the one described in [59], with the following additional properties: we allowed RXRα to dimerize with itself or with PPARγ, and we allocated two binding sites for RXRα (left and right, equal binding affinity), with one of them (left) also able to bind PPARγ. These considerations led to the definition of the following species: PPRE (X_0); PPARγ (X_1); RXRα (X_2); PPARγ:RXRα (X_{D1}); RXRα:RXRα (X_{D2}); PPARγ:PPRE (X_{10}); RXRα:PPRE (X_{20}); PPRE:RXRα (X_{02}); PPARγ:PPRE:RXRα (X_{12}); RXRα:PPRE:RXRα (X_{22}); PPARγ:RXRα:PPRE (X_{120}); PPRE:RXRα:PPARγ (X_{012}); PPRE:RXRα:RXRα (X_{022}); RXRα:RXRα:PPRE (X_{220}); and RXRα:PPARγ:PPRE (X_{210}); where the notation PPARγ:PPRE:RXRα (X_{12}) indicates that PPARγ binds to the left binding site of PPRE and RXRα to the right one. PPARγ:RXRα:PPRE (X_{120}) indicates that the PPARγ:RXRα heterodimer binds PPRE only via RXRα (see Figure 2.3).

All possible elementary interactions between PPARγ, RXRα and PPRE and are shown in Scheme 1. From the above relations, we define K_{DoD} (and K^*_{DoD}) as the total free energy leading to PPARγ:RXRα heterodimers (respectively RxRα:RxRα homodimers) bound on DNA. Each path must be equal in free energy hence we have the equalities:

$$K_{\mathrm{DoD}} = K_{00,10}K_{10,12} = K_{00,02}K_{02,12} = K_{\mathrm{D1}}K_{00,12} = \omega_{1,2}K_{00,10}K_{00,02} \tag{2.6}$$

$$K^*_{\mathrm{DoD}} = K_{00,20}K_{20,22} = K_{00,02}K_{02,22} = K_{\mathrm{D2}}K_{00,22} = \omega_{2,2}K_{00,20}K_{00,02} \tag{2.7}$$

where $K_{i,j}$ represent binding affinities involved in the transition from state i to j (Figure 2.3). K_{DoD} will be ultimately used to quantify the overall binding of the dimers on DNA, but cannot be measured directly. However, after applying the measurements of single interactions ($K_{00,10}$, $K_{00,20}$, K_{D1} and K_{D2}) to the above equalities, the unknowns $K_{10,12}$, $K_{02,12}$, $K_{00,12}$ and $\omega_{1,2}$ can be collapsed into one unknown, $\omega_{1,2}$. By doing the same with $K_{20,22}$, $K_{02,22}$, $K_{00,22}$ and $\omega_{2,2}$, the system is left with two unknowns, $\omega_{1,2}$ and $\omega_{2,2}$.

These two remaining parameters represent the cooperativity. They represent how much the binding of the second monomer is enhanced by the fact that the DNA is already bound.

2.2.3 Quantification of cooperativity

After the assignment of experimental values to $K_{00,10}$, $K_{00,02}$, K_{D1}, and K_{D2} measured in a previous experiment, the system remains with two independent parameters, $K_{00,12}$ and $K_{00,22}$. The experimental data are from MITOMI measurements of DNA bound to PPARγ in presence of RxRα, with different concentrations of DNA. PPARγ is immobilized on the chip and mixed with PPRE and RxRα. After equilibration, the mechanical trapping removes the elements that were not bound to PPARγ. The measurement reports the concentration of PPRE complexes bound to PPARγ for different PPRE concentrations, repeated for the whole PPRE library. The mathematical simulations were made by solving an ODE until the system equilibrates. We simulate the trapping by reporting the concentration of complexes involving PPARγ. The binding affinity of PPARγ and RxRα to each library member is adjusted by using the affinity measured independently. We solve the system at equilibrium, i.e. find the species concentrations such that all equilibrium relations are fulfilled. We calculate the fraction of PPRE involved in complexes with PPARγ and find the parameters $K_{00,12}$ and $K_{00,22}$ such that the simulation best fits the experimental measurements of PPRE bound to immobilized PPARγ using least square minimization. The accuracy of each fit was assessed through the residual sum of squares value (see RSS,

Figure 2.11). The simulations were performed with Matlab (Mathworks).

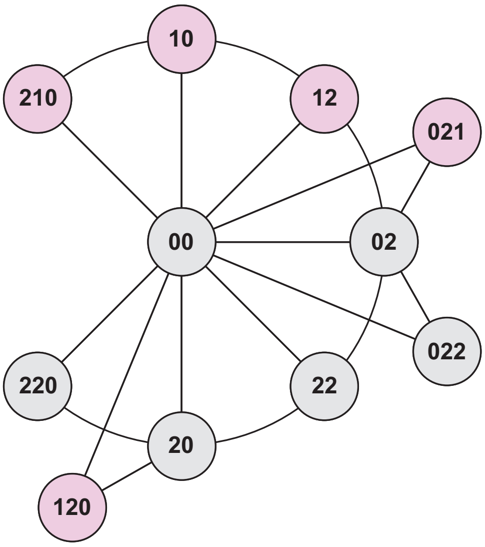


Figure 2.3. **States of the DNA target bound to PPARγ and RxRα proteins**. 0: free, 1: PPARγ, 2: RxRα. States are linked if they are separated by one binding event from each other. Pink nodes represent states involving PPARγ, which will mechanically trap the DNA to be measured as bound. States with 3 numbers mean that the outermost protein (labelled 1 or 2) is only attached to the other protein, not to the DNA site.

Scheme 1. Single reactions between PPARγ, RxRα and PPRE (see Figure 2.3), and their associated binding constants

$$\mathrm{PPAR}\gamma+\mathrm{PPRE}\overset{K_{00,10}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{PPRE} \qquad K_{00,10}=\frac{[\mathrm{PPAR}\gamma:\mathrm{PPRE}]}{[\mathrm{PPAR}\gamma][\mathrm{PPRE}]}$$

$$\mathrm{PPAR}\gamma+\mathrm{RxR}\alpha\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha \qquad K_{\mathrm{D1}}=\frac{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha]}{[\mathrm{PPAR}\gamma][\mathrm{RxR}\alpha]}$$

$$\mathrm{RxR}\alpha+\mathrm{RxR}\alpha\overset{K_{\mathrm{D2}}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{RxR}\alpha \qquad K_{\mathrm{D2}}=\frac{[\mathrm{RxR}\alpha:\mathrm{RxR}\alpha]}{[\mathrm{RxR}\alpha][\mathrm{RxR}\alpha]}$$

$$\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{00,20}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPRE} \qquad K_{00,20}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPRE}]}{[\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{PPRE}+\mathrm{RxR}\alpha\overset{K_{00,02}}{\rightleftharpoons}\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{00,02}=\frac{[\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{PPRE}][\mathrm{RxR}\alpha]}$$

$$\mathrm{PPAR}\gamma:\mathrm{PPRE}+\mathrm{RxR}\alpha\overset{K_{10,12}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{10,12}=\frac{[\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{PPAR}\gamma:\mathrm{PPRE}][\mathrm{RxR}\alpha]}$$

$$\mathrm{PPAR}\gamma+\mathrm{PPRE}:\mathrm{RxR}\alpha\overset{K_{02,12}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{02,12}=\frac{[\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{PPAR}\gamma][\mathrm{PPRE}:\mathrm{RxR}\alpha]}$$

$$\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{00,12}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{00,12}=\frac{[\mathrm{PPAR}\gamma:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{RxR}\alpha:\mathrm{PPRE}+\mathrm{RxR}\alpha\overset{K_{20,22}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{20,22}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{RxR}\alpha:\mathrm{PPRE}][\mathrm{RxR}\alpha]}$$

$$\mathrm{RxR}\alpha+\mathrm{PPRE}:\mathrm{RxR}\alpha\overset{K_{02,22}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{02,22}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{RxR}\alpha][\mathrm{RxR}\alpha:\mathrm{PPRE}]}$$

$$\mathrm{RxR}\alpha:\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{00,22}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha \qquad K_{00,22}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPRE}:\mathrm{RxR}\alpha]}{[\mathrm{RxR}\alpha:\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{PPAR}\gamma+\mathrm{RxR}\alpha:\mathrm{PPRE}\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha:\mathrm{PPRE} \qquad K_{\mathrm{D1}}=\frac{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha:\mathrm{PPRE}]}{[\mathrm{PPAR}\gamma][\mathrm{RxR}\alpha:\mathrm{PPRE}]}$$

$$\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{\mathrm{D2}}}{\rightleftharpoons}\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha:\mathrm{PPRE} \qquad K_{\mathrm{D2}}=\frac{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha:\mathrm{PPRE}]}{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{PPRE}+\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma \qquad K_{\mathrm{D1}}=\frac{[\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma]}{[\mathrm{PPRE}][\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha]}$$

$$\mathrm{PPRE}:\mathrm{RxR}\alpha+\mathrm{PPAR}\gamma\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma \qquad K_{\mathrm{D1}}=\frac{[\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma]}{[\mathrm{PPAR}\gamma][\mathrm{PPRE}:\mathrm{RxR}\alpha]}$$

$$\mathrm{PPRE}:\mathrm{RxR}\alpha+\mathrm{RxR}\alpha\overset{K_{\mathrm{D2}}}{\rightleftharpoons}\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{RxR}\alpha \qquad K_{\mathrm{D2}}=\frac{[\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{RxR}\alpha]}{[\mathrm{PPRE}][\mathrm{RxR}\alpha:\mathrm{RxR}\alpha]}$$

$$\mathrm{PPAR}\gamma:\mathrm{PPRE}+\mathrm{RxR}\alpha\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma:\mathrm{PPRE} \qquad K_{\mathrm{D2}}=\frac{[\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{RxR}\alpha]}{[\mathrm{PPRE}:\mathrm{RxR}\alpha][\mathrm{RxR}\alpha]}$$

$$\mathrm{RxR}\alpha:\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{00,20}}{\rightleftharpoons}\mathrm{PPRE}:\mathrm{RxR}\alpha:\mathrm{RxR}\alpha \qquad K_{00,20}=\frac{[\mathrm{RxR}\alpha:\mathrm{RxR}\alpha:\mathrm{PPRE}]}{[\mathrm{RxR}\alpha:\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{RxR}\alpha+\mathrm{RxR}\alpha:\mathrm{PPRE}\overset{K_{\mathrm{D2}}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{RxR}\alpha:\mathrm{PPRE} \qquad K_{\mathrm{D2}}=\frac{[\mathrm{RxR}\alpha:\mathrm{RxR}\alpha:\mathrm{PPRE}]}{[\mathrm{RxR}\alpha][\mathrm{RxR}\alpha:\mathrm{PPRE}]}$$

$$\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha+\mathrm{PPRE}\overset{K_{00,10}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma:\mathrm{PPRE} \qquad K_{00,10}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma:\mathrm{PPRE}]}{[\mathrm{PPAR}\gamma:\mathrm{RxR}\alpha][\mathrm{PPRE}]}$$

$$\mathrm{PPAR}\gamma:\mathrm{PPRE}+\mathrm{RxR}\alpha\overset{K_{\mathrm{D1}}}{\rightleftharpoons}\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma:\mathrm{PPRE} \qquad K_{\mathrm{D1}}=\frac{[\mathrm{RxR}\alpha:\mathrm{PPAR}\gamma:\mathrm{PPRE}]}{[\mathrm{PPAR}\gamma:\mathrm{PPRE}][\mathrm{RxR}\alpha]}$$

Cooperativity

We next use the values of the ternary complexes $K_{00,12}$ and $K_{00,22}$ derived from the model fits to assess the presence or absence of cooperative effects in heterodimer-DNA binding. Cooperativity effects can be quantified at the steady-state through the cooperativity factors shown in (2.8),

$$\omega_{1,2} = \frac{K_{10,12}}{K_{00,10}} = \frac{K_{\mathrm{D1}}\,K_{00,12}}{K_{00,10}\,K_{00,02}}, \quad \omega_{2,2} = \frac{K_{20,22}}{K_{00,02}} = \frac{K_{\mathrm{D2}}\,K_{00,22}}{{K_{00,02}}^{2}} \tag{2.8}$$

Where $\omega_{1,2}$ and $\omega_{2,2}$ are defined strictly as the ω coefficient presented in Ref. [43]). The cooperativity factors can take any value greater than 0; Cooperativity is positive when $\omega > 1$, and negative when $\omega < 1$. Note that this formulation quantifies the effect of cooperativity but does not elucidate its molecular nature, *i.e.* cooperativity can be due to direct ligand-ligand interactions or indirect communication between the ligands [60].

Motif enrichment in ChIP-seq data

ChIP-seq-based PPARγ:RXRα DNA binding regions in 3T3-L1 cells were retrieved from Nielsen et al. [61] and processed as in Raghav et al. [62]. Area under the characteristic curve (AUC) representing the binding site occupancy predicted by the binding model was calculated as described previously [63] in that a 200 bp region around the center of the peak was used as the positive binding region and a 200-bp-long genomic sequence 300 bp downstream of the peak center as the negative binding region. Three motifs were used in the search: 1) Position specific scoring matrix (PSSM) motif derived from K_d values; 2) PSSM motif derived from K_{DoD} values; 3) JASPAR motif (MA0065.2, JASPAR CORE database). PSSM contain the relative scores of nucleotide contributions derived from the binding affinities of PPARγ:RXRα to every PPRE from the single-point mutation library. PSSMs quantify the relative preferences for each bases of the PPRE, derived from K_d values (one-site binding model) and K_{DoD} values (two-site cooperative model). From the single-point mutation library, PSSMs can be used to predict the score of any given sequence. JASPAR is a curated database of TF-binding profiles in the form of PSSMs that can be used to predict TF binding sites [64].

2.3 Results

Benchmarking of MITOMI-based PPARγ:RXRα-DNA interaction analyses

Recent ChIP-seq [61], ChIP-chip [65], and ChIP-PET [66] analyses revealed that the PPRE is the main *cis*-acting element for high-affinity tethering of PPARγ:RXRα heterodimers to the DNA. The PPRE contains two copies of the 5'-AGGTCA-3' consensus half-site separated by one nucleotide, constituting the so-called DR1 element, as well as a 5'AAACT sequence that has been shown to be important for PPRE recognition by PPARγ [67]. To benchmark our MITOMI approach, we therefore first investigated the ability of *in vitro* expressed PPARγ, RXRα, and the heterodimer PPARγ:RXRα to preferentially bind to PPRE, as compared to other previously characterized nuclear receptor-binding sites such as the estrogen and glucocorticoid-response elements (ERE and GRE), canonical AGGTCA repeats separated by one or three nucleotides (DR1 and DR2 sites) and variants thereof, as well as the PAL3 element and variants thereof.

Because of the scalability of the MITOMI chips compared with traditional methods such as the gel shift assay, we were able to screen the entire library consisting of 192 sequences at four different DNA concentrations, against either PPARγ or RXRα alone or the PPARγ:RXRα dimer in a single MITOMI experiment. This is important because it allowed us to directly compare the relative DNA affinity of a certain TF for each sequence at uniform surface preparation, conditions, and sample handling. To evaluate the DNA binding preferences of PPARγ, RXRα and PPARγ:RXRα dimers within the queried nuclear receptor DNA binding site space, we quantified DNA bound to the TFs at the equilibrium state. (Figure 2.4A). We then estimated the relative DNA affinity of PPARγ, RXRα, and the heterodimer to given sequences as slopes of linear regression curves fitted to the data points (Figure 2.4B).

We found the binding preferences of PPARγ, RXRα or PPARγ:RXRα heterodimer detected within our MITOMI assay (Figure 2.4) to be consistent with previously identified DNA binding specificities for these TFs, both *in vitro* and *in vivo* [61], [68], thus validating our approach. For example, we observed that the affinity of RXRα to DR1-like sites is significantly greater than to glucorticoid- or estrogen-response element-like elements. In contrast, we found that PPARγ weakly binds to direct repeat sites but strongly to the PAL3 element, as reported previously [69], [70]. However, in the presence of RXRα, PPARγ shifted its specificity to DR1-like sites and no longer exhibited a preference for the PAL3 element. We confirmed this observation by performing independent MITOMI experiments in which we measured the amount of PPARγ that is interacting with RXRα in the presence of either PPRE or PAL3 sites (Figure 2.5A). We fixed the amount of RXRα molecules by immobilizing them on the surface of the chip and introduced PPARγ in amounts that were sufficient to saturate the binding to RXRα while varying the amount of accessible DNA. When using low DNA concentrations, the amount of formed heterodimer was similar for both PPRE and PAL3 elements. However, upon increasing the amount of PPRE target DNA, we observed an increase in heterodimer formation. In the presence of PAL3, we observed the opposite effect as the amount of formed heterodimer decreased, suggesting that PPARγ was bound by PAL3 and thus sequestered from the TF partner (Figure 2.5A). Together, our results clearly demonstrate that also in our MITOMI assay, PPRE is the site to which PPARγ:RXRα has the highest affinity. We therefore decided to use this site for an in-depth TF-TF-DNA binding characterization.

PPARγ and RXRα exhibit intrinsic affinity to the PPRE element prior to dimerization

We performed a detailed analysis of monomeric RXRα and PPARγ binding to the PPRE (Figure 2.5**,** B and D). To investigate the contribution of each nucleotide within the PPRE to the binding specificity of each tested monomeric TF, we generated a single base substitution library of PPRE whereby we substituted each base pair of the element, one nucleotide at a time. We then quantified the TF-bound amount of each mutated sequence at the equilibrium state in a range of input DNA concentrations. We fitted obtained binding curves with the model streamlined for monomeric TF-DNA binding (model fits and corresponding residuals are demonstrated in Figure 2.9 and Figure 2.10). Next, we derived the equilibrium binding constants of PPARγ-PPRE and RXRα-PPRE interactions after which we calculated the differences in binding energy between each sequence of the library and the canonical, non-mutated PPRE (Figure 2.5, B and D). Using these values, we subsequently derived the position specific scoring matrix for PPARγ and RXRα binding to the PPRE element and plotted corresponding enoLOGOS (Figure 2.5 B and D) [71]. This approach has been shown to accurately describe the DNA binding specificities of TFs, even though it assumes that each nucleotide of the element contributes to TF binding independently [54], [72]. We found the following: 1) RXRα binding to PPRE is highly specific such that even a single nucleotide substitution within the core DR1 motif causes a significant change in binding energy (Figure 2.5 B); 2) the 5’-AGGTCA-3’ is the energetically favorable hexameric motif for RXRα monomer binding (Figure 2.5, B and C) consistent with results from previous studies [69], [70], [73], [74]; 3) due to the symmetry of the DR1 element, RXRα can bind to either of the two hexameric half-sites (Figure 2.5, B and C); 4) the binding energy does not change significantly upon the addition of flanking bases up- or downstream of the AGGTCA sequence indicating that 6 bp are sufficient to accommodate an RXRα molecule (Figure 2.5C).

Interestingly, we observed that PPARγ, even without an RXRα partner, shows sequence-specific binding to the PPRE, with its target site located near the 5’-end of the element (Figure 2.5D). Unlike RXRα, sequence-specific DNA binding of PPARγ was not restricted to the 5’-AGGTCA-3’ half-site. The DNA binding energy of PPARγ also changed upon the substitution of bases that are located upstream of this core site and the 5’-AACT element of the DR1 half-site is required for a specific interaction (Figure 2.5**,** D and E). This observation supports the importance of this upstream element in mediating the stabilization of the C-terminal extension of the DNA binding domain of PPARγ, as reported previously [75].

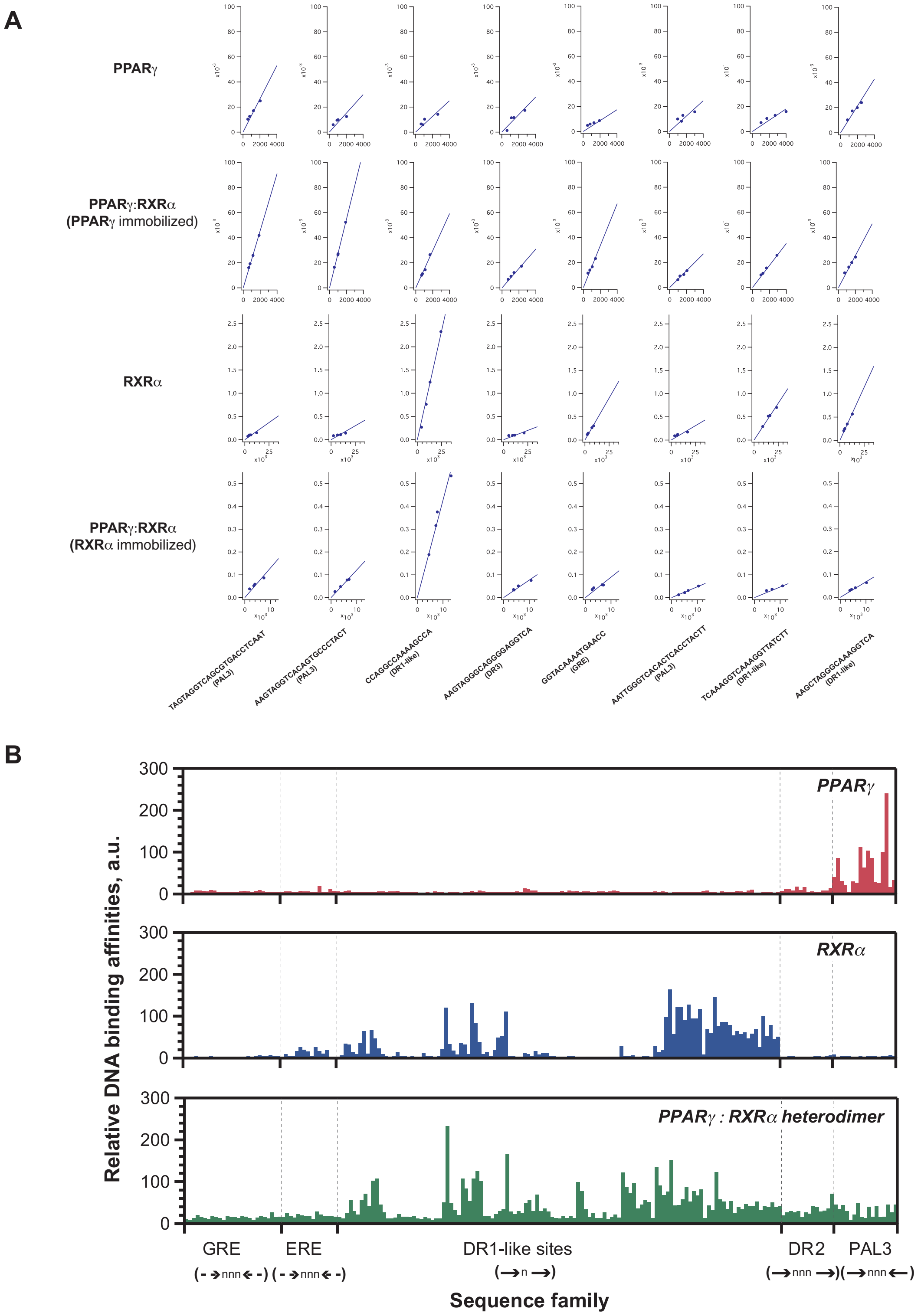


Figure 2.4 **DNA binding preferences of PPARγ, RXRα, as well as the PPARγ:RXRα heterodimer**. A, linear fits of binding data. Examples of binding curves (arbitrary units) and corresponding linear fits of PPARγ, RXRα and PPARγ:RXRα heterodimer interactions with DNA sequences containing putative nuclear receptor binding sites. B, Relative DNA binding affinities of PPARγ, RXRα, and the PPARγ:RXRα heterodimer to five putative nuclear receptor-binding sites and variants thereof. Each sequence family is defined by the orientation of the canonical hexameric sites (represented by arrows) and the spacing between them.

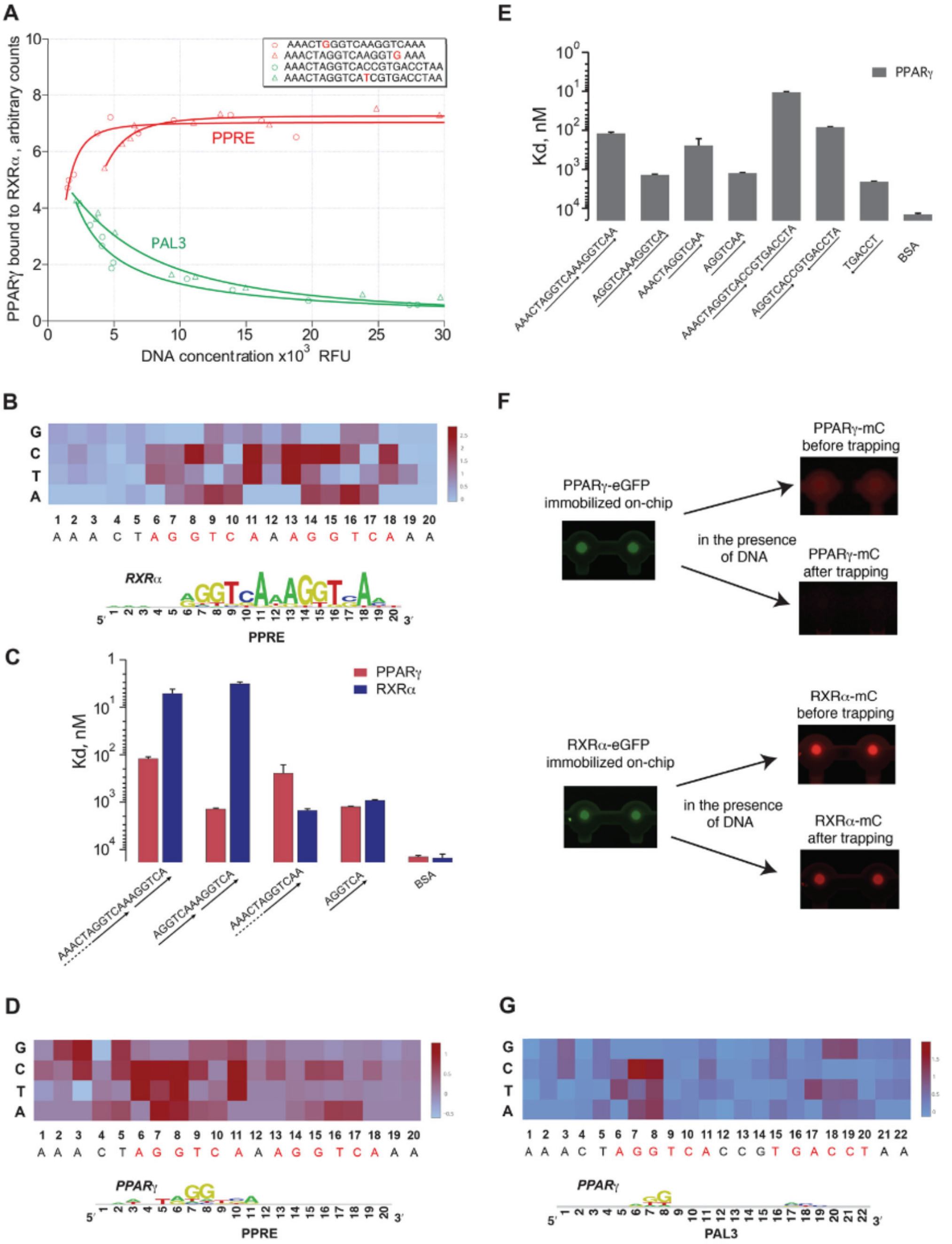


Figure 2.5 **The DNA binding behavior of PPARγ** and **RXRα on PPRE, PAL3, or variants thereof.** **A**: Heterodimer formation in the presence of PPRE and PAL3 DNA at different concentrations. **B**: The DNA binding landscape of RXRα monomer to single nucleotide variants of PPRE. The heatmap represents the mean of ddG values (the difference in Gibbs energy of RXRα binding to a mutant site compared to the energy of RXRα binding to canonical PPRE) derived from two independent MITOMI experiments. The sequence of the canonical PPRE is indicated along the x axis. Two core hexamer repeats, constituting the DR1, are highlighted in red. Below heatmap: nergy normalized sequence logo (enoLOGOS) [71] derived from the matrix of the binding energy contribution for each base at each position of PPRE. **C**: Binding affinities of PPARγ or RXRα to DR1 and PAL3 sites or truncated variants thereof. **D**: Same as for B, but for PPARγ instead of RXRα. **E**: Binding affinities of PPARγ to variants of DR1 and PAL3 sites. **F**: Visualization of on-chip assembly of putative PPARγ and RXRα dimers. **G:** DNA binding landscape of PPARγ monomer to PAL3 or single nucleotide variants thereof. Each bar represents the mean and standard deviation of ddG derived from two independent MITOMI experiments. Below heatmap : energy normalized sequence logo [76] derived from the matrix of the binding energy contribution for each base at each position in the PAL3 element.

PPARγ and RXRα bind PPRE in a cooperative fashion

To characterize the biophysical properties of PPARγ:RXRα binding to DNA, we implemented a similar approach as the one used for characterizing monomeric TF DNA binding. We measured the DNA occupancies of PPARγ:RXRα on each sequence belonging to the PPRE single base substitution library and derived equilibrium binding curves of the heterodimer with respect to different variants of the PPRE. However, a putatively confounding factor which may skew the quantification of heterodimer-bound DNA is the ability of RXRα to bind DNA as a homodimer [70] that can compete with the heterodimer PPARγ:RXRα for binding to PPRE (Figure 2.3A, step 3). To eliminate or at least reduce this bias, we opted to perform DNA binding experiments in which GFP-tagged PPARγ and not RXRα is immobilized on the surface of the chip such that mCherry-tagged RXRα is present at the “detection” area under the MITOMI button only when bound to PPARγ (Figure 2.3A). Nevertheless, we measured PPARγ:RXRα DNA binding in the two configurations (in which either PPARγ or RXRα is immobilized on chip) and obtained highly correlated relative affinity values ($R^2 = 0.84$) for heterodimer binding to each PPRE mutant, suggesting that the order bias may not be as large as initially hypothesized.

We first applied simple one-site equilibrium models for DNA binding [54], [77] to describe the heterodimer-DNA interactions, but these models failed to explain the MITOMI binding data of the PPARγ:RXRα heterodimer to PPRE and variants thereof (Figure 2.6A). Specifically, the experimental binding curves exhibited distinct behavioral modes depending on the composition of the DNA target site. The majority of the binding curves exhibited sigmoidal behavior suggesting that PPARγ and RXRα bind DNA in a cooperative manner (Figure 2.6A). Interestingly, certain substitutions within the AGGTCA repeat significantly affected the shape of the binding curves. For example, upon substitution of the guanines in the AGGTCA core, the DNA binding curve of the dimer did not display a sigmoidal behavior; rather, it followed the shape of a hyperbolic function that typically characterizes one-site binding curves (Figure 2.6A).

Next, we asked how much the DNA binding behavior of the heterodimer depends on the abundance of PPARγ given that we previously showed that RXRα is 4-5-fold more abundant than PPARγ in terms of nuclear protein copies in adipocytes [78]. To address this question, we analyzed binding of PPARγ:RXRα to several PPREs in the presence of different DNA and protein concentrations. We then represented the data obtained for each sequence as a three-dimensional scatter plot in which the DNA and PPARγ concentrations were projected onto the x and y axis, respectively, and the amount of DNA bound to an immobilized heterodimer on the z axis (Figure 2.6B). We observed that the DNA binding occupancy of the heterodimer depends both on the DNA concentration and on the concentration of PPARγ. Collectively, these observations led us to hypothesize that DNA binding of the PPARγ:RXRα heterodimer is achieved through a complex cooperative mechanism clarifying why standard equilibrium binding models may be inadequate to define the binding parameters of PPARγ:RXRα-DNA interactions.

Mechanistic model of cooperative PPARγ:RXRα DNA binding

We next asked whether the DNA binding behavior of the heterodimer could be explained by a single model of PPARγ:RXRα DNA binding based only on the knowledge of binding constants between each of the binding partners and PPRE. To address this question, we used the mass action reversible forms that were previously shown to mechanistically explain the binding of regulatory proteins to DNA [79]. As a first step, we described all the elementary reactions in the PPARγ:RXRα-PPRE binding process and generated the mass balance equations that describe the formation of the binding species (Figure 2.6C). Then, we used the knowledge on the energies of TF binding to DNA as single units as well as the energy of TF-TF interactions from the independent experiments introduced above to define corresponding parameters of the model. Solving the obtained mass balance equations for equilibrium binding, we estimated the affinity constants of ternary complexes to each PPRE mutant based on the best model fits to our data (Figure 2.6C).

