



**HAL**  
open science

# Towards biocatalyst selection using microfluidically distributed molecular programs

Vasily Shenshin

► **To cite this version:**

Vasily Shenshin. Towards biocatalyst selection using microfluidically distributed molecular programs. Biochemistry, Molecular Biology. Université Paris sciences et lettres, 2021. English. NNT : 2021UP-SLS098 . tel-04006306

**HAL Id: tel-04006306**

**<https://pastel.hal.science/tel-04006306>**

Submitted on 27 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à ESPCI

**Vers la sélection de biocatalyseurs à l'aide de programmes moléculaires distribués par microfluidique**

*Towards biocatalyst selection using microfluidically distributed molecular programs*

Soutenue par

**Vasily A Shenshin**

Le 29 janvier 2021

Ecole doctorale n° 563

**Médicaments,  
toxicologie, chimie,  
imageries**

Spécialité

**Biochimie**

Composition du jury :

Gilles, TRUAN DR, INSA Toulouse	<i>Président /Rapporteur</i>
Karen, POLIZZI Reader, Imperial College London	<i>Rapporteuse</i>
Philippe, NGHE MdC, ESPCI Paris – PSL	<i>Examineur</i>
Yannick, RONDELEZ DR, CNRS	<i>Directeur de thèse</i>

*To my parents, my family, my teachers  
And all those who were my living source  
Of inspiration and encouragement.*

## Résumé

Les circuits moléculaires *in vitro*, basés sur des chimies programmables par ADN, peuvent exécuter une gamme croissante de fonctions de haut niveau, telles que le calcul au niveau moléculaire, la reconnaissance d'images ou de motifs chimiques et la génération de motifs. La plupart des démonstrations rapportées, cependant, ne peuvent accepter que l'acide nucléique comme signaux d'entrée. Cela limite la portée de leur utilisation à la fois dans des contextes de recherche et appliqués, car la signalisation biologique ne repose souvent pas sur des acides nucléiques. De petites molécules sont utilisées à la place.

Nous abordons le problème en introduisant une stratégie générale pour interfacer les circuits basés sur l'ADN avec des signaux non ADN, basés sur des modules de traduction d'entrée. Cela se fait dans la boîte à outils PEN-DNA. Les modules de traduction contiennent une partie réponse ADN et une partie facteur de transcription allostérique, et utilisent une conception simple qui les rend entièrement réglables et modulaires. Ils peuvent être conçus pour transmettre ou inverser la réponse naturellement associée à la présence d'une entrée donnée. Nous démontrons la combinaison de la détection et du traitement au niveau de l'ADN en assemblant ces modules de traduction avec des motifs d'amplification robustes et sans fuite.

Tout d'abord, nous construisons des circuits de détection qui fournissent une réponse temporelle quantitative fluorescente à la concentration d'une entrée de petite molécule, avec une bonne spécificité et sensibilité. La programmabilité de la couche d'ADN est ensuite exploitée pour effectuer une opération de traitement de signal basée sur l'ADN. En particulier, une inversion logique, une modulation du signal et une tâche de classification sur deux entrées sont présentées.

Nous poursuivons pour démontrer que les circuits d'ADN sont également compatibles avec l'activité enzymatique de petites molécules et peuvent fournir des sorties chimiques. Plus précisément, nous montrons que nos circuits peuvent être utilisés dans le contexte du flux de travail d'évolution dirigée PEN-CSR *in vitro* et étendre sa portée aux enzymes n'agissant pas sur l'ADN. Nous détectons l'activité des enzymes immobilisées et amplifions leurs gènes en réponse dans une réaction en un seul pot. Ce travail a été réalisé avec la sous-unité B de Tryptophane synthase, nous présentons donc également certaines de nos observations spécifiques à cette enzyme.

Enfin, nous passons à l'exploration d'alternatives aux composants particuliers que nous avons utilisés dans les démonstrations du concept de détection. Nous montrons que notre architecture est généralisable en ce qui concerne l'ADN et les enzymes de traitement du signal, les facteurs de transcription et les analytes, et les différentes températures. Dans l'ensemble, cela ouvre des possibilités à la fois dans les applications futures du PEN-CSR *in vitro* et dans la création de programmes moléculaires évolutifs avec des entrées de petites molécules et des sorties chimiques.

## Abstract

*In vitro* molecular circuits, based on DNA-programmable chemistries, can perform an increasing range of high-level functions, such as molecular level computation, image or chemical pattern recognition and pattern generation. Most reported demonstrations, however, can only accept nucleic acid as input signals. This constrains the scope of their use in both research and applied contexts as biological signalling does not often rely on nucleic acids. Small molecules are used instead.

We address the issue by introducing a general strategy to interface DNA-based circuits with non-DNA signals, based on input-translating modules. This is done within the Polymerase-Nickase-Exonuclease (PEN)-DNA toolbox. The translating modules contain a DNA response part and an allosteric transcription factor part and use a simple design that renders them fully tunable and modular. They can be designed to either transmit or invert the response naturally associated with the presence of a given input. We demonstrate the combination of the sensing and the DNA-level processing by assembling these translating-modules together with robust and leak-free amplification motifs.

First, we build sensing circuits that provide a fluorescent quantitative time-response to the concentration of a small-molecule input, with good specificity and sensitivity. The programmability of the DNA layer is then leveraged to perform DNA-based signal processing operation. In particular, logical inversion, signal modulation and a classification task on two inputs is presented.

We move on to demonstrate that DNA circuits are also compatible with small molecule enzymatic activity and can provide chemical outputs. Specifically, we show that our circuits can be used in context of the *in vitro* PEN-CSR directed evolution workflow and expand its scope to non-DNA acting enzymes. We detect the activity of the immobilised enzymes and amplify their genes in response in a one-pot reaction. This work was done with Tryptophan synthase subunit B, so we also present some of our observations specific to this enzyme.

Finally, we move on to exploring alternatives to the particular components we have used in the demonstrations of the sensing concept. We show that our architecture is generalisable with respect the signal processing DNA and enzymes, the transcription factors and the analytes, and different temperatures. Overall, this opens possibilities both in the future applications of *in vitro* PEN-CSR as well as creating scalable molecular programs with small molecule inputs and chemical outputs.

*Mais les yeux sont aveugles. Il faut chercher avec le cœur.*

– *Antoine de Saint Exupéry*

## Acknowledgements

Gabriel Ragala undertook 2nd year BSc internship in the summer of 2018 and contributed to transcription factor protein production. Camille Lescanne took part in enzyme production and in-bulk activity detection during her 3<sup>d</sup> year ESPCI diploma internship in the summer of 2019. It was supervised by Yannick Rondelez with help from Guillaume Gines.

Expanding on this, I would like to start by thanking Yannick, whose lecture at Imperial over 3 years ago led to this great adventure. The abstract to the lecture was incomprehensible so I attended it without enthusiasm, yet what followed is something I colloquially describe as “love at first sight”. Beyond the most stimulating discussions, exciting ideas and words of wisdom, I express gratitude for the entire scope of opportunities I happened to receive. This includes Parisian life as whole – a lifechanging experience, the wonderful colleagues, both within the SPM team and broader Gulliver, as well as broader professional connections in Paris and beyond. I am also not going to overlook the luxurious working conditions I have been provided with.

I would like to express my appreciation to the reviewers, Drs Karen Polizzi and Gilles Truan, who kindly accepted to review my work and stood by their commitment even in face of a slipping submission date. Likewise, I would like to thank Philippe Nghe for participating in the defence as an examinateur.

My molecular programming adventure started together with Rocío Espada. Having arrived at the same time and having been placed on the neighbouring desks, I am thankful to her for remaining by my side (not only physically!) throughout what was sometimes rather emotional times. Being my closest advisor on all personal matters in Paris, you also served as an inextinguishable beacon of inspiration for learning and doing science. Our separation may be the most difficult page to turn, yet I wish you all the best in raising the next person arriving in April.

I would like to thank Guillaume Gines, the master of molecular programs and the CFXs for taking his time to introduce me to the art, having the patience to hold the discussions with me, and sharing his experience as well as contributing to the friendly running of the lab. At this point I would also like to mention Rémi Sieskind, who welcomed me to the lab before my actual arrival and helped me arrange the life I lived in Paris for over three years. This concerns both induction to the lab, in organisation of which he was central as well as setting me up in the ways of Navigo, Velib and Picard, without which my survival would be unlikely. Rémi’s work also laid foundations to this entire project and I am grateful to him for his time spent teaching me the art of microfluidics and IVTT as well as sharing his materials. I would like to add that Thibault Di Meo’s expertise on protein work was instrumental to the production of some of the critical components of this work.

I would like to thank the members of the SPM team past and present, who formed by family away from family. I would take the risk of misspelling or omitting someone and attempt to list them all: Roberta, Anne-Sophie, Thomas, Antoine, Vincent, Taro, Filippo, Adèle, Margarida. Thank you putting up with me, my taste of music and humour. I also appreciate all of the discussions and meetings formal and informal we had over the years.

The interns I had the pleasure of working with, Gabriel Ragala and Camille Lescanne, that are mentioned so dryly above were there at turning points of my project. I therefore would like to express

my gratitude to them for the fervour they worked with inspiring me to push the project further. In just 2 months with each of you we have achieved so much.

I would like to thank more broadly the Gulliver lab. As the name suggests, it is a magical and peculiar place that enriched me both academically and culturally. I am grateful for the ambience created by Élie Raphaël and consolidated by Olivier Dauchot. In particular both the social events proposed as well as my strange hour working habits were accommodated without a question. I was also supported by Solange Rogue, who helped me navigate the formal complexities of the CNRS when I arrived, the role is succeeded by Fée Sorrentino, who helped me finalise my documents during my departure.

Gulliver kitchen and corridor discussions were usually short but innumerable and varied in issues, so once again, I hope not to miss anyone off here and thank Matan, Pierre, Vincent Bertin, Paul, Juliana, Maitane, Matryna, Martina (hopefully I didn't mix you up), Samuel, Alice, Claire, Yann, Mengshi, Charlotte, Maziar, Michael Schindler, Ludwik Leibler, Tereza Lopez-Leon, David Lacoste, Anthony Maggs, Mathylde Reyssat and Joshua McGraw among others in no particular order.

Beyond Gulliver, I was privileged to participate in the journal club with members of the LBC lab with Adeline helping with the running. Inputs from Clément Nizak, Olivier Revoire and Philippe Nghe among others helped me grasp the concepts of directed evolution. Equally I would like to thank CRI for providing me with an opportunity to teach. I thank Virginie Chomier for putting her faith in me, the students for their time and passion for learning and the senior colleagues, from whom I learnt so much of the art of teaching.

I would like to thank Marko Storch for sharing with me his passion for simple nucleic acid-based engineering methods that defined my interests, motivated me to stay in science and get this far. Likewise, I would like to thank Matt Haines for sharing his experiences of going into adult academic life and being a friend.

My network of support and inspiration starts with my new friends in Paris but spans far beyond: North America, Japan, the UK, most of Europe... I thank you all.

Finally, there is family. This work in part was provoked by my late grandfather, who, when reviewing my previous project, in his last words to me said: "Well, it seems like none of the crap changed since my time." – referring to his work in the 90s. I hope he would have found novelty here. I would like to thank my grandmothers, who are always keen to hear my latest news and make every effort to understand my work. Above all, I thank my parents, without whom I would have not come to be. Their existence drives me every day to aspire to greater things. Through care and dedication to my education they provided me with all that I know and am able to do. This in turns opened me the exciting life I had, including the time of this PhD.

Acknowledgements may change in the final version.

# Table of Contents

RESUME .....	VII
ABSTRACT .....	IX
ACKNOWLEDGEMENTS .....	XI
TABLE OF CONTENTS .....	XIII
LIST OF FIGURES .....	XVI
LIST OF TABLES .....	XX
<b>INTRODUCTION .....</b>	<b>1</b>
0.1 AIMS .....	1
0.2 DEOXYRIBONUCLEIC ACID (DNA) .....	2
0.2.1 <i>Structure and physiochemical properties</i> .....	2
0.2.2 <i>Replication</i> .....	4
0.2.3 <i>Biological context</i> .....	8
0.2.4 <i>Proteins</i> .....	9
0.2.5 <i>Enzymes</i> .....	16
0.2.5.1 <i>Tryptophan synthase</i> .....	18
0.2.6 <i>Response to the environment through regulation of active protein quantities.</i> .....	18
0.2.6.1 <i>Response with protein already present</i> .....	19
0.2.6.2 <i>Response through control of protein production</i> .....	20
0.3 PROTEIN ENGINEERING .....	27
0.3.1 <i>Rational design</i> .....	27
0.3.1.1 <i>Sequence-based rational design</i> .....	28
0.3.1.2 <i>Structure inference-based methods</i> .....	29
0.3.2 <i>Directed evolution</i> .....	30
0.3.2.1 <i>A case of Tryptophan synthase subunit B directed evolution</i> .....	32
0.4 MOLECULAR PROGRAMMING .....	32
0.4.1 <i>PEN-DNA toolbox</i> .....	34
<b>CHAPTER 1: SENSING SMALL MOLECULES AS MULTIPLE MOLECULAR PROGRAM INPUTS. ....</b>	<b>37</b>
1.1 AIMS .....	37
1.2 MOTIVATION .....	38
1.3 DESIGN STRATEGY .....	39
1.3.1 <i>Circuit design and optimization</i> .....	40
1.3.2 <i>Sensing module concept</i> .....	43
1.4 SMALL MOLECULE SENSING USING POSITIVE AND NEGATIVE TF LOGICS .....	46
1.4.1 <i>Direct signal relay</i> .....	46
1.4.2 <i>Signal inversion</i> .....	48
1.4.3 <i>Design details</i> .....	51
1.5 MULTIPLE INPUT IMPLEMENTATION .....	53
1.5.1 <i>Multiple input processing variations</i> .....	55
1.6 SUMMARY .....	59
<b>CHAPTER 2 TOWARDS EVOLVING SMALL-MOLECULE ENZYMES IN EMUSIO. ....</b>	<b>61</b>
2.1 AIMS .....	61
2.2 INTRODUCTION .....	62
2.2.1 <i>Compartmentalised self-replication</i> .....	62
2.2.2 <i>In vitro PEN Compartmentalized Self-Replication</i> .....	63
2.2.3 <i>In vitro PEN CSR components</i> .....	66
2.2.3.1 <i>Plasmids</i> .....	66
2.2.4 <i>In vitro transcription translation IVTT</i> .....	68
2.2.5 <i>Primer producing templates</i> .....	68

2.3	RESULTS .....	69
2.3.1	<i>Combining the sensor with analyte synthesis</i> .....	69
2.3.2	<i>Detecting in situ synthesis</i> .....	70
2.3.2.1	Detection with conventional methods .....	70
2.3.2.2	Detection with a molecular program .....	73
2.3.3	<i>Optimising the sensor to work with IPA</i> .....	74
2.3.3.1	Bulk enzyme activity selection .....	76
2.3.3.2	Tryptophan synthase from <i>Pyrococcus furiosus</i> strongly binds DNA .....	83
2.3.3.3	rT hinders IPA PEN-DNA circuits .....	90
2.3.3.4	L-trp synthesis is suitable for in emulsion detection .....	93
2.3.3.5	Primer production time affects the quantity of the PCR product .....	95
2.3.4	<i>On-bead enzyme activity selection</i> .....	96
2.4	SUMMARY .....	103
<b>CHAPTER 3 EXPLORING BIOSENSOR IMPLEMENTATION. ....</b>		<b>104</b>
3.1	AIMS .....	104
3.2	ALTERNATIVE SENSING ARCHITECTURES .....	105
3.2.1	<i>T7 transcription system can start PEN circuits</i> .....	105
3.2.2	<i>Nb.BsrDI for molecular program sensing</i> .....	109
3.2.3	<i>Chemical optimisation of the pskT</i> .....	114
3.3	POTENTIAL FOR CIRCUIT EXPANSION .....	116
3.3.1	<i>B11b – an orthogonal switch</i> .....	116
3.3.1.1	An alternative strategy of inverting the signal .....	117
3.3.1.2	Sensing separately from direct actuation .....	120
3.3.2	<i>Nb.BssSI can be used to drive a circuit</i> .....	123
3.3.3	<i>Linear response to sensing</i> .....	125
3.3.4	<i>Using other transcription factors</i> .....	126
3.3.4.1	FadR .....	127
3.3.4.2	TetR .....	129
3.3.5	<i>Using different polymerases to drive the circuit at 37°C</i> .....	134
3.3.6	<i>TrpR-sensing is possible at higher temperatures</i> .....	135
3.3.6.1	Running the existing IPA circuit at higher temperatures .....	135
3.3.6.2	Transferring the architecture into a different switch. ....	140
3.3.7	<i>Other aspects to consider</i> .....	149
3.3.7.1	ppT design: trigger affinity impacts switch behaviour. ....	149
3.3.7.2	Transcription factor formulation strongly affects the behaviour of the sensor .....	153
3.4	SUMMARY .....	158
<b>DISCUSSION AND CONCLUSION .....</b>		<b>159</b>
<b>MATERIALS AND METHODS .....</b>		<b>162</b>
4.1	MOLECULAR BIOLOGY .....	162
4.1.1	<i>Transcription factor production</i> .....	162
4.1.2	<i>Enzyme purification</i> .....	163
4.1.3	<i>Site-directed mutagenesis</i> .....	163
4.1.4	<i>SNAP-tagging TrpB</i> .....	164
4.2	MOLECULAR PROGRAMMING .....	164
4.2.1	<i>The miR buffer</i> .....	164
4.2.2	<i>The IPA buffer</i> .....	165
4.2.3	<i>Oligos used</i> .....	165
4.2.4	<i>Running the reactions</i> .....	165
4.3	MICROFLUIDICS AND DIRECTED EVOLUTION WORKFLOW .....	166
4.3.1	<i>Hydrogel bead generation</i> .....	166
4.3.1.1	Hyaluronic acid thiolation .....	166
4.3.1.2	Microfluidic generation .....	166
4.3.1.3	Emulsion break and washing .....	168
4.3.2	<i>IVTT protein expression and immobilisation</i> .....	168

Table of Contents	XV
-------------------	----

---

4.3.3	<i>Vectors and genes</i> .....	169
<b>APPENDIX A STRUCTURE INFERENCE-BASED METHODS FOR RATIONAL DESIGN OF PROTEINS.</b> .....		<b>173</b>
<b>APPENDIX B L-TRP SYNTHESIS DETECTION THROUGH ABSORBANCE READINGS.</b> .....		<b>175</b>
<b>BIBLIOGRAPHY</b> .....		<b>178</b>

## List of Figures

Figure 1 Schematic comparison of natural evolution and the process considered in this work. ....	1
Figure 2 Schematic illustration of DNA. ....	3
Figure 3 Comparison of DNA and RNA monomers. ....	4
Figure 4 Schematic illustration of different biological forms of DNA. ....	4
Figure 5 Schematic of DNA unwinding for replication. ....	5
Figure 6 Schematic of DNA replication fork with Okazaki fragments. ....	6
Figure 7 Polymerase chain reaction (PCR) schematic. ....	7
Figure 8 Schematics of rolling circle amplification. ....	7
Figure 9 Nicking Enzyme Amplification Reaction (NEAR) schema. ....	8
Figure 10 Schematic of possible point mutations. ....	8
Figure 13 Schematic of transcription. ....	10
Figure 14 Codon table showing RNA triplet correspondence to amino acids. ....	11
Figure 15 Schematic of translation adapted from. ....	12
Figure 16 General amino acid structure ....	12
Figure 17 A table of the 20 common amino acids. ....	13
Figure 19 Polypeptide chain is constrained in 3D space. ....	14
Figure 20 Common protein secondary structures. ....	15
Figure 21 Helix-turn-helix motif binding DNA. ....	15
Figure 22 Tertiary and quaternary protein structure examples. ....	16
Figure 23 Enzyme mode of action examples. ....	17
Figure 25 Ribosomal traffic illustration. ....	21
Figure 26 Hok/sok system as an example of mRNA regulation in trans as in <sup>83</sup> . ....	22
Figure 27 Aptamer function in a biomedical setting schematic. ....	23
Figure 28 Transcriptional attenuation of the trp operon. ....	24
Figure 30 Summary of LacI and TrpR modes of action. ....	26
Figure 31 Rational engineering of TrpB example. ....	28
Figure 32 Site directed mutagenesis examples. ....	31
Figure 33 A toehold exchange reaction and its mechanism. ....	33
Figure 34 Genelets mode of action. ....	34
Figure 35 PEN-DNA toolbox elements. ....	35
Figure 36 Graphical abstract of chapter 1. ....	37
Figure 37 Aptamers used together to detect deoxycorticosterone (DC) and cortisol (CS). ....	38
Figure 38 Conversion template for miRNA detection in PEN ....	40
Figure 39 Source template (sT) concept. ....	41
Figure 40 Switch <sub>50</sub> performance of a range of temperatures. ....	42
Figure 41 Fluorescence time-trace of the switch with a range of pT concentrations. ....	43
Figure 42 Schematic of the protein sensing template (psT). ....	44
Figure 43 psT <sub>LacI</sub> detailed design. ....	44
Figure 44 psT <sub>LacI</sub> response to LacI. ....	45
Figure 45 Data processing workflow for Figure 44. ....	46
Figure 46 psT <sub>LacI</sub> -LacI-based analyte detection. ....	47
Figure 47 Design and implementation of psT <sub>TrpR</sub> . ....	48
Figure 48 Design and implementation of pskT <sub>LacI</sub> . ....	49

---

Figure 49 Design and implementation of $\text{pskT}_{\text{TrpR}}$ . .....	50
Figure 50 Effect of TrpR concentration on apparent $K_d$ of the L-trp $\text{pskT}$ -based sensor. ....	51
Figure 51 Detailed design of the source template (sT). .....	52
Figure 52 In detail representation of the sensing modules used in this work. ....	52
Figure 53 $\text{psT}_{\text{LacI}}$ and $\text{psT}_{\text{TrpR}}$ both driving the circuit. ....	53
Figure 54 Combinations of protein sensing (killer) templates with ranges of concentrations of two inputs. ....	54
Figure 55 Alternative plots for Figure 54D with altered input balance. ....	55
Figure 56 Alternative shapes for the positive/negative circuit with altered input balance. ...	56
Figure 57 Alternative plots for the positive/positive circuit with altered input balance. ....	57
Figure 58 Alternative plots for the negative/positive circuit with altered input balance. ....	58
Figure 59 Graphical abstract of chapter 2. ....	61
Figure 60 Original CSR workflow. ....	63
Figure 61 <i>In vitro</i> PEN-CSR scheme. ....	64
Figure 62 Generic plasmid map. ....	66
Figure 63 The His-tag purification process. ....	67
Figure 64 SNAP-tag mode of action schematic. ....	68
Figure 65 Primer producing template (ppT) design. ....	69
Figure 66 Effect of indole on miR-based reactions. ....	70
Figure 67 Characteristic tryptophan absorbance spectrum. ....	70
Figure 68 Sample chromatograms of filtered reaction mixtures. ....	72
Figure 69 Sensing of L-trp production <i>in situ</i> . ....	73
Figure 70 Agarose gel of PCR products in miR and IPA buffers with Vent and Q5 polymerases. ....	74
Figure 71 Agarose gel with PCR products of Vent polymerase in miR and IPA buffers with two template concentrations. ....	75
Figure 72 Detection and amplification of various concentrations of EcB in IPA. ....	77
Figure 73 Detection and amplification of various concentrations of PfB in IPA. ....	78
Figure 74 Agarose gel of PCR products retrieved from reactions shown in Figure 72, Figure 73. ....	79
Figure 75 Detection and amplification of a fine range of concentrations of EcB in IPA. ....	80
Figure 76 Agarose gel of PCR products retrieved from reaction shown in Figure 75. ....	81
Figure 77 EcB activity detection and gene amplification response. ....	82
Figure 79 Improved PEN enzyme activity detection and PCR amplification. ....	84
Figure 80 PCR amplification with various concentrations of PfB. ....	85
Figure 81 Electrophoretic mobility shift assay with PfB T292S at various concentrations and 59 ng/ 1.8 nM of plasmid DNA. ....	85
Figure 82 BSA as DNA competitor for PfB binding. ....	86
Figure 83 Range of template for PCR with PfB. ....	87
Figure 84 Heat shock purification of PfB protein. ....	88
Figure 85 IVTT expression and heat-shock purification of PfB. ....	89
Figure 86 Range of rT in IPA reaction. ....	90
Figure 87 Range of rT in an IPA reaction. ....	91
Figure 88 Range of rT in a miR reaction. ....	92
Figure 89 Effect of rT on switching gradient in miR. ....	93

---

Figure 91 Co-incubated L-trp and no-L-trp emulsions micrograph. ....	95
Figure 92 Impact of primer production time on PCR. ....	96
Figure 93 Range of L-thrombin in IPA reaction. ....	97
Figure 94 Investigating activity of EcB-SNAP. ....	98
Figure 95 Amplification in response to EcB-SNAP. ....	99
Figure 96 Determining limit of detection for EcB-SNAP in IPA. ....	100
Figure 97 Investigating activity of EcB-SNAP on beads. ....	101
Figure 98 Amplification in response to EcB-SNAP. ....	102
Figure 99 Graphical abstract of chapter 3. ....	104
Figure 100 Sequence-level representation of the T7 modules used as controls. ....	105
Figure 101 PEN circuit triggered by T7 RNA polymerase. ....	106
Figure 102 Operator placement around T7 promoter. ....	107
Figure 103 T7-based sensor design. ....	108
Figure 104 Simulation of secondary structures of transcripts from .Figure 103. ....	108
Figure 105 Nt. BstNBI could be modified. ....	109
Figure 106 Nb.BsrDI-driven $psT_{LacI}$ in detail. ....	110
Figure 107 Range of LacI with a Nb.BsrDI-driven $psT_{LacI}$ switch. ....	110
Figure 108 Range of pT with a Nb.BsrDI-driven $psT_{LacI}$ switch. ....	111
Figure 109 Nickase concentration effect on the switch. ....	112
Figure 110 Nb.BsrDI- $psT_{LacI}$ -based IPTG detection. ....	113
Figure 111 Natural and symmetric LacI operators. ....	114
Figure 112 $pskT_{LacI}$ variants tested in this work. ....	115
Figure 113 Chemical variants of $pskT_{LacI}$ detecting IPTG. ....	115
Figure 114 IPTG detection by $pskT_{LacI}$ “original S hpt”. ....	116
Figure 115 B11 and B11b are orthogonal. ....	117
Figure 116 Inverter template (inT) mode of action. ....	118
Figure 117 Inverter template design variants. ....	119
Figure 118 Inverter template variant tests. ....	119
Figure 119 B11topTB11b (a) test with pT. ....	120
Figure 120 B11b sensing circuit linked to B11 ppT. ....	121
Figure 121 B11btoB11 design and accompanying templates. ....	122
Figure 122 Range of B11btoB11 template concentrations in IPA buffer. ....	123
Figure 123 Nb.BssSI-driven switch components design. ....	124
Figure 124 aT Bss 2 can drive switching in miR circuits. ....	125
Figure 125 Linear response template designs. ....	125
Figure 126 Linear response to L-trp sensing. ....	126
Figure 127 Design of $psT_{FadR}$ . ....	127
Figure 128 FadR effect on $psT_{FadR}$ driven sensor. ....	128
Figure 129 Design of $psT_{TetR}$ . ....	129
Figure 130 Evaluation of $psT_{TetR}$ designs. ....	130
Figure 131 Range of TetR concentrations with $psT_{TetR}$ -driven circuit. ....	131
Figure 132 TetR regulation of a circuit at a range of temperatures. ....	132
Figure 133 Detection of tetracycline at a range of temperatures. ....	133
Figure 134 Range of tetracycline detection. ....	133
Figure 135 Detection of IPTG with a Klenow-driven $psT_{LacI}$ circuit. ....	135