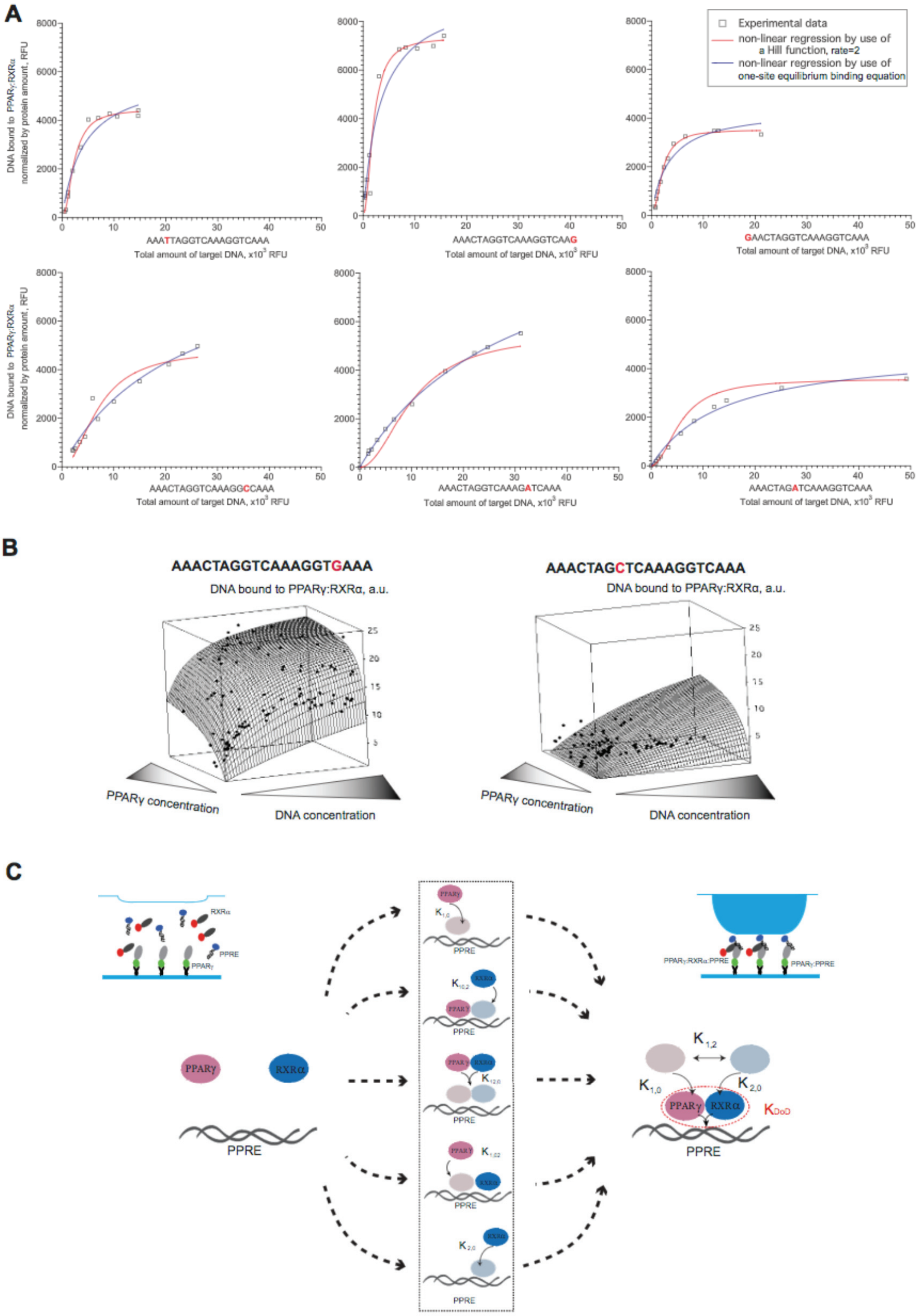


Figure 2.6 **Cooperative TF-DNA interactions. A**: Examples of binding curves representing PPARg:RXRα binding to PPRE or variants thereof. The nucleotide that was substituted in each sampled sequence is highlighted in red. **B**: Binding of the PPARγ:RXRα heterodimer to the DR1 element in function of different DNA and PPARγ concentrations. One example of respectively a strongly (left) and weakly (right) bound sequence is shown. Raw experimental data are represented by black dots and the surface plot represents the regression of the data using Voronoi interpolation. The amount of bound DNA is expressed in arbitrary units (a.u.). **C**: Schematic representation of various scenarios of heterodimer formation. We allow the heterodimer to be formed through either the monomer or dimer scenarios.

To determine the significance of cooperative effects in PPARγ:RXRα-PPRE binding, we quantified the cooperativity factor ω [43] of PPARγ:RXRα binding to each PPRE variant, which allowed us to profile the whole spectrum of cooperativity constant values within the PPRE mutant library (Figure 2.7A). We found that ω is much greater than 1 ($\omega >> 1$) for all tested sequences (Figure 2.7A). We also observed that single nucleotide changes within the PPRE do not equally affect the ability of the heterodimer to cooperate on the respective site. Specifically, we found that nucleotide changes in the first AGGTCA half-site tend to have a greater impact on ω (i.e. for the majority of nucleotide substitutions at PPRE positions 1-11, the value of $\omega_{1,2}$ varies more than for changes in the second half-site) (Figure 2.7A). As indicated above, this upstream PPRE region is bound by PPARγ through DNA binding domain-DNA contacts that are additionally stabilized by the interaction of a hinge region of the protein with a minor groove at the 5'-end of PPRE [75]. Thus, PPARγ does not only contribute to the specificity of the heterodimer, but our data indicate that it may also modulate the extent of cooperativity with RXRα on its target DNA sequence.

To investigate whether this cooperativity effect could also be observed when the heterodimer is bound to sites other than PPRE, we revisited our MITOMI data for 192 sequences representing various nuclear receptor response elements. However, for this DNA library, we were not able to directly quantify ω as we only measured relative affinities and did not generate the type of comprehensive binding data that we acquired for our single nucleotide substitution library. To resolve this issue, we estimated ω using the proxy value σ (with $\sigma \sim \omega$), which we defined here as the affinity change upon the addition of heterodimer partner for both PPARγ and RXRα as follows:

$\sigma_{PPAR\gamma \rightarrow PPAR\gamma:RXR\alpha}$ = affinity **PPARγ**:RXRα / affinity RXRα

$\sigma_{RXR\alpha \rightarrow PPAR\gamma:RXR\alpha}$ = affinity PPARγ:**RXRα** / affinity PPARγ

with the TF listed in bold being the one that was tethered to the surface of the MITOMI device.

We investigated the change of σ between different types of binding sites. Because estrogen- and glucocorticoid-response elements and PAL3 are essentially all palindromes separated by one nucleotide and some DR1 sequences are more similar to one another than to others, we first identified the similarity pattern between all 192 sequences. Using the multiple sequence alignment program MAFFT [80], we independently aligned all sequences, identifying 16 distinct target sequence clusters, and plotted the σ values for each of the sequences contained within each cluster (Figure 2.7, B and C). As expected, we found that the distribution of σ values for the majority of sequences is consistent with the clustering pattern. Interestingly, however, we also observed that for some sequence-homologous sites, the affinity of PPARγ to DNA significantly changes upon the presence of RXRα, as exemplified by PPRE-like type binding sites such as AATCTAGGANNNNNGTCA (Figure 2.7B). Similarly, we observed an RXRα affinity increase upon the presence of PPARγ for PPRE-like sites as well as for DR4-like sites (AAACTAGGTCANNNGAGGTCA)(Figure 2.7C). In both of these cases, we found that the affinity change could be different, even between very similar sequences (Figure 2.7, B and C, i.e. red and blue diamonds within the same sequence cluster). This result is in line with our observation described above in that not only the orientation and spacing between the half-sites appears to affect heterodimer-DNA binding cooperativity but also the nucleotide composition of the target sites themselves.

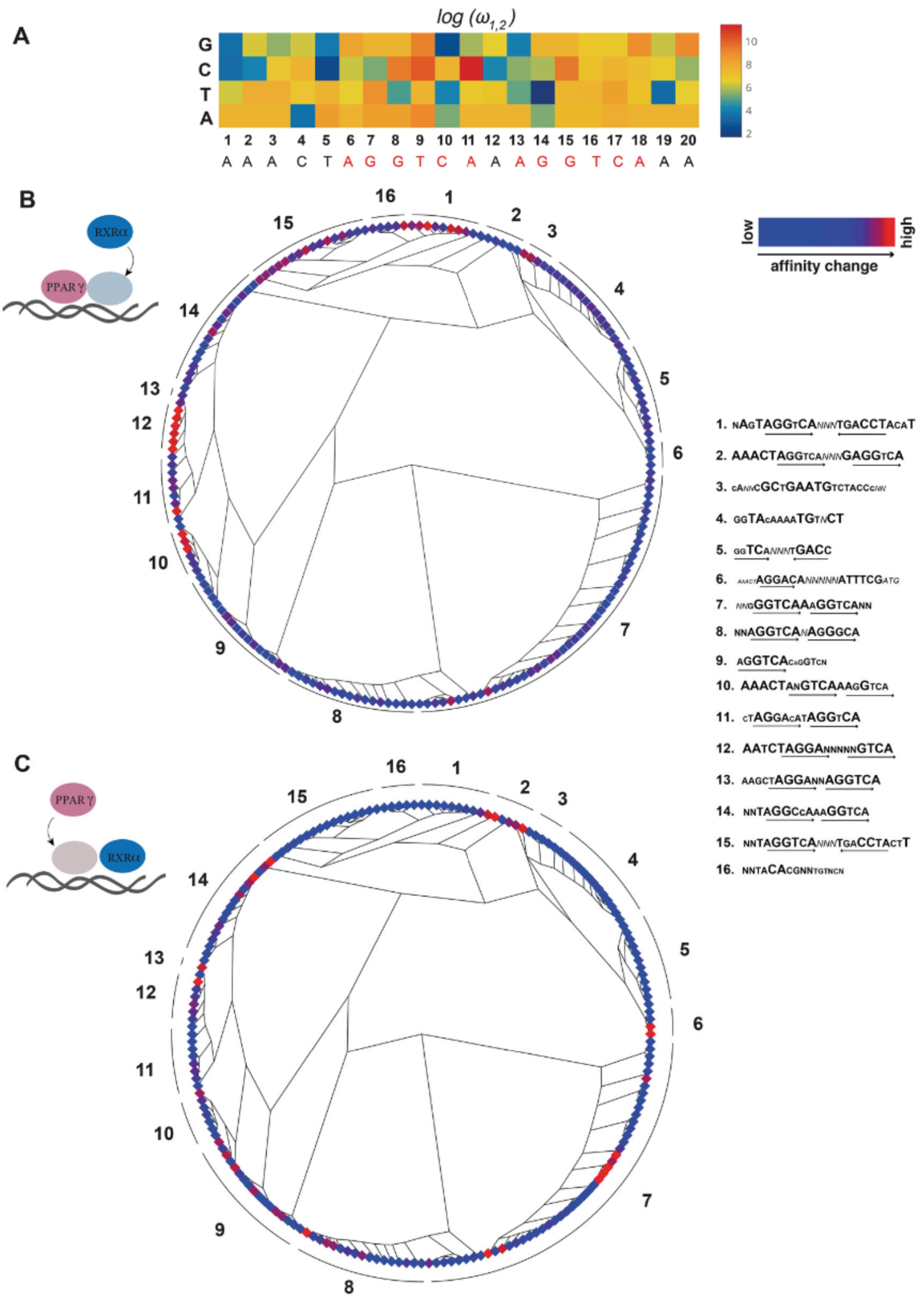


Figure 2.7 **Significance of cooperative effects in PPARγ:RXRα-DNA binding. A**: The cooperativity map represents log $\omega_{1,2}$ values calculated for each PPRE variant. **B**: DNA affinity change (σ) upon PPARγ heterodimerization with RXRα. 192 sequences were clustered using MAFFT, a multiple sequence alignment program, and plotted as a phylotree branching diagram. The representative sequence of each subtree is denoted outside of the tree circle. The values of occupancy change observed for each sequence are plotted as color plots at the terminal nodes of the phylotree. **C**: Same as B, but for RXRα heterodimerization with PPARγ.

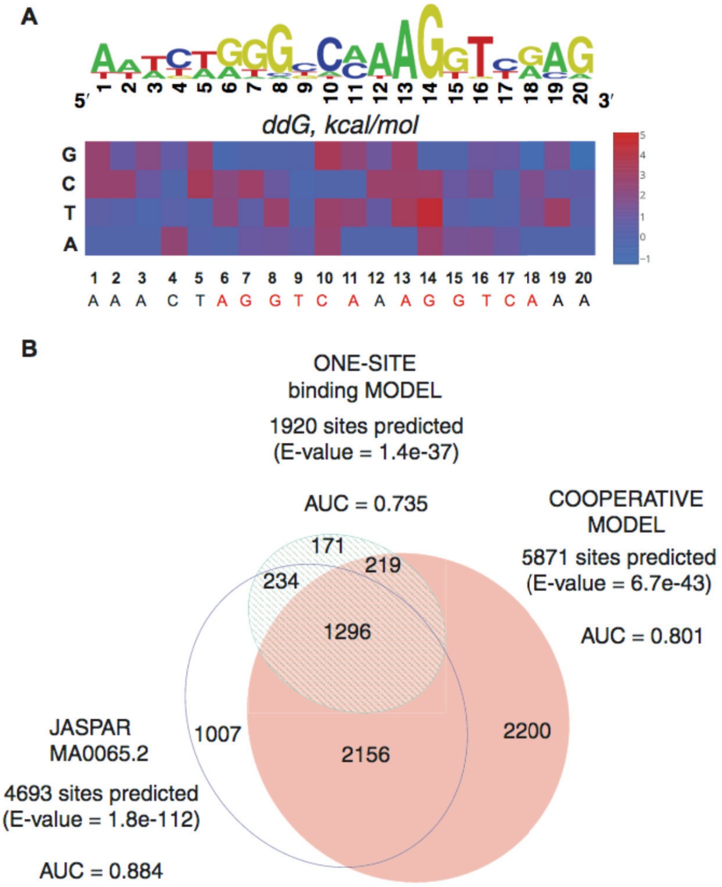


Figure 2.8 **Prediction of in vivo binding. A**: An affinity map as well as the corresponding sequence logo (enoLOGOS) [71] of PPARγ:RXRα heterodimer binding to PPRE. The affinity map represents the K_{DoD} values as calculated based on our cooperativity model. **B**: Venn diagram of the number of PPARγ:RXRα binding sites predicted by three different specificity models independently. The PPARγ:RXRα motif occurrence predicted within 200bp genomic regions identified through ChIP-seq at day 6 of 3T3-L1 adipocyte differentiation.

Apparent DNA binding affinity constant of a heterodimer

The above results emphasize the important role of cooperativity in defining specific heterodimer-DNA binding. To investigate whether incorporating cooperativity into quantitative DNA binding models could enhance the quality of the model and thus improve our ability to predict *in vivo* heterodimer DNA binding, we quantified the cooperativity-inclusive parameters of PPARγ:RXRα-PPRE binding. We defined the affinity of the heterodimer to PPRE through the apparent DNA binding affinity constant of a heterodimer (K_{DoD}) as the product of the binding affinities involved in each of the possible heterodimers on DNA formation pathway, and we estimated the K_{DoD} of PPARγ:RXRα for each single base pair substitution variant of PPRE from the experimental MITOMI data (Figure 2.8A). We next decided to investigate whether the K_{DoD} reflects heterodimer-DNA binding more accurately than a canonical K_d. To address this question, we fitted the experimental data with a one-site binding function, quantified corresponding K_d values and built a position specific scoring matrix of PPARγ:RXRα binding to PPRE (Table 2.2). We then assessed how well either the cooperativity model-based motif (derived from K_{DoD} values) or the motif generated from the one-site binding model (derived from K_d values) predicted *in vivo* PPARγ:RXRα binding in mature 3T3-L1 adipocytes (i.e. day 6 of adipogenesis, the time point of maximal PPARγ binding [61]), using as a reference the JASPAR motif that was derived from the PPARγ:RXRα ChIP-seq data itself. To do so, we computed the occurrence of either of the three motifs within previously published PPARγ:RXRα ChIP-seq data sets [61] and subsequently generated the area under a receiver operating characteristic (area under the receiver operating characteristic curve, a measure for the performance of classification models) scores for each motif [63]. Our results showed that although the JASPAR motif scored best, as expected, our cooperativity model predicts PPARγ:RXRα *in vivo* DNA binding more accurately than the single-site model (area under the receiver characteristic curve of 0.801 compared to 0.731 for the one-site binding model-derived motif and 0.884 for the JASPAR motif) (Figure 2.8B). In line with these results, we also found that the K_{DoD}-based motif predicts a larger number of PPARγ:RXRα ChIP-seq peaks compared to the K_d-based one: 5871 versus 1920 out of 10,114 total peaks (with the JASPAR motif predicting 4693 peaks). To confirm that the peaks predicted by our cooperativity model but not predicted by the JASPAR motif also contained the PPRE motif, we performed a MEME (Multiple Em for Motif Elicitation) [81] motif search on these peaks and identified the canonical AGGTCA repeat separated by one nucleotide as the main enriched motif (data not shown). Together, these results indicate that the accuracy of the specificity model of PPARγ:RXRα DNA binding increases when accounting for the cooperativity effects in heterodimer-DNA binding.

2.4 Discussion

Dimerization is an inherent property of metazoan TFs and plays an important role in transcriptional regulation underlying differential gene expression. Multiple studies showed that dimerization of TFs can influence the proximity and the orientation of the implicated DNA binding domains, and as a consequence, it forces TF complexes to recognize a specific DNA site that is distinct from those recognized by the individual TFs [82]–[85]. It has also been established that during the assembly of a heterodimer on DNA, the monomer-DNA intermediate tends to be kinetically less stable relative to the dimer-DNA complex [86]–[88]. However, none of these studies provided to our knowledge a quantitative link between cooperative dimer-DNA interactions and the respective binding specificity model.

To interrogate the complex DNA binding behavior of heterodimers in a quantitative manner, we implemented in this study a novel integrative framework in which we coupled an in-depth biophysical on-chip characterization of PPARγ:RXRα binding to DNA with *in silico* modeling of the dimer-DNA association process. The highly parallel on-chip measurements thereby allowed us to simultaneously probe the binding of our focal proteins to multiple DNA sites under uniform conditions. This in turn allowed us to directly determine and compare the relative affinities of PPARγ, RXRα, and PPARγ:RXRα to various target sites that have previously been demonstrated to be of great functional importance [67], [89]. These experiments revealed that RXRα binding is constrained to the AGGTCA hexamer such that even a single substitution within this site can cause a significant change in binding energy, consistent with data from previous studies [45], [90]. Because of the sequence symmetry in PPRE, we found that RXRα can bind to either of the two hexameric half-sites (Figure 2.5, B and C). In contrast, PPARγ alone did not have high affinity for

PPRE *in vitro* (Figure 2.5, C and D), but exhibited instead high affinity for the PAL3 element (Figure 2.5, E and G). Our results thereby suggest that PPARγ binds to PAL3 in monomeric rather than the previously proposed dimeric format [91], although further analyses will be required to formally validate this finding. These results raise the question as to why PPARγ is seldom associated with a PAL3 site *in vivo* [61] and why heterodimeric DNA binding by PPARγ and RXRα is preferred over the PPARγ-DNA or RXRα-DNA interactions. This question is especially relevant because the nuclear abundance of RXRα is much greater than that of PPARγ [78], which should theoretically favor the formation of RXRα-DNA complexes. Results from our analyses now indicate that the specificity of the heterodimer, even though somewhat dispersed among different response elements, is different from that identified for each partner independently (Figure 2.4B). We also found that the extent of DNA binding of the heterodimer depends on the concentration of PPARγ and that the two TF partners contribute to the total binding energy of the interaction in a non-linear and non-additive fashion (Figure 2.6, A and B). This significantly influences the shape of experimental binding curves such that it can no longer be explained with simple kinetic models (Figure 2.6A), implying complex cooperative effects between the implicated factors and DNA that may promote heterodimer DNA binding.

To further dissect the nature of these cooperative interactions and to characterize the strength of cooperative heterodimer DNA binding with respect to the composition of the target site, we built a mechanistic model that accounts for all possible intermediate and final complexes that can occur between the three focal components. Mechanistic modeling so far has been widely applied in various studies to describe the kinetics of enzymatic and metabolic pathways [92]–[94] and even to characterize the *lac* operon function in *E. coli* [79]. However, it has to our knowledge so far never been applied to comprehensively interpret high throughput heterodimer-DNA binding data. In contrast to the previously proposed quantitative models [95], the mechanistic approach did not require us to model the binding of a heterodimer to DNA as a one-step event nor to restrain the complex association to follow a monomer or a dimer pathway [86], [96]. Rather, we aimed to account for the cooperative nature of these interactions and determine comprehensive binding parameters (Figure 2.6C, Figure 2.7A, Figure 2.8A). As such, we were able to determine the apparent affinity constant of the heterodimer that does not depend on the order of binding events, providing a novel framework to quantitatively interrogate heterodimer-DNA interactions (Figure 2.6C, Figure 2.8A). Importantly, this affinity constant does account for cooperative heterodimer-DNA binding, which, we showed, increases the *in vivo* DNA binding predictive power of our binding specificity model compared with a regular one-site equilibrium binding model.

Experimental MITOMI data further showed that the extent of cooperative effects in PPARγ:RXRα DNA binding depends on the orientation and nucleotide composition of the target site (Figure 2.7B). Our model revealed that these patterns are associated more with PPARγ DNA binding rather than RXRα DNA interactions. Particularly, nucleotide alterations in the first part of the element resulted in greater variability of the cooperativity constant (as compared with the second part of PPRE) (Figure 2.7A), which serves as the principal PPARγ:DNA binding interface [75]. This observation implies that PPARγ plays an important role in mediating the specificity of the dimer as well as the strength of heterodimer DNA binding to a particular site.

2.5 Conclusion

Our model does not elucidate the molecular origin of cooperativity as it does not distinguish between direct protein-protein interaction effects or indirect effects involving, for example, conformational state changes of implicated molecules [60]. Nevertheless, the observed variability of the derived cooperative parameter ω as well as the K_{DoD} constant reveals the versatile nature of cooperative heterodimer-DNA binding at single base pair resolution. This finding clearly suggests that we need to account for this variation when aiming to accurately model the PPARγ:RXRα-DNA interactions and to subsequently derive a comprehensive specificity matrix. Achieving such a robustness requires a comprehensive training set of input parameters however, which in turn demands a rigorous quantification of the focal molecular interactions (i.e. the binding of each dimer partner to DNA) prior to model simulation. This exposes an important limitation of the utilized mechanistic model in that it requires extensive quantitative binding data to accurately predict the DNA binding behavior of heterodimers. However, given the increasing availability of powerful assays such as MITOMI enabling the systematic analysis of protein-protein and protein-DNA interactions, we think that our modeling approach has great potential to further unravel the complex nature of protein-DNA interactions and go beyond the mere evaluation of binding strength. This may apply not only to heterodimers, but also to even higher order complexes involving allosteric interactions between TFs, co-factors, ligands, and DNA [97], [98]. Nevertheless, despite our advance in deriving a DNA binding affinity constant of a heterodimer based on equilibrium-state measurements, our understanding of the kinetic mechanisms underlying the formation of heterodimers and their stabilization on DNA remains a challenging task. Follow-up studies may in this regard involve real time kinetic analyses of heterodimer-DNA complex formation for which the presented equilibrium binding data should prove highly valuable.

Table 2.1 **Model input parameters, K_{DoD} and cooperativity parameters of PPARγ:RXRα heterodimer binding to PPRE as estimated by the cooperativity model**

PPRE Sequence	$K_{00,10}$ M^{-1}	σ ($K_{00,10}$) M^{-1}	$K_{00,20}$ M^{-1}	σ ($K_{00,20}$) M^{-1}	$\omega_{1,2}$	$\omega_{2,2}$	K_{DoD} M^{-2}
AAACTAGGTCAAAGGTCAAA	2.08E+06	1.20E+05	1.98E+08	1.01E+07	3.56E+07	8.04E+07	1.46E+22
AAACTAGGTCAAAGGTCATA	1.23E+06	1.32E+05	3.78E+07	4.35E+06	2.23E+03	2.06E+04	1.04E+17
AAACTAGGTCAAAGGTCACA	2.33E+06	1.34E+05	1.22E+08	3.19E+07	4.59E+06	4.77E+06	1.31E+21
AAACTAGGTCAAAGGTCAGA	1.53E+06	7.23E+04	1.10E+07	2.10E+06	1.33E+06	1.08E+07	2.23E+19
AAACTAGGTCAAAGGTCAAT	2.38E+06	1.64E+05	2.79E+08	6.64E+07	3.80E+06	1.40E+05	2.52E+21
AAACTAGGTCAAAGGTCAAC	2.55E+06	2.78E+05	9.97E+08	3.88E+07	3.94E+05	1.50E+03	1.00E+21
AAACTAGGTCAAAGGTCAAG	1.65E+06	4.16E+04	1.00E+09	1.10E+08	1.01E+09	9.38E+06	1.67E+24
TAACTAGGTCAAAGGTCAAA	8.46E+06	2.43E+05	1.01E+08	1.91E+07	1.76E+06	1.17E+07	1.51E+21
CAACTAGGTCAAAGGTCAAA	1.21E+06	9.25E+04	1.17E+08	1.97E+07	2.80E+03	9.94E+03	3.95E+17
GAACTAGGTCAAAGGTCAAA	1.19E+06	1.84E+05	1.00E+08	1.16E+07	3.28E+03	1.57E+04	3.92E+17
ATACTAGGTCAAAGGTCAAA	5.83E+05	5.69E+04	1.45E+08	4.33E+07	6.04E+07	1.07E+08	5.12E+21
ACACTAGGTCAAAGGTCAAA	7.23E+05	4.34E+04	4.33E+07	4.92E+06	1.20E+04	5.95E+04	3.77E+17
AGACTAGGTCAAAGGTCAAA	8.27E+05	4.64E+04	2.05E+08	1.38E+07	2.03E+06	2.91E+06	3.45E+20
AATCTAGGTCAAAGGTCAAA	2.75E+06	2.28E+05	1.98E+08	5.25E+07	7.01E+07	1.52E+08	3.81E+22
AACCTAGGTCAAAGGTCAAA	2.86E+05	2.28E+05	1.25E+08	7.54E+07	7.10E+06	2.96E+06	2.53E+20
AAGCTAGGTCAAAGGTCAAA	2.46E+05	5.72E+03	5.27E+07	6.39E+06	3.26E+05	1.74E+05	4.22E+18
AAAATAGGTCAAAGGTCAAA	1.05E+06	1.06E+05	1.64E+08	1.52E+07	4.03E+03	6.25E+03	6.96E+17
AAATTAGGTCAAAGGTCAAA	1.70E+06	3.73E+05	2.59E+08	1.86E+07	7.34E+06	5.51E+06	3.22E+21
AAAGTAGGTCAAAGGTCAAA	3.52E+06	3.05E+05	1.38E+08	2.54E+07	1.61E+06	9.98E+06	7.84E+20
AAACAAGGTCAAAGGTCAAA	3.60E+05	2.76E+04	7.87E+07	2.41E+07	2.29E+08	2.13E+08	6.48E+21
AAACCAGGTCAAAGGTCAAA	3.82E+05	5.64E+04	3.82E+08	1.29E+07	2.13E+02	2.66E+02	3.12E+16
AAACGAGGTCAAAGGTCAAA	2.96E+05	8.56E+04	9.82E+07	7.68E+06	6.98E+03	1.83E+04	2.03E+17
AAACTTGGTCAAAGGTCAAA	1.16E+05	6.46E+03	4.68E+06	2.16E+05	3.58E+06	3.21E+07	1.95E+18
AAACTCGGTCAAAGGTCAAA	2.15E+05	5.91E+03	8.80E+06	7.57E+05	1.49E+06	1.04E+07	2.82E+18
AAACTGGGTCAAAGGTCAAA	2.27E+06	1.84E+05	1.73E+08	1.30E+07	1.40E+08	4.79E+08	5.48E+22
AAACTAAGTCAAAGGTCAAA	1.59E+05	5.62E+04	7.36E+06	9.64E+05	2.05E+08	1.26E+09	2.40E+20
AAACTACGTCAAAGGTCAAA	9.44E+04	3.32E+04	6.66E+06	1.54E+06	2.07E+05	3.43E+05	1.30E+17
AAACTATGTCAAAGGTCAAA	2.68E+05	1.26E+04	9.75E+06	6.34E+05	5.57E+08	1.86E+09	1.46E+21
AAACTAGATCAAAGGTCAAA	1.84E+05	4.90E+04	4.91E+06	4.72E+05	1.62E+08	1.10E+09	1.46E+20
AAACTAGCTCAAAGGTCAAA	1.06E+05	5.47E+04	9.80E+05	2.83E+04	2.04E+09	3.62E+10	2.13E+20
AAACTAGTTCAAAGGTCAAA	1.23E+05	3.22E+04	2.54E+07	3.83E+06	7.18E+04	6.85E+04	2.25E+17
AAACTAGGACAAAGGTCAAA	3.79E+05	2.29E+04	2.04E+06	1.41E+05	1.03E+09	9.14E+09	7.92E+20
AAACTAGGCCAAAGGTCAAA	6.22E+05	6.69E+04	5.86E+06	6.23E+05	1.21E+10	4.02E+11	4.41E+22
AAACTAGGGCAAAGGTCAAA	6.86E+05	1.02E+05	9.19E+06	8.21E+05	1.51E+09	1.13E+10	9.54E+21
AAACTAGGTAAAAGGTCAAA	5.15E+05	3.40E+04	3.42E+06	2.18E+05	2.01E+05	1.11E+07	3.55E+17

AAACTAGGTTAAAGGTCAAA	7.16E+05	9.70E+05	1.70E+07	1.34E+06	1.39E+04	7.87E+04	1.69E+17
AAACTAGGTGAAAGGTCAAA	2.23E+06	1.12E+05	2.66E+07	5.23E+06	3.12E+02	1.08E+03	1.85E+16
AAACTAGGTCTAAGGTCAAA	1.74E+05	4.81E+04	8.02E+05	2.72E+04	5.68E+06	5.28E+07	7.91E+17
AAACTAGGTCCAAGGTCAAA	1.44E+05	3.15E+04	3.68E+05	2.58E+04	2.40E+11	1.15E+13	1.27E+22
AAACTAGGTCGAAGGTCAAA	3.94E+05	7.31E+04	5.54E+06	1.42E+05	5.41E+05	1.50E+06	1.18E+18
AAACTAGGTCACAGGTCAAA	1.67E+06	6.23E+05	7.95E+06	4.61E+05	1.53E+04	2.94E+05	2.02E+17
AAACTAGGTCATAGGTCAAA	1.06E+06	2.59E+05	4.38E+07	5.05E+06	8.61E+06	9.90E+06	3.98E+20
AAACTAGGTCAGAGGTCAAA	1.27E+06	1.13E+05	2.69E+07	2.69E+06	3.40E+06	1.13E+07	1.16E+20
AAACTAGGTCAATGGTCAAA	7.91E+05	1.24E+05	3.98E+05	3.08E+04	1.15E+05	3.26E+07	3.63E+16
AAACTAGGTCAACGGTCAAA	6.16E+05	1.61E+05	1.05E+06	3.99E+04	2.93E+05	9.40E+07	1.90E+17
AAACTAGGTCAAGGGTCAAA	8.38E+05	2.25E+05	1.61E+07	1.68E+06	8.73E+03	5.58E+04	1.18E+17
AAACTAGGTCAAACGTCAAA	4.86E+05	1.44E+05	8.60E+05	4.27E+04	6.65E+05	1.34E+08	2.78E+17
AAACTAGGTCAAAAGTCAAA	4.99E+05	1.15E+05	2.41E+06	1.70E+05	2.19E+05	7.83E+06	2.63E+17
AAACTAGGTCAAATGTCAAA	1.32E+06	2.81E+05	5.14E+06	3.14E+05	4.26E+01	8.63E+02	2.88E+14
AAACTAGGTCAAAGCTCAAA	4.51E+05	1.34E+05	2.18E+05	5.69E+03	2.44E+09	1.47E+10	2.40E+20
AAACTAGGTCAAAGATCAAA	5.93E+05	3.42E+04	2.73E+06	1.58E+05	2.61E+07	1.86E+07	4.23E+19
AAACTAGGTCAAAGTTCAAA	1.57E+06	6.33E+04	4.44E+07	2.30E+06	3.08E+07	1.01E+06	2.14E+21
AAACTAGGTCAAAGGACAAA	9.98E+05	8.48E+04	9.06E+05	3.85E+04	2.45E+07	4.33E+07	2.22E+19
AAACTAGGTCAAAGGCCAAA	6.04E+05	1.42E+05	3.59E+06	2.31E+05	1.15E+07	7.53E+05	2.50E+19
AAACTAGGTCAAAGGGCAAA	1.57E+06	1.20E+05	1.33E+07	8.40E+05	6.73E+06	7.41E+05	1.40E+20
AAACTAGGTCAAAGGTGAAA	3.08E+06	7.80E+04	1.32E+07	5.04E+05	6.25E+06	8.50E+05	2.55E+20
AAACTAGGTCAAAGGTAAAA	1.09E+06	9.83E+04	4.39E+06	3.43E+05	6.82E+07	1.70E+08	3.28E+20
AAACTAGGTCAAAGGTTAAA	2.12E+06	1.59E+05	1.79E+07	2.14E+06	1.85E+08	7.64E+07	7.05E+21
AAACTAGGTCAAAGGTCTAA	8.43E+05	1.23E+05	4.55E+06	2.34E+05	1.20E+07	5.38E+07	4.61E+19
AAACTAGGTCAAAGGTCCAA	7.31E+05	3.54E+04	1.56E+06	8.43E+04	1.39E+07	5.22E+07	1.58E+19
AAACTAGGTCAAAGGTCGAA	1.55E+06	1.13E+05	2.70E+08	8.29E+07	6.63E+08	4.54E+07	2.79E+23

PPRE Sequence: Mutation library of the PPRE element

$K_{00,10}$: binding affinity of PPARg to PPRE

$K_{00,20}$: binding affinity of RxRa to PPRE

K_{D1}: PPARg-RxRa proteins affinity

K_{D2}: RxRa-RxRa proteins affinity

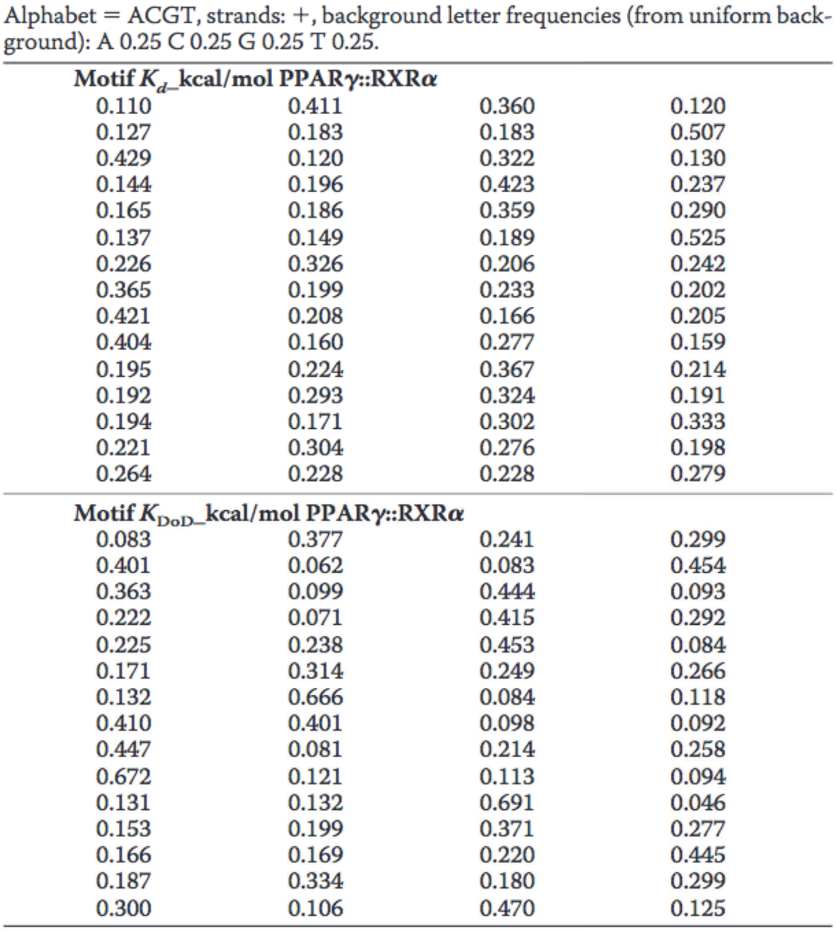
$\omega_{1,2}$: PPARg-RxRa cooperativity

$\omega_{2,2}$: RxRa-RxRa cooperativity

The effect of the mutations on the binding affinity has been independently measured for the two proteins and inserted in the model as input parameter.