---

---

Figure 136 TrpR incubated at a range of temperatures senses L-trp in IPA. ....	137
Figure 137 IPA reaction running over a range of temperatures. ....	137
Figure 138 Cy5 fluorescent traces of IPA reaction at 40 °C. ....	138
Figure 139 Fluorescent traces of IPA reaction at 42 °C. ....	139
Figure 140 Bo switch components used in this work. ....	140
Figure 141 Bo-based switch detecting L-trp at different temperatures in IPA buffer. ....	141
Figure 142 Range of sT Bo at different temperatures in IPA buffer. ....	142
Figure 143 Range of psT <sub>TrpR</sub> Bo at different temperatures in IPA buffer. ....	143
Figure 144 Range of Bst polymerase added to Bo-based switch. ....	144
Figure 145 Range of L-trp detected by psT <sub>TrpR</sub> Bo-driven switch. ....	145
Figure 146 Ranges of pskT <sub>TrpR</sub> Bo. ....	146
Figure 147 FAM fluorescent traces for some samples in Figure 146. ....	147
Figure 148 pskT <sub>TrpR</sub> Bo detecting L-trp over a range of temperatures. ....	148
Figure 149 Optimisation step of the pskT <sub>TrpR</sub> -based circuit used in 1.3.2. ....	149
Figure 150 Examples of different ppT designs. ....	151
Figure 151 L-trp detection and amplification response with different ppTs in IPA. ....	152
Figure 152 Effect of change of TrpR stock on detection capacity of pskT-based switch. ....	154
Figure 153 Effect of Lacl concentration on IPTG resolution. ....	155
Figure 154 Loss of IPTG detection of low Lacl concentrations. ....	155
Figure 155 Cy5 fluorescence for a range of TrpR protein affecting psT-based switch. ....	157
Figure 156 An example of a gel used to evaluate TF purification. ....	163
Figure 157 Characteristic absorbance readings at 290 nm (blue) and 320 nm (orange) during L-trp synthesis reactions. ....	176
Figure 158 Absorbance spectra of L-tryptophan used in this work. ....	176
Figure 159 UV–vis absorption spectra of pyruvic acid over a range of pH. ....	177

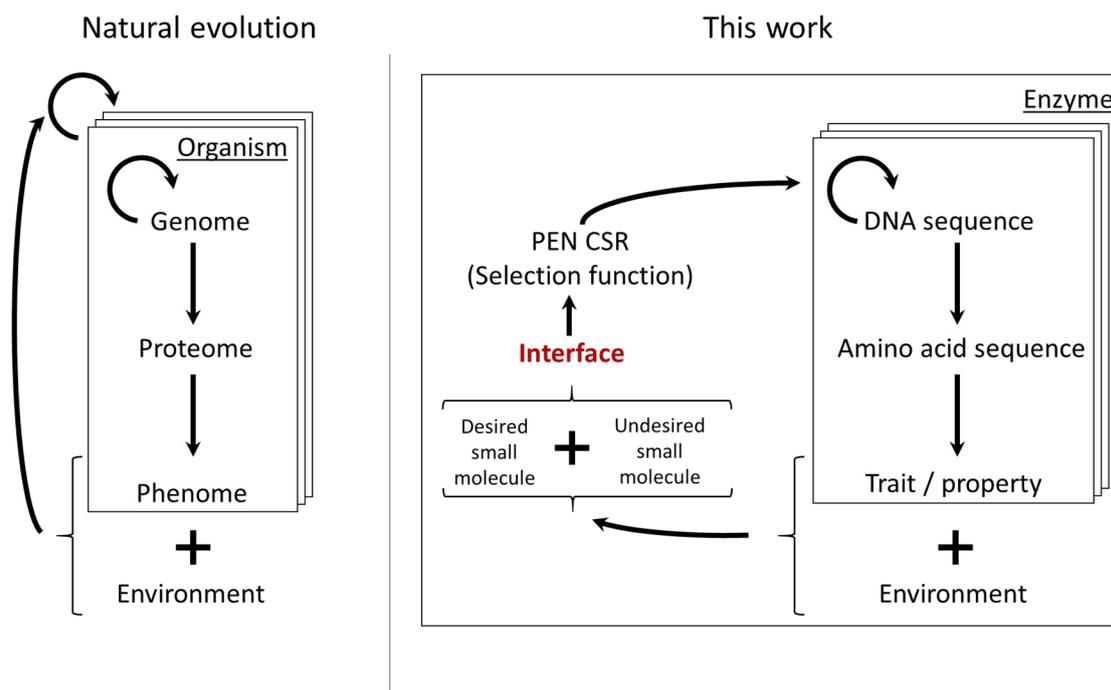
## List of Tables

Table 1. Oligos used in the switch <sub>50</sub> . .....	41
Table 2 Trigger sequences of B11 and B11b. ....	117
Table 3 Primers used for transcription factor production. ....	162
Table 4 Primers used for enzyme cloning. ....	163
Table 5 Primers used for site directed mutagenesis. ....	163
Table 6 Primers, specifically used for SNAP-tag addition. ....	164
Table 7 Oligos used in molecular programs. ....	165
Table 8 Cell extract-based in vitro transcription/translation (IVTT) mixture used. ...	168
Table 9 Vectors and genes used in this work unavailable in public databases. ....	169

# Introduction

## 0.1 Aims

This work aims to explore transcription factors as a potential interface between PEN-DNA molecular programming toolbox and small-chemical molecules. A goal we set ourselves to demonstrate the power of this approach concerns the use of *in vitro* CSR to evolve tryptophan synthase subunit B (**Figure 1**). We therefore feel it is important to outline below an introduction to some concepts the work is based on. In particular, we will start with some useful deoxyribonucleic acid (DNA) properties, followed by introduction to molecular programming. It is an approach for rational creation of chemical reaction networks, based on these properties. We shall also cover a potential area of application of this work in the field of directed evolution and briefly introduce the model target we used. The three chapters that follow will build on this introduction. The first chapter will concern the demonstration of a functional transcription factor-based interface, whilst the second will look at a potential application within directed evolution context. We will finish with a chapter where we discuss transcription factor-based approaches alternative to the ones reported in prior chapters as well as some other considerations regarding the work presented.



**Figure 1 Schematic comparison of natural evolution and the process considered in this work.** The concepts shown will be covered in the introduction. On the left: the genome replicates within the organism as it grows and/or divides; the genome determines the proteome, which in turn gives rise to a sum of traits (the phenome). These, together with the environment determine the success of the reproduction of an organism as a whole. Multiple organisms go through this cycle, some being more successful than others and rising in relative numbers. On the right: one possible variant of an artificial molecular evolution process. An enzyme of choice is encoded by a particular DNA sequence. The genetic variant determines the amino acid sequence composition of the enzyme and thus its traits/properties. In the context of a particular environment, these manifest in catalytic production of desired and/or undesired small molecules. Through the interface we envisage (red), information about the product generated is relayed to PEN CSR – a selection tool that determines the degree to which the particular DNA sequence, responsible for the particular enzyme variant is replicated, according to a pre-encoded selection function.

## 0.2 Deoxyribonucleic acid (DNA)

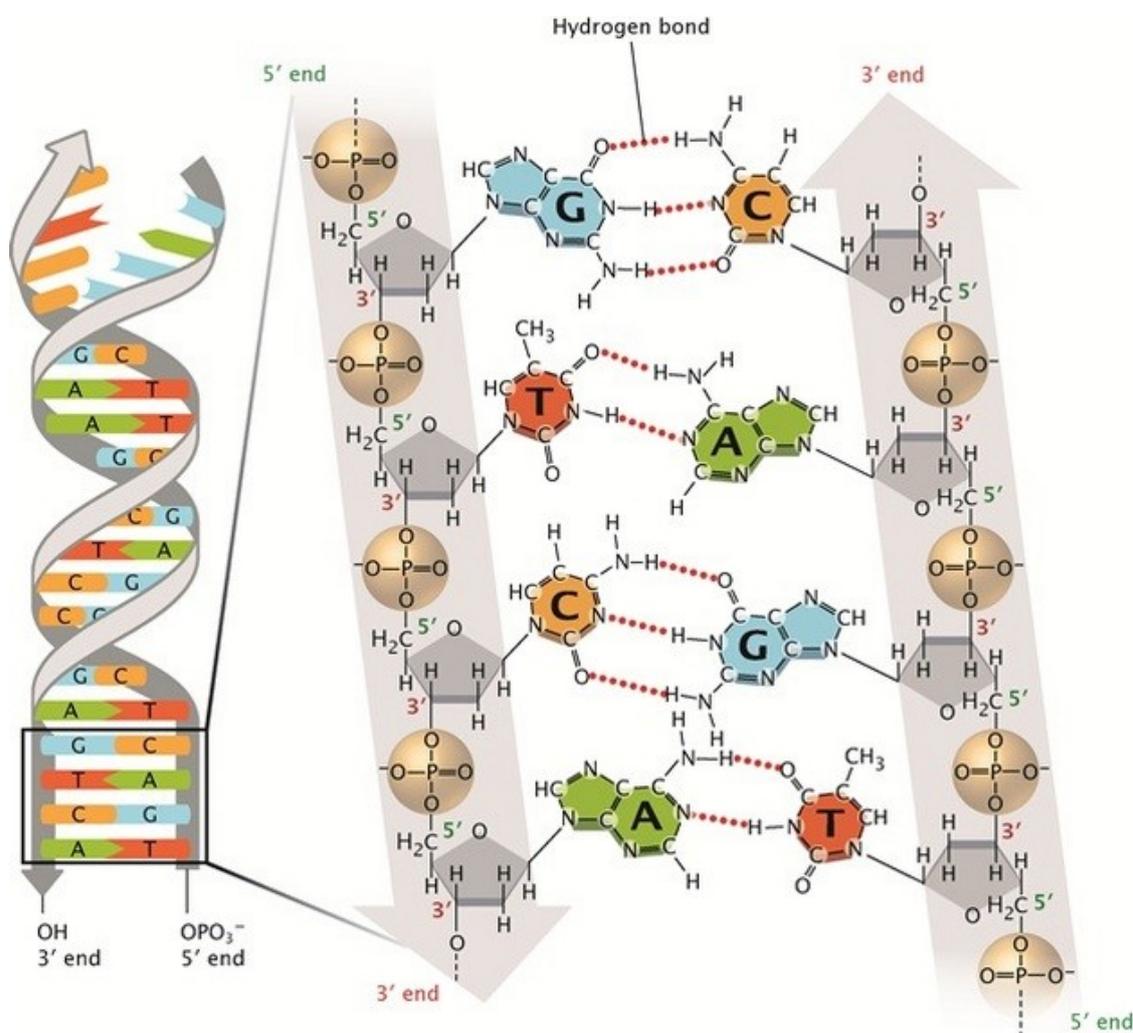
### 0.2.1 Structure and physiochemical properties

Deoxyribonucleic acid commonly known as DNA, is frequently referred to as the “molecule of life”. Indeed, a series of experiments<sup>1-3</sup> have shown that it is specifically the DNA, that is responsible for introducing genetic traits when added to a new host. We shall give a more detailed example of how this can be achieved artificially in the following section.

A single strand (ssDNA) consists of deoxyribose-phosphate backbone and nitrogenous bases. The backbone is responsible for acidic properties, and negative charge at physiologically relevant pH due to phosphate groups present ( $pK_a \sim 1^4$ ). This negative charge, proportional to molecule size frequently determines DNA interactions with DNA-binding molecules. The latter typically have positive patches responsible for sequence independent (see below), non-specific DNA binding<sup>5</sup>. Negative charge also means that presence of monovalent and divalent cations can strongly affect DNA interactions with itself and proteins<sup>7</sup> by competing with other charge-charge interactions. Attachment of the phosphate groups at the 3<sup>rd</sup> and the 5<sup>th</sup> carbons of asymmetric deoxyribose molecule allows for designation of directionality: ssDNA molecule can have the 3' and the 5' end. By convention, the 5' to 3' directionality of a sequence presented is assumed in line with typical DNA polymerisation *in vivo*. This direction is sometimes emphasised by an arrow or half-arrow head, when DNA is drawn schematically.

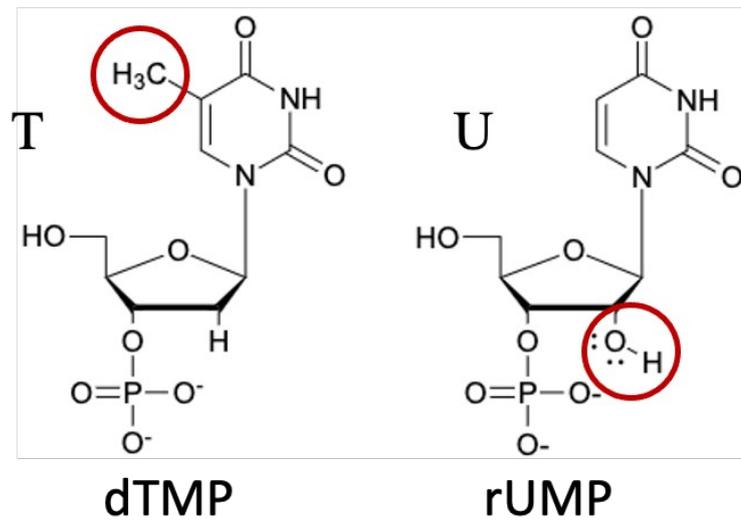
The backbone carries an arbitrary sequence of nitrogenous bases: cytosine [C], guanine [G], adenine [A] or thymine [T]<sup>8,9</sup>. These represent genetic information. A particular sequence of nucleotides can be targeted for binding as we will later show. This is based on specific, sequence-determined spatial arrangement of the nitrogenous bases that collectively stabilise their binding to a complementary biomolecule. As noted by Chargaff, the stoichiometry of about 1:1 tends to hold for the A-T pair and G-C pair within a biological DNA duplex and even individual strands<sup>10-12</sup>. The former observation in turn formed some of the basis for the Watson-Crick DNA molecular structure model, postulating DNA to generally be present in living organisms as an antiparallel double helix with nitrogenous bases on the inside, paired as Chargaff's rule would suggest (**Figure 2**). Hence, the unit of length of DNA commonly used would be termed a “base-pair” (bp) or kilobase-pair (kbp), representing 1000 bp.

Crick cautiously pointed out that each strand of the double stranded DNA (dsDNA) can separately serve as a template to re-produce the complement. This immediately suggested the semiconservative model of DNA replication: strands separate to each serve as template for one new complement, thus resulting in two new dsDNA molecules each containing one strand of the original duplex. This was later elegantly confirmed by Meselson and Stahl<sup>13</sup>. Some of the molecular programs we will discuss also rely on this property.



**Figure 2 Schematic illustration of DNA.** Double helical structure of DNA with a close-up view of some nucleotides as found in educational materials<sup>14</sup>. The arrows denote the 5' to 3' directionality. Cytosine, guanine, adenine and thymine are represented by C, G, A and T accordingly.

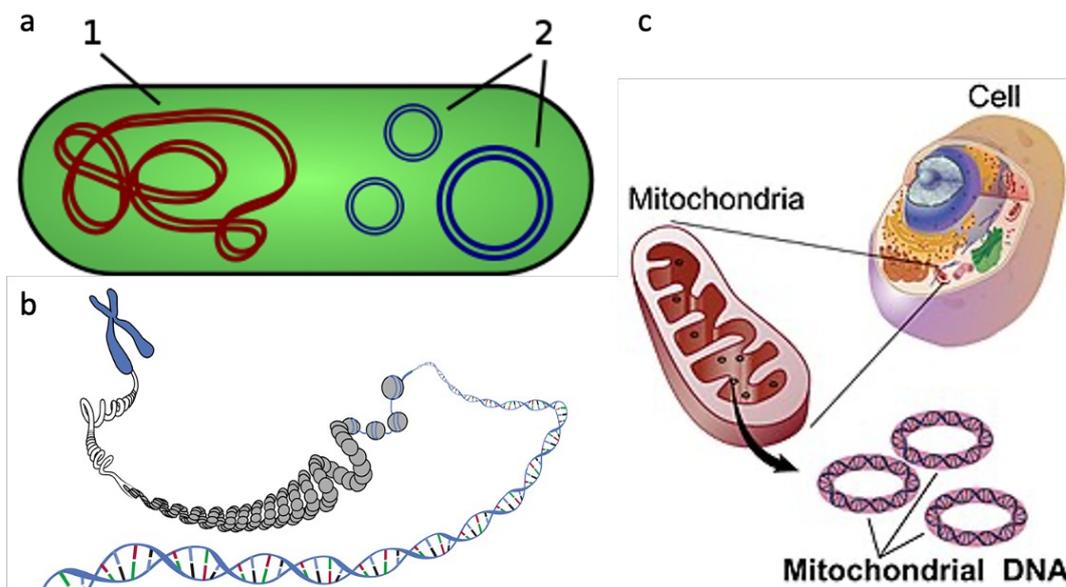
A small alteration to the structure of DNA, namely replacement of deoxyribose by ribose yields a drastically different polymer – ribonucleic acid, known as RNA. Additional hydroxyl group with 2 free electron pairs on the oxygen allow for intramolecular interactions. These include both instantaneous intra-strand hydrolysis in absence of free floating water molecules<sup>15</sup> and greater potential for intramolecular hydrogen bonding. In part, this is why RNA is typically present in single stranded form *in vivo*, often assuming intricate conformations<sup>16</sup>. In RNA, thymine is replaced by uracil, that is lacking the methyl group (**Figure 3**).



**Figure 3 Comparison of DNA and RNA monomers.** On the left deoxythymidine monophosphate. On the right Uridine monophosphate. Differing parts are highlighted by red circles: the methyl group of thymine and hydroxyl group of ribose.

## 0.2.2 Replication

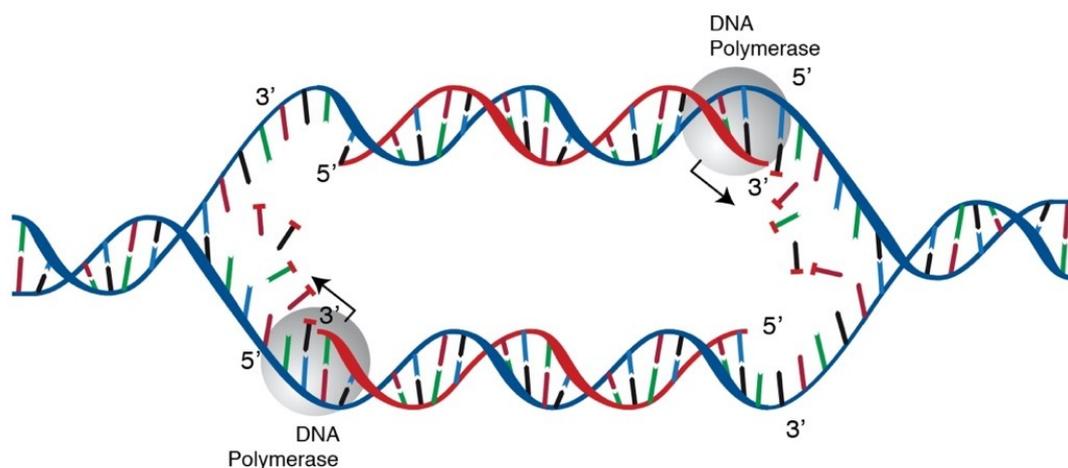
One of the key properties of living beings is the ability to grow and procreate. DNA encoding genetic information of every cell thus reproduces to sustain itself and the organism it is carried by. In nature, essential organismal DNA is present in form of one or more chromosomes<sup>17</sup>. These can be single copy and circular in case of prokaryotes, ranging in size from around 130 kbp to 14000 kbp<sup>18,19</sup>. Eukaryotes in turn, are characterised by multiple linear chromosomes sometimes present in near-identical copies (ploidy) and are usually over tens of Mbps in length<sup>20</sup>. Extrachromosomal DNA also exists among others in eukaryotic organelles<sup>21</sup> and in form of plasmids<sup>22</sup> (see 2.2.2) (**Figure 4**).



**Figure 4 Schematic illustration of different biological forms of DNA.** (a) Bacterial chromosome (1) and plasmid (2)<sup>23</sup>. (b) Eukaryotic chromosome partially unwound<sup>24</sup>. (c) Mitochondrial DNA, inside mitochondria, residing in a eukaryotic cell<sup>25</sup>.

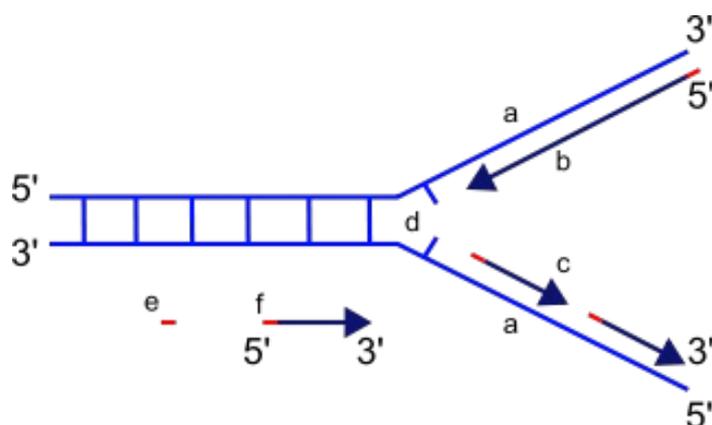
The exact process of replication varies and can be complex to the extent of not being fully unravelled for some organisms. In general, replication initiation, strand separation, replication frequency as well as final copy number are intricately controlled by the cellular machinery<sup>26</sup>. In brief, an origin of replication (or several) – a specific type of a DNA sequence – along with associated proteins are usually responsible for the initiation of these events. A certain form of feedback loop allows to monitor the copy number<sup>27</sup>, whilst synchronicity with overall cell division cycle is also possible for chromosomal DNA. When recruited, DNA-replication complex can unwind dsDNA, separate the strands and initiate strand replication<sup>28</sup>.

In addition to the given single stranded templates, several components are involved: a DNA polymerase – an enzyme (see 0.2.5) responsible for extending a DNA strand<sup>29</sup>; RNA primers – short pieces of RNA, complementary to the 3' ends of the region of interest on the template; deoxynucleoside triphosphates (dNTPs) – nucleotides containing triphosphate groups; suitable cellular environment, containing appropriate amount of divalent and monovalent cations. As we mentioned before, DNA polymerase can extend a nascent ssDNA strand in the 5' to 3' direction, however, only if a primer serves as an anchor, a template is provided to define the new sequence and dNTPs act as fuel and building material (**Figure 5**). In vivo, primers are typically generated by RNA primase that is part of the replication



units in grey.

The reader may notice that for synchronous replication of the antiparallel strands, through unidirectional complex progression,  $3' \rightarrow 5'$  DNA polymerase is seemingly required. Okazaki fragments are used as a consequence of the unidirectionality of DNA polymerase ( $5' \rightarrow 3'$ ). As the DNA replication complex advances along the dsDNA, one DNA polymerase unit extends a single primer in the  $5' \rightarrow 3'$  as usual. The other strand is replicated through extension of multiple primers constantly produced<sup>32,33</sup>. This allows extension of the second strand in the direction, opposite to that of the replication complex progression (**Figure 6**).



**Figure 6 Schematic of DNA replication fork with Okazaki fragments.** From <sup>34</sup>. The replication fork (d), contains partially unwound template strands (a), with one RNA primer (e) being extended in the direction of the replication fork motion (b), whilst multiple primers and hence Okazaki fragments (f) are formed on the antiparallel strand (c).

Artificial approaches have been developed, inspired by the process described. Based on the natural DNA amplification, several general requirements are apparent:

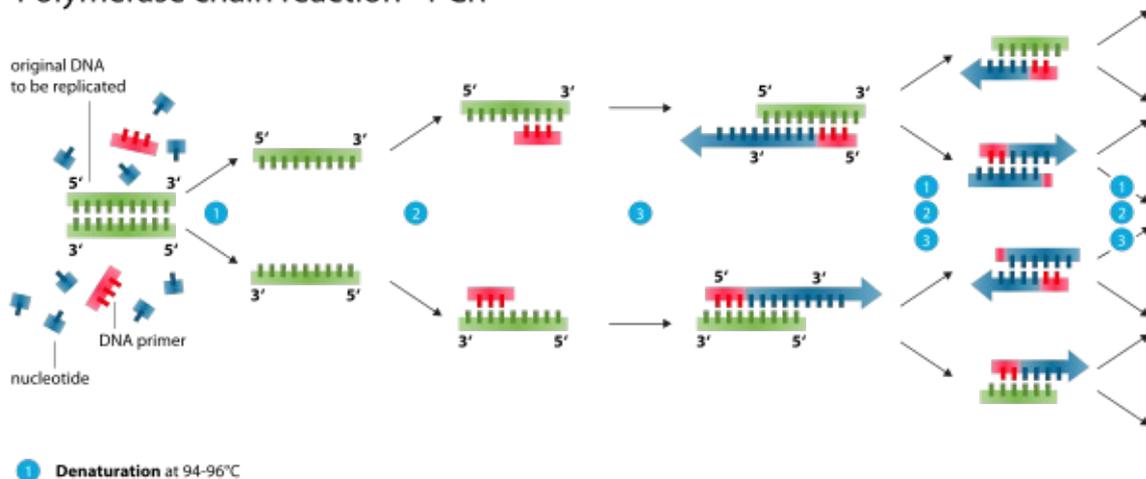
1. A template is needed to define the nascent strand to be produced.
2. DNA polymerase – a catalyst is required for DNA extension to occur at reasonable rate.
3. DNA polymerase requires a primer – a short nucleic acid sequence that binds template first and acts as an anchor for nucleotides to be added to as it is unable to do so from scratch.
4. dNTPs are used as fuel and material for the synthesis.
5. The buffer should allow for dsDNA stability as well as for DNA polymerase–DNA complex formation. Magnesium ions are also required as a cofactor for DNA polymerase.

Artificial DNA replication approaches can be divided into those requiring temperature cycling during the amplification reaction and those that do not (isothermal). Below we provide examples of both used in this work.

Polymerase chain reaction (PCR) is an amplification method widely used both for detection of specific DNA fragments (through specific primers) and production of large quantities of desired fragments for future use. The reaction somewhat captures the biological process. However, DNA primers are usually added to the solution and replication is controlled through temperature cycling we explain below. This imposes a requirement for a thermocycler – a machine capable of quickly altering reaction temperature in 40 – 100 °C range – and a DNA polymerase capable of withstanding such temperatures.

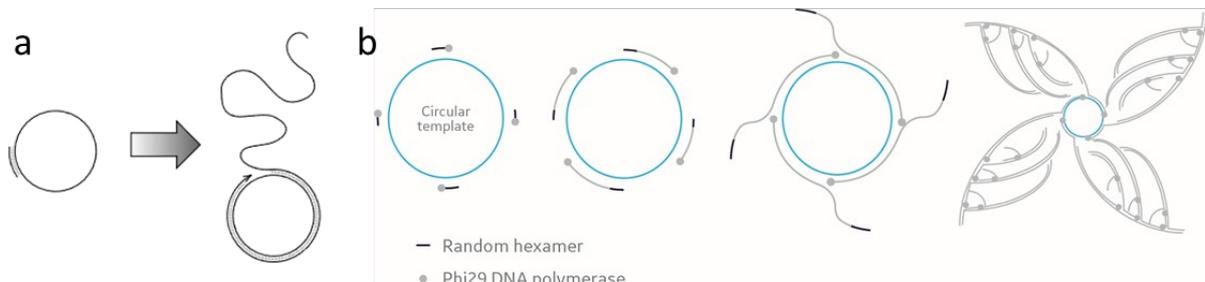
PCR centres around a cycle of 3 steps <sup>35,36</sup>: template denaturation, primer annealing and primer extension. During denaturation the reaction is heated to 95+ °C, disrupting weak forces stabilising the double-stranded helix. The reaction is then rapidly cooled to allow primers to anneal (~45-72 °C). The primers are extended by the polymerase and the temperature is shifted to favour polymerase processivity (~72 °C). Each cycle results in doubling of the template sequence building up the chain reaction (**Figure 7**).

## Polymerase chain reaction - PCR



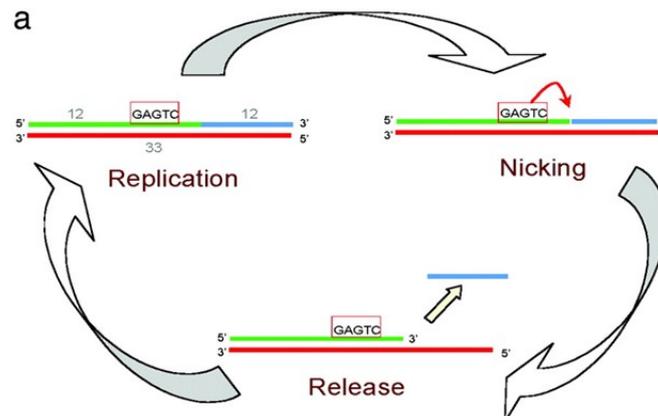
**Figure 7 Polymerase chain reaction (PCR) schematic.** Adapted from <sup>37</sup>. DNA template strands (green) are shown to be replicated using primers (red) and dNTPs (blue). 1. Denaturation. 2. Annealing. 3. Elongation. 2 cycles are shown.