$K_{D1} \pm \sigma$, M^{-1}	$K_{D2} \pm \sigma$, M^{-1}
$(6.13 \pm 0.13) \cdot 10^7$	$(5.13 \pm 0.09) \cdot 10^7$

Table 2.2 **Position specific scoring matrices of PPARγ:RxRα towards PPRE.**



Alphabet = ACGT, strands: +, background letter frequencies (from uniform background): A 0.25 C 0.25 G 0.25 T 0.25.

Motif K_d_kcal/mol PPARγ::RXRα			
0.110	0.411	0.360	0.120
0.127	0.183	0.183	0.507
0.429	0.120	0.322	0.130
0.144	0.196	0.423	0.237
0.165	0.186	0.359	0.290
0.137	0.149	0.189	0.525
0.226	0.326	0.206	0.242
0.365	0.199	0.233	0.202
0.421	0.208	0.166	0.205
0.404	0.160	0.277	0.159
0.195	0.224	0.367	0.214
0.192	0.293	0.324	0.191
0.194	0.171	0.302	0.333
0.221	0.304	0.276	0.198
0.264	0.228	0.228	0.279
Motif K_{DoD}_kcal/mol PPARγ::RXRα			
0.083	0.377	0.241	0.299
0.401	0.062	0.083	0.454
0.363	0.099	0.444	0.093
0.222	0.071	0.415	0.292
0.225	0.238	0.453	0.084
0.171	0.314	0.249	0.266
0.132	0.666	0.084	0.118
0.410	0.401	0.098	0.092
0.447	0.081	0.214	0.258
0.672	0.121	0.113	0.094
0.131	0.132	0.691	0.046
0.153	0.199	0.371	0.277
0.166	0.169	0.220	0.445
0.187	0.334	0.180	0.299
0.300	0.106	0.470	0.125

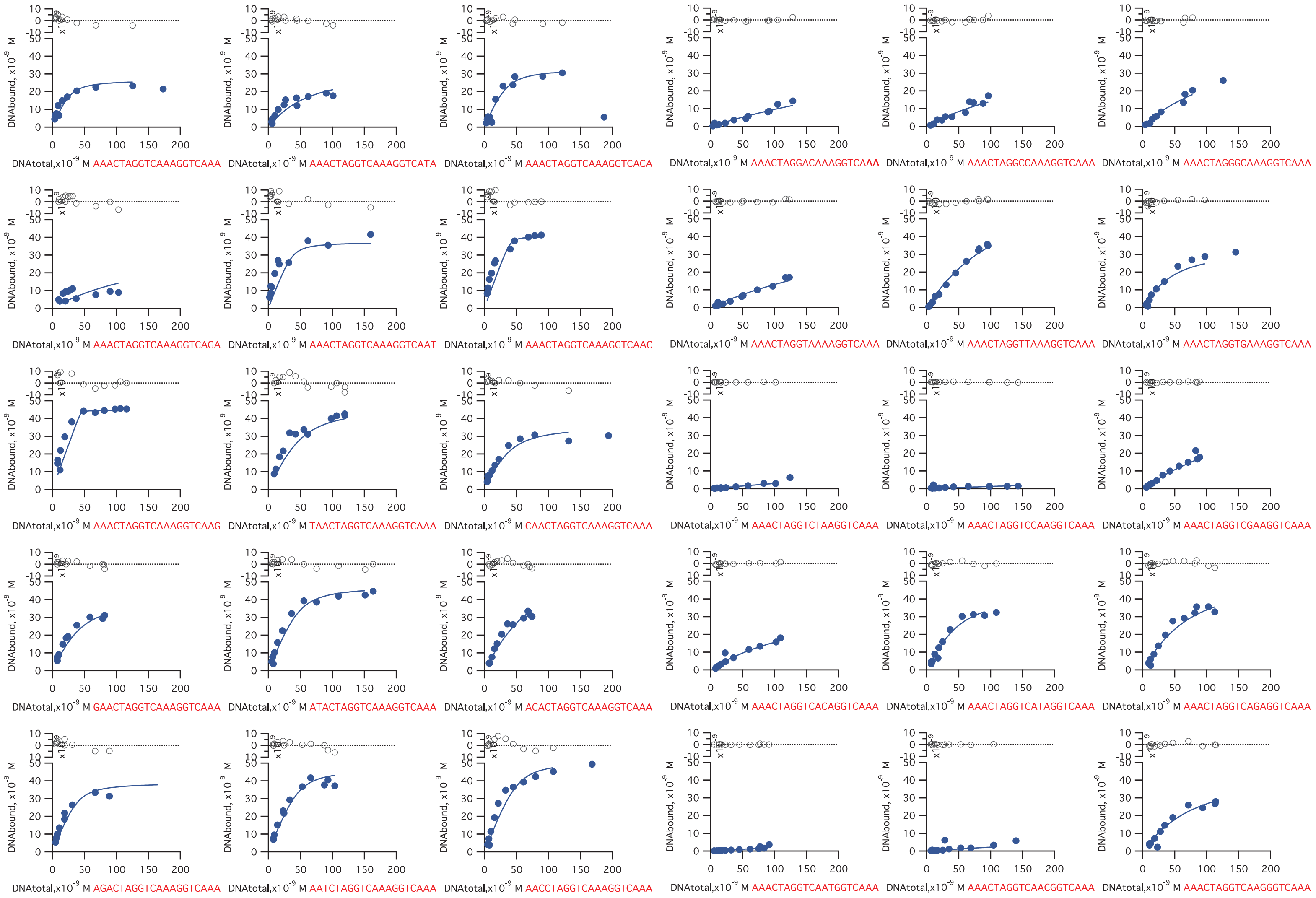


Figure 2.9 **Model fits of RXRα-PPRE experimental binding curves**. The mechanistic model is streamlined for monomeric TF-DNA binding in this case and thus restrained to consider RXRα, PPRE, RXRα-PPRE interactions. Residuals, calculated for each sequence, are represented on each plot above the respective sequence fit and are randomly scattered around zero, indicating an accurate model fit.

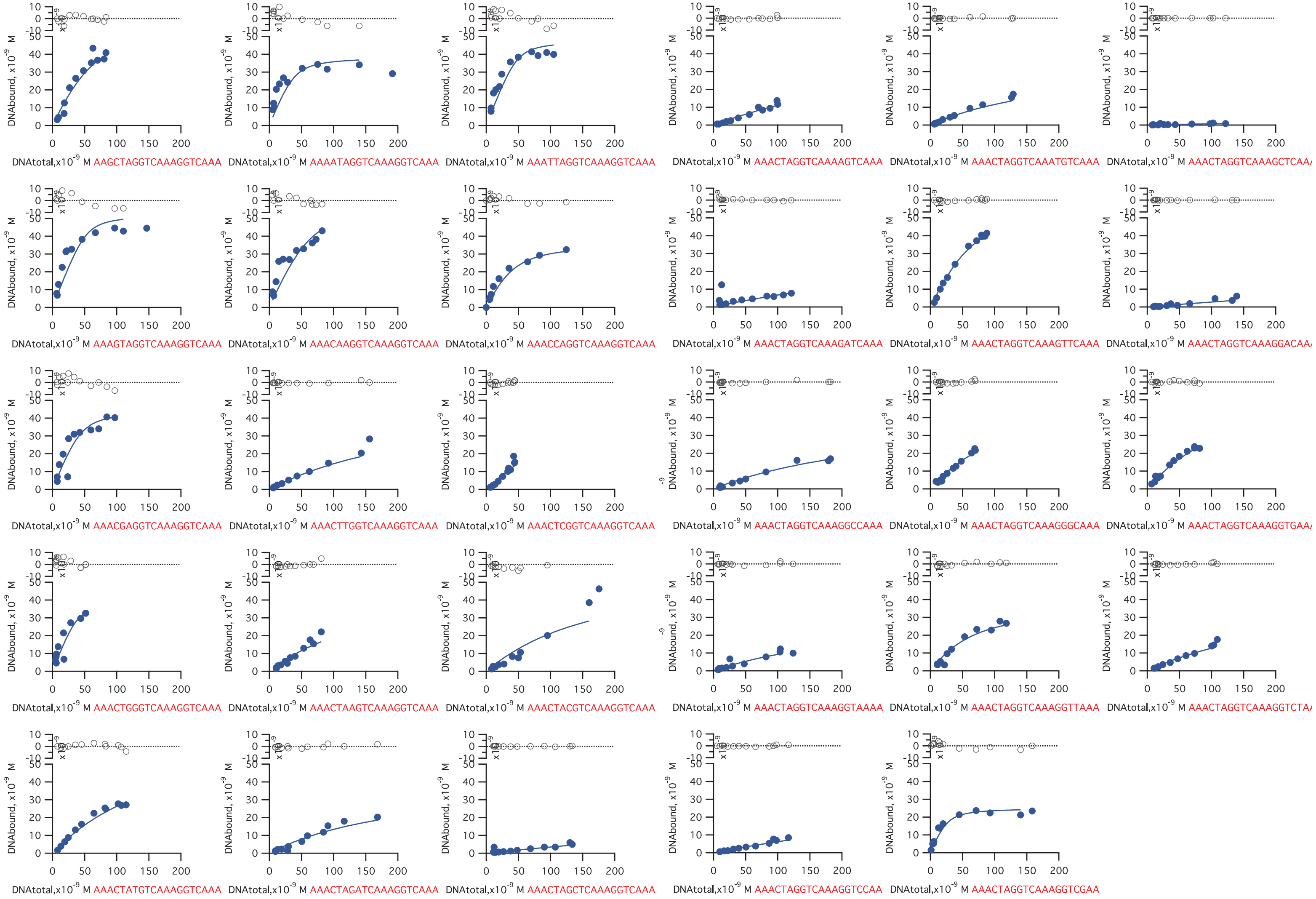


Figure 2.9 (continued) **Model fits of RXRα-PPRE experimental binding curves**. The mechanistic model is streamlined for monomeric TF-DNA binding in this case and thus restrained to consider RXRα, PPRE, RXRα-PPRE interactions. Residuals, calculated for each sequence, are represented on each plot above the respective sequence fit and are randomly scattered around zero, indicating an accurate model fit.

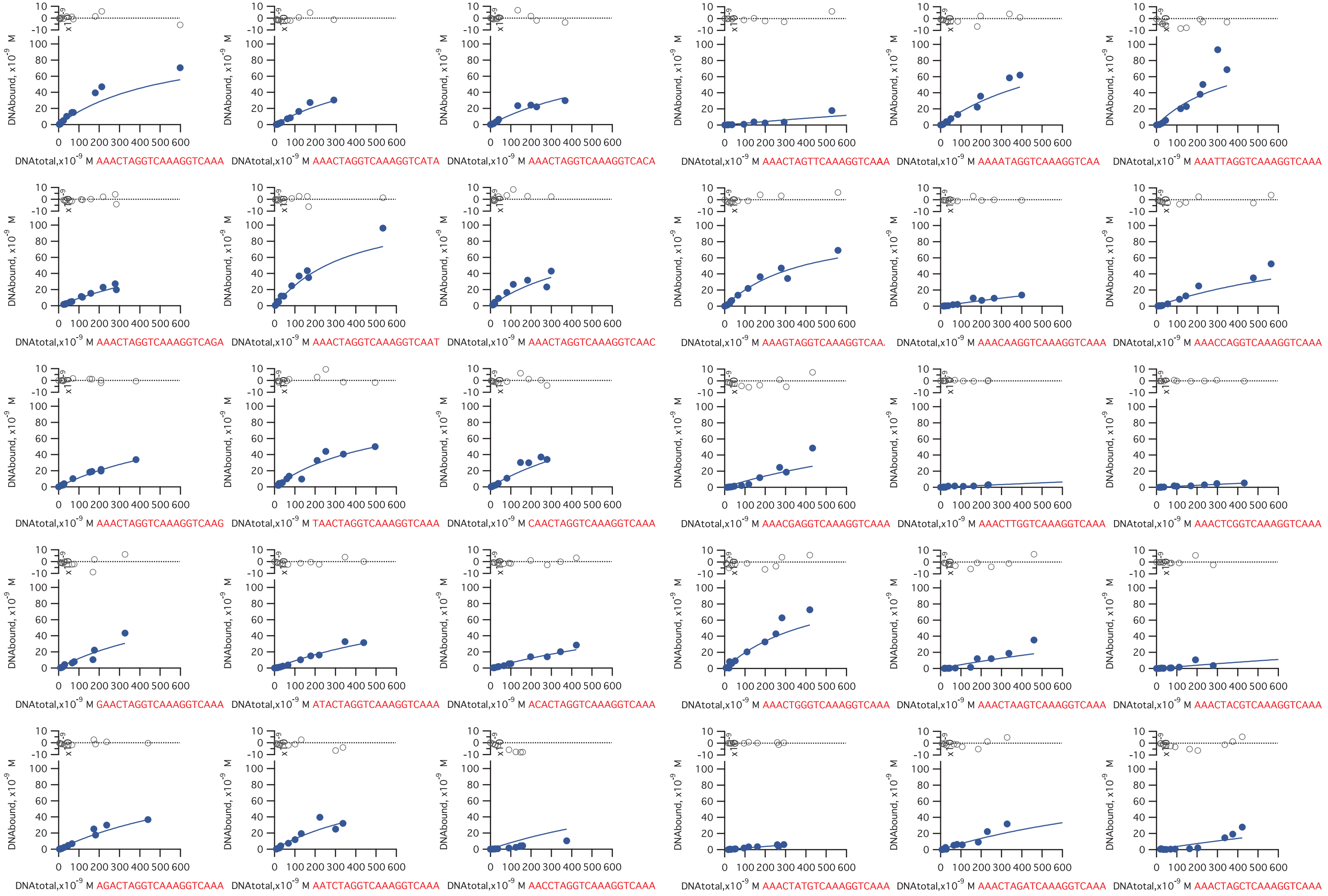


Figure 2.10 **Same as Figure 2.9 but for PPARγ.**

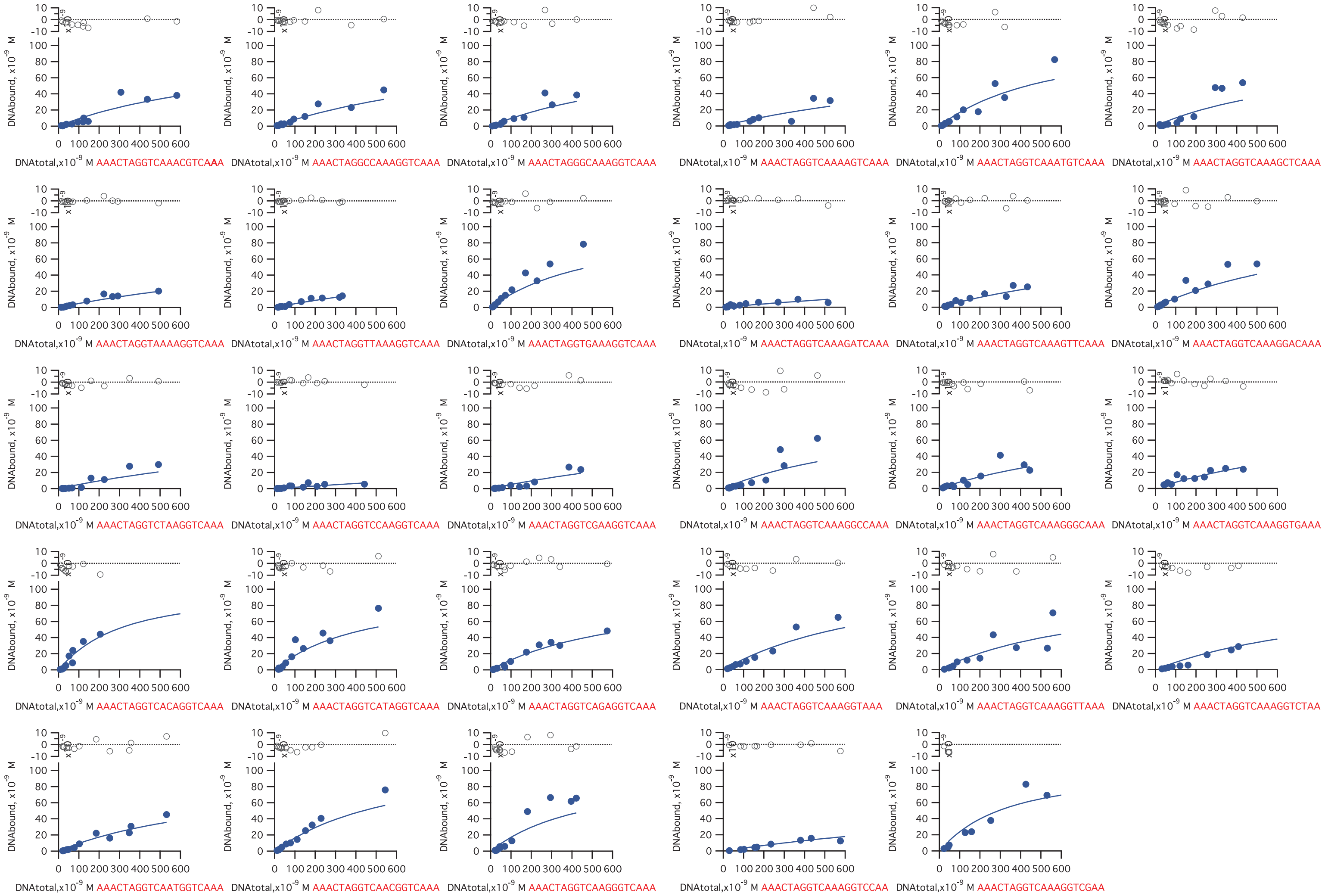


Figure 2.10 (continued) **Same as Figure 2.9 but for PPARγ.**

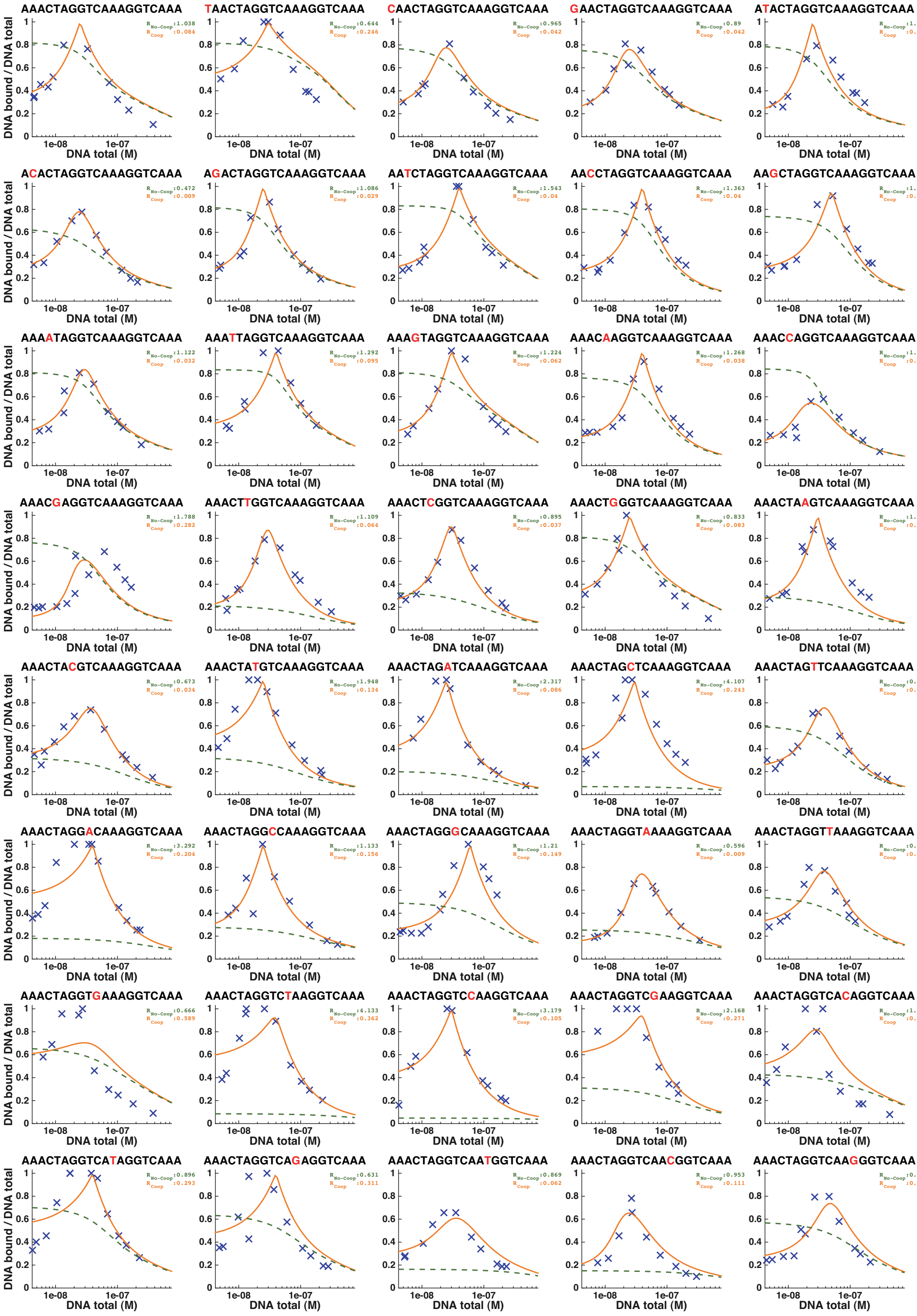


Figure 2.11 **Performance of the mechanistic model solved for equilibrium**. Examples of experimental data corresponding to each tested PPRE variant and corresponding binding curves as predicted by the model when ether accounting for cooperativity (red curves) or not (green curves).

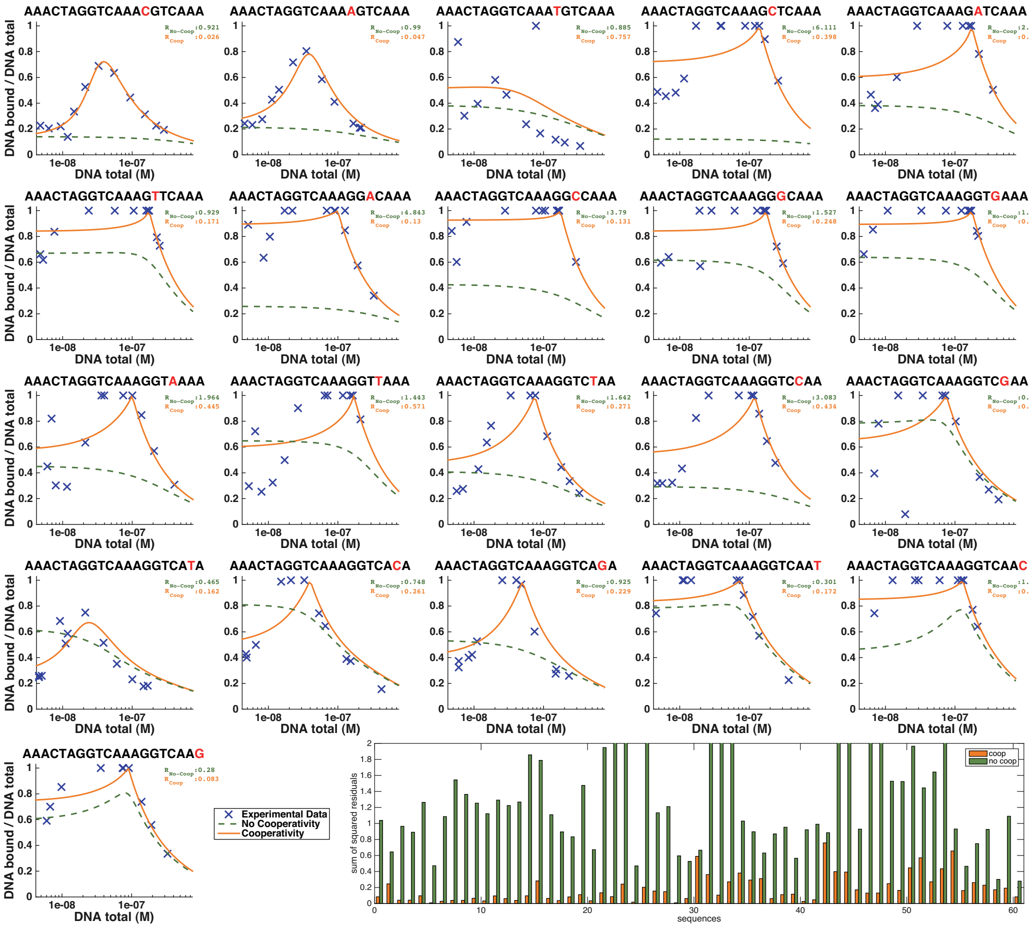


Figure 2.11 (continued) **Performance of the mechanistic model solved for equilibrium**. Examples of experimental data corresponding to each tested PPRE variant and corresponding binding curves as predicted by the model when ether accounting for cooperativity (red curves) or not (green curves). The sum of squared residuals of both cooperativity and “no cooperativity” model fits are indicated for each PPRE mutant (in red and green respectively) and plotted for each sequence as a bar plot at the bottom of the figure for a direct comparison.

Chapter 3 Model of gene regulatory circuits for arsenic bioreporters

Reprinted (adapted) with permission from (Berset, Yves, Davide Merulla, Aurélie Joublin, Vassily Hatzimanikatis, and Jan R. van der Meer. "Mechanistic Modeling of Genetic Circuits for ArsR Arsenic Regulation." *ACS Synthetic Biology* 6, no. 5 (May 19, 2017): 862–74. doi:10.1021/acssynbio.6b00364.). Copyright (2017) American Chemical Society.

Author contributions: Y.B. and JR. vdM. designed the study and wrote the paper, D.M. and JR. vdM. Designed the bioreporter strains, D.M. and A.J. performed the assays, Y.B. and V.H. performed mechanistic modeling.

Abstract

Bioreporters are living cells that generate an easily measurable signal in the presence of a chemical compound. They acquire their functionality from synthetic gene circuits, whose configuration defines the response signal and signal-to-noise ratio. Bioreporters based on the *E. coli* ArsR system have raised significant interest for quantifying arsenic pollution, but they need to be carefully optimized to accurately work in the required low concentration range (1–10 µg arsenite L^{-1}). In order to better understand the general functioning of ArsR-based genetic circuits, we developed a comprehensive mechanistic model that was empirically tested and validated in *E. coli* carrying different circuit configurations. The model accounts for the different elements in the circuits (proteins, DNA, chemical species), and their detailed affinities and interactions, and predicts the (fluorescent) output from the bioreporter cell as a function of arsenite concentration. The model was parametrized using existing ArsR biochemical data, and then complemented by parameter estimations from the accompanying experimental data using a scatter search algorithm. Model predictions and experimental data were largely coherent for feedback and uncoupled circuit configurations, different ArsR alleles, promoter strengths and presence or absence of arsenic efflux in the bioreporters. Interestingly, the model predicted a particular useful circuit variant having steeper response at low arsenite concentrations, which was experimentally confirmed and may be useful as arsenic bioreporter in the field. From the extensive validation, we expect the mechanistic model to further be a useful framework for detailed modeling of other synthetic circuits.

3.1 Introduction

One of the immediate application areas for synthetic biology are bioreporters, living cells with simple designed genetic circuits that permit detection of a specific chemical or group of chemicals, under the concomitant production of an easily but accurately quantifiable reporter signal [76]–[79]. Bioreporters have attracted considerable interest because they offer cheap alternatives for chemical analysis in remote areas where high-end instruments are unavailable [103] and can potentially be embedded in automated microfluidics devices [104]–[107]. Genetic circuits of interest for bioreporters consist of a limited number of interacting elements, which are placed in a specific DNA configuration to obtain a functioning circuit with the desired output [102]. Notably, bioreporter designs require one or more elements that can act as the primary cellular sensor for the target (for example, a transcription factor), and a series of relays to transmit the sensory perception to an actuator protein, which produces the signal to be measured (Figure 3.1). Numerous examples of reporter circuits have been produced, some more based on trial-and-error approaches [108], others on high-throughput screenings [109] or on combinations of logic gate modeling and subsequent experimental testing [110].

Synthetic biology approaches can become particularly powerful when experimental trials are combined with computational methods that can explain observations and correctly predict trends for variants or variant conditions that cannot all be tested experimentally. Computational methods base on model conceptions, which surprisingly enough, vary widely even for simple gene reporter circuits. As examples, circuit designs involving elements such as illustrated in Figure 3.1, can be conceptualized using Boolean logic gates, or with continuous and stochastic models, among others. Boolean models provide a qualitative analysis of gene regulatory networks, where genes and proteins are in an active or inactive state depending on their associated Boolean functions. Boolean logic gates can be used to automatize the design of gene regulatory networks [110]. Continuous models are ordinary differential equations that represent exact, time dependent molecular concentrations, which can be compared with experimental results. Such models are used for parameter estimation [111] or for network inference [112]. Stochastic models operate at single molecule level and generate intrinsic noise. At low copy numbers, which is typical of gene regulatory networks, stochastic models increase the network variability and can lead to bimodal distributions [113]. Finally, mechanistic models can be developed, which are based on assumptions of true molecular interactions between the circuit elements and cellular components, in contrast to most of the other models that make some *ad hoc* assumptions for model reduction and simplifications. In pioneering work, Lee and Bailey [114] developed a mechanistic model for the *lac* operon, which was later adapted to explain the tunable response of bioreporters based on uncoupled circuits [23]. In mechanistic models, parameters relate to mechanistic steps and are quantitatively relevant, whether they are incorporated from experimental data or inferred from parameter estimation.

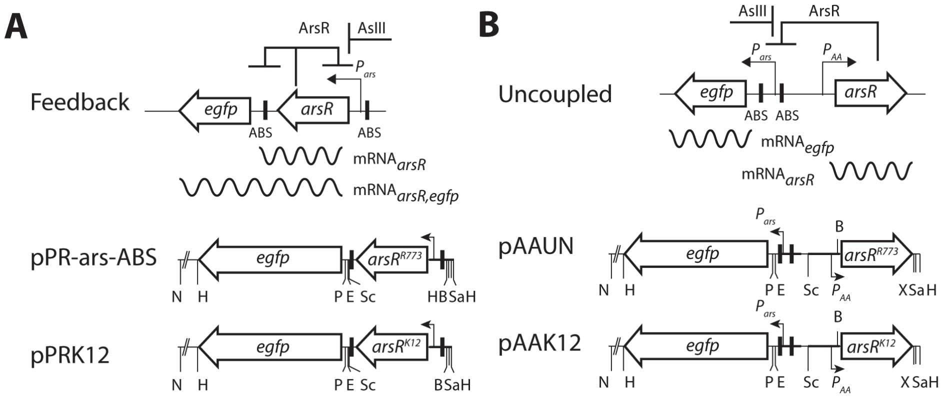


Figure 3.1 **Constructs and configurations of the arsenite bioreporter circuits**. **A**: Feedback configuration, with *arsR* and *egfp* transcription under control of ArsR from the *ars*-promoter. Derepression in presence of arsenite (As_{III}). Note the formation of two mRNAs, depending on the occupancy of the ArsR operator (ABS) in between *arsR* and *egfp*. **B**: Uncoupled configuration, with *arsR* transcription controlled by the constitutive P_{AA} promoter and *egfp* under control of the *ars*-promoter, ArsR and arsenite. Note the different *arsR* alleles in pPR-ars-ABS/pAAUN versus pPRK12/pAAK12. N, NheI; H, HindIII; P, PseI; E, EcoRI; Sc, ScaI; X, XhoI; Sa, SalI; B, BamHI.