Rolling circle replication (RCA) is an example of an isothermal method. In simplest form, it allows for linear DNA replication by following one strand of a circular piece of DNA by extending a single primer <sup>38</sup>. This can be modified in multiple ways. For example, if multiple primers are added, the process can become exponential<sup>39</sup> (**Figure 8**).



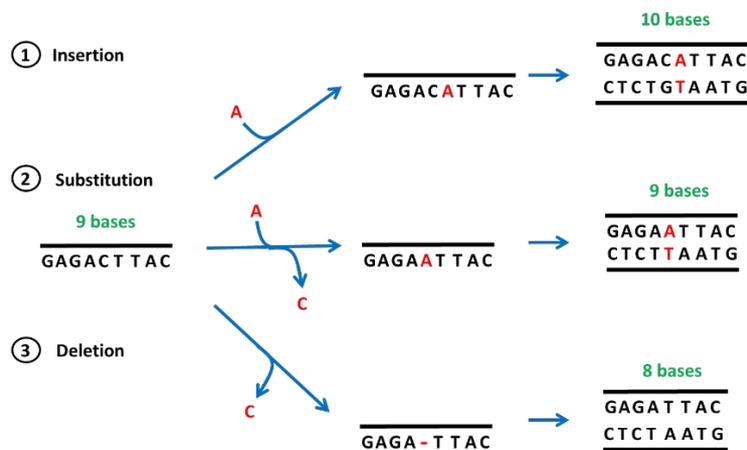
**Figure 8 Schematics of rolling circle amplification.** (a) Linear amplification. ssDNA is bound by a primer. Primer is then continuously extended along the circular template. (b) Exponential amplification through branched RCA. Multiple random primers are used giving rise to a highly branched structure. From<sup>40</sup>.

A variety of commercially available alternatives exists<sup>41</sup>. We would only briefly mention strand displacement amplification<sup>42</sup> (SDA) and related nicking enzyme amplification reaction (NEAR)<sup>43</sup>. As the names would suggest, they rely on strand displacement and nicking enzymes for amplification. Such group of methods was recently reviewed<sup>44</sup>. They rely on the fact, that DNA can be nicked – a single strand break introduced – at a particular position by an enzyme – a nickase. A DNA polymerase would then consider the 5' fragment as a primer, extend and displace the 3' fragment. Upon re-formation of the initial dsDNA the process can be repeated, thus continuously generating defined ssDNA fragments (**Figure 9**).



**Figure 9 Nicking Enzyme Amplification Reaction (NEAR) schema.** As in<sup>42</sup>. Starting top left. Template strand (red) is bound to priming strand (green) and produced oligonucleotide (blue). The GAGTC site, Nt.BstNBI nickase recognition domain attracts the nickase to nick 4 bases downstream (red arrow). The priming strand is elongated and the short oligonucleotide is displaced. The cycle repeats.

We should note, that despite templating, in all cases a random incorporation of a nucleotide outside of the canonical pair is possible<sup>45</sup>. This is a source of mutations – alterations to the DNA sequence. In this case, either a transition – substitution for a like-ringed nucleobase, or a transversion – substitution for an unlike-ringed nucleobase are possible<sup>46</sup>. Likewise, single base insertions or deletions are possible if an extra random base is inserted or omitted, typically along a homopolymer region of the template<sup>47</sup> (**Figure 10**).



**Figure 10 Schematic of possible point mutations.** Insertion (1), substitution (2) or deletion (3) are possible.

### 0.2.3 Biological context

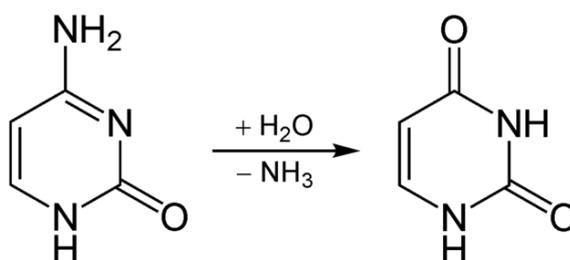
The physicochemical properties of DNA and RNA reflect their fundamental biological functions. Postulated by Crick as the central dogma of molecular biology<sup>48</sup>, in general, DNA is the primary information carrier. It is sustained through replication with information being transferred to messenger RNA (mRNA), which is then translated into proteins (**Figure 11**). These typically assume some structural or catalytic role and hence, give rise to observable traits (some examples will be provided in the following sections).



**Figure 11 Simplified schematic of the central dogma of molecular biology.** Information flow represented by the arrows. Information is replicate in form of DNA. It also flows into RNA and then proteins.

As mentioned, there are two major chemical differences between RNA and DNA: the additional hydroxyl group in the RNA and replacement of thymine with uracil. The properties of ribose hydroxyl groups outlined above (0.2.1) present a trade-off satisfactory for short-lived RNA: functional versatility at the cost of stability. Such versatility is not required of DNA, as *in vivo* it acts solely as information carrier. However, spontaneous hydrolysis of the DNA strand, frequently only present in one or two copies is clearly of great disadvantage. This trade-off between the two materials is frequently considered when one is chosen for artificial systems.

The less relevant difference we are compelled to mention is that the methyl group is not present in uracil of RNA. Such modification is thought to impose additional energetic cost in living organisms. However, the methyl group in the thymine of DNA is essential: cytosine can spontaneously undergo oxidative deamination to form uracil<sup>49</sup> (**Figure 12**). If uracil were naturally present in DNA, such mutation would go unnoticed. The genetic sequence though would gradually change, replacing a CG pairs with UG, hence UA and finally TA, resulting in a transition mutation. As uracil should not be present in DNA, it is excised by DNA repair machinery and replaced by cytosine, using the template strand as a guide.



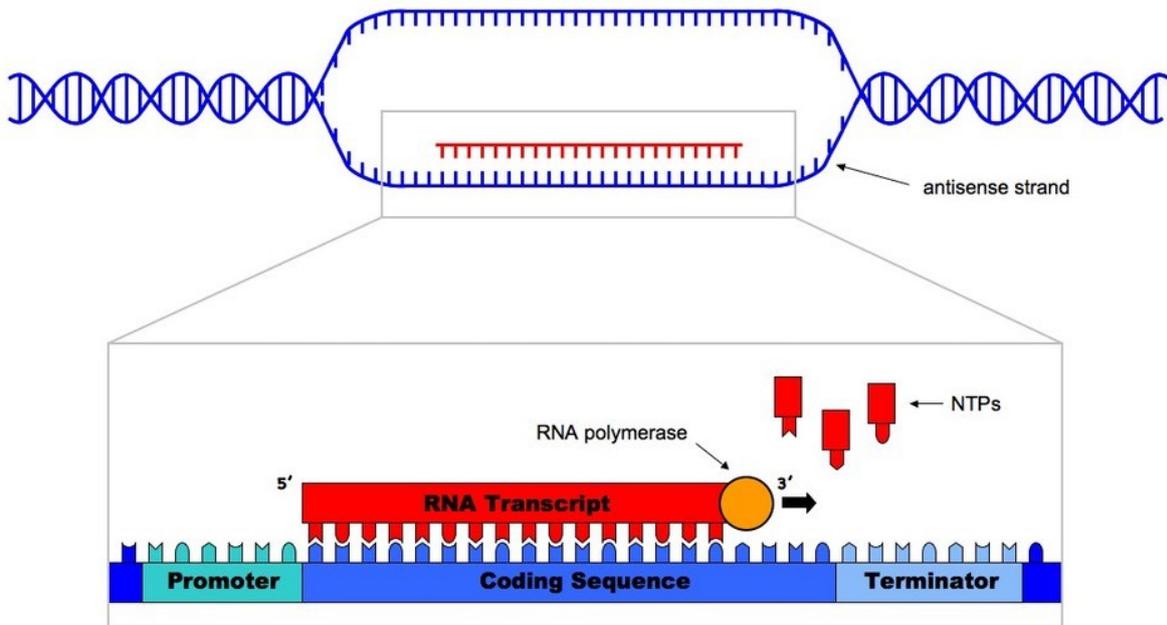
**Figure 12 Oxidative deamination of uracil.** Briefly, the water molecule carries out nucleophilic attack on C4. Subsequently ammonia molecule leaves.

## 0.2.4 Proteins

DNA is transcribed into RNA as an intermediate in protein synthesis pathway. The process is similar to templated ssDNA generation: RNA polymerase arrives at an accessible transcription initiation site known as promoter and starts adding ribonucleotides to a short RNA primer or the first nucleotide recruited to the template<sup>50</sup>. RNA polymerase proceeds until it encounters a terminator sequence at which point it dissociates, terminating transcription<sup>51</sup> (**Figure 13**).

RNA produced in cells can carry out a multitude of functions, the molecule of interest to us here is mRNA mentioned before. As the name suggests, it encodes a sequence of

polymerised amino acids forming a protein<sup>52</sup>. The term “translation” is used to describe conversion of the signal from NA-based code to amino acid one. 20 standard amino acids are usually used (see **Figure 17** below), requiring at least 3 nucleotides ( $4^3 = 64$ ) to uniquely code for each one. Nucleotide triplets are also known as codons and represent a specific amino acid (with some redundancy). 3 codons – the “stop” codons – also represent the end of translation<sup>53,54</sup> (**Figure 14**).



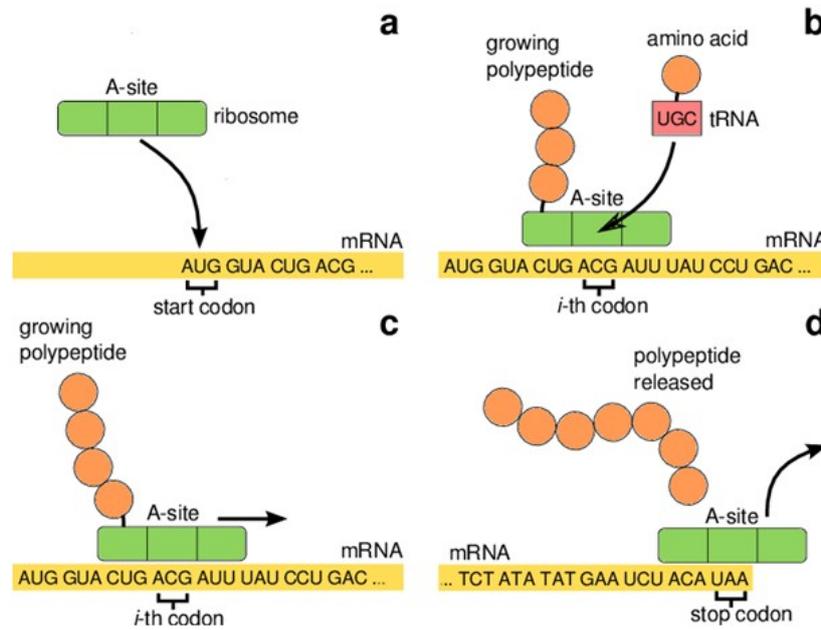
**Figure 13 Schematic of transcription.** From<sup>55</sup>. Unwound DNA section is zoomed in on. Elongation stage of transcription is shown: RNA polymerase is proceeding along the DNA coding sequence adding appropriate nucleotide triphosphates based on the DNA template. Coding sequence is preceded by the promoter region, used for RNA polymerase recruitment and followed by the terminator region, that will cause RNA polymerase unbinding and hence end of transcription.

		1st base								
		U		C		A		G		
2nd base	U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
		UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine	C
		UUA	Leucine	UCA	Serine	UAA	Stop	UGA	Stop	A
		UUG	Leucine	UCG	Serine	UAG	Stop	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U	
	CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine	C	
	CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine	A	
	CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine	G	
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U	
	AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine	C	
	AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine	A	
	AUG	Methionine (Start)	ACG	Threonine	AAG	Lysine	AGG	Arginine	G	
G	GUU	Valine	GCU	Alanine	GAU	Aspartic Acid	GGU	Glycine	U	
	GUC	Valine	GCC	Alanine	GAC	Aspartic Acid	GGC	Glycine	C	
	GUA	Valine	GCA	Alanine	GAA	Glutamic Acid	GGA	Glycine	A	
	GUG	Valine	GCG	Alanine	GAG	Glutamic Acid	GGG	Glycine	G	

Nonpolar, aliphatic  
 Polar, uncharged  
 Aromatic  
 Positively charged  
 Negatively charged

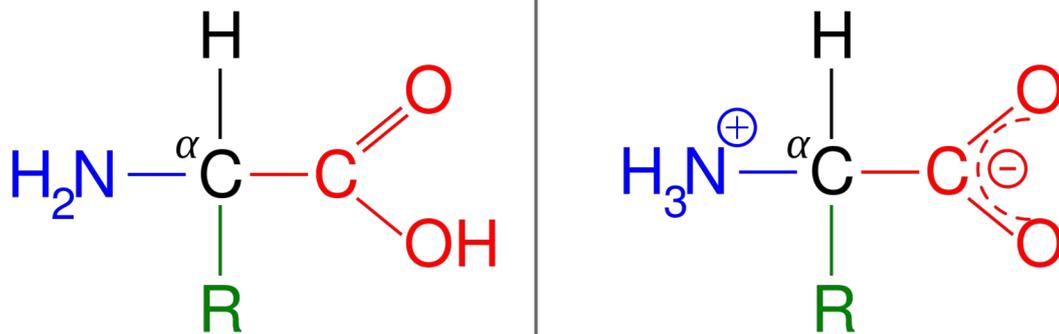
**Figure 14 Codon table showing RNA triplet correspondence to amino acids.** From<sup>56</sup>. 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> indices are shown for ease of interpretation. Amino acids are coloured according to their chemical properties.