Here we focus on building an accurate conceptual mechanistic model of a genetic circuit based on the ArsR regulatory protein, the *ars* promoter and the ArsR operator or binding site (ABS) from *E. coli*. Circuits based on ArsR can be used to quantify arsenic, and arsenic bioreporter assays have proven useful in measuring arsenic contamination in potable water sources in exposed areas of the world where chemical analytics is cumbersome [103], [115], [116]. Elements of the arsenic bioreporter circuitry are derived from a natural bacterial arsenic resistance element, notably the *ars*^{R773} operon from plasmid R773. The main control is achieved at the transcriptional level by ArsR, a *trans*-acting As(III)/Sb(III)-responsive repressor [117], [118]. ArsR homodimers bind to an operator directly upstream of the *arsR* promoter (ABS, Figure 3.1A), which inhibits RNA polymerase access, reducing expression of the operon to a basal level in absence of inducer [118]. Arsenite (and antimonate) binding to the ArsR-dimer reduces its affinity to the operator. Hence, when cells are exposed to arsenite, the *arsR* promoter is on average more or more often accessible for RNA polymerase and the rate of transcription of the operon increases [118]. Bioreporter circuits for arsenic have exploited the ArsR protein and its transcriptional control, by fusing the P_{ars} promoter to genes for suitable reporter proteins either directly [22], [119], or in more complex circuitry [120]. One of the drawbacks of the ArsR-P_{ars} system is the relatively high background expression as a result of the natural feedback loop (Figure 3.1A), which can be reduced by inclusion of a secondary ABS [121]. The major challenge for the exploitation of the arsenic circuitry is to be able to measure at very low arsenic concentrations (1–10 µg As l^{-1}), which might be achieved by optimization on variants with mutations in the ABS that resulted in up to 12-fold better signal-to-noise ratios upon induction with arsenite[122], whereas recent work on the *Chromobacterium violaceum ars* operon indicates a potentially tighter binding ArsR-P_{ars} variant that might be interesting to exploit further [123].

The main goal of this work was thus to construct an accurate model for the arsenic circuit, which would be sufficiently generally exploitable to predict changes in the circuitry elements that would be favorable to achieve better performance at low arsenic concentrations. We base the model on a mechanistic conceptualization of ArsR operator binding, molecular affinities of ArsR binding to DNA and to arsenic, in conjunction with general cellular transcription, translation and protein maturation or degradation rates. The model is parameterized on existing biochemical data of the ArsR system and further by reiteration on experimental observations. Predictions from the model suggesting improvements on the arsenic response were then reconstructed in the laboratory for verification. Notably, these included a variant allele of *arsR* and presence or absence of the arsenic efflux system. We expect that the detailed generalized mechanistic model of ArsR-P_{ars} will be useful for optimization of other genetic circuits.

3.2 Results

3.2.1 Mechanistic Model for ArsR-P_{ars} Regulation

A mechanistic model was built for a feedback-controlled and an uncoupled ArsR-controlled genetic circuit, which predicts EGFP fluorescence as a function of arsenite concentration, taking affinity constants of ArsR for its operator and for arsenite, as well as other cellular processes (arsenite influx, transcription and translation rates, protein and mRNA degradation rates, promoter strength and EGFP maturation) into account (Figure 3.2). Initial basic key parameters for the model derive from experimental values for (i) the background EGFP concentration in *E. coli* cells carrying reporter plasmid pPR-ars-ABS measured by cross-correlation spectroscopy (~700 nM, or 400 molecules per cell [124]), (ii), the reported background ArsR^{K12} protein level in *E. coli* K12 in minimal medium (24 copies per cell [125]), and (iii) the affinity of purified ArsR^{R773} for its operator measured by DNA binding assays (~10^7 M^{-1} [126]). In addition, we measured EGFP fluorescence by flow cytometry in *E. coli* reporter cells carrying either the feedback (pPR-ars-ABS) or uncoupled circuit (pAAUN, Figure 3.1), exposed for 3 h to 0, 5, 10, 50 and 150 µg arsenite-As l^{-1}. These *E. coli* cells were further devoid of the chromosomal *arsRBC* operon. Unknown model parameters were then fitted by a scatter search (heuristic global optimization method) algorithm [127], from which we extracted the 100 best fitted parameter sets (Supporting Information Figure 3.9). Distributions of the 100 best-fitted parameter sets show limited variation (Supporting Information Figure 3.9, standard deviations) and have representative average parameter values that are reported in Table 3.1 and Table 3.2. The resulting 100 best data, top and mean simulations fall relatively close to the measured values with slight deviations for the highest arsenite concentrations (Figure 3A,B), which is due to the fitting algorithm that equally values experimental observations at low and high arsenite concentrations (e.g., using eq. 10).

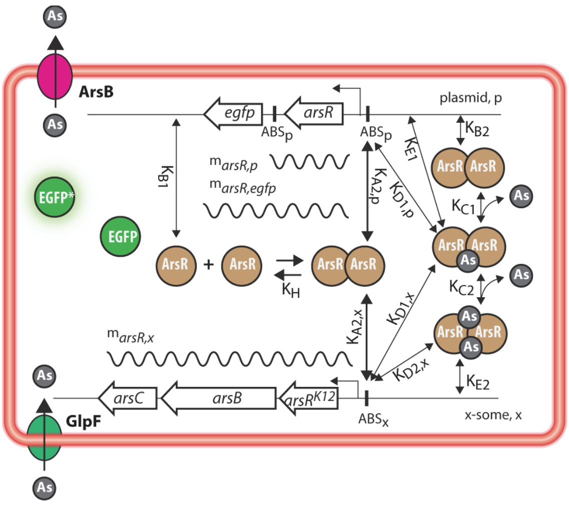


Figure 3.2 **Conceptualization of the ArsR arsenic reporter circuitry**. Cells carry the *egfp* reporter circuitry on a plasmid (as example here the feedback circuit) in presence or absence of an additional chromosomal *arsRBC* operon. Straight lines and open arrows depict DNA and location of open reading frames, respectively, with hooked arrows showing the *ars*-promoters. Sinusoid lines represent relevant transcribed mRNAs. Colored circles or ovals represent relevant proteins with their names. Two-facing line arrows indicate binding equilibria between relevant partners and their corresponding affinity constant (see Table 3.2). Transcription, translation, maturation and degradation rate constants not indicated (see Table 3.1). As, arsenite; ABS_x, chromosomal ArsR operator; ABS_p, plasmid ArsR operator (from the R773 system). Note that, depending on the plasmid construct, ArsR can be $ArsR^{K12}$ or $ArsR^{R773}$, and that the modeling further allows formation of heterodimers (i.e., $ArsR^{K12}$-$ArsR^{R773}$). Further note that the host strain can have a deletion of the full *arsRBC* operon or only of the *arsBC* genes.

Sensitivity analysis of the parameters in those two circuits suggested, among others, that changes in the DNA binding affinity constant of ArsR (K_A, Figure 3.4A-B) would have a major impact on EGFP output, particularly at low arsenite concentrations (Figure 3.4A-B). Furthermore, changes in the affinity of ArsR-dimers to arsenite (K_C, Figure 3.2) would affect output primarily at higher arsenite concentrations (Figure 3.4A-B). Finally, the circuits are, as expected, sensitive to changes in transcription, translation and EGFP maturation rates (Figure 3.4A, B), whereas the uncoupled circuit is sensitive to the strength of the promoter driving *arsR* expression (i.e., P_{AA}). To confirm this, we measured EGFP fluorescence as a function of arsenite exposure from a bioreporter strain where P_{AA} had been replaced by the P_{LTetO1} promoter [23]. We forced the model to estimate the strength of the P_{LTetO1} promoter with normalized least-square fitting of the EGFP output, keeping the other parameters unchanged. As expected, and in agreement with the experimental data, EGFP fluorescence is decreased across the complete range of arsenic concentrations (2.5–150 µg l^{-1}) by increasing the activity of the uncoupled promoter (Figure 3.3A, blue lines). The average modeled activity of P_{LTetO1} (1.15) was 14 times higher than that of P_{AA} (0.0844, Table 3.2), which is three times as high as reported previously based on mRNA synthesis measurements [128].

Table 3.1 **Model species, synthesis, and degradation rates**.

Reaction	Rate constant(s)	Value[a]	Units	Rate description	Circuits
$\emptyset \rightarrow m_{EGFP}$	P $k_{s,mEGFP}$[b,c]	$1.50 \cdot 10^{-2}$	s^{-1}	*egfp* transcription	UN[d]
$m_{EGFP} \rightarrow \emptyset$	$k_{d,mEGFP}$	$7.62 \cdot 10^{-3}$	s^{-1}	m_{egfp} degradation	UN
$m_{EGFP} \rightarrow EGFP$	$k_{s,EGFP}$	$1.84 \cdot 10^{-1}$	s^{-1}	EGFP translation	UN
$\emptyset \rightarrow m_{ArsR\text{-}EGFP}$	P $k_{s,m\text{-}arsR\text{-}EGFP}$	$1.50 \cdot 10^{-2}$	s^{-1}	*arsR-egfp* transcription	FB
$m_{ArsR\text{-}EGFP} \rightarrow \emptyset$	$k_{d,m\text{-}arsR\text{-}EGFP}$	$7.62 \cdot 10^{-3}$	s^{-1}	$m_{ArsR\text{-}EGFP}$ degradation	FB
$m_{ArsR\text{-}EGFP} \rightarrow EGFP$	$k_{s,EGFP}$	$1.84 \cdot 10^{-1}$	s^{-1}	EGFP translation	FB
$EGFP \rightarrow EGFP^*$	$k_{s,EGFP^*}$[e]	$1.20 \cdot 10^{-3}$	s^{-1}	EGFP maturation	UN,FB
$EGFP^{(*)} \rightarrow \emptyset$	$k_{d,EGFP}$[f]	$1.0 \cdot 10^{-5}$	s^{-1}	EGFP degradation	UN,FB
$m_{ArsR\text{-}EGFP} \rightarrow ArsR$[g]	$k_{s,ArsR}$	$7.33 \cdot 10^{-2}$	s^{-1}	ArsR translation	FB
$\emptyset \rightarrow m_{ArsR\text{-}K12}$	P $k_{s,m\text{-}arsR\text{-}K12}$	$1.50 \cdot 10^{-2}$	s^{-1}	$arsR^{K12}$ transcription	UN,FB
$m_{ArsR\text{-}K12} \rightarrow \emptyset$	$k_{d,m\text{-}arsR\text{-}K12}$	$4.81 \cdot 10^{-3}$	s^{-1}	$m_{ArsR\text{-}K12}$ degradation	UN,FB
$m_{ArsR\text{-}K12} \rightarrow ArsR^{K12}$	$k_{s,ArsR\text{-}K12}$	$7.33 \cdot 10^{-2}$	s^{-1}	$ArsR^{K12}$ translation	UN,FB
$ArsR^{K12} \rightarrow \emptyset$	$k_{d,ArsR\text{-}K12}$	$3.07 \cdot 10^{-3}$	s^{-1}	$ArsR^{K12}$ degradation	UN,FB
$\emptyset \rightarrow m_{ArsR\text{-}R773}$	P $k_{s,m\text{-}arsR\text{-}R773}$	$1.50 \cdot 10^{-2}$	s^{-1}	$m_{ArsR\text{-}R773}$ transcription	UN,FB
$m_{ArsR\text{-}R773} \rightarrow \emptyset$	$k_{d,m\text{-}arsR\text{-}R773}$	$4.81 \cdot 10^{-3}$	s^{-1}	$m_{ArsR\text{-}R773}$ degradation	UN,FB
$m_{ArsR\text{-}R773} \rightarrow ArsR^{R773}$	$k_{s,ArsR\text{-}R773}$	$7.33 \cdot 10^{-2}$	s^{-1}	$ArsR^{R773}$ translation	UN,FB
$ArsR^{R773} \rightarrow \emptyset$	$k_{d,ArsR\text{-}R773}$	$3.07 \cdot 10^{-3}$	s^{-1}	$ArsR^{R773}$ degradation	UN,FB
$ArsR_2 \rightarrow \emptyset$	$1/2 \cdot k_{d,ArsR}$	$1.53 \cdot 10^{-3}$	s^{-1}	ArsR dimer degradation	UN,FB
$ArsR_2{:}As \rightarrow As + \emptyset$	$1/2 \cdot k_{d,ArsR}$	$1.53 \cdot 10^{-3}$	s^{-1}	ArsR – arsenite complex degradation	UN,FB
$ArsR_2{:}As_2 \rightarrow 2\ As + \emptyset$	$1/2 \cdot k_{d,ArsR}$	$1.53 \cdot 10^{-3}$	s^{-1}	ArsR – arsenite complex degradation	UN,FB
$As_{ext} \rightarrow As_{int}$	$v_{max\text{-}GlpF}$	0.0332	$\mu M\ s^{-1}$	Maximal arsenic influx	UN,FB
	$K_{M\text{-}GlpF}$	2.0	μM	Michaelis constant of arsenic influx	UN,FB
$As_{int} \rightarrow As_{ext}$	$k_{cat\text{-}ArsB}$	1.34	s^{-1}	ArsB catalytic constant	UN,FB
	$K_{M\text{-}ArsB}$	1.68	μM	Michaelis constant of arsenic efflux	UN,FB

a) Values reported are the averages of the 100 best fitted parameter sets (Figure 3.9), except if noted otherwise.
b) P, promoter activity. Promoter activities are circuit-dependent, defined by eqs. (3.5) and (3.6). In uncoupled circuits, the value of the constitutive promoter P_{AA}=0.0844
c) Maximal transcription rate.
d) UN, specific for uncoupled circuit only. FB, feedback circuit only. UN, FB, both circuits.
e) Value taken from ref [129].
f) Value taken from ref [130].
g) Either $arsR^{K12}$ or $arsR^{R773}$, since both alleles share transcription, translation and degradation rates (see Assumptions).

Table 3.2 **Binding affinities and reactions.**

Binding constant symbol	**Reaction**[a]	**Estimated binding constant**[b] **(M^{-1})**	
		$ArsR^{K12}$	**$ArsR^{R773}$**
K_{A2}[c]	$ABS^{R773} + ArsR_2 \rightleftharpoons ABS^{R773}:ArsR_2$	$3.33 \cdot 10^7$	$6.90 \cdot 10^7$
K_{B1}	$U + ArsR \rightleftharpoons U:ArsR$	$3.51 \cdot 10^1$	$4.70 \cdot 10^2$
K_{B2}	$U + ArsR_2 \rightleftharpoons U:ArsR_2$	$1.68 \cdot 10^1$	$6.47 \cdot 10^1$
K_{C1}	$ArsR_2 + As \rightleftharpoons As:ArsR_2$	$1.66 \cdot 10^7$	$2.58 \cdot 10^6$
K_{C2}	$As:ArsR_2 + As \rightleftharpoons As_2:ArsR_2$	$1.66 \cdot 10^7$	$2.58 \cdot 10^6$
K_{D1}	$ABS^{R773} + As:ArsR_2 \rightleftharpoons ABS^{R773}:As:ArsR_2$	$1.26 \cdot 10^4$	$3.38 \cdot 10^4$
K_{D2}	$ABS^{R773} + As_2:ArsR_2 \rightleftharpoons ABS^{R773}:As_2:ArsR_2$	$1.26 \cdot 10^4$	$3.38 \cdot 10^4$
K_{E1}	$U + As:ArsR_2 \rightleftharpoons U:As:ArsR_2$	$3.41 \cdot 10^2$	$1.20 \cdot 10^3$
K_{E2}	$U + As_2:ArsR_2 \rightleftharpoons U:As:ArsR_2$	$3.41 \cdot 10^2$	$1.20 \cdot 10^3$
K_H	$ArsR + ArsR \rightleftharpoons ArsR_2$	$1.59 \cdot 10^8$	$2.59 \cdot 10^8$

a) For species symbols, see Table 3.1.

b) The estimated binding affinities for ArsR heterodimers are obtained from the average free energies of binding from the 100 best fitting parameter sets (Figure 3.9) for $ArsR^{K12}$ and $ArsR^{R773}$ (see Assumption 2).

c) Binding constant is specified by ArsR type, regardless of binding chromosomal or plasmid ABS. Thus, only one value applies for a binding of ABS^{K12} or ABS^{R773}, reported as K_{A2}.

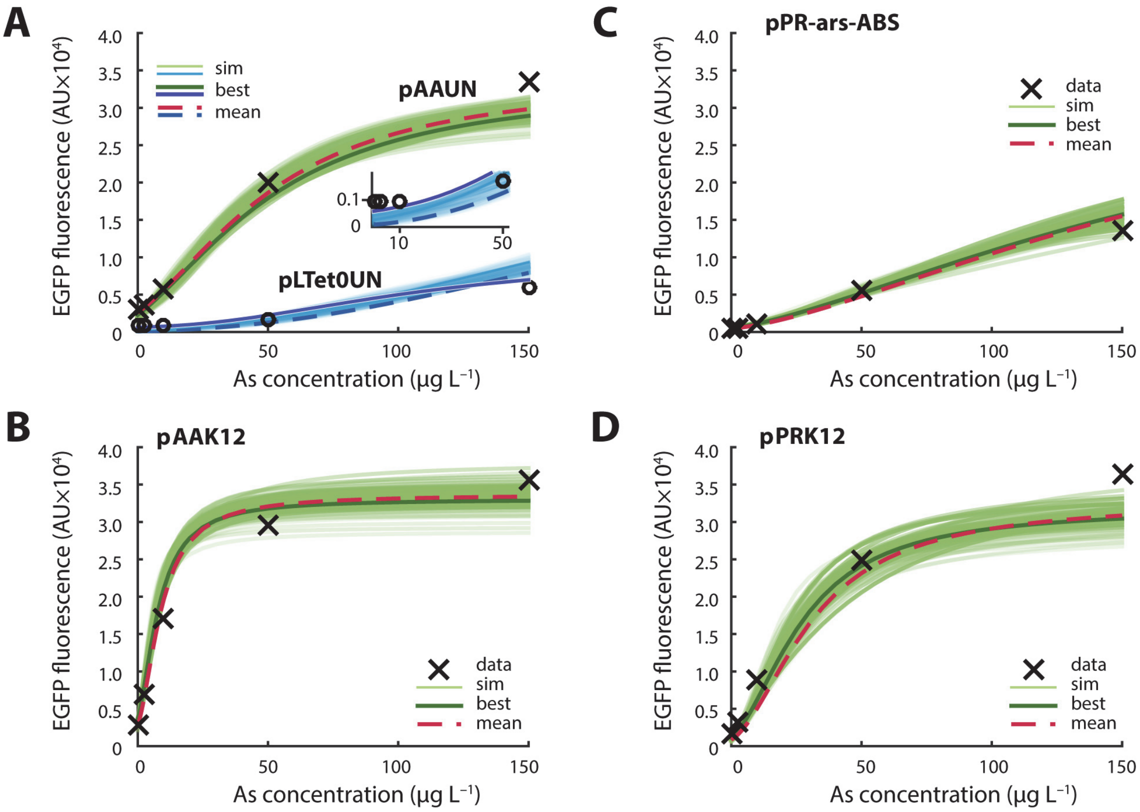


Figure 3.3 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As$_{III}$ concentrations after 3h induction for the four plasmid bioreporter configurations in an *E. coli* host without chromosomal *arsRBC* operon**. Lines (*sim*) show results of the best 100 fits from the parameter estimation, simulated and plotted in all different configurations. Line darkness increases with the overall fitting score, the overall best fit (*best*) being represented with the darkest, and the average of all parameter sets (*mean*) with a dashed line. Data points show the mean of independent biological triplicates. Error bars are smaller than the used symbol size and are therefore not indicated.

3.2.2 Effect of ArsR Binding Affinity on Circuit Output

In order to validate whether the circuits would react to changes in ArsR binding affinity with a different EGFP output, we replaced the *arsR*^{R773} by the *arsR*^{K12} gene, which we expected might have different binding affinities. Both *arsR* alleles (ArsR^{R773}, Genbank accession number P15905; ArsR^{K12}, Genbank accession number AAC76526) have 74% nucleotide and 77% amino acid identity (Supporting Information Figure 3.10). Interestingly, both feedback and uncoupled constructs carrying the *arsR*^{K12} gene produce "steeper" reaction curves to arsenite than their *arsR*^{R773} equivalents, meaning they produce more EGFP at the same arsenite concentration (Figure 3.3C-D). Fitting the equivalent affinity parameters in the model for ArsR^{K12} showed that the binding affinity to the operator site (which still originates from the R773 system) is two-fold less but the binding affinity for arsenite is six-fold higher than for ArsR^{R773} (Table 3.2). Less binding affinity to the operator would lead to release of the promoter at lower arsenite concentrations and thus higher EGFP output, as observed (Figure 3.3C-D). Simulations of the best 100 sets of parameters showed slightly more variation of the ArsR^{K12} circuit predictions than for the ArsR^{R773} circuit (e.g., compare Figure 3.3A, B with the corresponding Figure 3.3C and D). Sensitivity analysis of fitted parameters in the ArsR^{K12} circuits showed essentially the same contributing factors, except that these circuits were less sensitive to variations in the binding affinities of ArsR to arsenite (Figure 3.4C-D), which may be due to the six-fold higher binding constant predicted for ArsR^{K12} than ArsR^{R773} (Table 3.2).

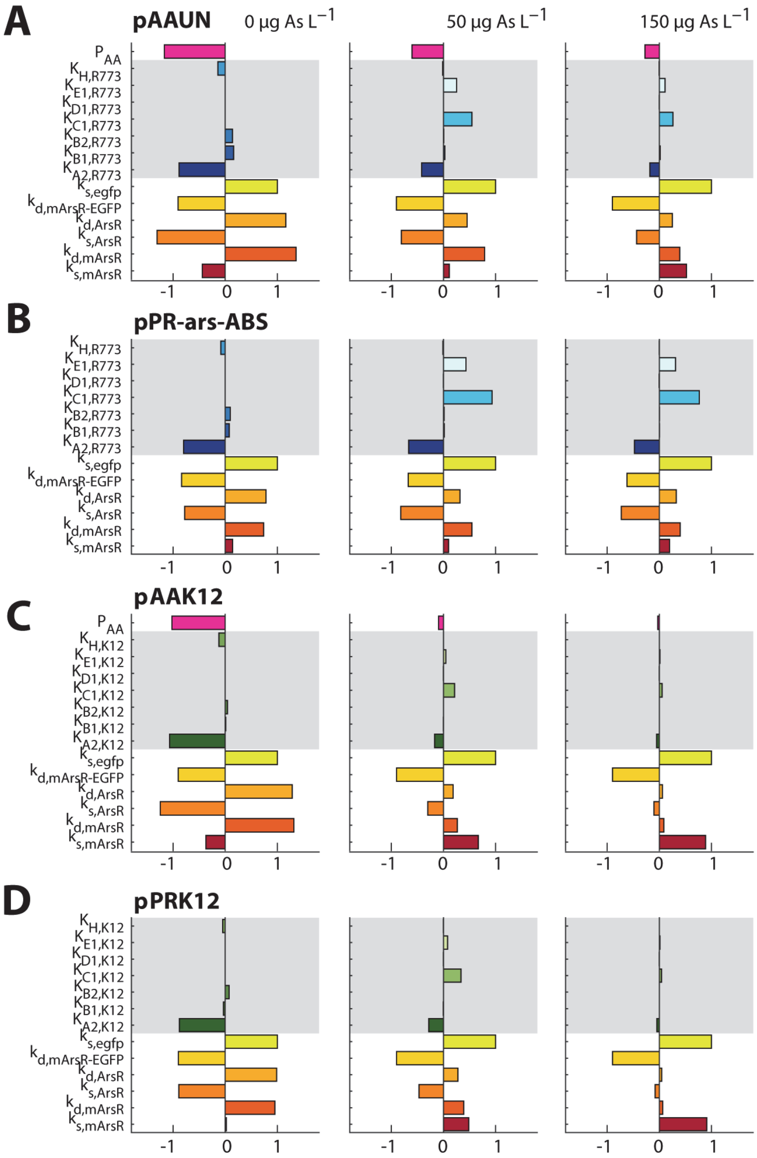


Figure 3.4 **Sensitivity analysis of key parameters in the model.** Horizontal bars show the relative deviation in EGFP fluorescence output from the respective circuit (i.e., pAAUN etc.) in *E. coli* MG1655 Δ*arsRBC* at a 5% difference of the mean parameter value (*mean*, as in Figure 3.3). Scenarios calculated for circuit output in absence of arsenite (0 µg As l^{-1}), at 50 µg As l^{-1} and at 150 µg As l^{-1}). Note the relatively large sensitivities to overall transcription, translation and degradation rates of mRNA and protein (which the model implicitely assumes constant), and the large sensitivities to promoter strength, the specific binding affinities to the ABS (K$_A$) and the binding affinity of ArsR to As (K$_C$).

3.2.3 Including an Arsenite Efflux System Decreases Sensitivity of the Circuit

Arsenite is transported into the reporter cells by the action of the GlpF aquaglyceroporin facilitator and is effluxed by the cells through the ArsB transporter system [131] (Figure 3.2). In the host cells for the experiments in Figures 3 and 4, the arsenite efflux system was removed by deletion of the chromosomal *arsRBC* genes, while keeping the influx intact. Inclusion of an ArsB arsenite efflux system displaying Michaelis-Menten kinetics into the model showed that the circuit output (i.e. EGFP fluorescence) would decrease at the same arsenite concentration (Figure 3.5, Figure 3.6). The different reporter circuits were thus reintroduced into *E. coli* MG1655 wild-type carrying the full *arsRBC* operon. Both measurements and predictions of EGFP fluorescence in these strains as a function of 3 h arsenite exposure were largely in agreement, and showed an overall reduction in EGFP levels in strains with the efflux system compared to those without, exposed to the same arsenite concentration (Figure 3.6). The overall type of response (more "linear", e.g., with pPR-ars-ABS, or "steep saturation", with pAAK12) was maintained between *E. coli* MG1655 Δ*arsRBC* and *E. coli* MG1655 as hosts (compare Figure 3.3 and Figure 3.6). These results also suggested that the introduced Michaelis-Menten parameters for the ArsB efflux system were sufficient to predict the circuit behavior in *E. coli* MG1655 and that no major other biochemical reactions would have been necessary to explain the system's behavior. Evidently, the circuit performance is sensitive to small variations in the kinetic parameters of the efflux system (k_{cat} and K_M, Figure 3.5).

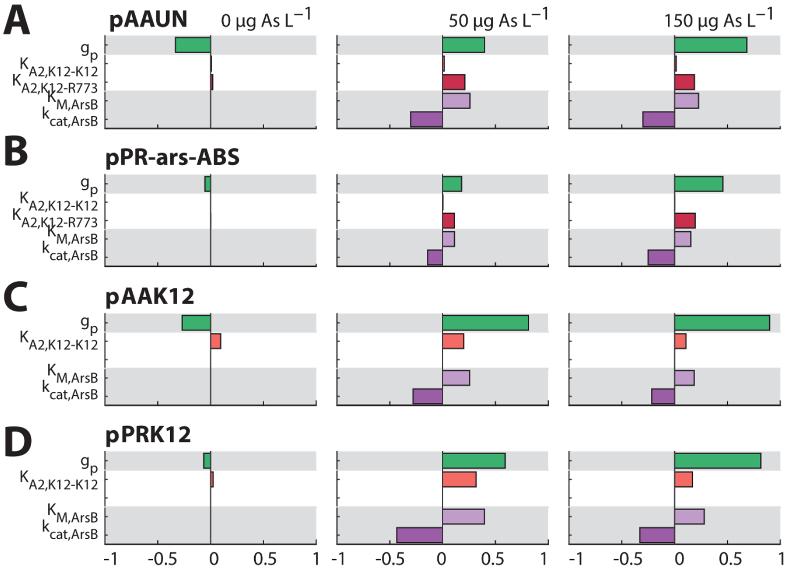


Figure 3.5 **Sensitivity analysis of the effect of further model parameters on EGFP output from the four reporter circuits in *E. coli* MG1655** (**A-D**, as indicated). g_p, plasmid copy number; $K_{A2,\ K12\text{-}K12}$, DNA binding constant for having both chromosomal and plasmid *arsR*^{K12} allele; $K_{A2,\ K12\text{-}R773}$, DNA binding constant for having chromosomal *arsR*^{K12} but plasmid *arsR*^{773} allele; $K_{M,\ ArsB}$, Michaelis-Menten constant of the ArsB efflux pump; $k_{cat,ArsB}$, catalysis rate of the ArsB efflux pump. Horizontal bars show the relative deviation in EGFP fluorescence output from the respective circuit (i.e., pAAUN etc.) at a 5% difference of the mean parameter value (*mean*, as in Figure 3.3). Scenarios calculated for circuit output in absence of arsenite (0 µg As l^{-1}), at 50 µg As l^{-1} and at 150 µg As l^{-1}.

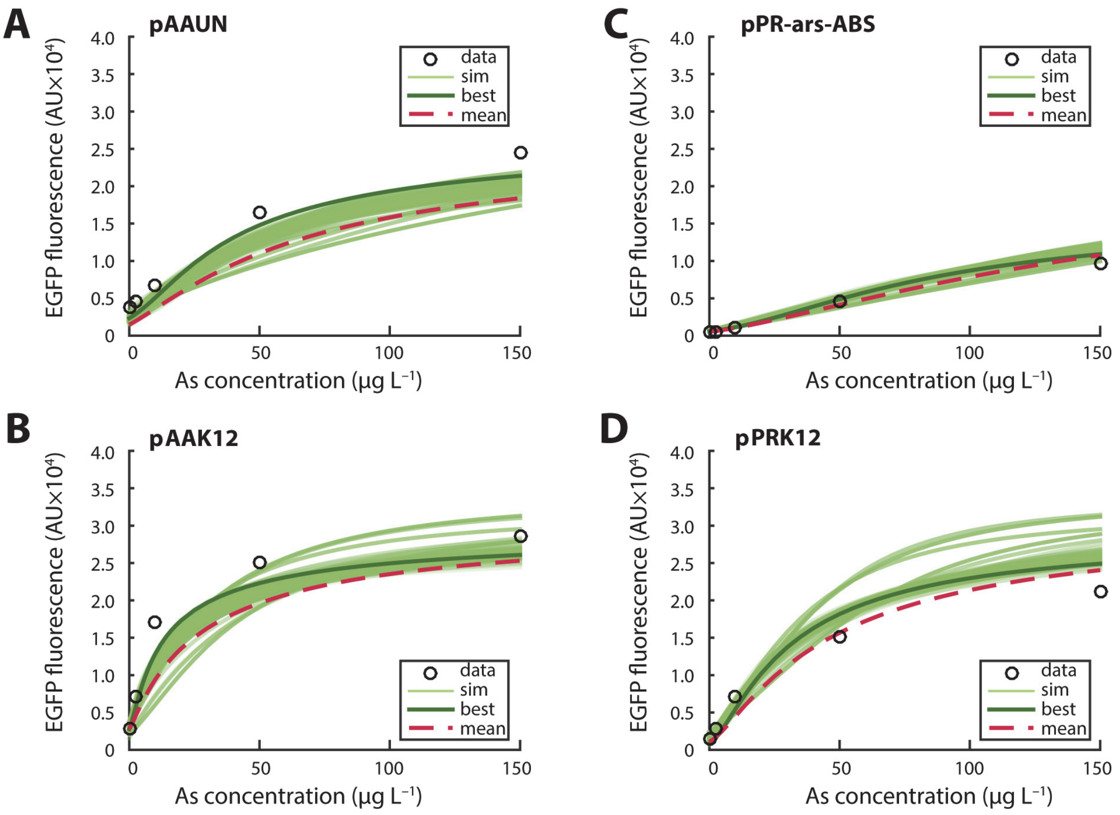


Figure 3.6 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As_{III} concentrations after 3 h induction for the four plasmid bioreporter configurations in *E. coli* MG1655 with the arsenite ArsBC efflux pump.** See further legend to Figure 3.3.

3.2.4 Possible Cross-Binding Effects of Double ArsR Alleles

Since the host strain expressing the ArsBC efflux pump also carries the chromosomal *arsR*^{K12} allele, it would be conceivable that some cross-binding or formation of ArsR-heterodimers (between ArsR^{K12} and ArsR^{R773}) is occurring, influencing the outcome of the ArsR-dependent circuit. When assuming that the properties of heterodimers are the sum of half the properties of each individual homodimer, the model predicts that circuits composed of the ArsR^{R773} allele (i.e., pAAUN and pPR-ars-ABS) in a host background with chromosomal *arsR*^{K12} would slightly increase EGFP output at the same arsenite-As exposure (Figure 3.5, K_{A2} sensitivity). To test this prediction, we transformed the respective plasmid circuits in *E. coli* MG1655 with *arsR*^{K12} but without *arsBC*. Results showed that indeed there was a slight difference of EGFP output at the same arsenite concentration in *E. coli* MG1655 Δ*arsBC* compared to Δ*arsRBC,* but experimental and modeling data were not completely in agreement (Figure 3.7A, B, Supporting Information Figure 3.11). This suggested that the behavior of ArsR^{K12}-ArsR^{R773} heterodimers is not correctly predicted by the model. In contrast, strains carrying both a chromosomal and reporter circuit *arsR*^{K12} allele cannot form heterodimers (i.e., pAAK12 and pPRK12 in *E. coli* MG1655 Δ*arsBC*). The models predict that these strains would not change EGFP output despite a higher *arsR*^{K12} copy number (Figure 3.7C-D). This is in agreement with experimental data, within the variation of both models and experiments (Supporting Information Figure 3.11). Finally, the sensitivity analysis suggests that variations in copy numbers of the plasmids carrying the reporter circuits would also have an effect on the EGFP signal (g_p, Figure 3.5), but this was not tested experimentally. All reporter plasmids in the tested strains are based on the same replicon (pPROBE-tagless [132]) and, therefore, are assumed to have the same copy number (Assumption 6, see below).

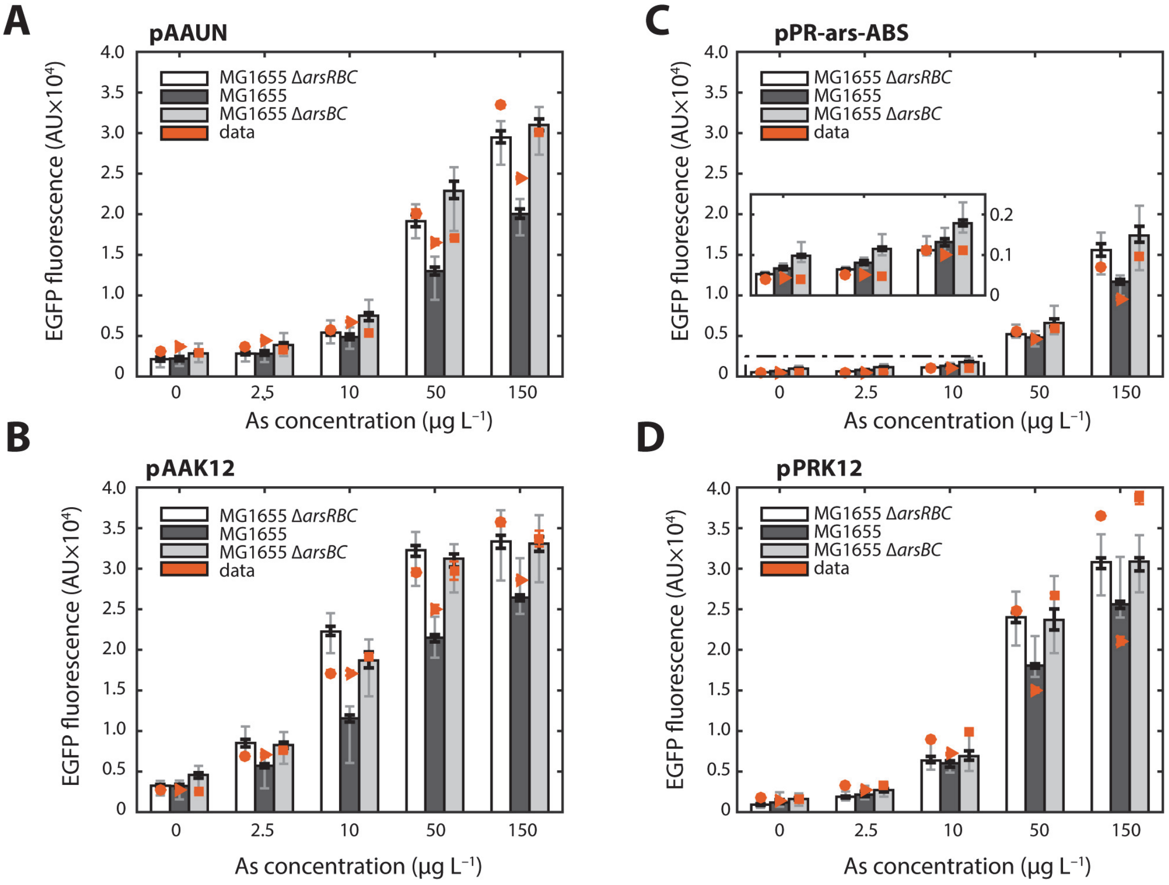


Figure 3.7 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As_{III} concentration after 3 h induction for the three series of four plasmid bioreporter configurations in *E. coli* MG1655 with Δ*arsRBC*, wild-type and Δ*arsBC* background**. Data points (round, triangle and square markers) and error bars, when larger than the markers, show the mean and standard deviation of independent biological triplicate measurements, respectively. Bars show the median output of the best 100 fits from the parameter estimation, with lower and upper quartiles (black) and minimum and maximum output (gray). When larger than symbol sizes, error bars indicate the standard deviation.

Table 3.3 **Figures of merit for the different arsr reporter circuits and genetic backgrounds as a function of arsenite exposure.**

Reporter circuit	*E. coli* Strain	Basal GFP expression (AU)[a]	Fold induction[b] at 2.5 µg As l^{-1}	10 µg As l^{-1}	50 µg As l^{-1}	150 µg As l^{-1}
pAAUN	MG1655	3768 ± 227	1.2	1.8	4.4	6.5
	MG1655 Δ*arsRBC*	3042 ± 61	1.2	1.9	6.6	11
	MG1655 Δ*arsBC*	3007 ± 105	1.1	1.8	5.7	10
pAAK12	MG1655	2811 ± 61	2.5	6.1	8.9	10.2
	MG1655 Δ*arsRBC*	2765 ± 32	2.5	6.2	10.7	12.9
	MG1655 Δ*arsBC*	2460 ± 51	3.1	7.8	12.1	13.7
pPR-arsR-ABS	MG1655	429 ± 8	1.2	2.3	10.7	22.4
	MG1655 Δ*arsRBC*	394 ± 9	1.3	2.8	14.1	34.4
	MG1655 Δ*arsBC*	393 ± 33	1.2	2.8	15.1	37.8
pPRK12	MG1655	1380 ± 39	2	5.2	10.9	15.3
	MG1655 Δ*arsRBC*	1721 ± 19	1.9	5.2	14.4	21.2
	MG1655 Δ*arsBC*	1698 ± 99	1.9	5.8	15.7	22.8

a) Values are averages from triplicate measurements in flow cytometry (FITC-a channel) of 10,000 cells, after 180 min of incubation in absence of arsenite.

b) Calculated as the ratio between the average signal (from triplicates) at the indicated arsenite concentration, divided by the average basal expression.

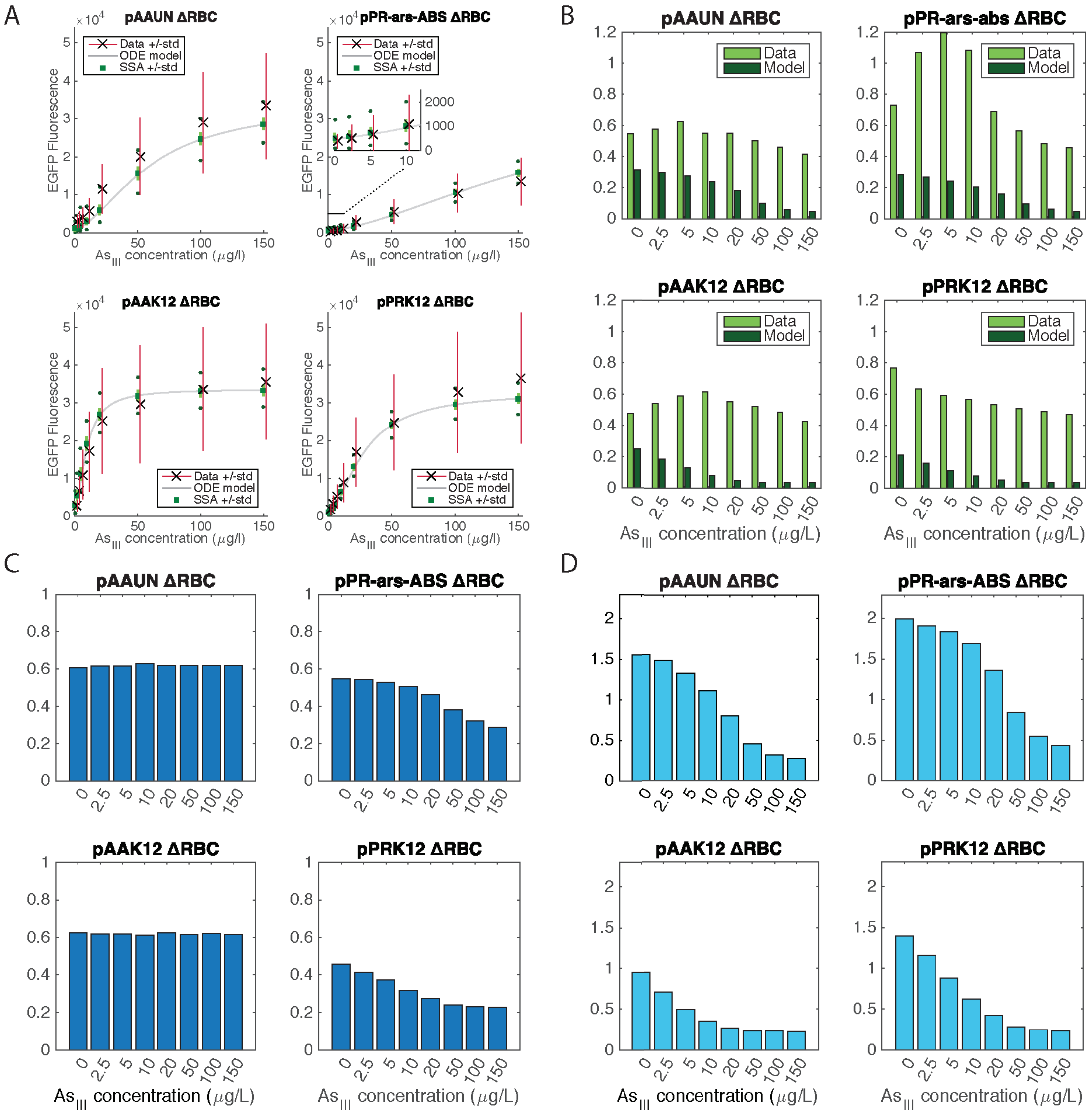


Figure 3.8 **A: Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-AsIII concentrations.** The results are obtained after 3 hours of induction for four plasmid bioreporter configurations in an *E. coli* host without chromosomal arsRBC operon (left column: uncoupled circuits, right column: feedback circuits; upper row: carrying *arsR*^{R773} allele, lower row: carrying *arsR*^{K12} allele). Crosses and error bars (Data) show average fluorescence and standard deviations from single cell measurements. Solid lines (ODE) were simulated from ordinary differential equations and squares with error bars (SSA) result from stochastic simulations of the model (10^4 realizations). Dots denote the minimum and maximum values across the stochastic simulations. **B**: Coefficient of variation of fluorescence output at different arsenite-AsIII concentrations for the four bioreporter configurations from experimental measurements (Data) and stochastic simulations (Model). **C**: Coefficient of variation of short mRNA strands, *i.e.* encoding only ArsR, measured in the stochastic simulations at different arsenite-AsIII concentrations for the four bioreporter configurations. **D**: Same as **C** but for long mRNA strands encoding for EGFP. In the feedback circuits pPR-ars-ABS and pPRK12, long mRNA strands encode for both EGFP and ArsR proteins.

3.2.5 Variability analysis of the $\Delta arsRBC$ bioreporter strains

In order to compare the variability of experimental cell-to-cell measurements (Figure 3.8A) with the mathematical model, we transformed the system of ODE into a stochastic simulation algorithm (SSA) as described in Section 3.3.10. We applied the Gillespie algorithm to select series of reactions that happen in the bioreporter cell. The probability of a reaction to happen depends on the state of the system, e.g. the probability for a protein to be translated is higher for a higher level of mRNA. After each reaction, levels of chemical species as well as the probabilities for choosing the next reaction are updated. Every realization of a stochastic simulation is a possible outcome of the biochemical system. By running multiple realizations of the algorithm, different outcomes are generated and the variability of the system can be computed as shown in Figure 3.8B, the coefficient of variation of the fluorescent signal is much higher in experimental data than in stochastic simulations. This suggests that neglected aspects of the model such as cell growth, partitioning in daughter cells or molecular crowding, among others, have a large contribution in the variability. In stochastic simulations, the coefficient of variation is consistently decreasing with the arsenic concentration, whereas it tends to have a peak in the range of 5-10 µg/L of arsenic in experimental data.

The model gives further information on the variability of species that have not been measured experimentally, in particular for the mRNA species. As expected, the coefficients of variation of the short arsR-mRNA strands (Figure 3.8**C**) in uncoupled circuits are arsenic-independent, as the *arsR* transcription is placed under a constitutive promoter. In the feedback constructs, the variability decreases with an increasing arsenic concentration, and is slightly lower in the construct carrying the stronger $arsR^{K12}$ repressor. The long mRNA strands only carry *egfp* in uncoupled constructs (Figure 3.8**D**, pAAUN and pAAK12) and carry both *arsR* and *egfp* in feedback constructs (pPR-ars-ABS, pPRK12). For a given *arsR* allele, the uncoupled circuits have less variability than their feedback counterpart. Overall, circuits with the stronger $arsR^{K12}$ repressor have a lower mRNA variability. This shows that both the circuit topology and the strength of the repressor have a contribution in mRNA variability.

3.3 Methods

3.3.1 Strains, Culturing procedures, Cloning and Molecular Techniques

Strains and primers used in this study are listed in Supporting Information Table 3.5 and Table 3.6, respectively. Standard procedures in molecular biology were followed for cloning in *E. coli,* for DNA isolation and manipulations, or for DNA amplifications by polymerase chain reaction (PCR)[133], [134]. For standard growth conditions, *E. coli* was cultured in liquid or on agar-solidified (1.5% *w/v*) Luria Broth (LB) at 37°C, under inclusion of the appropriate antibiotics to select for the presence of plasmids carrying the reporter circuits. Assays for arsenic reporter induction are described below.

3.3.2 Deletion of the Chromosomal *ars* Genes in *E. coli*

The complete *arsRBC* operon or *arsBC* only were deleted from *E. coli* MG1655 using homologous recombination and I-SceI counterselection [135], [136]. In short, boundary regions were produced by PCR amplification (Supporting Information Figure 3.12), which were cloned into R6K-based plasmid pJP5603-ISceIv2 [135]. Since this plasmid cannot replicate in *E. coli* without the lambda-pir protein, selection for its antibiotic resistance marker (kanamycin, Km) allows recovery of single cross-over events between plasmid and chromosome at the *ars* locus. Subsequent introduction into *E. coli* MG1655 with the single cross-over events of the pSW(I-SceI) plasmid expressing I-SceI [135] leads to double-strand breakage. Recovered *E. coli* MG1655 colonies sensitive to Km were screened by PCR for the absence of *arsRBC* or *arsBC*, and subsequently cultured for multiple batch transfers in absence of ampicillin to retrieve those having lost pSW(I-SceI). Final candidates were verified again by PCR for correct loss of *arsRBC* or *arsBC*.

3.3.3 Construction of Arsenic Reporter Plasmids

Two reporter circuit configurations were tested with both *arsR* alleles: (i) a feedback circuit, in which ArsR controls both its own expression and that of the *egfp* reporter gene (Figure 3.1A) and (ii) an uncoupled circuit, in which expression of *arsR* is constitutive, but ArsR regulates *egfp* expression from P_{ars} (Figure 3.1B). All plasmids contained an extra copy of the ArsR binding site (ABS) to reduce reporter gene expression in absence of arsenite [22], [121]. All plasmids were assembled in pPROBE-tagless, which carries a pBRR1 origin of replication and has an estimated mean copy number in the cell of 10 (Ref[34]). Plasmid pPR-ars-ABS contains the *arsR*^{R773} allele in the feedback situation (Figure 3.1A) and its construction has been described previously [22]. To replace the *arsR*^{R773} allele in the feedback reporter circuit we ordered the *arsR*^{K12} gene with appropriate restriction sites by gene synthesis (DNA2.0 Inc., Menlo Park, CA) (Supporting Information Figure 3.13). This fragment was recovered from the production plasmid by digestion with PsiI and SalI, and ligated with the vector part of pPR-arsR-ABS cut with the same enzymes. After transformation into *E. coli* and verification of the plasmid content, this resulted in plasmid pPRK12 (Figure 3.1B).

Plasmid pAAUN contains the *arsR*^{R773} gene under control of the P_{AA} promoter (Figure 3.1B), and its construction was described previously [23]. To produce an equivalent plasmid as pAAUN with the *arsR*^{K12} allele, the gene was amplified from the genome of *E. coli* MG1655 by PCR, while adding the same ribosome binding site as upstream of *arsR*^{R773} and two flanking restriction sites (BamHI and XbaI, Supporting Information Table 3.6). The PCR fragment was purified from agarose gel and ligated to prepared pGEM-T-easy vector (Promega). After transformation into *E. coli* this resulted in plasmid pGEM_arsR-K12. The *arsR*^{K12} insert was validated by DNA sequencing. Plasmid pAAK12 was assembled by recovering the *arsR*^{K12} fragment from pGEM_arsR_K12 by XbaI-BamHI digestion, which was ligated with the BamHI-NheI digested P_{ars}-*egfp* fragment from pAAUN (Figure 3.1B) and the linearized vector pPROBE (cut with NheI-XbaI). The final plasmid was validated for the correct sequence of the *arsR*^{K12}-P_{ars}-*egfp* fragment. Note that the P_{ars} promoter on pAAK12 is still of R773 origin. All plasmids were subsequently transformed into *E. coli* MG1655, MG1655Δ*arsRBC* and MG1655Δ*arsBC* (Supporting Information Table 3.5).

3.3.4 Arsenic Bioreporter Assays

Induction of EGFP fluorescence from the arsenic reporter circuits in single cells of the various strains was measured by flow cytometry on assays prepared in 96 well microplates (Greiner bio-one). Assays consisted of 180 µl aliquots of exponentially growing bioreporter cells suspended to a culture turbidity of 0.1 at 600 nm in pre-warmed (37 °C) MOPS-glucose-medium [124], mixed with 20 µl aqueous solution containing between 0 and 1.5 mg l^{-1} arsenite-As, prepared by serial dilution of a 0.05 M solution of $NaAsO_2$ (Merck) in arsenic-free tap water. Bioreporter assays were prepared in triplicate and incubated at 30°C under mixing at 500 rpm for 3 h in a thermostated shaker (THERMOstar, BMG Labtech). After the indicated incubation times 5 µl of each assay was removed, twice diluted by mixing with 195 µl of distilled water, after which 3 µl was aspired and immediately analyzed on a Becton Dickinson LSR-Fortessa flow cytometer (BD Biosciences, Erembodegem, Belgium). EGFP-fluorescence was excited at 488 nm and detected using the 'FITC' channel (530 ± 15 nm). We report the average of the population mean fluorescence across biological triplicates.

3.3.5 ArsR Circuit Description and Assumptions

The developed mechanistic model describes the rate of EGFP reporter protein synthesis from the *ars* promoter as a function of the intracellular arsenite concentration in various configurations of the bioreporter circuit (e.g., Figure 3.1 and Figure 3.2). The *E. coli* bioreporters can carry two heterologous *ars* systems: one originating from the R773 plasmid (i.e., *arsR*^{R773}) [117], [118], and the other the chromosomally located *arsRBC* operon (the *arsR* allele of which is denoted as *arsR*^{K12}) [137]. The bioreporter circuit is assembled on a plasmid-derivative of pPROBE [132] with transcription of either *arsR*^{R773} or *arsR*^{K12} allele occurring either from the native P_{ars}-promoter of R773 origin (feedback configuration, Figure 3.1A) or from the constitutive P_{AA}-promoter [128] (uncoupled configuration, Figure 3.1B). The ArsR protein binds the operator (ABS, of R773 origin) in the *ars*-promoter, thereby repressing transcription. The functional DNA binding proteins are ArsR dimers [138], which carry two arsenite binding pockets [139]. The affinity of ArsR^{R773} for its plasmid operator (ABS^{R773}) can be estimated from DNA binding studies as ~10^7 M^{-1} [126]. Interaction of ArsR dimers with arsenite diminishes DNA binding affinity and thus, on average, RNA polymerase more frequently starts transcription from P_{ars}. Reporter expression (EGFP) in both feedback and uncoupled circuits is driven from the *ars*-promoter. To diminish high background EGFP expression in absence of arsenite, a secondary operator (originating from the R773 plasmid system, ABS^{R773}) is placed directly upstream of the *egfp* gene (Supporting Information Figure 3.13). The feedback circuit therefore can form two overlapping mRNAs: one encompassing only *arsR* and the other *arsR-egfp* [121]. In the uncoupled circuit, an *egfp*-mRNA transcript is formed from the P_{ars}-promoter, and a separate *arsR*-mRNA from P_{AA} (Figure 3.1, Figure 3.2).

ArsR dimers are allowed to bind non-specifically to DNA (*U*, unspecific DNA) other than its cognate operator (ABS). In case of a cell having both *arsR* alleles, heterodimers (ArsR^{K12}-ArsR^{R773}) may form. All possible dimer forms are allowed to bind the ABS^{R773} on the plasmid DNA (ABS^{R773}) and/or the ABS^{K12} operator on the chromosome (ABS^{K12}). EGFP maturates at a rate of 0.0012 s^{-1} according to [140]. Arsenite is imported in cells through aquaglyceroporin facilitators (encoded by *glpF*)[131], and can be effluxed through the ArsB ATP-dependent pump [141], which is described with reversible Michaelis-Menten kinetics. Transcription, translation and degradation rates of mRNA and proteins are governed by first order rate constants (see below). All the circuits differ from one another in the number of incorporated elements (mathematical "species"), in their structure (feedback or uncoupled), in the types of ArsR alleles, and in the presence or absence of the arsenic efflux pump plus reductase.

3.3.6 Mathematical Model

All ArsR reporter circuits are described by a set of ordinary differential equations (ODE), where the concentrations of mRNA (m_i) and protein (X_i) of ArsR and GFP can be expressed by the following general equations:

$$\frac{d[m_i]}{dt} = P\, k_{s,m_i} - k_{d,m_i}[m_i] \tag{3.1}$$

$$\frac{d[X_i]}{dt} = k_{s,X_i}[m_i] - k_{d,X_i}[X_i] - \sum_{j \in Q_i} q\left(K_{i,j},[X_i],[X_j]\right) \tag{3.2}$$

where P is the promoter activity (dimensionless), k_s and k_d are the synthesis and degradation constants (s^{-1}), respectively, described in Table 3.1 **Model species, synthesis, and degradation rates**.. The interaction of protein X_i with another species X_j forming a complex X_i:X_j is governed by forward and backward reactions with rate constants, k_f and k_b, respectively. We denote Q_i the ensemble of species interacting with X_i, and q the net rate of reaction ($M\ s^{-1}$)

$$q\left(K_{i,j},[X_i],[X_j]\right) = k_{\mathrm{f}}\left([X_i][X_j] - [X_i : X_j]/K_{i,j}\right) \tag{3.3}$$

$K_{i,j} = \frac{k_f}{k_b}$ is the equilibrium constant of the X_i:X_j complex formation (Table 3.2). EGFP does not bind to any further species, hence its synthesis has the form of equations (3.1) and (3.2) but without the summation term. In contrast, we add a maturation term, following equation (3.4):

$$\frac{d[EGFP^*]}{dt} = k_{s,\mathrm{EGFP}^*}[EGFP] - k_{d,\mathrm{EGFP}^*}[EGFP^*] \tag{3.4}$$

where *EGFP* is the immature EGFP protein and *EGFP** the mature (fluorescent) form.

The full circuit with all elements contains 6 primary interacting species forming 38 complexes (see Figure 3.2). The primary species are ArsR^{K12}, ArsR^{R773}, ABS^{K12}, ABS^{R773}, As (arsenite-As$_{III}$), and U (non-specific binding sites on the DNA).

Given that P_{ars} is controlled by ArsR, its activity P in the model is determined by the state of its upstream ABS. If the ABS is free, transcription is allowed to occur at its maximal rate; otherwise, P_{ars} is repressed (set to 0). The second ABS placed downstream acts as a gate that either allows further transcription when it is unoccupied or stops it when occupied by ArsR [121]. In the feedback circuits, P follows equation (3.5):

$$P = \begin{cases} f_{ABS}\left(1 - f_{ABS}\right) & \text{for } m_{arsR} \\ f_{ABS}{}^{2} & \text{for } m_{arsR-EGFP} \end{cases} \tag{3.5}$$

where $f_{\mathrm{ABS}} = \frac{[\mathrm{ABS}]}{[\mathrm{ABS}]_{\mathrm{T}}}$ is the ratio of free ABS over the total, and where the quadratic form f_{ABS}^{2} for $\mathrm{m}_{\text{arsR-EGFP}}$ comes from the conditional probability that both ABS must be unoccupied in order for RNA polymerase to read through and produce *egfp*-mRNA (Figure 3.1A).

In the uncoupled circuits, promoter activities can be described by:

$$P = \begin{cases} P_{AA} & \text{for } m_{arsR} \\ f_{ABS}{}^{2} & \text{for } m_{EGFP} \end{cases} \tag{3.6}$$

with P_{AA} being the activity of the constitutive P_{AA} promoter.

In- and efflux of arsenite are modeled using reversible Michaelis-Menten kinetics. Arsenite is taken up by passive transport through the GlpF aquaglyceroporin [131]. Because it is a passive transport, we model the uptake rate v_{IN} by reversible Michaelis-Menten kinetics with symmetric properties, which has the form:

$$v_{\mathrm{IN}} = v_{\mathrm{max,IN}} \frac{[As_{\mathrm{ext}}] - [As_{\mathrm{int}}]}{[As_{\mathrm{ext}}] + [As_{\mathrm{int}}] + K_{\text{M-GlpF}}} = [\mathrm{GlpF}]\, k_{\mathrm{cat,GlpF}} \frac{[As_{\mathrm{ext}}] - [As_{\mathrm{int}}]}{[As_{\mathrm{ext}}] + [As_{\mathrm{int}}] + \dfrac{2k_{\mathrm{cat,GlpF}}}{k_{\mathrm{f}}}} \tag{3.7}$$

Where [GlpF] is the concentration of aquaglycerol facilitator proteins [125], $\mathrm{K}_{\text{M-GlpF}}$ is the Michaelis constant of the reaction that can be expressed with the forward rate constant k_{f} (see Constraint 1) and catalysis rate $\mathrm{k}_{\mathrm{cat,GlpF}}$, and $\mathrm{As}_{\mathrm{ext}}$ and $\mathrm{As}_{\mathrm{int}}$ are the concentration of extracellular and intracellular arsenite, respectively. Arsenite is effluxed by the ArsB pump by a rate $\mathrm{v}_{\mathrm{OUT}}$, which follows irreversible Michaelis-Menten kinetics:

$$v_{\mathrm{OUT}} = v_{\mathrm{max,OUT}} \frac{[As_{\mathrm{int}}]}{[As_{\mathrm{int}}] + K_{\text{M-ArsB}}} = [\mathrm{ArsB}]\, k_{\mathrm{cat,ArsB}} \frac{[As_{\mathrm{int}}]}{[As_{\mathrm{int}}] + K_{\text{M-ArsB}}} \tag{3.8}$$

Where [ArsB] is the concentration of extrusion pumps, $\mathrm{k}_{\mathrm{cat,ArsB}}$ the catalysis rate, $\mathrm{K}_{\text{M-ArsB}}$ the Michaelis constant of the pump, and $\mathrm{As}_{\mathrm{int}}$ the cytoplasmic concentration of arsenite. Production of ArsB, like ArsR, is dependent on the arsenite concentration, and for simplicity the concentration of ArsB is linked as one-third of the total ArsR concentration, following steady-state experimental measurements [125] (see Assumptions). In the initial cell conditions (t=0), ArsR proteins are present only in form of monomers, with copy numbers depending on the circuit; ABS and non-specific binding sites are free, all the other species are set to zero. In feedback (FB) circuits, there are initially 24 ArsR per plasmid (either $\mathrm{ArsR}^{\mathrm{R773}}$ or $\mathrm{ArsR}^{\mathrm{K12}}$), and 24 more $\mathrm{ArsR}^{\mathrm{K12}}$ in strains where the chromosomal copy is intact. In uncoupled (UN) circuits, the initial number ($\mathrm{ArsR}_{\mathrm{ini}}$) depends on the promoter strength, and the synthesis and degradation rates of mRNA and ArsR, according to:

$$ArsR_{\mathrm{ini}} = g_{\mathrm{p}}\, P_{AA}\, \frac{k_{\mathrm{s,mArsR}}}{k_{\mathrm{d,mArsR}}}\, \frac{k_{\mathrm{s,ArsR}}}{k_{\mathrm{d,ArsR}}} \tag{3.9}$$

The full set of species equations and all circuit configurations are listed in Supplementary Material Table 3.7 and Table 3.9, respectively. The systems of ODE were solved in Matlab (version R2014b, Mathworks).

3.3.7 Model Assumptions and Constraints

Assumptions

Assumption 1. Transcription, translation and degradation rates of m_{arsR} and ArsR are allele-independent, and independent of the location of the gene (plasmid or chromosome).

Assumption 2. ArsR dimers can occur in three types: two homodimers ($\mathrm{ArsR}^{\mathrm{R773}}$ or $\mathrm{ArsR}^{\mathrm{K12}}$) and one heterodimer ($\mathrm{ArsR}^{\mathrm{R773}}$-$\mathrm{ArsR}^{\mathrm{K12}}$). Heterodimer binding energies are obtained from the average free energy of binding of the homodimers.

Assumption 3. The binding affinity of ArsR dimers is the same for both operators in the system ($\mathrm{ABS}^{\mathrm{R773}}$ or $\mathrm{ABS}^{\mathrm{K12}}$).

Assumption 4. Transcription, translation and degradation rates are the same for $m_{arsR,egfp}$ and m_{egfp} (feedback and uncoupled circuits).

Assumption 5. Degradation rates are the same for mature and immature EGFP protein.

Assumption 6. A bioreporter cell has ten plasmid copies (each with two ABS operators), one chromosome (with one ABS operator), and 10^7 non-specific DNA binding sites (5 Mb genome, non-specific sites occurring on both strands).

Assumption 7. ArsR dimers with one or two bound arsenite molecules have the same affinities for the operator or for non-specific sites.

Assumption 8. Only ArsR dimers can bind the operator and interact with arsenite. ArsR monomers can only bind non-specific sites or dimerize.

Assumption 9. Any ArsR-DNA complex is protected from degradation.

Assumption 10. The degradation rate of an ArsR dimer is one-half of the monomer degradation rate.

Assumption 11. Degradation of an ArsR dimer bound to arsenite will lead to that arsenite being added to the intracellular pool of free arsenite.

Assumption 12. The background EGFP fluorescence of a single cell in flow cytometry after 3 h induction in absence of arsenite is equivalent to 400 EGFP molecules per cell [124].

Assumption 13. ArsR binding to $\mathrm{ABS}^{\mathrm{K12}}$ is neglected in the case of *E. coli* Δ*arsRBC*.

Assumption 14. ArsB and ArsC are formed with the same rate as $\mathrm{ArsR}^{\mathrm{K12}}$ in case of host cells having the *arsRBC* operon. Their degradation rates, however, are different.

Assumption 15. Influx of As by GlpF is unregulated and amounts of GlpF remain constant at 20 copies per cell (33 nM). Efflux of As by ArsB is regulated, and concentration of ArsB is assumed to be three times less than ArsR [125].

Constraints

Constraint 1. Forward rate constants are fixed at $k_\mathrm{f} = 10^6\ \mathrm{M^{-1}s^{-1}}$, equivalent to the diffusion rate constant.

Constraint 2. All other parameters must remain within upper and lower boundaries (as defined in Supporting Information Figure 3.9).

Constraint 3. Cells do not divide during the induction period (and no heterogeneity among cells is allowed due to the nature of the model).

Constraint 4. Induction of EGFP formation in the modeled circuits by arsenite is determined by the configuration and initial state of the circuit, which is estimated from synthesis and degradation rate constants, and promoter activity, in absence of arsenite.

3.3.8 Parameter Estimation

Background EGFP from the reporter circuit in *E. coli* was estimated from literature (~700 nM), which is equivalent to 400 molecules per cell [124]. Basal ArsR^{K12} protein levels in *E. coli* K12 are between 24 and 82 copies per cell [125] (in minimal and complex medium, respectively). The affinity of ArsR^{R773} for its operator is ~10^7 M^{-1}, which we deduced from published DNA binding experiments [126]. Further parameters were fit to experimental data sets of EGFP fluorescence expressed by *E. coli* reporter strains as a function of arsenite concentration; two configurations with the uncoupled circuits and without the chromosomal *arsRBC* operon: *E. coli ΔarsRBC* (pAAUN) or (pAAK12), and two configurations with the feedback circuits: *E. coli ΔarsRBC* (pPR-ArsR-ABS) or (pPRK12, Figure 3.1). Each set contained 21 parameters, whose starting values are uniformly distributed between their set boundaries. 20,000 parameter sets satisfying the basal ArsR^{K12} requirement were used as initial sets for the Levenberg-Marquardt scatter search algorithm [127] in order to find the best 100 parameter sets that fit the observed mean EGFP fluorescence after 3 h simulated induction with all four above-mentioned experimental observations simultaneously.

The fitness *F(p)* of the parameter set *p* was evaluated as the normalized least-square to *N* experimental data points, according to equation:

$$F(p) = \sum_{i=1}^{N} \left(\frac{y_i(p) - y_i^*}{y_i^*} \right)^2 \tag{3.10}$$

where y_i are the fluorescence levels simulated by *p* and y^*_i are the corresponding experimental values.

3.3.9 Sensitivity Analysis

The sensitivity measures the influence of each parameter on the output. The scaled sensitivity coefficient (*C*) of the fluorescence output y with respect to each parameter p_i is defined by equation (3.11):

$$C_{p_i}^{y} = \frac{dy}{dp_i}\frac{p_i}{y} \approx \frac{y(p_i+\Delta p_i)-y(p_i)}{\Delta p_i}\frac{p_i}{y(p_i)} \tag{3.11}$$

where the derivative is calculated under the finite difference approximation, using $\Delta p = 0.05\times \mathrm{p}$. The sensitivity coefficient is dimensionless and characterizes the ratio between the relative perturbations of the output and the parameter. The variability of a species is measured by its coefficient of variation, the ratio of its standard deviation over its mean

$$C = \frac{\sigma}{\mu} \tag{3.12}$$

3.3.10 Stochastic formulation Gillespie Algorithm

The ODE equations used to simulate the chemical species in the system can be written in the form

$$\frac{d\mathbf{X}}{dt} = \mathbf{S}\cdot\mathbf{v}\,, \tag{3.13}$$

where the left-hand side of the equation represents the time evolution of the vector **X** of chemical species, **S** is the stoichiometric matrix and **v** is the flux vector of the reactions (Table 3.8 and Table 3.9). The Gillespie algorithm uses the same information, *i.e.* the stoichiometry and the fluxes of the reaction, but does not provide a continuous expression for the concentrations through time. Instead, the species concentrations are discretized into copy number, using the cell volume, and the reactions occur one at a time. These discontinuities add some variability to the system, especially if regulatory species are at low copy number. In the implementation of the Gillespie algorithm, fluxes are turned into propensities to first determine the time until the next reaction happens, and then which of the reactions happened [26].