In bacteria, similar to RNA synthesis, the protein synthesis is initiated, when a ribosome, a macromolecular machine, assembles over the Shine-Dalgarno sequence<sup>57</sup>. This sequence is also known as the ribosomal binding site (RBS) and is adjacent to the start codon, coincidentally coding for methionine. The amino acids are delivered to the ribosome on transfer RNA (tRNA)<sup>58</sup>. Hence, the ribosome recruits methionine-loaded tRNA and then proceeds along the RNA in 5' to 3' direction in steps of three nucleotides and polymerises appropriate amino acids into a chain by aligning appropriate tRNAs. Once the stop codon is reached, the complex dissociates and the amino acid chain is released<sup>59</sup> (**Figure 15**). Frequently, alternative start codons exist still coding methionine at the start position and it is actually *N*-formylmethionine that is delivered at the starting position for prokaryotes. Likewise, other alternative codon usages are possible that are beyond the scope of this introduction<sup>60</sup>.



**Figure 15 Schematic of translation adapted from.** From <sup>61</sup>. (a) Translation initiation at the methionine start codon. (b-c) Elongation through amino acid addition to the growing chain in accordance to mRNA. (d) Translation termination at the stop codon.

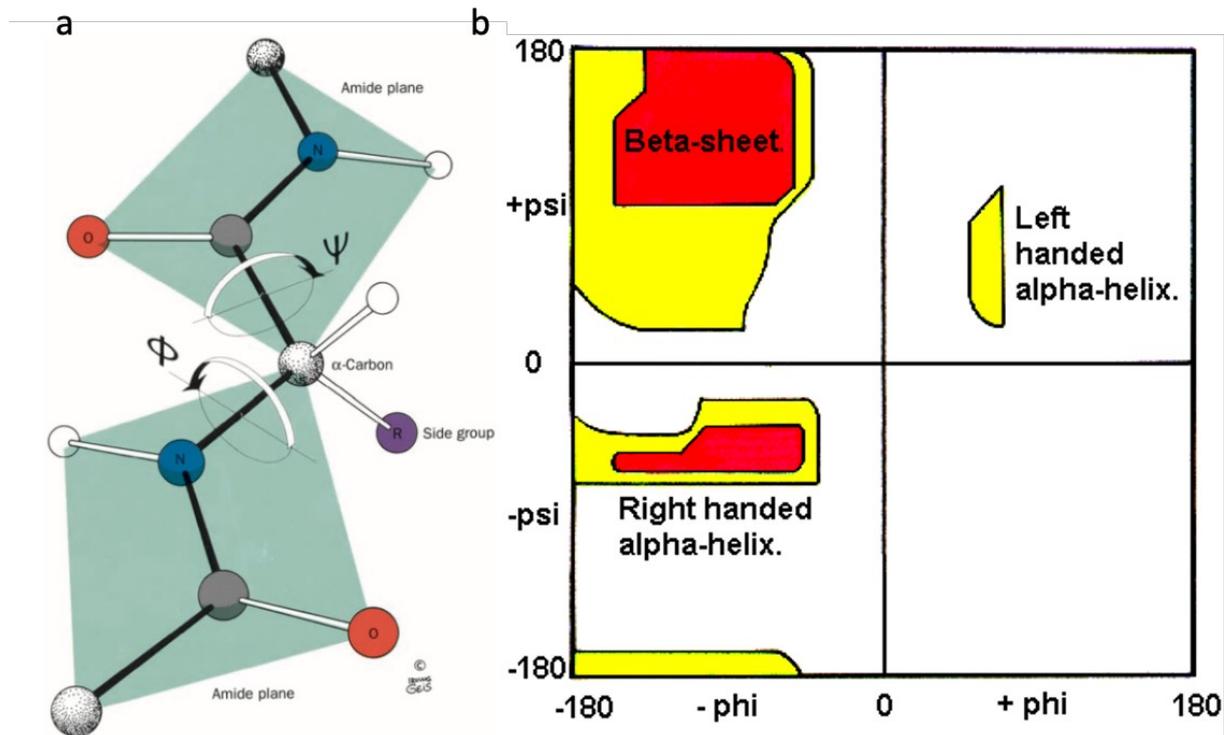
The amino acid chain (sequence) represents a primary structure of a protein. The name – amino acid – arises from a common amine and carboxyl groups joined by the alpha-carbon. As it also carries a hydrogen (apart from proline), and a functional group the molecule is chiral (Figure 16). Typically, only L-isomers of the amino acids are present in biological settings.



**Figure 16 General amino acid structure.** From<sup>62</sup>. On the left, in molecular, on the right in zwitterionic form. R – denotes the variable functional group.

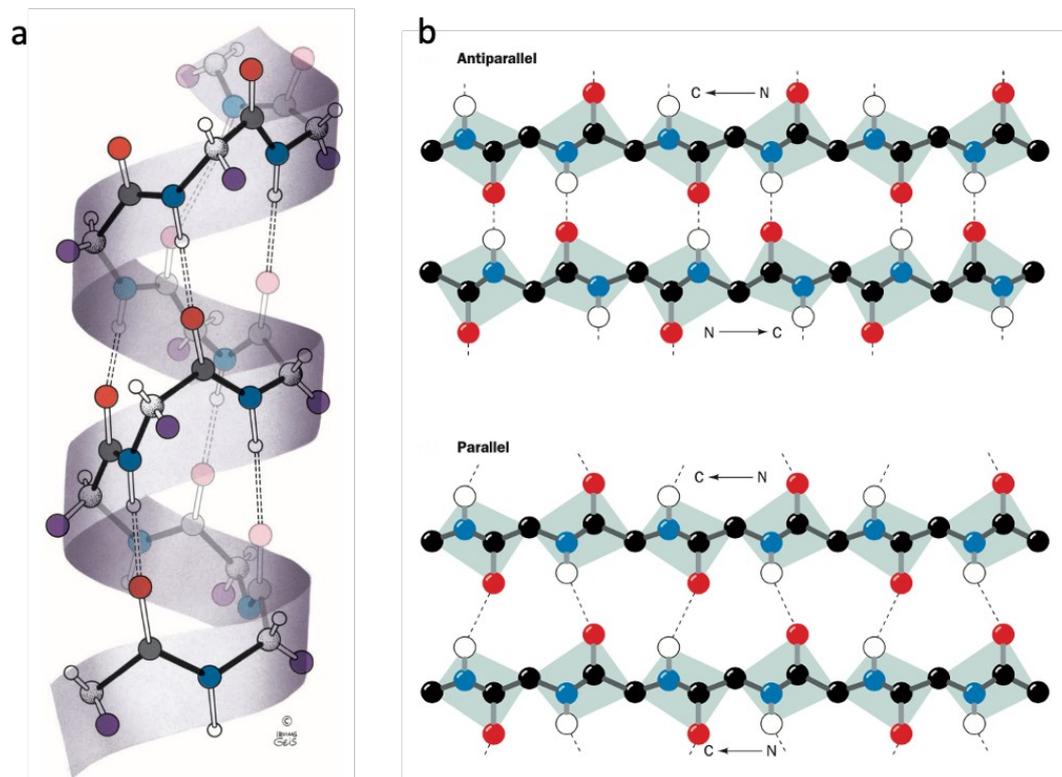
Amino acids can be grouped due to various properties and generally vary in size, charge and hydrophobicity, among others (Figure 17).



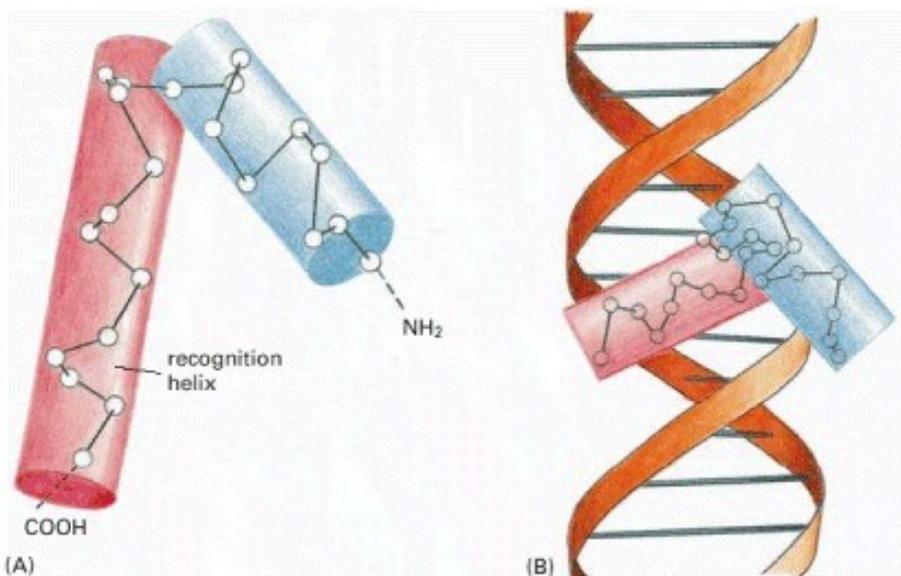


**Figure 19** Polypeptide chain is constrained in 3D space. (a)  $\phi$  (phi) and  $\psi$  (psi) angles are shown between two different planes<sup>66</sup>. (b) Ramachandran plot showing angles favouring certain configurations. White regions are disallowed, stricter criteria are applied for allowed red regions<sup>67</sup>.

The potential for hydrogen-bond formation between the amino hydrogen and the hydroxyl oxygen give rise to another constraint. Together these lead to frequent formation of certain 3D motifs: alpha-helices and beta-pleated sheets among others (**Figure 20**). These form spontaneously and represent secondary structure. Such structures can already give rise to functions: for example, alpha-helices with a certain arrangement of positively charged residues on the outside can facilitate DNA-binding (**Figure 21**).



**Figure 20 Common protein secondary structures.** From <sup>66</sup>. Atoms of the polypeptide are coloured according to the convention: carbon in black, nitrogen in blue, oxygen in red and hydrogen in white. Hydrogen bonds are shown in dotted lines. (a) Alpha helix. An amino acid residue is hydrogen-bonded to one, four positions away. (b) Beta-pleated sheets in antiparallel (top) and parallel (bottom) arrangement. Note the regular staggering of the backbone forming a “pleat”.

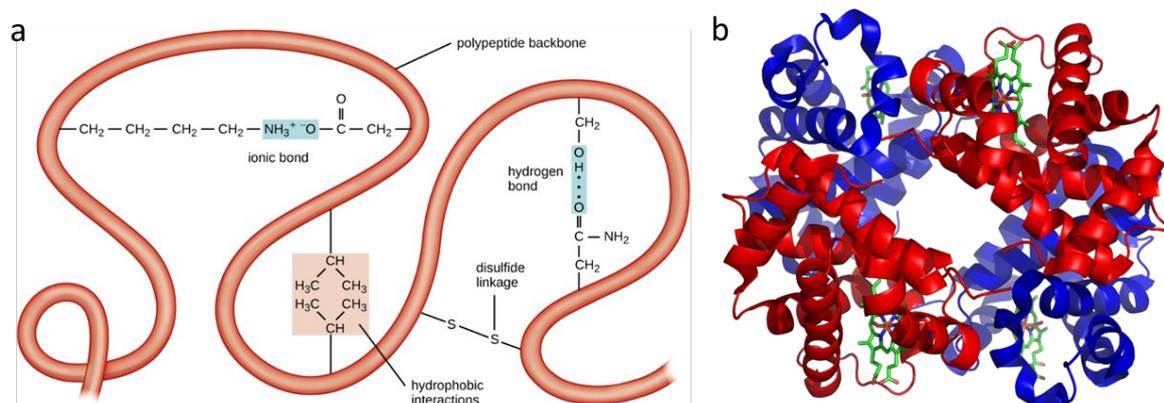


**Figure 21 Helix (A)**

**(B)**

DNA. Based off.

The motifs are constrained further by individual properties and interactions of the amino acids: a hydrophobic core can form, as well as charge-charge interactions, disulphide bonds and Van der Waal’s interactions (**Figure 22a**). This gives rise to tertiary structure. In some cases, several folded chains can come together to form a multimeric protein, possessing a quaternary structure – a specific arrangement of the subunits (**Figure 22b**). Tertiary and quaternary structure gives rise to the overall fold and functions.



**Figure 22 Tertiary and quaternary protein structure examples.** (a) Polypeptide is shown, constrained by ionic bond, hydrophobic effect, disulphide bridge and hydrogen bonds<sup>69</sup>. (b) Haemoglobin heterotetramer is shown consisting of 2  $\alpha$  and 2  $\beta$  chains, held together by non-covalent bonds<sup>70</sup>.

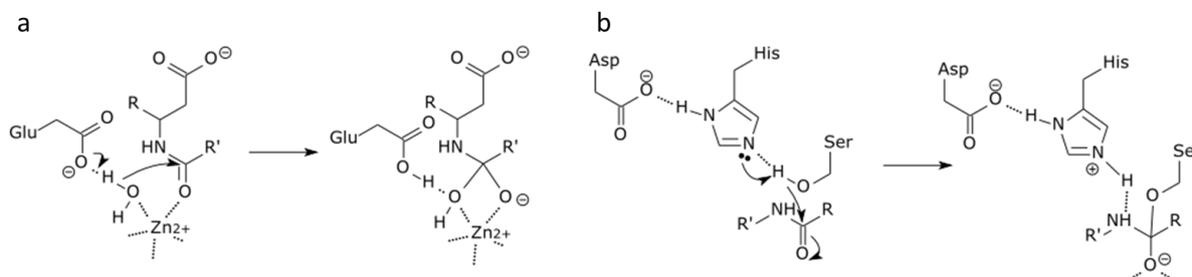
## 0.2.5 Enzymes

The combination of individual amino-acid properties, their arrangement and overall shape give rise to protein properties. Some are more trivial, such as mass, charge,  $\text{pK}_a$  and isoelectric point, calculated through aggregating properties of individual amino acids. Some are less so: similar to mentioned before, a patch of same-charge amino acids on protein surface can favour charge-based binding to particular molecules. More crucially for us, a particular spatial composition of amino acids, typically in a pocket-like arrangement can facilitate specific catalysis – significant increase in particular chemical reaction rate without being consumed. Proteins capable of catalysis are called enzymes. In most cases, in absence of an enzyme the rate of a given biological reaction is slow. Therefore the presence or absence of the enzyme can be synonymised with the reaction taking place or not (assuming no other constraints). This, in turn defines the metabolic processes of an organism: the collection of life-sustaining chemical reactions taking place. Metabolism can be broken down into metabolic pathways: a series of chemical reactions carrying out a specific biotransformation, with enzymes catalysing individual steps.