For example, a system of reactions given by

$$\begin{aligned} P &\rightarrow P + mR \\ mR &\rightarrow mR + R \\ R + R &\rightleftharpoons R_2 \\ mR &\rightarrow \varnothing \\ R &\rightarrow \varnothing \\ R_2 &\rightarrow \varnothing \end{aligned} \tag{3.14}$$

represents a promoter P that transcribes a gene into a mRNA (mR) and a protein (R) that can dimerize (R_2). The mRNA and protein have degradation. The stoichiometry of this system is given by

$$S = \left(\begin{array}{ccccccc} 1 & & & & -1 & & \\ & 1 & -2 & 2 & & -1 & \\ & & 1 & -1 & & & -1 \end{array}\right) \begin{array}{c} mR \\ R \\ R_2 \end{array} \tag{3.15}$$

and the fluxes associated with each reaction (3.14) are given by

$$v = \begin{pmatrix} k_{\mathrm{s,mR}} \mathrm{P} \\ k_{\mathrm{s,R}} [mR] \\ k_{\mathrm{f,H}} [R]^2 \\ k_{\mathrm{b,H}} [R_2] \\ k_{\mathrm{d,mR}} [mR] \\ k_{\mathrm{d,R}} [R] \\ k_{\mathrm{d,R_2}} [R_2] \end{pmatrix} . \tag{3.16}$$

The time until the next reaction depends on the flux vector. It is chosen by a random draw with the associated probability density:

$$P(\tau) = a \exp(-a\tau) \tag{3.17}$$

where $a = \sum_i v_i$. Next, a second random draw selects the reaction that occurred, proportionally to its flux.

Hence, from two numbers drawn from a uniform distribution, we obtain the time that has elapsed τ and the index of the reaction that occurred j :

$$\begin{array}{l} \tau = \dfrac{1}{\ln(r_1)} \dfrac{1}{a} \\ \dfrac{\sum_{i=1}^{j-1} v_i}{a} \leq r_2 < \dfrac{\sum_{i=1}^{j} v_i}{a} \end{array} \qquad \text{with} \begin{cases} r_1, r_2 \sim \mathrm{unif}(0,1) \\ a = \sum_i v_i \end{cases} \tag{3.18}$$

Once the elapsed time and reaction are determined, the system is updated by adding the column j of the stoichiometric matrix to the system state.

Table 3.4 **Stochastic model reactions**

$\varnothing \rightarrow m_{\text{ArsR-EGFP}}$

$\varnothing \rightarrow m_{\text{ArsR}}$

$m_{\text{ArsR}} \rightarrow m_{\text{ArsR}} + ArsR$

$m_{\text{ArsR-EGFP}} \rightarrow m_{\text{ArsR-EGFP}} + \text{EGFP}$

$ArsR + ArsR \rightleftharpoons ArsR_2$

$O_1 + ArsR_2 \rightleftharpoons O_1{:}ArsR_2$

$O_2 + ArsR_2 \rightleftharpoons O_2{:}ArsR_2$

$ArsR_2 + As \rightleftharpoons ArsR_2{:}As$

$ArsR_2{:}As + As \rightleftharpoons ArsR_2{:}As_2$

$O_1 + ArsR_2{:}As \rightleftharpoons O_1{:}ArsR_2{:}As$

$O_2 + ArsR_2{:}As \rightleftharpoons O_2{:}ArsR_2{:}As$

$O_1 + ArsR_2{:}As_2 \rightleftharpoons O_1{:}ArsR_2{:}As_2$

$O_2 + ArsR_2{:}As_2 \rightleftharpoons O_2{:}ArsR_2{:}As_2$

$U + ArsR \rightleftharpoons U{:}ArsR$

$U + ArsR_2 \rightleftharpoons U{:}ArsR_2$

$U + ArsR_2{:}As \rightleftharpoons U{:}ArsR_2{:}As$

$U + ArsR_2{:}As_2 \rightleftharpoons U{:}ArsR_2{:}As_2$

$m_{\text{ArsR}} \rightarrow \varnothing$

$m_{\text{ArsR-EGFP}} \rightarrow \varnothing$

$ArsR \rightarrow \varnothing$

$ArsR_2 \rightarrow \varnothing$

$ArsR_2{:}As \rightarrow As$

$ArsR_2{:}As_2 \rightarrow 2As$

$EGFP \rightarrow \varnothing$

$EGFP \rightarrow EGFP^*$

$EGFP^* \rightarrow \varnothing$

$As_{\text{ext}} \rightarrow As$

3.4 Conclusion

We built and validated a mechanistic model of different feedback and uncoupled gene circuits for arsenic reporting in *E. coli*, based on the ArsR-P$_{\text{ars}}$ topology (Figure 3.2). The main model parameters were defined and estimated from a subset of experimental data with reporter strains carrying a deletion of the chromosomal *arsRBC* operon (Figure 3.3). The parameter values were then set and used to simulate bioreporter strains carrying a full *arsRBC* operon, which allowed estimating global Michaelis-Menten constants of the pumps (Figure 3.6). The simulations are in good agreement with experimental data, suggesting that no major processes or interactions are missing from the model (Figure 3.2, Table 3.1), and that the parametrized values have biological relevance (Table 3.2). Finally, we used the model to simulate the bioreporter response to a partial deletion of the *arsRBC* operon, leaving the chromosomal *arsR*$^{\text{K12}}$ intact. The model predicts a slight increase of signal when both *arsR* alleles are present (e.g., reporter plasmids pAAUN and pPR-ars-ABS), but this could not be confirmed by experimental observations, although both experimental and model output variations may mask these effects (Figure 3.7). Hence, if heterodimers are produced in these strains, their behaviour is different than assumed in the model.

Within every host background, bioreporters based on *arsR*$^{\text{K12}}$ instead of the original *arsR*$^{\text{R773}}$ were more sensitive to arsenite. The model suggests that the reason for this is the weaker interaction of ArsR$^{\text{K12}}$ than ArsR$^{\text{R773}}$ protein with the operator site (ABS). In contrast, ArsR$^{\text{K12}}$ dimers would have higher affinity for arsenite than ArsR$^{\text{R773}}$, which explains the higher response at low arsenite concentrations. The chromosomal *arsRBC* operon may thus entail *E. coli* with a specific response at lower arsenite concentrations than the R773 plasmid *arsRDABC* operon.

The advance that our model is making is that it describes in realistic detail all the pairwise reactions among the chemical and biological species (i.e., DNA, proteins, arsenite), as well as translation, transcription and transport rates. The model is a numerical extension of a previous algebraic attempt, which could describe general trends but without specific correct parameter values [23]. In most gene circuit models, nonessential mechanisms are neglected (e.g. non-specific binding) [142] or simplified (e.g. Hill kinetics [129], [143], constant intracellular effector level [144]) for practical reasons. In the model proposed and validated here, the obtained parameter values correspond to and characterize the impact of the mechanical steps in a more quantitative manner. We do acknowledge that the drawback of a large model is that various combinations of parameter sets will satisfy the observed data, as a consequence of the large number of estimated parameters. For this reason, we plot the 100 best fitting sets, the overall best fit and the average of all parameter sets (e.g., Figure 3.3 and Figure 3.6) from the parameter scatter search algorithm. Despite variation in the 100 best fitting parameter distributions (Supporting Information Figure 3.9), the simulations performed with the average parameter set as reported in Tables 1 and 2 are representative of the best fitting sets.

In addition, the model produces quantitatively correct EGFP fluorescence values. This makes the model able to take into account for allelic variations, gene deletions, promoter strengths or rewiring of the circuit. Furthermore, the output can be recalculated if specific parameters (e.g., transcription rates) need to be adjusted from experimental measurements. By changing parameter values a model should be able to capture the output of similar gene circuit architectures, which could be useful for optimization of other bioreporters.

Because of the used fitting algorithm that equally values experimental observations at low and high arsenite concentrations the predictions of the model are slightly poorer at higher arsenite concentrations (e.g., Figure 3.7). In contrast, the background fluorescence obtained by the model is overall in good agreement with experimental data. The reason for the model not being able to perfectly fit both low and high arsenite concentration ends may come from Assumptions 1 and 4 that neglect the effects of the secondary DNA and mRNA structure on transcription and translation [145], [146]. The current formulation of the model is deterministic, which is appropriate for averaged triplicate assays and less computationally demanding to perform parameter estimation. As a consequence, the model does not account for cell-to-cell variability, which is an important aspect of gene regulatory networks. The current model is further constrained by absence of cell division of the reporter cells, which may have to be adjusted for applications requiring cellular growth.

At a more practical point of view, the model allowed a number of important improvements to existing arsenic bioreporters. Most importantly, the reporter circuits incorporating the *arsR*^{K12} allele (e.g., pAAK12 and pPRK12) improve detection at low arsenite concentrations, reaching 2.5 and 3.1 fold inductions at 2.5 μg As l^{-1} with Δ*arsRBC* and Δ*arsBC* hosts, respectively, instead of between 1.1-1.3 for the *arsR*^{R773} allele (Table 3.3). The different circuits also offer more flexibility to tune responses to the expected range of concentration measurements. For example, the shapes of the calibration curves in the different circuit configurations can be applied either for broader concentration ranges or more sensitive detection at low concentrations. Both pAAUN and pPR-arsR-ABS circuits in all hosts show a broad linear response over the range of 0–150 μg l^{-1}, whereas the pAAK12 and pPRK12 circuits display a steeper response. The steeper response under maintenance of the absolute fluorescence signal output results in better measurements in the range of 2.5–20 μg As l^{-1} (Figure 3.3, Table 3.3). The increase in absolute signal intensity is crucial when detecting the EGFP signals with cheap low sensitivity detectors in portable field-scale instruments [147]. In conclusion, the thorough modeling and experimental validations of a seemingly uncomplicated system such as ArsR-P$_{ars}$ yielded an intrinsically robust system that can be exploited for optimization and fine-tuning. The broader outcome of this model is that provides a toolbox that can now be further deployed and adapted for the optimization of other gene circuits of application interest.

3.5 Supplementary Material

Table 3.5 **Bacterial strains used in the study**

Strain number	**Name**	**Plasmid**[a]	**Reference**
3307	*E. coli* MG1655	pAAUN	[23]
3316	*E. coli* MG1655 Δ*arsRBC*	pPR-arsR-ABS	This study
3328	*E. coli* MG1655	pPR-arsR-ABS	This study
3333	*E. coli* MG1655 Δ*arsBC*	pPR-arsR-ABS	This study
3391	*E. coli* MG1655 Δ*arsRBC*	pAAUN	This study
3392	*E. coli* MG1655 Δ*arsBC*	pAAUN	This study
4758	*E. coli* MG1655 Δ*arsRBC*	pPRK12	This study
4759	*E. coli* MG1655 Δ*arsBC*	pPRK12	This study
4760	*E. coli* MG1655	pPRK12	This study
4761	*E. coli* MG1655 Δ*arsRBC*	pAAK12	This study
4762	*E. coli* MG1655 Δ*arsBC*	pAAK12	This study
4763	*E. coli* MG1655	pAAK12	This study

Table 3.6 **Primers used in this study**.

Number	Length (nt)	Tm (°C)	%GC	Sequence 5'-3'	Remark
100107	27	59.9	40.7	GAATTCTTGGTATGGACGAAATGTTGC	EcoRI, furthest upstream *arsR*^{K12}, *arsRBC* deletion
100109	26	61.9	42.3	GAATTCCTTTGAAAGCGTTTATGCGC	EcoRI, *arsBC* deletion
100108	30	65.2	50.0	ACTAGTCGCTTCTGACATATTGCGCTCCTG	SpeI, *arsRBC* deletion up
100110	30	62.0	46.7	ACTAGTCGCTTCAGTAACATAATGCCTCCC	SpeI, *arsBC* deletion
100111	27	60.8	48.1	GAAGCGACTAGTCGCCTGAAATAAAGC	SpeI, downstream fragment
100112	30	59.2	40.0	GGGATCCCATATTGATCAGAGATATATCCT	BamHI, downstream fragment
120101				cttGGATCCaaaggagaggggaaATGTCATTTCTGTTACCCATCCAATTG	BamHI site and the *arsR*^{R773} ribosome binding site
120102				tctagaTTAACTGCAAATGTTCTTACTGTCCC	XbaI site

For restriction positions, see Figure 3.1 and Figure 3.12

Construction of the mathematical model

The systems of ordinary differential equations (ODE) $\frac{df}{dt}$ for the different strains are obtained by multiplication of the stoichiometric matrix S (Table 3.9) by the fluxes $v = [v_1 \dots v_{104}]^{\mathrm{T}}$ (Table 3.8), *i.e.* $\frac{df}{dt} = Sv$. The different circuits and deletions are obtained by applying the Table 3.7. Heterodimer binding energies are obtained from the average free energy of binding of the homodimers, e.g. $K_{A2}^{\mathrm{xp}} = \exp\left[\left(\ln(K_{A2}^{\mathrm{x}}) + \ln(K_{A2}^{\mathrm{p}})\right)/2\right]$.

Table 3.7 **Constraints on fluxes, plasmid and chromosomal types for construction of bioreporter models.**

Feedback (FB)				**Plasmid (p)**	**Chromosome (x)**
	pPR-arsR-ABS	$\Delta arsRBC$	$v_{36}, \dots, v_{104} = 0$	R773	
		$\Delta arsBC$	$v_{102}, \dots, v_{104} = 0$	R773	K12
		WT		R773	K12
	pPRK12	$\Delta arsRBC$	$v_{36}, \dots, v_{104} = 0$	K12	
		$\Delta arsBC$	$v_{102}, \dots, v_{104} = 0$	K12	K12
		WT		K12	K12
Uncoupled (UN)					
In UN circuits: $v_1 = \mathrm{k}_{\mathrm{s,mArsR}} P_{AA} [g_p]$ and $v_6 = 0$.	pAAUN	$\Delta arsRBC$	$v_{36}, \dots, v_{104} = 0$	R773	
		$\Delta arsBC$	$v_{102}, \dots, v_{104} = 0$	R773	K12
		WT		R773	K12
	pAAK12	$\Delta arsRBC$	$v_{36}, \dots, v_{104} = 0$	K12	
		$\Delta arsBC$	$v_{102}, \dots, v_{104} = 0$	K12	K12
		WT		K12	K12

Table 3.8 **Reaction fluxes used in the models.**

$$v_{1} = \mathrm{k}_{\mathrm{s,mArsR}}\; f_{\mathrm{ABS}^{\mathrm{p}}}\left(1-f_{\mathrm{ABS}^{\mathrm{p}}}\right)[g_{\mathrm{p}}]$$

$$v_{2} = \mathrm{k}_{\mathrm{s,mArsR}}\left(f_{\mathrm{ABS}^{\mathrm{p}}}\right)^{2}[g_{\mathrm{p}}]$$

$$v_{3} = \mathrm{k}_{\mathrm{s,EGFP}}[\mathrm{m}_{\mathrm{ArsR\text{-}EGFP}}]$$

$$v_{4} = \mathrm{k}_{\mathrm{mat,EGFP}}[\mathrm{EGFP}]$$

$$v_{5} = \mathrm{k}_{\mathrm{s,ArsR}}[\mathrm{m}_{\mathrm{ArsR}^{\mathrm{p}}}]$$

$$v_{6} = \mathrm{k}_{\mathrm{s,ArsR}}[\mathrm{m}_{\mathrm{ArsR\text{-}EGFP}}]$$

$$v_{7} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{p}}]^{2}$$

$$v_{8} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}]/\mathrm{K}_{\mathrm{H}}^{\mathrm{p}}$$

$$v_{9} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{p}}][\mathrm{As}]$$

$$v_{10} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{C1}}^{\mathrm{p}}$$

$$v_{11} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{p}}{:}\mathrm{As}][\mathrm{As}]$$

$$v_{12} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{C2}}^{\mathrm{p}}$$

$$v_{13} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}][\mathrm{ABS}^{\mathrm{p}}]$$

$$v_{14} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{p}}$$

$$v_{15} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}][\mathrm{ABS}^{\mathrm{p}}]$$

$$v_{16} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{p}}$$

$$v_{17} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}][\mathrm{ABS}^{\mathrm{p}}]$$

$$v_{18} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D2}}^{\mathrm{p}}$$

$$v_{19} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{p}}][\mathrm{U}]$$

$$v_{20} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}^{\mathrm{p}}]/\mathrm{K}_{\mathrm{B1}}^{\mathrm{p}}$$

$$v_{21} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}][\mathrm{U}]$$

$$v_{22} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}]/\mathrm{K}_{\mathrm{B2}}^{\mathrm{p}}$$

$$v_{23} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}][\mathrm{U}]$$

$$v_{24} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{E1}}^{\mathrm{p}}$$

$$v_{25} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}][\mathrm{U}]$$

$$v_{26} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{E2}}^{\mathrm{p}}$$

$$v_{27} = \mathrm{k}_{\mathrm{d,mArsR}}[\mathrm{m}_{\mathrm{ArsR}^{\mathrm{p}}}]$$

$$v_{28} = \mathrm{k}_{\mathrm{d,m\text{-}ArsR\text{-}EGFP}}[\mathrm{m}_{\mathrm{ArsR\text{-}EGFP}}]$$

$$v_{29} = \mathrm{k}_{\mathrm{d,EGFP}}[\mathrm{EGFP}]$$

$$v_{30} = \mathrm{k}_{\mathrm{d,EGFP}}[\mathrm{EGFP}^{*}]$$

$$v_{31} = \mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}^{\mathrm{p}}]$$

$$v_{32} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{p}}]$$

$$v_{33} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]$$

$$v_{34} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]$$

$$v_{35} = [\mathrm{GlpF}]\frac{\mathrm{k}_{\mathrm{cat,GlpF}}\left(\mathrm{As}_{\mathrm{ext}}-\mathrm{As}\right)}{\mathrm{As}_{\mathrm{ext}}+\mathrm{As}+\dfrac{2\,\mathrm{k}_{\mathrm{cat,GlpF}}}{\mathrm{k}_{\mathrm{f}}}}$$

$$v_{36} = \mathrm{k}_{\mathrm{s,mArsR}}\; f_{\mathrm{ABS}^{\mathrm{x}}}$$

$$v_{37} = \mathrm{k}_{\mathrm{s,ArsR}}[\mathrm{m}_{\mathrm{ArsR}^{\mathrm{x}}}]$$

$$v_{38} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{x}}]^{2}$$

$$v_{39} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}]/\mathrm{K}_{\mathrm{H}}^{\mathrm{x}}$$

$$v_{40} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}][\mathrm{As}]$$

$$v_{41} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{C1}}^{\mathrm{x}}$$

$$v_{42} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{x}}{:}\mathrm{As}][\mathrm{As}]$$

$$v_{43} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{C2}}^{\mathrm{x}}$$

$$v_{44} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}][\mathrm{ABS}^{\mathrm{x}}]$$

$$v_{45} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{x}}$$

$$v_{46} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}][\mathrm{ABS}^{\mathrm{x}}]$$

$$v_{47} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{x}}$$

$$v_{48} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}][\mathrm{ABS}^{\mathrm{x}}]$$

$$v_{49} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D2}}^{\mathrm{x}}$$

$$v_{50} = \mathrm{k}_{\mathrm{d,mArsR}}[\mathrm{m}_{\mathrm{ArsR}^{\mathrm{x}}}]$$

$$v_{51} = \mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}^{\mathrm{x}}]$$

$$v_{52} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{x}}]$$

$$v_{53} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]$$

$$v_{54} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]$$

$$v_{55} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{p}}]$$

$$v_{56} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{p}}$$

$$v_{57} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]$$

$$v_{58} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{p}}$$

$$v_{59} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]$$

$$v_{60} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{p}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D2}}^{\mathrm{p}}$$

$$v_{61} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{x}}]$$

$$v_{62} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{x}}$$

$$v_{63} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]$$

$$v_{64} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{x}}$$

$$v_{65} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]$$

$$v_{66} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{x}}$$

$$v_{67} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{x}}][\mathrm{ArsR}^{\mathrm{p}}]$$

$$v_{68} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}]/\mathrm{K}_{\mathrm{H}}^{\mathrm{xp}}$$

$$v_{69} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}][\mathrm{As}]$$

$$v_{70} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{C1}}^{\mathrm{xp}}$$

$$v_{71} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}][\mathrm{As}]$$

$$v_{72} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{C2}}^{\mathrm{xp}}$$

$$v_{73} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}]$$

$$v_{74} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{xp}}$$

$$v_{75} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]$$

$$v_{76} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{xp}}$$

$$v_{77} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]$$

$$v_{78} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{x}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D2}}^{\mathrm{xp}}$$

$$v_{79} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}]$$

$$v_{80} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}]/\mathrm{K}_{\mathrm{A2}}^{\mathrm{xp}}$$

$$v_{81} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]$$

$$v_{82} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{D1}}^{\mathrm{xp}}$$

$$v_{83} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}][\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]$$

$$v_{84} = \mathrm{k}_{\mathrm{f}}[\mathrm{ABS}^{\mathrm{p}}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{D2}}^{\mathrm{xp}}$$

$$v_{85} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}][\mathrm{U}]$$

$$v_{86} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}]/\mathrm{K}_{\mathrm{B2}}^{\mathrm{xp}}$$

$$v_{87} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}][\mathrm{U}]$$

$$v_{88} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{E1}}^{\mathrm{xp}}$$

$$v_{89} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}][\mathrm{U}]$$

$$v_{90} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{E2}}^{\mathrm{xp}}$$

$$v_{91} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}^{\mathrm{x}}][\mathrm{U}]$$

$$v_{92} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}^{\mathrm{x}}]/\mathrm{K}_{\mathrm{B1}}^{\mathrm{x}}$$

$$v_{93} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}][\mathrm{U}]$$

$$v_{94} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}]/\mathrm{K}_{\mathrm{B2}}^{\mathrm{x}}$$

$$v_{95} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}][\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]$$

$$v_{96} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}]/\mathrm{K}_{\mathrm{E1}}^{\mathrm{x}}$$

$$v_{97} = \mathrm{k}_{\mathrm{f}}[\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}][\mathrm{U}]$$

$$v_{98} = \mathrm{k}_{\mathrm{f}}[\mathrm{U}{:}\mathrm{ArsR}_{2}^{\mathrm{x}}{:}\mathrm{As}_{2}]/\mathrm{K}_{\mathrm{E2}}^{\mathrm{x}}$$

$$v_{99} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}]$$

$$v_{100} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}]$$

$$v_{101} = 1/2\,\mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{ArsR}_{2}^{\mathrm{xp}}{:}\mathrm{As}_{2}]$$

$$v_{102} = 1/3\,\mathrm{k}_{\mathrm{s,ArsR}}[\mathrm{mArsR}^{\mathrm{x}}]$$

$$v_{103} = \mathrm{k}_{\mathrm{d,ArsR}}[\mathrm{mArsR}^{\mathrm{x}}]$$

$$v_{104} = [\mathrm{ArsB}]\mathrm{k}_{\mathrm{cat,ArsB}}\frac{\mathrm{As}}{\mathrm{As}+\mathrm{K}_{\mathrm{M\text{-}ArsB}}}$$

Table 3.9 **Stoichiometric matrix used for the construction of the models.**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
mArsR	1																									
mArsR-EGFP		1																								
EGFP			1	-1																						
EGFP*				1																						
ArsRp					1	1	-2	2											-1	1						
ArsRp2							1	-1	-1	1			-1	1							-1	1				
ArsRp2:As									1	-1	-1	1			-1	1							-1	1		
ArsRp2:As2											1	-1					-1	1							-1	1
As									-1	1	-1	1														
ABSp													-1	1	-1	1	-1	1								
ABSp:ArsRp2													1	-1												
ABSp:ArsRp2:As															1	-1										
ABSp:ArsRp2:As2																	1	-1								
U																			-1	1	-1	1	-1	1	-1	1
U:ArsRp																			1	-1						
U:ArsRp2																					1	-1				
U:ArsRp2:As																							1	-1		
U:ArsRp2:As2																									1	-1
mArsRx																										
ArsRx																										
ArsRx2																										
ArsRx2:As																										
ArsRx2:As2																										
ABSx																										
ABSx:ArsRx2																										
ABSx:ArsRx2:As																										
ABSx:ArsRx2:As2																										
ABSx:ArsRp2																										
ABSx:ArsRp2:As																										
ABSx:ArsRp2:As2																										
ABSp:ArsRx2													1	-1												
ABSp:ArsRx2:As															1	-1										
ABSp:ArsRx2:As2																	1	-1								
ArsRxp																										
ArsRxp:As																										
ArsRxp:As2																										
ABSx:ArsRxp																										
ABSx:ArsRxp:As																										
ABSx:ArsRxp:As2																										
ABSp:ArsRxp																										
ABSp:ArsRxp:As																										
ABSp:ArsRxp:As2																										
U:ArsRxp																										
U:ArsRxp:As																										
U:ArsRxp:As2																										
U:ArsRx																										
U:ArsRx2																										
U:ArsRx2:As																										
U:ArsRx2:As2																										
ArsB																										

	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
mArsR	-1																									
mArsR-EGFP		-1																								
EGFP			-1																							
EGFP*				-1																						
ArsRp					-1																					
ArsRp2						-1																				
ArsRp2:As							-1																			
ArsRp2:As2								-1																		
As							1	2	1					-1	1	-1	1									
ABSp																										
ABSp:ArsRp2																										
ABSp:ArsRp2:As																										
ABSp:ArsRp2:As2																										
U																										
U:ArsRp																										
U:ArsRp2																										
U:ArsRp2:As																										
U:ArsRp2:As2																										
mArsRx										1														-1		
ArsRx											1	-2	2												-1	
ArsRx2												1	-1	-1	1			-1	1							-1
ArsRx2:As														1	-1	-1	1			-1	1					
ArsRx2:As2																1	-1					-1	1			
ABSx																		-1	1	-1	1	-1	1			
ABSx:ArsRx2																		1	-1							
ABSx:ArsRx2:As																				1	-1					
ABSx:ArsRx2:As2																						1	-1			
ABSx:ArsRp2																										
ABSx:ArsRp2:As																										
ABSx:ArsRp2:As2																										
ABSp:ArsRx2																										
ABSp:ArsRx2:As																										
ABSp:ArsRx2:As2																										
ArsRxp																										
ArsRxp:As																										
ArsRxp:As2																										
ABSx:ArsRxp																										
ABSx:ArsRxp:As																										
ABSx:ArsRxp:As2																										
ABSp:ArsRxp																										
ABSp:ArsRxp:As																										
ABSp:ArsRxp:As2																										
U:ArsRxp																										
U:ArsRxp:As																										
U:ArsRxp:As2																										
U:ArsRx																										
U:ArsRx2																										
U:ArsRx2:As																										
U:ArsRx2:As2																										
ArsB																										

	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
mArsR																										
mArsR-EGFP																										
EGFP																										
EGFP*																										
ArsRp															-1	1										
ArsRp2			-1	1																						
ArsRp2:As					-1	1																				
ArsRp2:As2							-1	1																		
As	1	2															-1	1	-1	1						
ABSp									-1	1	-1	1	-1	1												
ABSp:ArsRp2																										
ABSp:ArsRp2:As																										
ABSp:ArsRp2:As2																										
U																										
U:ArsRp																										
U:ArsRp2																										
U:ArsRp2:As																										
U:ArsRp2:As2																										
mArsRx																										
ArsRx															-1	1										
ArsRx2									-1	1																
ArsRx2:As	-1										-1	1														
ArsRx2:As2		-1										-1	1												
ABSx			-1	1	-1	1	-1	1													-1	1	-1	1	-1	1
ABSx:ArsRx2																										
ABSx:ArsRx2:As																										
ABSx:ArsRx2:As2																										
ABSx:ArsRp2			1	-1																						
ABSx:ArsRp2:As					1	-1																				
ABSx:ArsRp2:As2							1	-1																		
ABSp:ArsRx2									1	-1																
ABSp:ArsRx2:As											1	-1														
ABSp:ArsRx2:As2													1	-1												
ArsRxp															1	-1	-1	1			-1	1				
ArsRxp:As																	1	-1	-1	1			-1	1		
ArsRxp:As2																			1	-1					-1	1
ABSx:ArsRxp																					1	-1				
ABSx:ArsRxp:As																							1	-1		
ABSx:ArsRxp:As2																									1	-1
ABSp:ArsRxp																										
ABSp:ArsRxp:As																										
ABSp:ArsRxp:As2																										
U:ArsRxp																										
U:ArsRxp:As																										
U:ArsRxp:As2																										
U:ArsRx																										
U:ArsRx2																										
U:ArsRx2:As																										
U:ArsRx2:As2																										
ArsB																										

	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	##	101	102	103	104
mArsR																										
mArsR-EGFP																										
EGFP																										
EGFP*																										
ArsRp																										
ArsRp2																										
ArsRp2:As																										
ArsRp2:As2																										
As																						1	2			-1
ABSp	-1	1	-1	1	-1	1																			
ABSp:ArsRp2																										
ABSp:ArsRp2:As																										
ABSp:ArsRp2:As2																										
U							-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1						
U:ArsRp																										
U:ArsRp2																										
U:ArsRp2:As																										
U:ArsRp2:As2																										
mArsRx																										
ArsRx													-1	1												
ArsRx2															-1	1										
ArsRx2:As																	-1	1								
ArsRx2:As2																			-1	1						
ABSx																										
ABSx:ArsRx2																										
ABSx:ArsRx2:As																										
ABSx:ArsRx2:As2																										
ABSx:ArsRp2																										
ABSx:ArsRp2:As																										
ABSx:ArsRp2:As2																										
ABSp:ArsRx2																										
ABSp:ArsRx2:As																										
ABSp:ArsRx2:As2																										
ArsRxp	-1	1					-1	1													-1					
ArsRxp:As			-1	1					-1	1												-1				
ArsRxp:As2					-1	1					-1	1										-1			
ABSx:ArsRxp																										
ABSx:ArsRxp:As																										
ABSx:ArsRxp:As2																										
ABSp:ArsRxp	1	-1																							
ABSp:ArsRxp:As			1	-1																					
ABSp:ArsRxp:As2					1	-1																			
U:ArsRxp							1	-1																	
U:ArsRxp:As									1	-1															
U:ArsRxp:As2											1	-1													
U:ArsRx													1	-1											
U:ArsRx2															1	-1									
U:ArsRx2:As																	1	-1							
U:ArsRx2:As2																			1	-1					
ArsB																								1	-1	

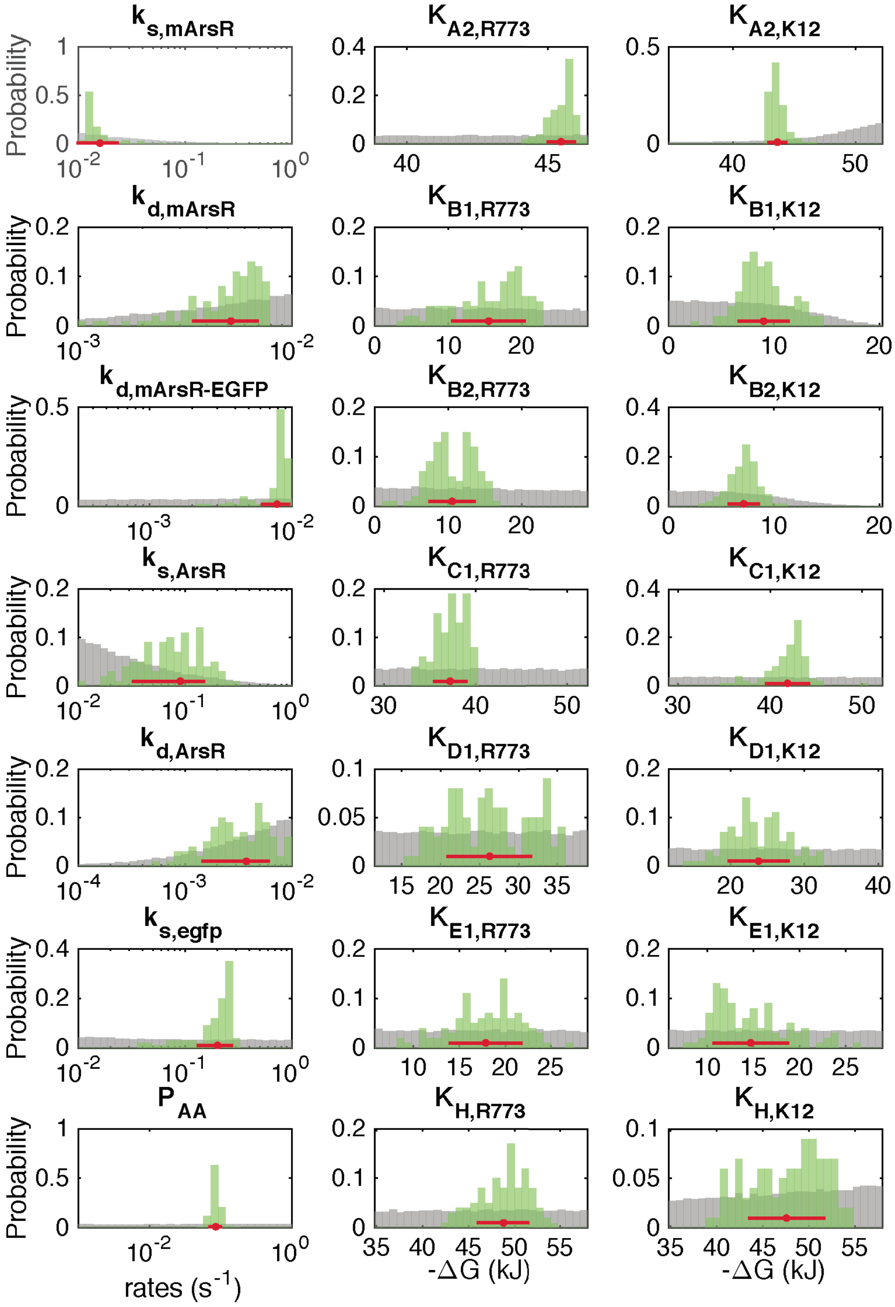


Figure 3.9 **Distribution of starting population (gray) and top 100 best individual parameter values after the scatter search parameter estimation algorithm (green).** Mean values and standard deviations are indicated by the red dot and lines, respectively.