Enzymes are therefore essential for life but can also be exploited as catalysts for arbitrary reactions desired by the user. Unlike some inorganic catalysts, based on precious and/or rare metals, proteins synthesis relies on substances less scarce. Later on we will also demonstrate examples, where enzymes are a more attractive option due to stereospecificity or natural predisposition to catalyse a specific reaction that is biomedically relevant.

Enzymes typically accept a specific substrate and convert it into a specific product. This is facilitated by the active site. Like with conventional catalysts, the activation energy is lowered in variety of ways, such as stabilising the transition state (through creating suitable local environment); providing a reaction pathway with lower activation energies (through transient reaction with the substrate); or destabilising the ground state of the substrate (through distorting the bound substrate or orienting substrates into a productive arrangement, thus increasing their effective concentration) (**Figure 23**). Enzymes frequently

utilize co-factors – small chemical compounds or metallic ions that are also not consumed during catalysis but immediately facilitate it.



**Figure 23 Enzyme mode of action examples.** As in <sup>71,72</sup>. (a) The tetrahedral intermediate (left) is stabilised by the ionic bond of the Zn<sup>2+</sup> and the negatively charged oxygen in a carboxypeptidase reaction step. (b) The enzyme transiently reacts with the peptide bond through histidine accepting serine proton, preparing it for nucleophilic attack on the bond in a serine protease reaction step.

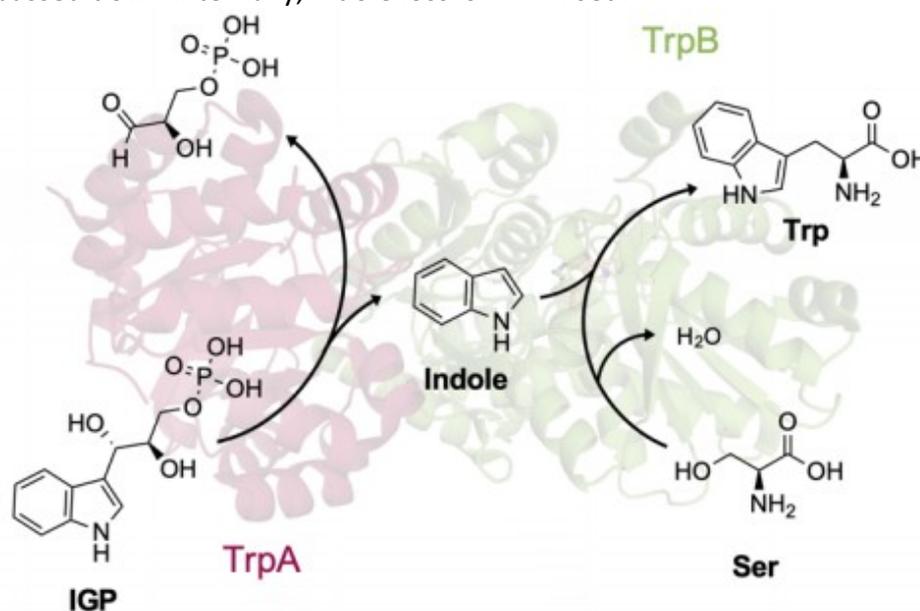
The specificity is insured through particular charge, shape and hydrophilicity of the active site, only allowing the complementary substrate to enter and bind. Sometimes, a group of similar molecules fits the active site and can undergo catalysis. In this case, the enzyme is considered promiscuous.

Although only a limited number of amino acids immediately participates in the reaction (around 3-4), the rest of the structure (reaching thousands of residues in size) is essential for maintaining the overall orientation of the said amino acids. The enzymes are not rigid but rather dynamic. This is reflected in the “Induced fit” model. As the name suggests, the enzyme, and sometimes the substrate, alter their shape for a more specific fit also essential for catalysis. Although protein primary structure can usually be deduced from the genetic code, the complexity of three-dimensional structure of an enzyme, coupled with dynamical nature of catalysis frequently makes function prediction difficult. We will discuss this problem below.

The dynamic nature of proteins in general means they can possess allosteric sites. These can be affected by small-molecule or other protein binding to induce overall conformational change. The change in conformation in turn affects function, typically upregulating or downregulating it. As tertiary and quaternary structures are stabilised by weak interactions (in addition to disulphide bridges), even slight changes in temperature and pH can affect stability of the conformation, disrupting function. Complete loss of fold is termed denaturation and can be reversible but most often not. This can be, for example, due to initial folding happening in particular order during initial protein synthesis, which is not replicated during whole chain refolding.

### 0.2.5.1 Tryptophan synthase

An exemplar enzyme used in this work is Tryptophan synthase (TrpS), in particular, its subunit B (TrpB). We selected it as it has been a subject of many investigations over the molecular biology age and therefore seemed to be a convenient target. TrpS is found in all domains of life and as the name would suggest catalyses L-tryptophan (L-trp) production with pyridoxal 5'-phosphate (PLP) as a co-factor. TrpS forms a heterodimeric complex with 2  $\alpha$  subunits catalysing indole release from indole glycerol phosphate (IGP), which then diffuses down an internal tunnel to a  $\beta$  domain, where indole is added to L-serine (L-ser) to form L-trp (**Figure 24**). The process is tightly orchestrated between the subunits through allostery: once IGP is bound at the  $\alpha$  subunit, formation of L-ser based PLP-bound amino-acrylate intermediate is favoured at the  $\beta$  subunit. This, in turn stimulates indole release at the TrpA. It is thought that such tight coordination was evolved to preserve indole, that easily diffuses away from the enzymes and out of a cell across the membrane; if it is produced at the last moment and passed down internally, indole loss is minimised.



**Figure 24 TrpS native catalysis.** As shown *iff.* TrpA (left) produces indole from indole glycerol phosphate (IGP), which diffuses to TrpB (right), where it is added to L-serine with water being released.

### 0.2.6 Response to the environment through regulation of active protein quantities.

In this work we built a molecular reaction network capable of responding to its environmental cues through pre-programmed responses relying on protein activity. This was inspired by the living systems. We will therefore discuss here some examples of natural protein activity regulation. We feel we should present all of the major strategies used in the biological systems to provide a reasonable overview and evidence why we drew inspiration only from some of them. They will therefore be presented in the order of increasing relevance to our work.

Most of the DNA is not continuously (constitutively) expressed. That is, only a small fraction of the information is transcribed and translated at any given time. This fraction varies

to respond adequately to the surrounding environment. For example, as growth conditions change to those more abundant in nutrients it may be beneficial to express more proteins for nutrient import and growth. The opposite would be true if nutrients are scarce. In finer detail, the nutrient chemistry can also change, requiring different metabolic pathways for successful assimilation into central metabolism. Therefore, biological systems have to balance the cost of DNA expression with the associated benefits through expression regulation.

Considering the central dogma of molecular biology, we can easily infer that all of the steps of the information flow from DNA to protein we have outlined above can be used as a point of control. We would also remark that the closer the method of control is to the final protein, the faster the effect on the active protein concentration will be, yet more of the prior steps would happen “in vain”. This represents a balance of cost against response rate. Below we describe some common approaches. We start with a broad category of the costliest response with protein already present (post-translational regulation), followed by response through protein production control that can happen at post-translational and finally, transcriptional levels.

#### 0.2.6.1 Response with protein already present

Protein modification, transport and degradation allow the fastest active protein concentration adjustment within a cell. However, such mode of action is costly: a protein is first produced for its local activity over lifetime to then be decreased from maximum possible, thus decreasing efficiency. Degradation, in particular, incurs the highest energy cost, because, although amino acids can be rationally re-used, the protein has to be resynthesized.

Post-translational modifications (PTMs), as the name suggests, encompass modifications occurring to proteins after they are synthesised. This implies modification to the existing functional groups or the termini of the protein. For example, phosphorylation – the addition of a phosphate group – is the most common PTM according to Floudas and colleagues<sup>74</sup>. Phosphorylation typically causes structure alteration leading to activation, deactivation or function modification. This is done through an action of a protein kinase and can be reverted by protein phosphatases via dephosphorylation. For example, glycogen phosphorylase *b* (GPb) is activated to glycogen phosphorylase *a* (GPa) isoform through phosphorylation. In brief, GPb is typically inactive, when ATP and glucose 6-phosphate are present and is activated by Phosphorylase kinase (PhK) at Ser14 to increase its activity related to glucose 6-phosphate production<sup>75</sup>.

Protein transport can concern displacement within a cell, for example, towards a particular organelle, responsible for a specific environment or a set of reactions. This allows maximising local concentration in a sub-compartment. Alternatively, in case of secretion, proteins can be rapidly ejected in large quantities at once, when required<sup>76</sup>. An example of such process involves pathogenic behaviour exhibited by bacteria. For instance, expression of type 6 secretion system (T6SS) employed by Proteobacteria is upregulated upon contact with the host cell. However, this part of the response is slow. Therefore, in case of *Pseudomonas aeruginosa*, a series of phosphorylation events by serine–threonine kinase (PpkA) allow to rapidly activate hemolysin-coregulated protein (Hcp1) protein secretion<sup>77</sup>. Hcp1 is implicated in both T6SS assembly and chaperoning its pathogenic substrates<sup>78</sup>, that include Tse1-3 proteins. Tse1 and Tse3, for instance are lytic enzymes<sup>79</sup>.

Regarding protein degradation, the process, known as proteolysis, is facilitated by the proteases. Two means exist: proteolysis within a lysosome, which is spatially differentiated or ubiquitin-dependent targeting towards a proteasome (a protease complex). Lysosomes are animal cell organelles (known as vacuoles in other cells), bound by a membrane and containing a range of hydrolytic enzymes at pH of  $\sim 4.5-5.0$ . During the process of autophagy, cellular components are directed to an available lysosome for degradation and recycling to form autolysosome. In brief, this can concern: large cellular components delivered in autophagosome (macroautophagy); random direct engulfment of cargo (microautophagy); or specific proteins, directed by chaperones (chaperone-mediated autophagy). Alternatively, select proteins can be tagged with ubiquitin – a 8.6 kDa protein, which can direct them for degradation by 26S proteasome. Ubiquitination results in degradation if linkage is done through particular ubiquitin residues (K48 and K29), sometimes concatenated. The process of delivery to the proteasome is not fully understood<sup>80</sup>. Lon protease is one of 4 ATP-dependent proteases prominent in *E. coli*, responsible for protein turnover<sup>81</sup>. As for eukaryotes, precise regulation is still under investigation although it is clear that efficient amino acid resource reallocation can be critical for survival.

We did not discuss regulation by allostery in this section, as we will talk about it later.

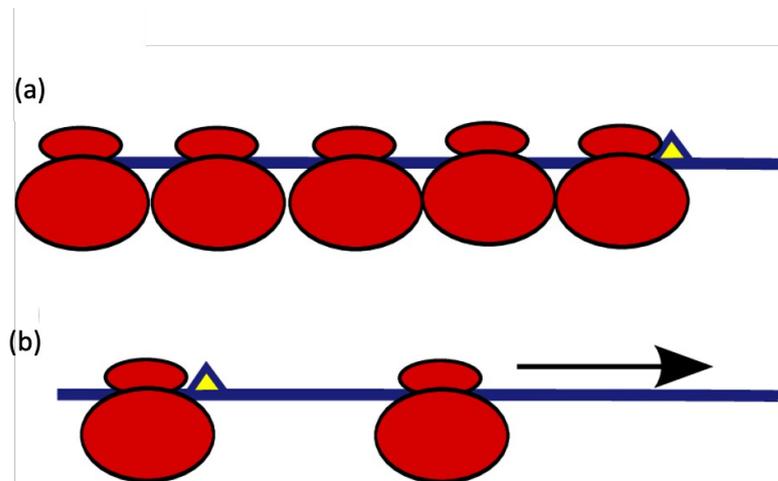
### 0.2.6.2 Response through control of protein production

In contrast to the approach above, producing proteins only when necessary can strongly decrease the energetical costs associated. As a trade-off, rapid response becomes problematic as the protein has to be produced first. Post-transcriptional regulation concerns the rate at which the proteins are produced based on mRNA templates, already present. In turn, transcriptional regulation we will describe last determines the rate, at which mRNA itself is produced.

#### 0.2.6.2.1 Post-transcriptional regulation

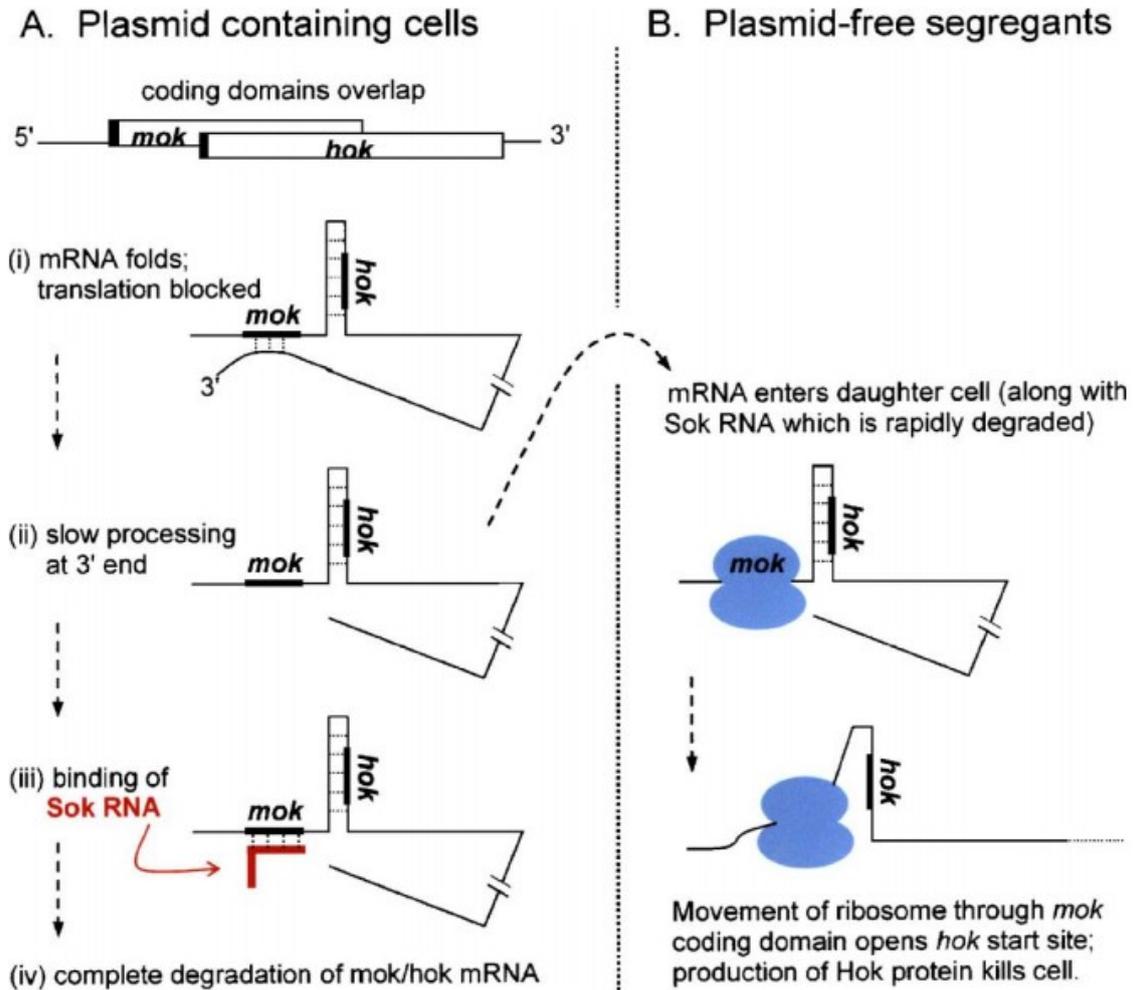
Considering post-transcriptional regulation, if mRNA is already produced, translation of the coded proteins can be regulated. Similar to the protein level, mRNA can be degraded either through pre-encoded sequences defining mRNA stability or specific targeting, for example, through RNA interference (RNAi) via micro- or small interfering RNA (miRNA or siRNA). However, a more energy-preserving approach is possible through control of the amount of protein translated from a given mRNA. Overall, this is affected by the frequency of translation complex formation, in part determined by the RBS sequence and its context. Computational tools exist to predict this based on comparing Gibb's free energy of folded mRNA and free ribosomal subunit against the assembled complex<sup>82</sup>.