BLASTP

```
Query: P15905, ArsR R773
Sbjct: AAC76526, ArsR Escherichia coli str. K-12 substr. MG1655

Score =   185 bits (469),  Expect = 1e-65, Method: Compositional matrix adjust.
 Identities = 87/113 (77%), Positives = 99/113 (88%), Gaps = 0/113 (0%)

Query  4    LTPLQLFKNLSDETRLGIVLLLREMGELCVCDLCMALDQSQPKISRHLAMLRESGILLDR  63
            L P+QLFK L+DETRLGIVLLL E+GELCVCDLC ALDQSQPKISRHLA+LRESG+LLDR
Sbjct  4    LLPIQLFKILADETRLGIVLLLSELGELCVCDLCTALDQSQPKISRHLALLRESGLLLDR  63

Query  64   KQGKWVHYRLSPHIPSWAAQIIEQAWLSQQDDVQVIARKLASVNCSGSSKAVC  116
            KQGKWVHYRLSPHIP+WAA+II++AW  +Q+ VQ I R LA  NCSG SK +C
Sbjct  64   KQGKWVHYRLSPHIPAWAAKIIDEAWRCEQEKVQAIVRNLARQNCSGDSKNIC  116
```

BLASTN + manual upstream start site

Alignment statistics for match #1

Score	**Expect**	**Identities**
230 bits(254)	2e-60	272/367(74%)

```
>gi|42716|emb|X16045.1|:1-512 E. coli R-factor R773 arsR gene
>gi|545778205|gb|U00096.3|:3648400-3648914  Escherichia coli str. K-12 substr.
MG1655, complete genome

Query  1        ----GAATT-CCAAGTTA--TCTCACCTACCT-TAAGGTAATAGTGT  39
                    ::::  :::: : :    ::::  :::: : :  :
Sbjct  3648400  AAATGAATAGCCAACTCAAAATTCAC--ACCTATTACCTTCCTCT


                     ArsR binding site           -35                  -10
Query  40       GATTAATCATATGCGTTTTTGGTTATGTGTTGTTTGACTTAATATCAGAGCCGAGAGATA  99
                : :    :: :: ::::      ::: :::: :::::::::    :   :  :::::: :
Sbjct  3648442  GCACTTACACATTCGTTAAGTCATATATGTT-TTTGACTTATCCGCTTCGAAGAGAGACA

Query  100      CTTGTTTTCTACAAA--GGAGAGGGAAATGTTGCAACTAACACCACTTCAGTTATTTAAA
145
                ::        ::::   :::: :  : |||             || ||| ||||   |
Sbjct  3648502  CTACCTGC-AACAATCAGGAGCGCAATATG------------TCATTTCTGTTACCCATC

Query  146      CAGTTATTTAAAAACCTGTCCGATGAAACCCGTTTGGGTATCGTGTTGTTGCTCAGGGAG
205
```

```
                || || || ||||  ||  | |||||||||||| |||| ||||| ||  ||||||| ||
Sbjct  3648549  CAATTGTTCAAAATTCTTGCTGATGAAACCCGTCTGGGCATCGTTTTACTGCTCAGCGAA

Query  206      ATGGGAGAGTTGTGCGTGTGTGATCTTTGCATGGCACTGGATCAATCACAGCCCAAAATA
265
                 |||||||||| ||||| || ||||| ||||  || || || || || |||||||| ||
Sbjct  3648609  CTGGGAGAGTTATGCGTCTGCGATCTCTGCACTGCTCTCGACCAGTCGCAGCCCAAGATC

Query  266      TCCCGTCATCTGGCGATGCTACGGGAAAGTGGAATCCTTCTGGATCGTAAACAGGGAAAA
325
                ||||| || |||||  |||| || ||||| ||  |  | ||||| || || || || ||
Sbjct  3648669  TCCCGCCACCTGGCATTGCTGCGTGAAAGCGGGCTATTGCTGGACCGCAAGCAAGGTAAG

Query  326      TGGGTTCACTACCGCTTATCACCGCATATTCCTTCATGGGCTGCCCAGATTATTGAGCAG
385
                |||||||| |||||||||||||||||||||||  ||||||| ||  | ||||||||  ||
Sbjct  3648729  TGGGTTCATTACCGCTTATCACCGCATATTCCAGCATGGGCGGCGAAAATTATTGATGAG

Query  386      GCCTGGTTAAGCCAACAGGACGACGTTCAGGTCATCGCACGCAAGCTGGCTTCAGTTAAC
445
                ||||||  | |  ||||||| | |||||||  || |  ||||| ||||||  |   |||
Sbjct  3648789  GCCTGGCGATGTGAACAGGAAAAGGTTCAGGCGATTGTCCGCAACCTGGCTCGACAAAAC

Query  446      TGCTCCGGTAGCAGTAAGGCTGTCTGCATCTAAAAAATTTGCCTGAACATATATGTTTTA
505
                || |||||   |||||||    | ||||  | ||||||||  || |||| |||||  ||
Sbjct  3648849  TGTTCCGGGGACAGTAAGAACATTTGCA-GTTAAAAATTTAGCTAAACACATATGAATTT

Query  506      TCAAATG
                ||| |||
Sbjct  3648908  TCA ATG
```

Figure 3.10 **Sequence alignments of the ArsR^{K12} and the ArsR^{R773} proteins (upper part) and of the corresponding nucleotide sequences, including the promoter regions (lower part).**

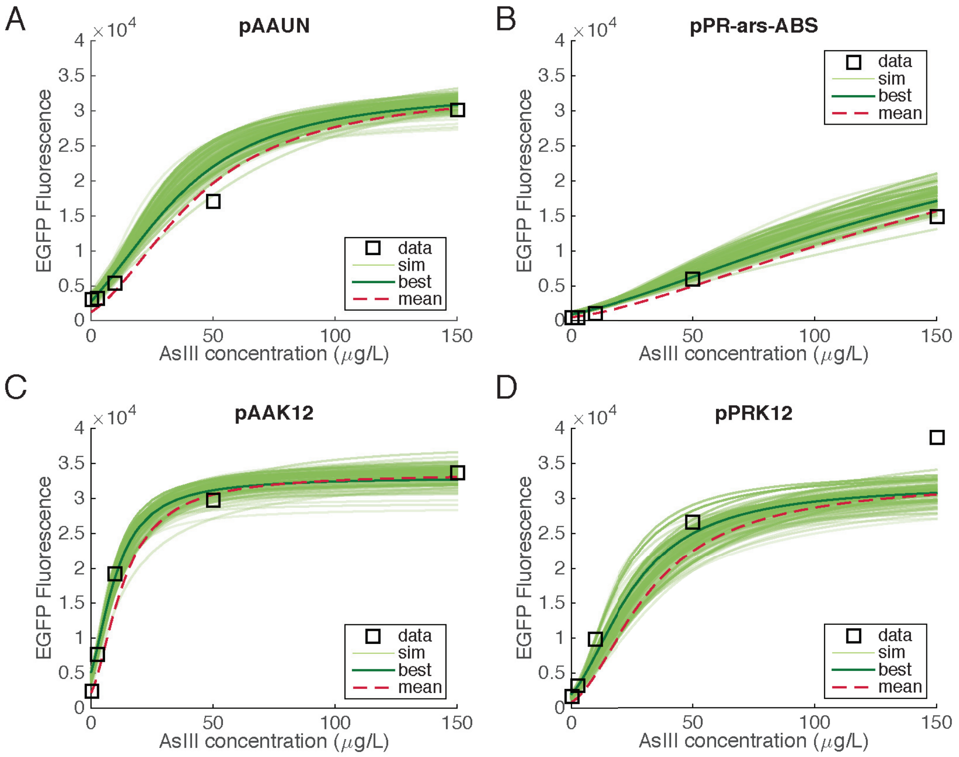


Figure 3.11 **Experimentally observed and modeled EGFP fluorescence output as a function of arsenite-As$_{III}$ concentrations after 3 h induction for the four plasmid bioreporter configurations in *E. coli* MG1655 Δ*arsBC* (including the chromosomal *arsR* gene).** Lines (*sim*) show results of the best 100 fits from the parameter estimation, simulated and plotted in all different configurations. Line darkness increases with the overall fitting score, the overall best fit (*best*) being represented with the darkest, and the average of all parameter sets (*mean*) with a dashed line. Data points show the mean of independent biological triplicates. Error bars are smaller than the used symbol size and are therefore not indicated.

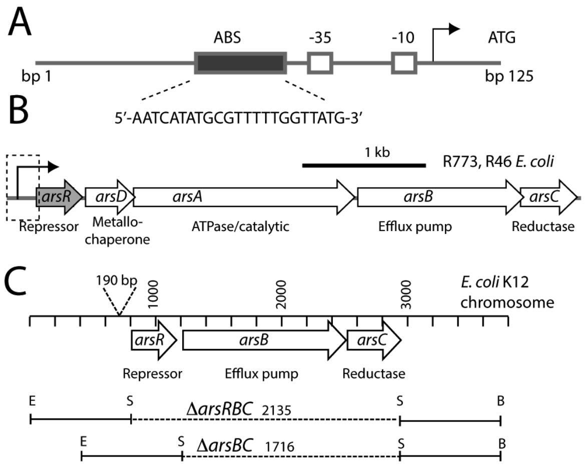
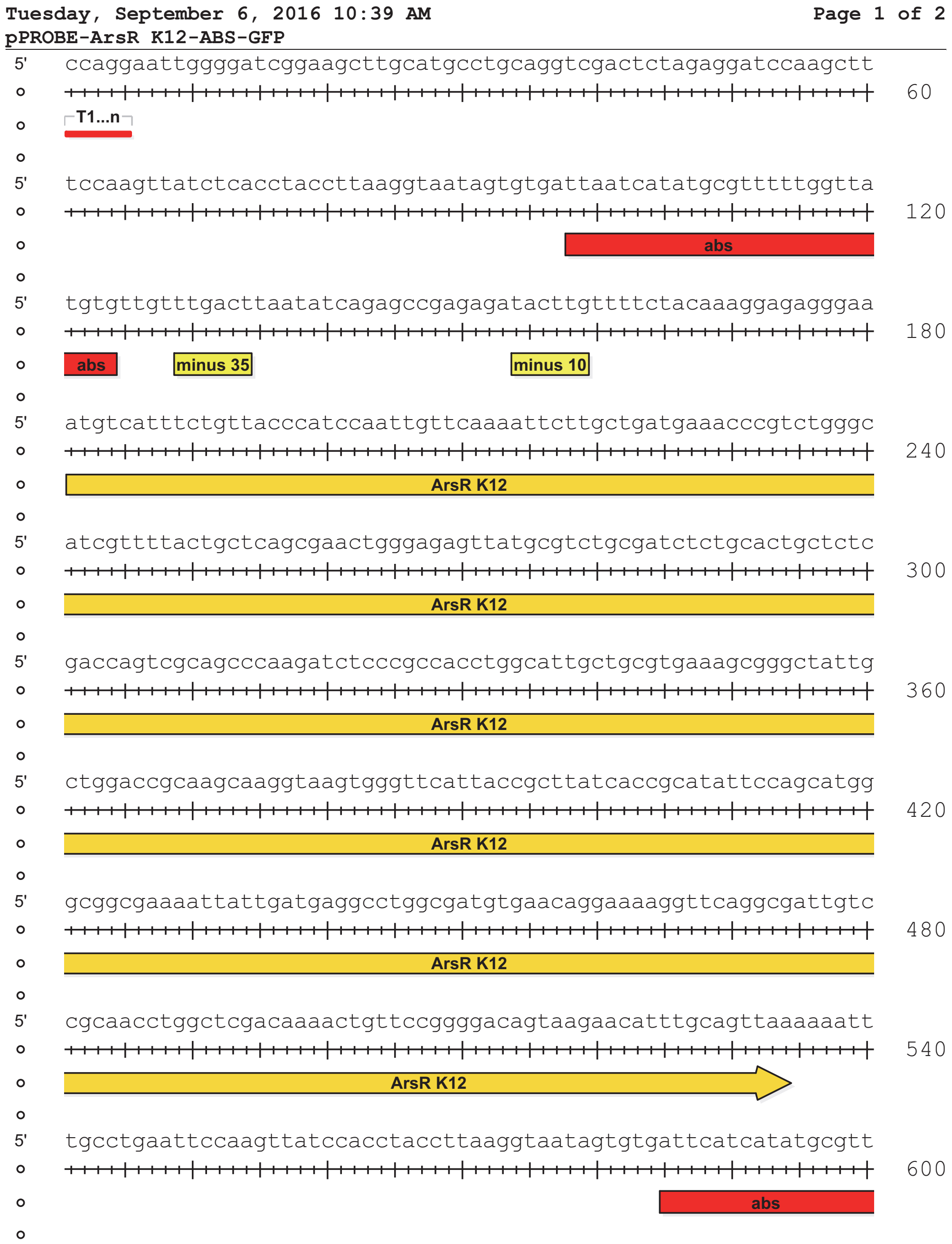


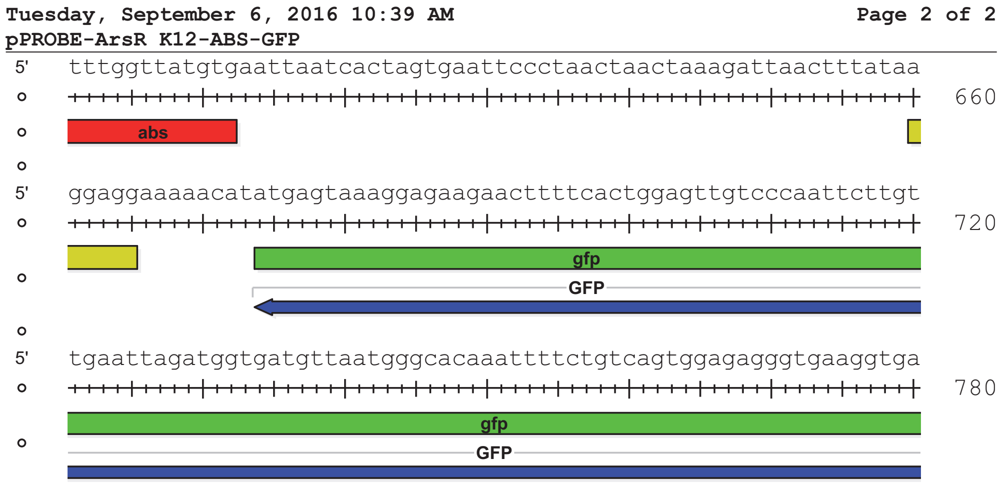
Figure 3.12 **Graphical overview of the *ars* operons and the constructed deletions**. **A**: Schematic detail of the regulatory elements in the *ars* promoter from plasmid R773 with the ArsR binding site (ArsR b.s.), the -35 and -10 promoter and the transcription (arrow facing to the right) and translation starts (ATG). Full sequence shown in Figure 3.10. **B**: Organization of the R773 *ars* operon. Dotted region corresponds to panel **A**. **C**: Organization of the chromosomal *ars* operon of *E. coli* K12 and the location of the produced deletions (dotted lines, number indicating the length). Regions amplified for the I-SceI based recombination selection system are shown as solid lines flanked by (introduced) restriction sites. The 190 bp insert was a sequence difference found in comparison to the published K12 sequence (Accession number U00096.3).



Tuesday, September 6, 2016 10:39 AM **Page 1 of 2**

pPROBE-ArsR K12-ABS-GFP

```
5'  ccaggaattggggatcggaagcttgcatgcctgcaggtcgactctagaggatccaagctt   60
5'  tccaagttatctcacctaccttaaggtaatagtgtgattaatcatatgcgtttttggtta  120
5'  tgtgttgtttgacttaatatcagagccgagagatacttgttttctacaaaggagagggaa  180
5'  atgtcatttctgttacccatccaattgttcaaaattcttgctgatgaaacccgtctgggc  240
5'  atcgttttactgctcagcgaactgggagagttatgcgtctgcgatctctgcactgctctc  300
5'  gaccagtcgcagcccaagatctcccgccacctggcattgctgcgtgaaagcgggctattg  360
5'  ctggaccgcaagcaaggtaagtgggttcattaccgcttatcaccgcatattccagcatgg  420
5'  gcggcgaaaattattgatgaggcctggcgatgtgaacaggaaaaggttcaggcgattgtc  480
5'  cgcaacctggctcgacaaaactgttccggggacagtaagaacatttgcagttaaaaaatt  540
5'  tgcctgaattccaagttatccacctaccttaaggtaatagtgtgattcatcatatgcgtt  600
```



```
Tuesday, September 6, 2016 10:39 AM                    Page 2 of 2
pPROBE-ArsR K12-ABS-GFP
5'  tttggttatgtgaattaatcactagtgaattccctaactaactaaagattaactttataa  660
5'  ggaggaaaaacatatgagtaaaggagaagaacttttcactggagttgtcccaattcttgt  720
5'  tgaattagatggtgatgttaatgggcacaaattttctgtcagtggagagggtgaaggtga  780
```

Figure 3.13 **Sequence of the resynthesized *arsR*^{K12} gene with the *ars* promoter from R773, the ArsR binding site (abs, in red) and the downstream linkage to the *egfp* gene as in plasmid pPRK12**.

Chapter 4 Genetic toggle switch to sharpen arsenic bioreporters

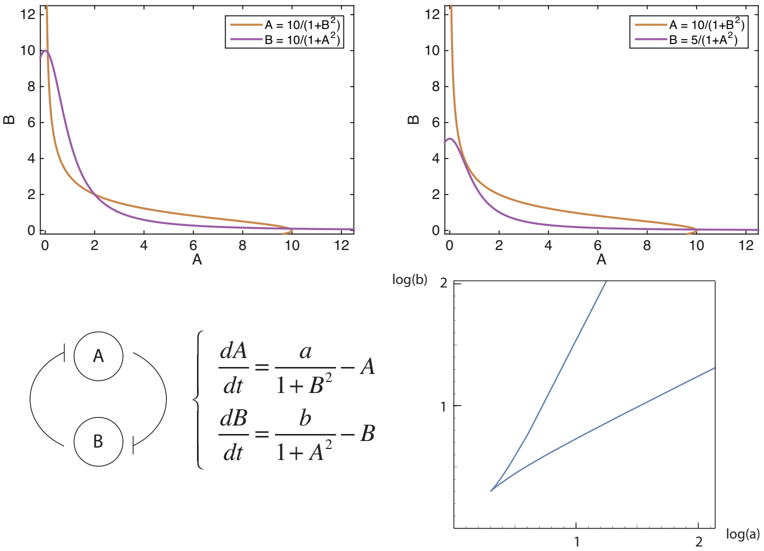
The similarities between the building blocks of synthetic biology (promoter, operator, genes, terminator) and electrical engineering allowed fundamental synthetic biology to develop devices inspired from electronics, such as the repressilator [148] – a genetic ring-oscillator – and the genetic toggle switch [142]. The repressilator consists of a network of three regulatory genes that successively repress in a loop, creating an oscillatory output signal. The genetic toggle switch is a similar construct of two genes that repress each other. Under particular conditions, it has the particularity to show a bistable behavior, *i.e.* in a culture of cells, two subpopulations of cells will differentiate and stay in stable states.

Bistability analysis have already been done through a simple model by Gardner et al. [142], which has been used to define the bistability regions. We would like to investigate the feasibility of a genetic toggle switch for the detection of arsenic. Because we estimated the binding affinities of the ArsR-P_{ars} system (see Chapter 3), we can build a model and predict the conditions for the detection of arsenic at low concentration. In this chapter, we start from the toggle switch model described by Gardner et al., expand it to take into account the induction of arsenic and apply the parameters previously estimated. Ultimately, we performed SSA to observe the bimodal distribution of the signals in the bistable region.

4.1 Methods

4.1.1 Gardner toggle switch model

The toggle switch is formulated into a system of two coupled equations. In order to look for solutions at steady-state, we set both derivatives to zero, and a solution is found if both equations are satisfied at the same time. In order to help the visualization of the solutions, we can plot the curves satisfying the steady-state for each equation, called nullclines (Figure 4.1, top). When the nullclines cross, both equations are satisfied and we have a solution. Depending on the parameters a and b, the nullclines intersect only once (stable steady-state), or three times (two stable steady-states and one unstable steady-state, giving rise to bistability). Using this information, there is a parametric solution that determines for each pair of parameters a and b if the system is bistable or not (Figure 4.1, bottom right).

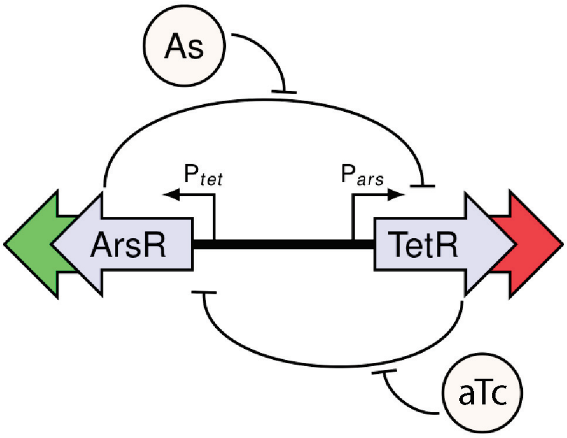


$$\begin{cases} \dfrac{dA}{dt} = \dfrac{a}{1+B^2} - A \\ \dfrac{dB}{dt} = \dfrac{b}{1+A^2} - B \end{cases}$$

Figure 4.1 **Gardner model of toggle switch**. Bottom left: Model in coupled equations, with species A and B being mutual repressors. Top: depending on the parameters a and b, nullclines intersect to give bistable states (two stable steady-states and one unstable steady-state, left) or monostable states (one stable steady-state, right). Bottom right: region of bistability on the (a,b) plane. The “wing” shape delimits the region of bistability, while the outside of the shape corresponds to monostable regions.

4.1.2 ArsR-TetR toggle switch model

Our model of ArsR-TetR toggle switch is very similar to the model of Gardner et al. [142], at the exception that we take into account of the influence of the effectors – arsenic (As) and anhydrotetracycline (aTc) – to make it suitable for arsenic detection. We added two parameters K_a and K_b to the coupled system of equations (Figure 4.2, right) that affect the repressory strength of each species, using the binding affinities from the literature and from the parameter estimation of Chapter 3 (Table 4.1). Addition of As and aTc decrease the parameters K_a and K_b, which translates into a weaker repressive effect of ArsR and TetR, respectively. $ArsR_{max}$ and $TetR_{max}$ were set to 500 copies per cell (50 copies per plasmid), *i.e.* approximately 800 nM.



$$\begin{array}{ll} ArsR & \left\{ \begin{array}{l} \dfrac{dA}{dt} = \dfrac{ArsR_{\mathrm{max}}}{1+K_b B^2} - A, \\ \dfrac{dB}{dt} = \dfrac{TetR_{\mathrm{max}}}{1+K_a A^2} - B, \end{array} \right. \\ TetR & \end{array}$$

$$K_b = K_{\mathrm{H,TetR}} \frac{K_{\mathrm{A,TetR}} + K_{\mathrm{C1,TetR}}[aTc]\left(K_{\mathrm{D1,TetR}} + K_{\mathrm{C2,TetR}} K_{\mathrm{D2,TetR}}[aTc]\right)}{1+K_{\mathrm{C1,TetR}}[aTc]\left(1+K_{\mathrm{C2,TetR}}[aTc]\right)}$$

$$K_a = K_{\mathrm{H,ArsR}} \frac{K_{\mathrm{A,ArsR}} + K_{\mathrm{C1,ArsR}}[As]\left(K_{\mathrm{D1,ArsR}} + K_{\mathrm{C2,ArsR}} K_{\mathrm{D2,ArsR}}[As]\right)}{1+K_{\mathrm{C1,ArsR}}[As]\left(1+K_{\mathrm{C2,ArsR}}[As]\right)}$$

$ArsR_{\mathrm{max}}$: stationary unrepressed level of ArsR

$TetR_{\mathrm{max}}$: stationary unrepressed level of TetR

Figure 4.2 **Model of genetic toggle switch for detection of arsenic.** Left: ArsR and TetR are mutual repressors, weakened in presence of Arsenic (As) and anhydrotetracycline (aTc), respectively. Right: Formulation of the model incorporating the effectors.

Table 4.1 **Binding affinities used**

	Binding affinities (M^{-1})		
	TetR	Ref.	**ArsR**^{K12} (Ref. [149])
K_A	10^{11}	[150]	$3.33 \cdot 10^7$
K_{C1}	$9.8 \cdot 10^{11}$	[151]	$1.66 \cdot 10^7$
K_{C2}	$9.8 \cdot 10^{11}$	[151]	$1.66 \cdot 10^7$
K_{D1}	$3.2 \cdot 10^7$	[150]	$1.26 \cdot 10^4$
K_{D2}	$3.2 \cdot 10^4$	[150]	$1.26 \cdot 10^4$
K_H	10^7	[152]	$1.59 \cdot 10^8$

4.2 Results and discussion

4.2.1 Analysis of bistability regions

By incorporating the parameters K_a and K_b, the parametric equation defining the bistable regions can be expressed as function of As and aTc concentrations. We also show the influence of other parameters on the bistability regions, such as maximal expressions of ArsR and TetR and the binding affinity of TetR towards aTc.

Results show that by reducing $ArsR_{max}$, the system can be sensitive to low levels of As, instead of originally having a bistable state that spans on the range 0 – 150 µg/L of As.

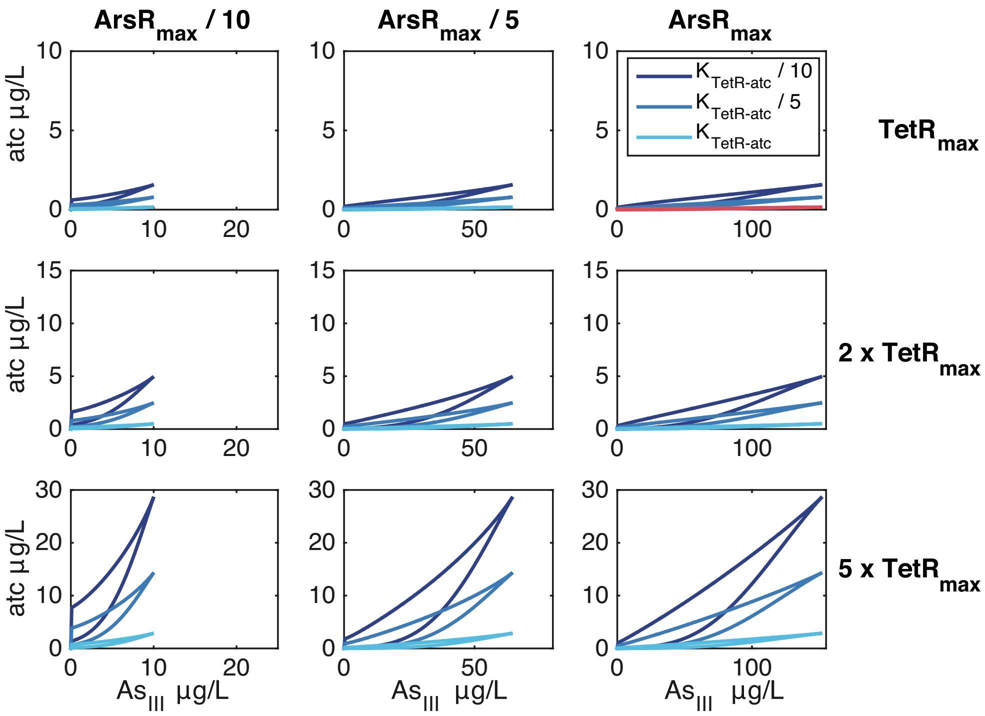


Figure 4.3. **Bistability regions in the ArsR$_{max}$,TetR$_{max}$ and K$_{TetR\text{-}atc}$ parameter space as functions of As and aTc.** The reference parameters provide a barely visible bistability region (top right panel, red line). The bistable regions are shown as function of the unrepressed levels of ArsR and TetR, and on each panel with the influence of a reduced TetR-aTc interaction (K$_C$ in Table 4.1)

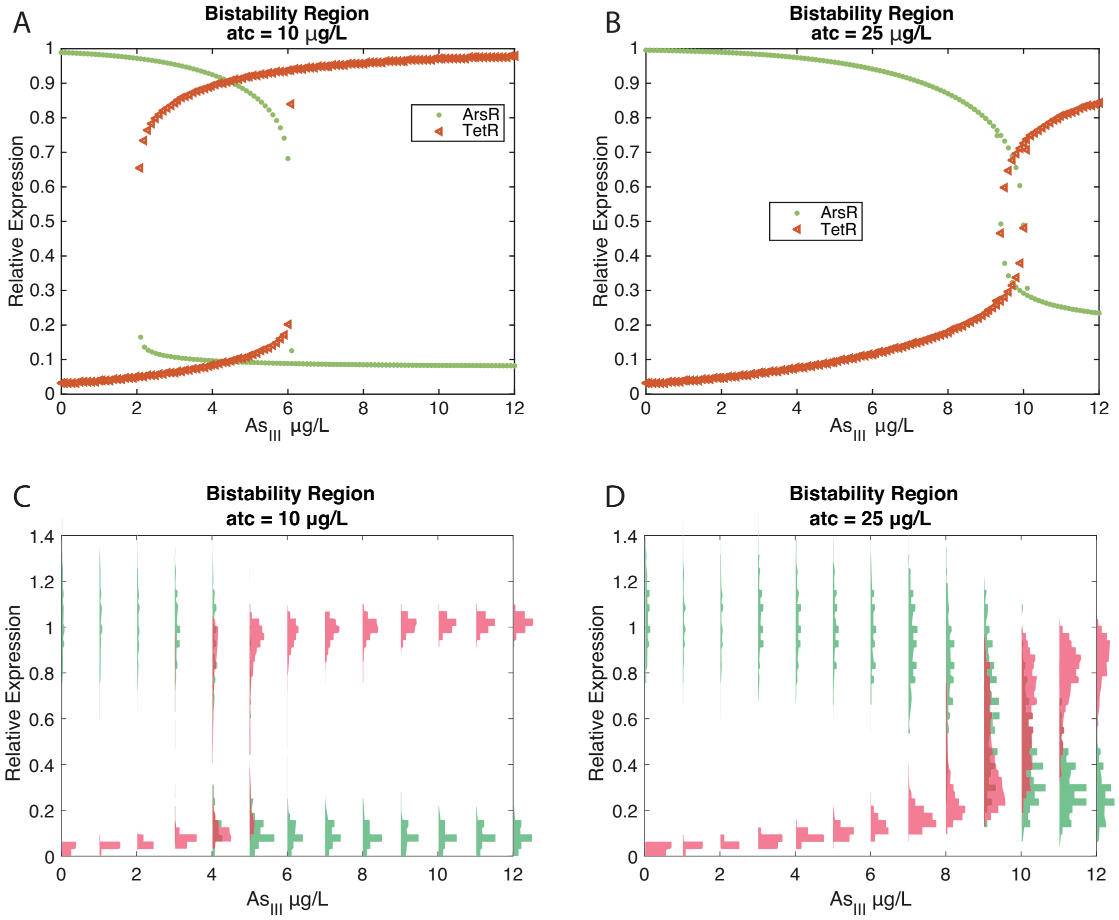


Figure 4.4. **Switch behavior of the ArsR and TetR concentrations by induction of arsenic. A:** Relative levels of ArsR and TetR at steady state as function of increasing and decreasing arsenic concentrations. **B**: Same as **A** with 25 µg/L aTc. **C**: Distributions of ArsR and TetR expressions based on 1000 stochastic simulations. **D**: same as **C** with 25 µg/L aTc.

Simulations were performed on the bistability region depicted on the lower-left panel of Figure 4.3. In agreement with the predicted bistability regions, the system with 10 µg aTc/L is bistable for arsenic values ranging from 2 – 6 µg/L, and the bistability region is smaller for an aTc concentration of 25 µg/L. However, a sharper bistability region implies a smaller switch in intensity, until the bistable region vanishes and the signal resumes to a continuous induction curve (not shown). Populations of ArsR tend to be more spread around their maximal expressions because of their lower copy number (in this case, $ArsR_{max} = 50$, $TetR_{max} = 2500$).

On top of being *per se* of interest as a device for turning gene on/off in cells, this ArsR-TetR toggle switch could be potentially used as a sharper detection device for arsenic, showing an all or nothing response around a certain arsenic threshold.

Chapter 5 Conclusion and perspectives

Using information on the molecular interactions between single TF-DNA and TF-TF binding, we built a cooperativity model of heterodimer and homodimer protein binding to a two-site DNA target. Using the high-throughput experimental data, we could estimate the cooperativity coefficients for each protein type and on the whole mutation library of the DNA element. The binding motif resulting from our two-site model with cooperativity, in the form of a PSSM, predicts TF-TF interaction with DNA more accurately than the binding motif derived from a single-site model. Because our two-site model is more detailed, it requires extensive experimental measurements of single TF-DNA interactions. However, the ongoing development of high-throughput methods for the measurements of molecular interactions can potentially make more uses of detailed mechanistic models for the understanding of cooperative TF regulation.

On a more applied angle, we were able to build a model for arsenic bioreporters, calibrate it with experimental data and improve the detection at low concentration by changing the repressor allele. The model describes different feedback and uncoupled circuits integrated in *E. coli.* We again used a detailed model that accounts for pairwise molecular interactions and provides parameters of biological relevance. By changing parameter values, the model should be able to capture the output of similar gene circuit architectures, which could be useful for optimization of other bioreporters.

In a try to further improve the arsenic detection, we investigated the feasibility of a bioreporter based on a toggle switch. Using the parameter values from the ArsR-P_{ars} circuit, we could build a model for a ArsR-TetR toggle switch for the detection of arsenic and estimate the regions of bistability as function of different parameters. The model suggest to tune the maximal level of ArsR and TetR, e.g. by ribosome binding site editing, to optimize the switch for low arsenic concentrations.

References

[1] F. Crick, “Central dogma of molecular biology,” *Nature*, vol. 227, no. 5258, pp. 561–563, Aug. 1970.

[2] A. J. Griffiths, W. M. Gelbart, J. H. Miller, and R. C. Lewontin, “Regulation of the Lactose System,” 1999.

[3] G. M. Cooper, “Regulation of Transcription in Eukaryotes,” 2000.

[4] B. D. Strahl and C. D. Allis, “The language of covalent histone modifications,” *Nature*, vol. 403, no. 6765, pp. 41–45, Jan. 2000.

[5] J. S. Mattick, R. J. Taft, and G. J. Faulkner, “A global view of genomic information--moving beyond the gene and the master regulator,” *Trends Genet. TIG*, vol. 26, no. 1, pp. 21–28, Jan. 2010.

[6] R. J. Jackson, C. U. T. Hellen, and T. V. Pestova, “The mechanism of eukaryotic translation initiation and principles of its regulation,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 2, p. 113, Feb. 2010.

[7] H. Li, C. K. Tsang, M. Watkins, P. G. Bertram, and X. F. S. Zheng, “Nutrient regulates Tor1 nuclear localization and association with rDNA promoter,” *Nature*, vol. 442, no. 7106, pp. 1058–1061, Aug. 2006.

[8] Y. L. Deribe, T. Pawson, and I. Dikic, “Post-translational modifications in signal integration,” *Nat. Struct. Mol. Biol.*, vol. 17, no. 6, p. 666, Jun. 2010.

[9] T. I. Lee and R. A. Young, “Transcriptional Regulation and its Misregulation in Disease,” *Cell*, vol. 152, no. 6, pp. 1237–1251, Mar. 2013.

[10] M. Grzmil and B. A. Hemmings, “Translation regulation as a therapeutic target in cancer,” *Cancer Res.*, vol. 72, no. 16, pp. 3891–3900, Aug. 2012.

[11] C. G. Proud, “Signalling to translation: how signal transduction pathways control the protein synthetic machinery,” *Biochem. J.*, vol. 403, no. 2, pp. 217–234, Apr. 2007.

[12] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, “A census of human transcription factors: function, expression and evolution,” *Nat. Rev. Genet.*, vol. 10, no. 4, p. 252, Apr. 2009.

[13] M. Ashburner *et al.*, “Gene ontology: tool for the unification of biology. The Gene Ontology Consortium,” *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.

[14] H. Salgado *et al.*, “RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions,” *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. D394-397, Jan. 2006.

[15] X. Fang *et al.*, “Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 38, pp. 10286–10291, Sep. 2017.

[16] S. J. Maerkl and S. R. Quake, “A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors,” *Science*, vol. 315, no. 5809, pp. 233–237, Jan. 2007.

[17] S. Rockel, M. Geertz, and S. J. Maerkl, “MITOMI: a microfluidic platform for in vitro characterization of transcription factor-DNA interaction,” *Methods Mol. Biol. Clifton NJ*, vol. 786, pp. 97–114, 2012.

[18] A. Isakova, Y. Berset, V. Hatzimanikatis, and B. Deplancke, “Quantification of Cooperativity in Heterodimer-DNA Binding Improves the Accuracy of Binding Specificity Models,” *J. Biol. Chem.*, vol. 291, no. 19, pp. 10293–10306, May 2016.

[19] J. R. van der Meer and S. Belkin, “Where microbiology meets microengineering: design and applications of reporter bacteria,” *Nat. Rev. Microbiol.*, vol. 8, no. 7, p. 511, Jul. 2010.

[20] C. Roggo and J. R. van der Meer, “Miniaturized and integrated whole cell living bacterial sensors in field applicable autonomous devices,” *Curr. Opin. Biotechnol.*, vol. 45, pp. 24–33, Jun. 2017.

[21] S. Melamed, T. Elad, and S. Belkin, “Microbial sensor cell arrays,” *Curr. Opin. Biotechnol.*, vol. 23, no. 1, pp. 2–8, Feb. 2012.

[22] J. Stocker *et al.*, “Development of a Set of Simple Bacterial Biosensors for Quantitative and Rapid Measurements of Arsenite and Arsenate in Potable Water,” *Environ. Sci. Technol.*, vol. 37, no. 20, pp. 4743–4750, Oct. 2003.

[23] D. Merulla, V. Hatzimanikatis, and J. R. van der Meer, “Tunable reporter signal production in feedback-uncoupled arsenic bioreporters: Uncoupled ArsR genetic circuits,” *Microb. Biotechnol.*, vol. 6, no. 5, pp. 503–514, Sep. 2013.

[24] N. Xuan, M. Chetty, R. Coppel, and P. P. Wangikar, “Gene regulatory network modeling via global optimization of high-order dynamic Bayesian network,” *BMC Bioinformatics*, vol. 13, p. 131, Jun. 2012.

[25] A. Saadatpour and R. Albert, “Boolean modeling of biological regulatory networks: a methodology tutorial,” *Methods San Diego Calif*, vol. 62, no. 1, pp. 3–12, Jul. 2013.

[26] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,” *J. Comput. Phys.*, vol. 22, no. 4, pp. 403–434, 1976.

[27] E. H. Davidson and D. H. Erwin, “Gene regulatory networks and the evolution of animal body plans,” *Science*, vol. 311, no. 5762, pp. 796–800, Feb. 2006.

[28] B. Deplancke, “Experimental advances in the characterization of metazoan gene regulatory networks,” *Brief. Funct. Genomic. Proteomic.*, vol. 8, no. 1, pp. 12–27, Jan. 2009.

[29] G. Badis *et al.*, “Diversity and Complexity in DNA Recognition by Transcription Factors,” *Science*, vol. 324, no. 5935, pp. 1720–1723, May 2009.

[30] P. M. Fordyce *et al.*, “Basic leucine zipper transcription factor Hac1 binds DNA in two distinct modes as revealed by microfluidic analyses,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 45, pp. E3084-3093, Nov. 2012.

[31] D. J. Galas and A. Schmitz, “DNAase footprinting a simple method for the detection of protein-DNA binding specificity,” *Nucleic Acids Res.*, vol. 5, no. 9, pp. 3157–3170, 1978.

[32] A. Jolma *et al.*, “DNA-Binding Specificities of Human Transcription Factors,” *Cell*, vol. 152, no. 1–2, pp. 327–339, Jan. 2013.

[33] C. A. Grove *et al.*, “A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors,” *Cell*, vol. 138, no. 2, pp. 314–327, Jul. 2009.

[34] T. Ravasi *et al.*, “An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man,” *Cell*, vol. 140, no. 5, pp. 744–752, Mar. 2010.

[35] R. Gordân *et al.*, “Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape,” *Cell Rep.*, vol. 3, no. 4, pp. 1093–1104, Apr. 2013.

[36] A. Hoffmann, T. H. Leung, and D. Baltimore, “Genetic analysis of NF-kappaB/Rel transcription factors defines functional specificities,” *EMBO J.*, vol. 22, no. 20, pp. 5530–5539, Oct. 2003.

[37] A. Jolma and J. Taipale, “Methods for Analysis of Transcription Factor DNA-Binding Specificity In Vitro,” *Subcell. Biochem.*, vol. 52, pp. 155–173, 2011.

[38] J. D. Klemm, S. L. Schreiber, and G. R. Crabtree, “Dimerization as a regulatory mechanism in signal transduction,” *Annu. Rev. Immunol.*, vol. 16, pp. 569–592, 1998.

[39] F. Rastinejad, “Retinoid X receptor and its partners in the nuclear receptor family,” *Curr. Opin. Struct. Biol.*, vol. 11, no. 1, pp. 33–38, Feb. 2001.

[40] J. S. Reece-Hoyes, B. Deplancke, J. Shingles, C. A. Grove, I. A. Hope, and A. J. M. Walhout, “A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks,” *Genome Biol.*, vol. 6, no. 13, p. R110, 2005.

[41] T. Siggers *et al.*, “Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding,” *Nat. Immunol.*, vol. 13, no. 1, pp. 95–102, Nov. 2011.

[42] C. Zechel, X. Q. Shen, P. Chambon, and H. Gronemeyer, “Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements,” *EMBO J.*, vol. 13, no. 6, pp. 1414–1424, Mar. 1994.

[43] C. K. L. Ng, N. X. Li, S. Chee, S. Prabhakar, P. R. Kolatkar, and R. Jauch, “Deciphering the Sox-Oct partner code by quantitative cooperativity measurements,” *Nucleic Acids Res.*, vol. 40, no. 11, pp. 4933–4941, Jun. 2012.

[44] M. Slattery *et al.*, “Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins,” *Cell*, vol. 147, no. 6, pp. 1270–1282, Dec. 2011.

[45] C. K. Glass, “Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers,” *Endocr. Rev.*, vol. 15, no. 3, pp. 391–407, Jun. 1994.

[46] M. J. Reginato, J. Zhang, and M. A. Lazar, “DNA-independent and DNA-dependent mechanisms regulate the differential heterodimerization of the isoforms of the thyroid hormone receptor with retinoid X receptor,” *J. Biol. Chem.*, vol. 271, no. 45, pp. 28199–28205, Nov. 1996.

[47] J. Barcroft and A. V. Hill, “The nature of oxyhaemoglobin, with a note on its molecular weight,” *J. Physiol.*, vol. 39, no. 6, pp. 411–428, Mar. 1910.

[48] A. V. Hill, “The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves,” *J. Physiol.*, vol. 40, pp. i–vii, Jan. 1910.

[49] G. S. Adair, “A Comparison of the Molecular Weights of the Proteins,” *Biol. Rev.*, vol. 1, no. 2, pp. 75–78, Apr. 1924.

[50] I. M. Klotz, F. M. Walker, and R. B. Pivan, “The binding of organic ions by proteins,” *J. Am. Chem. Soc.*, vol. 68, pp. 1486–1490, Aug. 1946.

[51] J. N. Weiss, “The Hill equation revisited: uses and misuses.,” *FASEB J.*, vol. 11, no. 11, pp. 835–841, Sep. 1997.

[52] G. D. Stormo, T. D. Schneider, L. Gold, and A. Ehrenfeucht, “Use of the ‘Perceptron’ algorithm to distinguish translational initiation sites in E. coli,” *Nucleic Acids Res.*, vol. 10, no. 9, pp. 2997–3011, May 1982.

[53] Y. Zhao, S. Ruan, M. Pandey, and G. D. Stormo, “Improved Models for Transcription Factor Binding Site Identification Using Nonindependent Interactions,” *Genetics*, vol. 191, no. 3, pp. 781–790, Apr. 2012.

[54] S. J. Maerkl and S. R. Quake, “A systems approach to measuring the binding energy landscapes of transcription factors,” *Science*, vol. 315, no. 5809, pp. 233–237, Jan. 2007.

[55] R. Siersbaek, R. Nielsen, and S. Mandrup, “PPARgamma in adipocyte differentiation and metabolism--novel insights from genome-wide studies,” *FEBS Lett.*, vol. 584, no. 15, pp. 3242–3249, Aug. 2010.

[56] P. Tontonoz, E. Hu, and B. M. Spiegelman, “Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor,” *Cell*, vol. 79, no. 7, pp. 1147–1156, Dec. 1994.

[57] S. J. Maerkl and S. R. Quake, “Experimental determination of the evolvability of a transcription factor,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 44, pp. 18650–18655, Oct. 2009.

[58] C. Gubelmann *et al.*, “A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks,” *Mol. Syst. Biol.*, vol. 9, Aug. 2013.

[59] T. L. Hill, *Cooperativity Theory in Biochemistry: Steady-state and Equilibrium Systems*. Springer-Verlag, 1985.

[60] A. Ben-Naim, *Cooperativity and Regulation in Biochemical Processes*, Softcover reprint of hardcover 1st ed. 2001. Boston, MA: Springer US, 2010.

[61] R. Nielsen *et al.*, “Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis,” *Genes Dev.*, vol. 22, no. 21, pp. 2953–2967, Nov. 2008.

[62] S. K. Raghav *et al.*, “Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPβ and KAISO,” *Mol. Cell*, vol. 46, no. 3, pp. 335–350, May 2012.

[63] Y. Orenstein and R. Shamir, “A comparative analysis of transcription factor binding models learned from PBM, HT-SELEX and ChIP data,” *Nucleic Acids Res.*, Feb. 2014.

[64] A. Khan *et al.*, “JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework,” *Nucleic Acids Res.*, Nov. 2017.

[65] Y. Nakachi *et al.*, “Identification of novel PPARγ target genes by integrated analysis of ChIP-on-chip and microarray expression data during adipocyte differentiation,” *Biochem. Biophys. Res. Commun.*, vol. 372, no. 2, pp. 362–366, Jul. 2008.

[66] M. S. Hamza *et al.*, “De-Novo Identification of PPARγ/RXR Binding Sites and Direct Targets during Adipogenesis,” *PLoS ONE*, vol. 4, no. 3, p. e4907, Mar. 2009.

[67] W. Wahli, O. Braissant, and B. Desvergne, “Peroxisome proliferator activated receptors: transcriptional regulators of adipogenesis, lipid metabolism and more...,” *Chem. Biol.*, vol. 2, no. 5, pp. 261–266, May 1995.

[68] A. IJpenberg *et al.*, “In vivo activation of PPAR target genes by RXR homodimers,” *EMBO J.*, vol. 23, no. 10, pp. 2083–2091, May 2004.

[69] D. J. Mangelsdorf and R. M. Evans, “The RXR heterodimers and orphan receptors,” *Cell*, vol. 83, no. 6, pp. 841–850, Dec. 1995.

[70] J. Osz *et al.*, “Structural Basis of Natural Promoter Recognition by the Retinoid X Nuclear Receptor,” *Sci. Rep.*, vol. 5, p. 8216, Feb. 2015.

[71] C. T. Workman, Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, and P. V. Benos, “enoLOGOS: a versatile web tool for energy normalized sequence logos,” *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W389-392, Jul. 2005.

[72] M. T. Weirauch *et al.*, “Evaluation of methods for modeling transcription factor sequence specificity,” *Nat. Biotechnol.*, vol. 31, no. 2, pp. 126–134, Jan. 2013.

[73] T. Perlmann, K. Umesono, P. N. Rangarajan, B. M. Forman, and R. M. Evans, “Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors,” *Mol. Endocrinol. Baltim. Md*, vol. 10, no. 8, pp. 958–966, Aug. 1996.

[74] F. Rastinejad, T. Wagner, Q. Zhao, and S. Khorasanizadeh, “Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1,” *EMBO J.*, vol. 19, no. 5, pp. 1045–1054, Mar. 2000.

[75] V. Chandra *et al.*, “Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA,” *Nature*, vol. 456, no. 7220, pp. 350–356, Nov. 2008.

[76] C. T. Workman, Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, and P. V. Benos, “enoLOGOS: a versatile web tool for energy normalized sequence logos,” *Nucleic Acids Res.*, vol. 33, no. suppl_2, pp. W389–W392, Jul. 2005.

[77] R. K. Shultzaberger, S. J. Maerkl, J. F. Kirsch, and M. B. Eisen, “Probing the Informational and Regulatory Plasticity of a Transcription Factor DNA–Binding Domain,” *PLoS Genet.*, vol. 8, no. 3, p. e1002614, Mar. 2012.

[78] J. Simicevic *et al.*, “Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics,” *Nat. Methods*, vol. 10, no. 6, pp. 570–576, Apr. 2013.

[79] S. B. Lee and J. E. Bailey, “Genetically structured models forlac promoter-operator function in the Escherichia coli chromosome and in multicopy plasmids: Lac operator function,” *Biotechnol. Bioeng.*, vol. 26, no. 11, pp. 1372–1382, Nov. 1984.

[80] K. Katoh and D. M. Standley, “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability,” *Mol. Biol. Evol.*, vol. 30, no. 4, pp. 772–780, Apr. 2013.

[81] T. L. Bailey and C. Elkan, “Fitting a mixture model by expectation maximization to discover motifs in biopolymers,” *Proc. Int. Conf. Intell. Syst. Mol. Biol. ISMB Int. Conf. Intell. Syst. Mol. Biol.*, vol. 2, pp. 28–36, 1994.

[82] A. S. Ethayathulla, H. T. Nguyen, and H. Viadiu, “Crystal structures of the DNA-binding domain tetramer of the p53 tumor suppressor family member p73 bound to different full-site response elements,” *J. Biol. Chem.*, vol. 288, no. 7, pp. 4744–4754, Feb. 2013.

[83] A. P. W. Funnell and M. Crossley, “Homo- and Heterodimerization in Transcriptional Regulation,” in *Protein Dimerization and Oligomerization in Biology*, vol. 747, J. M. Matthews, Ed. New York, NY: Springer New York, 2012, pp. 105–121.

[84] J. Zhou, L. Pérès, N. Honoré, R. Nasr, J. Zhu, and H. de Thé, “Dimerization-induced corepressor binding and relaxed DNA-binding specificity are critical for PML/RARA-induced immortalization,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 24, pp. 9238–9243, Jun. 2006.

[85] T. Siggers and R. Gordan, “Protein-DNA binding: complexities and multi-protein codes,” *Nucleic Acids Res.*, vol. 42, no. 4, pp. 2099–2111, Feb. 2014.

[86] J. J. Kohler and A. Schepartz, “Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization,” *Biochemistry (Mosc.)*, vol. 40, no. 1, pp. 130–142, Jan. 2001.

[87] S. J. Metallo and A. Schepartz, “Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA,” *Nat. Struct. Biol.*, vol. 4, no. 2, pp. 115–117, Feb. 1997.

[88] F. Rastinejad, T. Perlmann, R. M. Evans, and P. B. Sigler, “Structural determinants of nuclear receptor assembly on DNA direct repeats,” *Nature*, vol. 375, no. 6528, pp. 203–211, May 1995.

[89] M. Okuno, E. Arimoto, Y. Ikenobu, T. Nishihara, and M. Imagawa, “Dual DNA-binding specificity of peroxisome-proliferator-activated receptor gamma controlled by heterodimer formation with retinoid X receptor alpha,” *Biochem. J.*, vol. 353, no. Pt 2, pp. 193–198, Jan. 2001.

[90] H. Castelein, A. Janssen, P. E. Declercq, and M. Baes, “Sequence requirements for high affinity retinoid X receptor-α homodimer binding,” *Mol. Cell. Endocrinol.*, vol. 119, no. 1, pp. 11–20, May 1996.

[91] V. T. Todorov, M. Desch, T. Schubert, and A. Kurtz, “The Pal3 promoter sequence is critical for the regulation of human renin gene transcription by peroxisome proliferator-activated receptor-gamma,” *Endocrinology*, vol. 149, no. 9, pp. 4647–4657, Sep. 2008.

[92] A. J. Griggs, J. J. Stickel, and J. J. Lischeske, “A mechanistic model for enzymatic saccharification of cellulose using continuous distribution kinetics I: depolymerization by EGI and CBHI,” *Biotechnol. Bioeng.*, vol. 109, no. 3, pp. 665–675, Mar. 2012.

[93] A. Chakrabarti, L. Miskovic, K. C. Soh, and V. Hatzimanikatis, “Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints,” *Biotechnol. J.*, vol. 8, no. 9, pp. 1043–1057, Sep. 2013.

[94] G. J. Boggy and P. J. Woolf, “A Mechanistic Model of PCR for Accurate Quantification of Quantitative PCR Data,” *PLoS ONE*, vol. 5, no. 8, p. e12355, Aug. 2010.

[95] S. Park *et al.*, “Determination of binding constant of transcription factor myc-max/max-max and E-box DNA: the effect of inhibitors on the binding,” *Biochim. Biophys. Acta*, vol. 1670, no. 3, pp. 217–228, Feb. 2004.

[96] O. Ecevit, M. A. Khan, and D. J. Goss, “Kinetic analysis of the interaction of b/HLH/Z transcription factors Myc, Max, and Mad with cognate DNA,” *Biochemistry (Mosc.)*, vol. 49, no. 12, pp. 2627–2635, Mar. 2010.

[97] J. A. Lefstin and K. R. Yamamoto, “Allosteric effects of DNA on transcriptional regulators,” *Nature*, vol. 392, no. 6679, pp. 885–888, Apr. 1998.

[98] T. H. Leung, A. Hoffmann, and D. Baltimore, “One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers,” *Cell*, vol. 118, no. 4, pp. 453–464, Aug. 2004.

[99] J. R. van der Meer, “Bacterial Sensors: Synthetic Design and Application Principles,” *Synth. Lect. Synth. Biol.*, vol. 2, no. 1, pp. 1–167, Dec. 2010.

[100] S. K. Checa, M. D. Zurbriggen, and F. C. Soncini, “Bacterial signaling systems as platforms for rational design of new generations of biosensors,” *Curr. Opin. Biotechnol.*, vol. 23, no. 5, pp. 766–772, Oct. 2012.

[101] M. Park, S.-L. Tsai, and W. Chen, “Microbial biosensors: engineered microorganisms as the sensing machinery,” *Sensors*, vol. 13, no. 5, pp. 5777–5795, 2013.

[102] A. S. Khalil and J. J. Collins, “Synthetic biology: applications come of age,” *Nat. Rev. Genet.*, vol. 11, no. 5, pp. 367–379, May 2010.

[103] K. Siegfried *et al.*, “Field testing of arsenic in groundwater samples of Bangladesh using a test kit based on lyophilized bioreporter bacteria,” *Environ. Sci. Technol.*, vol. 46, no. 6, pp. 3281–3287, Mar. 2012.

[104] N. Buffi *et al.*, “Development of a microfluidics biosensor for agarose-bead immobilized Escherichia coli bioreporter cells for arsenite detection in aqueous samples,” *Lab. Chip*, vol. 11, no. 14, pp. 2369–2377, Jul. 2011.

[105] P. J. Lee, N. Ghorashian, T. A. Gaige, and P. J. Hung, “Microfluidic System for Automated Cell-Based Assays,” *JALA J. Assoc. Lab. Autom.*, vol. 12, no. 6, pp. 363–367, Dec. 2007.

[106] A. Rothert, S. K. Deo, L. Millner, L. G. Puckett, M. J. Madou, and S. Daunert, “Whole-cell-reporter-gene-based biosensing systems on a compact disk microfluidics platform,” *Anal. Biochem.*, vol. 342, no. 1, pp. 11–19, Jul. 2005.

[107] N. Buffi *et al.*, “An automated microreactor for semi-continuous biosensor measurements,” *Lab. Chip*, vol. 16, no. 8, pp. 1383–1392, Apr. 2016.

[108] S. Yagur-Kroll, B. Bilic, and S. Belkin, “Strategies for enhancing bioluminescent bacterial sensor performance by promoter region manipulation,” *Bioeng. Bugs*, vol. 1, no. 2, pp. 151–153, Mar. 2010.

[109] S. Yagur-Kroll, C. Lalush, R. Rosen, N. Bachar, Y. Moskovitz, and S. Belkin, “Escherichia coli bioreporters for the detection of 2,4-dinitrotoluene and 2,4,6-trinitrotoluene,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 2, pp. 885–895, Jan. 2014.

[110] A. A. K. Nielsen *et al.*, “Genetic circuit design automation,” *Science*, vol. 352, no. 6281, p. aac7341, Apr. 2016.

[111] T. R. Rieger, R. I. Morimoto, and V. Hatzimanikatis, “Mathematical Modeling of the Eukaryotic Heat-Shock Response: Dynamics of the hsp70 Promoter,” *Biophys. J.*, vol. 88, no. 3, pp. 1646–1658, Mar. 2005.

[112] T. Chen, V. Filklov, and S. S. Skiena, “Identifying Gene Regulatory Networks from Experimental Data,” *Parallel Comput*, vol. 27, no. 1–2, pp. 141–162, Jan. 2001.

[113] R. Karmakar and I. Bose, “Stochastic model of transcription factor-regulated gene expression,” *Phys. Biol.*, vol. 3, no. 3, p. 200, 2006.

[114] S. B. Lee and J. E. Bailey, “Genetically structured models for lac promoter–operator function in the chromosome and in multicopy plasmids: Lac promoter function,” *Biotechnol. Bioeng.*, vol. 26, no. 11, pp. 1383–1389, 1984.

[115] D. Merulla *et al.*, “Bioreporters and biosensors for arsenic detection. Biotechnological solutions for a world-wide pollution problem,” *Curr. Opin. Biotechnol.*, vol. 24, no. 3, pp. 534–541, Jun. 2013.

[116] P. T. K. Trang, M. Berg, P. H. Viet, N. Van Mui, and J. R. Van Der Meer, “Bacterial bioassay for rapid and accurate analysis of arsenic in highly variable groundwater samples,” *Environ. Sci. Technol.*, vol. 39, no. 19, pp. 7625–7630, Oct. 2005.

[117] J. Wu and B. P. Rosen, “The ArsR protein is a trans-acting regulatory protein,” *Mol. Microbiol.*, vol. 5, no. 6, pp. 1331–1336, Jun. 1991.

[118] J. Wu and B. P. Rosen, “Metalloregulated expression of the ars operon.,” *J. Biol. Chem.*, vol. 268, no. 1, pp. 52–58, Jan. 1993.

[119] D. L. Scott, S. Ramanathan, W. Shi, B. P. Rosen, and S. Daunert, “Genetically engineered bacteria: electrochemical sensing systems for antimonite and arsenite,” *Anal. Chem.*, vol. 69, no. 1, pp. 16–20, Jan. 1997.

[120] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, L. S. Tsimring, and J. Hasty, “Sensing array of radically coupled genetic biopixels,” *Nature*, vol. 481, no. 7379, pp. 39–44, Dec. 2011.

[121] D. Merulla and J. R. van der Meer, “Regulatable and Modulable Background Expression Control in Prokaryotic Synthetic Circuits by Auxiliary Repressor Binding Sites,” *ACS Synth. Biol.*, Sep. 2015.

[122] L. Li *et al.*, “Evolved Bacterial Biosensor for Arsenite Detection in Environmental Water,” *Environ. Sci. Technol.*, vol. 49, no. 10, pp. 6149–6155, May 2015.

[123] L. M. Arruda, L. M. O. Monteiro, and R. Silva-Rocha, “The Chromobacterium violaceum ArsR Arsenite Repressor Exerts Tighter Control on Its Cognate Promoter Than the Escherichia coli System,” *Front. Microbiol.*, vol. 7, Nov. 2016.

[124] M. Wells, M. Gösch, R. Rigler, H. Harms, T. Lasser, and J. R. van der Meer, “Ultrasensitive reporter protein detection in genetically engineered bacteria,” *Anal. Chem.*, vol. 77, no. 9, pp. 2683–2689, May 2005.

[125] G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, “Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources,” *Cell*, vol. 157, no. 3, pp. 624–635, Apr. 2014.

[126] W. Shi, J. Dong, R. A. Scott, M. Y. Ksenzenko, and B. P. Rosen, “The Role of Arsenic-Thiol Interactions in Metalloregulation of the ars Operon,” *J. Biol. Chem.*, vol. 271, no. 16, pp. 9291–9297, Apr. 1996.

[127] J. A. Egea, E. Balsa-Canto, M.-S. G. García, and J. R. Banga, “Dynamic Optimization of Nonlinear Processes with an Enhanced Scatter Search Method,” *Ind. Eng. Chem. Res.*, vol. 48, no. 9, pp. 4388–4401, May 2009.

[128] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, “Tuning genetic control through promoter engineering,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 36, pp. 12678–12683, Sep. 2005.

[129] Y. Cai, B. Davidson, H. Ma, and C. French, “Modeling the arsenic biosensor system,” *BMC Syst. Biol.*, vol. 1, no. 1, p. P83, 2007.

[130] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin, “New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria,” *Appl. Environ. Microbiol.*, vol. 64, no. 6, pp. 2240–2246, Jun. 1998.

[131] Y.-L. Meng, Z. Liu, and B. P. Rosen, “As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in Escherichia coli,” *J. Biol. Chem.*, vol. 279, no. 18, pp. 18334–18341, Apr. 2004.

[132] W. G. Miller, J. H. Leveau, and S. E. Lindow, “Improved gfp and inaZ broad-host-range promoter-probe vectors,” *Mol. Plant-Microbe Interact. MPMI*, vol. 13, no. 11, pp. 1243–1250, Nov. 2000.

[133] F. M. Ausubel, *Current protocols in molecular biology*. New York: John Wiley & Sons, 1994.

[134] J. Sambrook, *Molecular Cloning: A Laboratory Manual, Third Edition*, 3rd edition. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press, 2001.

[135] E. Martínez-García and V. de Lorenzo, “Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440,” *Environ. Microbiol.*, vol. 13, no. 10, pp. 2702–2716, Oct. 2011.

[136] G. Pósfai, V. Kolisnychenko, Z. Bereczki, and F. R. Blattner, “Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome,” *Nucleic Acids Res.*, vol. 27, no. 22, pp. 4409–4415, Nov. 1999.

[137] C. Diorio, J. Cai, J. Marmor, R. Shinder, and M. S. DuBow, “An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria.,” *J. Bacteriol.*, vol. 177, no. 8, pp. 2050–2056, Apr. 1995.

[138] C. Xu and B. P. Rosen, “Dimerization Is Essential for DNA Binding and Repression by the ArsR Metalloregulatory Protein of Escherichia coli,” *J. Biol. Chem.*, vol. 272, no. 25, pp. 15734–15738, Jun. 1997.

[139] B. P. Rosen, H. Bhattacharjee, and W. Shi, “Mechanisms of metalloregulation of an anion-translocating ATPase,” *J. Bioenerg. Biomembr.*, vol. 27, no. 1, pp. 85–91, Feb. 1995.

[140] R. Iizuka, M. Yamagishi-Shirasaki, and T. Funatsu, “Kinetic study of de novo chromophore maturation of fluorescent proteins,” *Anal. Biochem.*, vol. 414, no. 2, pp. 173–178, Jul. 2011.

[141] H. Bhattacharjee, T. Zhou, J. Li, D. L. Gatti, A. R. Walmsley, and B. P. Rosen, “Structure-function relationships in an anion-translocating ATPase,” *Biochem. Soc. Trans.*, vol. 28, no. 4, pp. 520–526, 2000.

[142] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in Escherichia coli,” *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[143] H. De Jong, “Modeling and simulation of genetic regulatory systems: a literature review,” *J. Comput. Biol.*, vol. 9, no. 1, pp. 67–103, 2002.

[144] S. B. Lee and J. E. Bailey, “Genetically structured models forlac promoter–operator function in the Escherichia coli chromosome and in multicopy plasmids: Lac operator function,” *Biotechnol. Bioeng.*, vol. 26, no. 11, pp. 1372–1382, Nov. 1984.

[145] J. Pérez-Martín, F. Rojo, and V. de Lorenzo, “Promoters responsive to DNA bending: a common theme in prokaryotic gene expression,” *Microbiol. Rev.*, vol. 58, no. 2, pp. 268–290, Jun. 1994.

[146] A. Ay and D. N. Arnosti, “Mathematical modeling of gene expression: a guide for the perplexed biologist,” *Crit. Rev. Biochem. Mol. Biol.*, vol. 46, no. 2, pp. 137–151, Apr. 2011.

[147] F. Truffer *et al.*, “Compact portable biosensor for arsenic detection in aqueous samples with Escherichia coli bioreporter cells,” *Rev. Sci. Instrum.*, vol. 85, no. 1, p. 015120, Jan. 2014.

[148] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators,” *Nature*, vol. 403, no. 6767, pp. 335–338, Jan. 2000.

[149] Y. Berset, D. Merulla, A. Joublin, V. Hatzimanikatis, and J. R. van der Meer, “Mechanistic Modeling of Genetic Circuits for ArsR Arsenic Regulation,” *ACS Synth. Biol.*, Feb. 2017.

[150] T. Lederer, M. Takahashi, and W. Hillen, “Thermodynamic Analysis of Tetracycline-Mediated Induction of Tet Repressor by a Quantitative Methylation Protection Assay,” *Anal. Biochem.*, vol. 232, no. 2, pp. 190–196, Dec. 1995.

[151] O. Scholz, P. Schubert, M. Kintrup, and W. Hillen, “Tet repressor induction without Mg2+,” *Biochemistry (Mosc.)*, vol. 39, no. 35, pp. 10914–10920, Sep. 2000.

[152] W. Hillen, C. Gatz, L. Altschmied, K. Schollmeier, and I. Meier, “Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions,” *J. Mol. Biol.*, vol. 169, no. 3, pp. 707–721, Sep. 1983.

Curriculum Vitae

Yves Berset
Rte de Bertigny 35
1700 Fribourg
Switzerland

Tél : +41 (0) 79 257 8525
yberset@gmail.com

PhD student in Systems Biotechnology, MSc in Physics

Education

2013 – 2017 Doctoral Assistant, EPF Lausanne
2009 – 2012 MSc in Physics, Unifr (Minor: Mathematics)
2006 – 2009 BSc in Physics, Unifr
2001 – 2005 Collège Ste-Croix, Fribourg

Professional Qualifications

Productivity: Microsoft Office, Latex, Illustrator

Programming Languages: Matlab, Python, R, C.

Publications

Berset, Yves, and Matus Medo. “The Effect of the Initial Network Configuration on Preferential Attachment.” *The European Physical Journal B* 86, no. 6 (June 2013). doi:10.1140/epjb/e2013-30998-1.

Berset, Yves, Davide Merulla, Aurélie Joublin, Vassily Hatzimanikatis, and Jan R. van der Meer. “Mechanistic Modeling of Genetic Circuits for ArsR Arsenic Regulation.” ACS Synthetic Biology 6, no. 5 (May 19, 2017): 862–74. doi:10.1021/acssynbio.6b00364.

Isakova, Alina, Yves Berset, Vassily Hatzimanikatis, and Bart Deplancke. “Quantification of Cooperativity in Heterodimer-DNA Binding Improves the Accuracy of Binding Specificity Models.” Journal of Biological Chemistry 291, no. 19 (May 6, 2016): 10293–306. doi:10.1074/jbc.M115.691154.

Interests

Music, Running