Progression of the ribosome along the mRNA, and hence, translation rate, is also affected by the abundance of loaded tRNA matching particular codons. If an amino acid is encoded by a codon, complementary to a rare tRNA, delivery of such amino acid would be slow despite potential abundance. The ribosome will stall until the loaded tRNA arrives and block upstream ribosomes. This can lead to ribosome sequestration and is usually avoided by natural or artificial codon optimisation: codon usage is optimised to match tRNA occurrence in the cell in general and “slower” codons can be found at the beginning of the transcript to optimise ribosomal traffic (**Figure 25**).



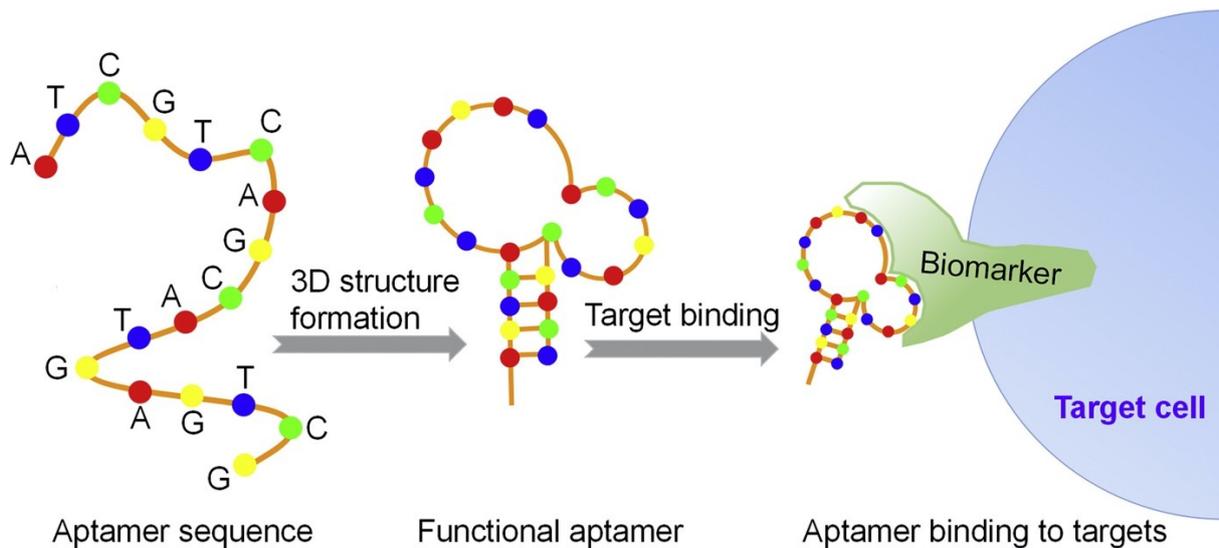
**Figure 25 Ribosomal traffic illustration.** As in<sup>83</sup>. Slow codons illustrated by a triangle. (a) Slow codons towards the end of an mRNA result in ribosomal jamming. (b) Slow codons at the beginning of mRNA prevent excess ribosome binding and increase efficiency.

More dynamic processes allow for translation to take place or not, depending on specific inputs. This is done through particular folds at the 5' end of the mRNA. These can occur intramolecularly for regulation in *cis*. Alternatively, additional RNA strands and proteins can be involved. An example of the latter is a frequently used *hok/sok* system where host-killing (*Hok*) protein expression is controlled by *Sok* antisense RNA. In short, a hairpin structure occludes *hok* RBS. This can be unfolded if overlapping upstream *mok* gene is expressed. However, by default *Sok* RNA bound to *mok* prevents such an event. Once *Sok* RNA is no longer expressed, *mok* gene is, leading to *hok* also being expressed (**Figure 26**)<sup>84</sup>. This process allows to enforce plasmid retention in a cell culture: if the plasmid is lost, *Hok* is expressed and kills the rogue cell, avoiding non-plasmid-containing culture to arise.



**Figure 26** Hok/sok system as an example of mRNA regulation in trans as in <sup>84</sup>. Here, cells constitutively express the mRNA from a plasmid. (A) Construct containing overlapping *mok* and *hok* coding domains is transcribed. (i) mRNA fold prevents *mok* and *hok* translation through obstructing the RBS. (ii) mRNA is naturally degraded from the 3'-end making *mok* gene available for translation initiation. (iii) antisense Sok RNA coded by the same plasmid binds *mok* region, making it unavailable again and leaving RNA to be degraded over time. (B) If the plasmid is lost, Sok RNA and *mok-hok* mRNA synthesis ceases. Sok RNA is degraded faster, leaving *mok* available as in Aii. When *mok* translation is initiated, *hok* also becomes available due to hairpin disruption and is expressed.

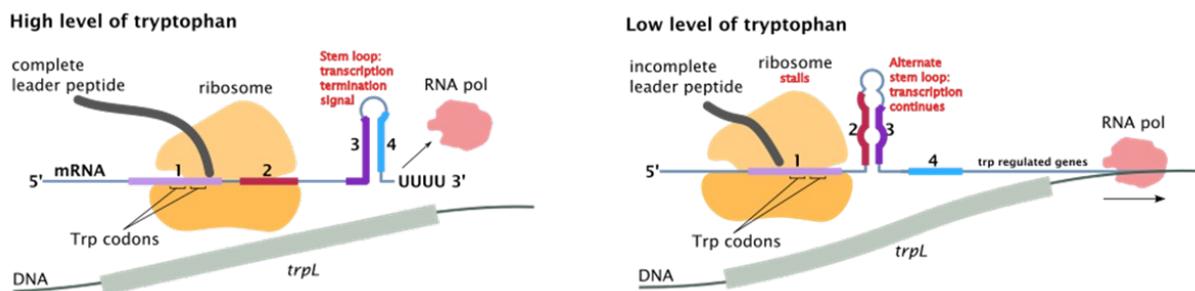
Regulation in *cis* relies solely on intramolecular conformational changes involving structures called riboswitches. Formation of such structures can affect the way RNA is processed (spliced) in fungi and plants, block RBS (similar to above) or result in premature transcriptional termination (discussed below). Riboswitches are a case of aptamers – oligonucleotides or peptides<sup>85</sup> that bind a specific molecule. Often, this induces a conformational change. Usually made of RNA they are frequently used due to availability of custom selection methods catering for a broad range of targets as well as biocompatibility (**Figure 27**)<sup>86</sup>. Specific affinity, coupled with simple tagging allows for applications in bioimaging, diagnostics and treatments, whilst potential conformational changes allow for sensing and controlled substance delivery.



**Figure 27 Aptamer function in a biomedical setting schematic.** As in <sup>87</sup>. RNA sequence folds and is then able to bind with high affinity and specificity to its target, which is a cell-surface biomarker.

Premature transcriptional termination is a way of dynamically controlling mRNA synthesis. Transcription is constitutively initiated, however, transcriptional terminator structure forms in the 5' region of RNA resulting in RNA polymerase dissociation, in turn leading to a short non-functional fragment that is degraded. This process can be rescued if terminator formation is prevented either through riboswitch action or by that of translational machinery. In the former case, riboswitch tends to bind a part of the nascent transcriptional terminator.

The latter case is exemplified by a process known as transcription attenuation. This is specific to bacteria, where translation occurs concurrently with transcription. One mean of controlling transcription of L-trp synthesis-related genes is transcription attenuation. The transcript starts with a specific leader region (*trpL*) containing an attenuator towards its end. We shall describe the leader sequence as containing regions 1, 2, 3 and 4. Regions 1 and 2 are in immediate proximity as are 3 and 4. Regions 3 and 4 form the aforementioned transcriptional terminator, however if region 2 is available, it preferentially binds region 3, preventing terminator formation and hence, RNA polymerase dissociation. The switching is determined by ribosomal progression rate along region 1, coding the leader peptide containing two trp codons. If L-trp is abundant in the cell, so will be trp-loaded tRNA and ribosome will quickly reach the end of region 1, affecting ability of region 2 to interact with region 3. Region 3 will bind region 4 eventually leading to trp synthesis related genes suppression. If L-trp is not abundant, the ribosome will stall, giving region 2 time to rescue further transcription through binding region 3 (**Figure 28**). Both in case of L-trp and in general, transcription can also be controlled by DNA-binding proteins called transcription factors.



**Figure 28** Transcriptional attenuation of the *trp* operon. Adapted from<sup>88</sup>. Left, tryptophan level is high, allowing the ribosome to proceed rapidly through sequence 1 and onto sequence 2, preventing formation of the 2-3 sequence complex. 3-4 sequence is formed resulting in transcription termination. Right, tryptophan level is low. Ribosome stalls, allowing 2-3 complex to form, preventing 3-4 complex formation and allowing RNA polymerase to proceed with full-open reading frame transcription.

#### 0.2.6.2.2 Transcriptional regulation and transcription factors

This subsection concerns transcription factors, the novel element we incorporated into PEN DNA toolbox. They serve as an interface between small molecules and the molecular programs. Hence, we provide below a brief introduction to the concept as well as some details about the particular transcription factors we have focused on: LacI and TrpR.

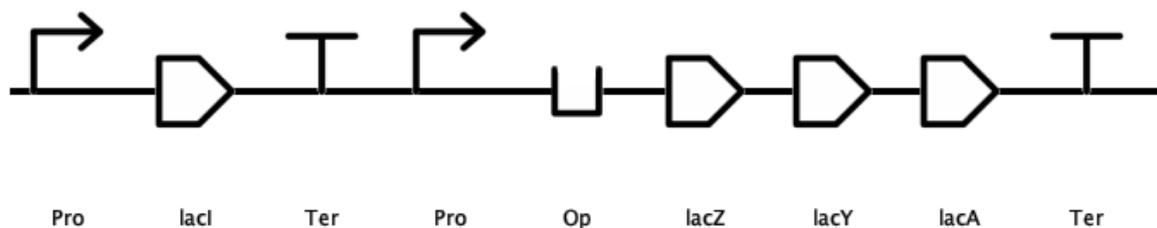
Transcription factors (TFs) are protein molecules, capable of modulating gene expression. They do so through DNA binding resulting in transcription control. The mechanisms vary in complexity but can be broadly categorised into: immediately affecting RNA polymerase ability to bind DNA (either improving or impeding it); indirectly affecting RNA polymerase through recruiting coactivator or corepressor proteins; affecting DNA availability through eukaryotic chromosome arrangement modifications. The last phenomenon concerns eukaryotic chromosomes, where DNA is complexed with histone proteins in a “beads-on-a-string” fashion. Direct or indirect modification of such beads affects DNA packing and its availability for transcription<sup>89</sup>. Complex 3-dimensional rearrangements can be involved, presenting some challenges for studies, which we shall not delve into.

Elaborate coactivator and corepressor complexes are also characteristic of eukaryotic cells<sup>90</sup>. These allow complex computations, that at the end affect TF ability to bind DNA and recruit RNA polymerase. In bacteria, a similar mode of action is usually simplified to use of allosteric transcription factors (aTFs) – these immediately bind DNA to recruit/obstruct RNA polymerase and, at the same time, sense specific-molecule binding. Such activator/repressor directly changes aTF conformation, affects its ability to bind DNA and hence, has effect on transcription. Within this work, we only used aTFs and therefore would occasionally refer to them simply as TFs.

A large selection of aTFs has been reported and is used in biotechnological applications<sup>91</sup>. We will focus on simple examples important for this work: classical LacI and TrpR<sup>92</sup> proteins. These form dimers to strongly bind a defined DNA sequence – an operator site. Interestingly, the term “allosteric inhibition” itself was first used in the context of early *trp*-related studies<sup>93</sup>.

LacI – the *lac* repressor protein is probably the most well-known regulatory protein<sup>94–97</sup> first studied in *Escherichia coli*. Part of the exemplar *lac* operon (**Figure 29**), it contributes to lactose metabolism regulation of bacteria. Natively, it represses expression of lactose-

metabolising proteins until lactose appears in the cell. In such case, trace amounts of beta-galactosidase (beta-gal) mostly hydrolyse lactose into glucose and galactose. However, during degalactosylating catalytic step, if water molecule does not happen to act as an acceptor, D-glucose does so, resulting in transgalactosylation and hence, allolactose is formed<sup>98</sup>. Allolactose, serving as a proxy of lactose, can then bind Lacl to drastically decrease its affinity for the operator site. Unobstructed, the operator site then becomes available for transcription of lactose-metabolising genes to occur.

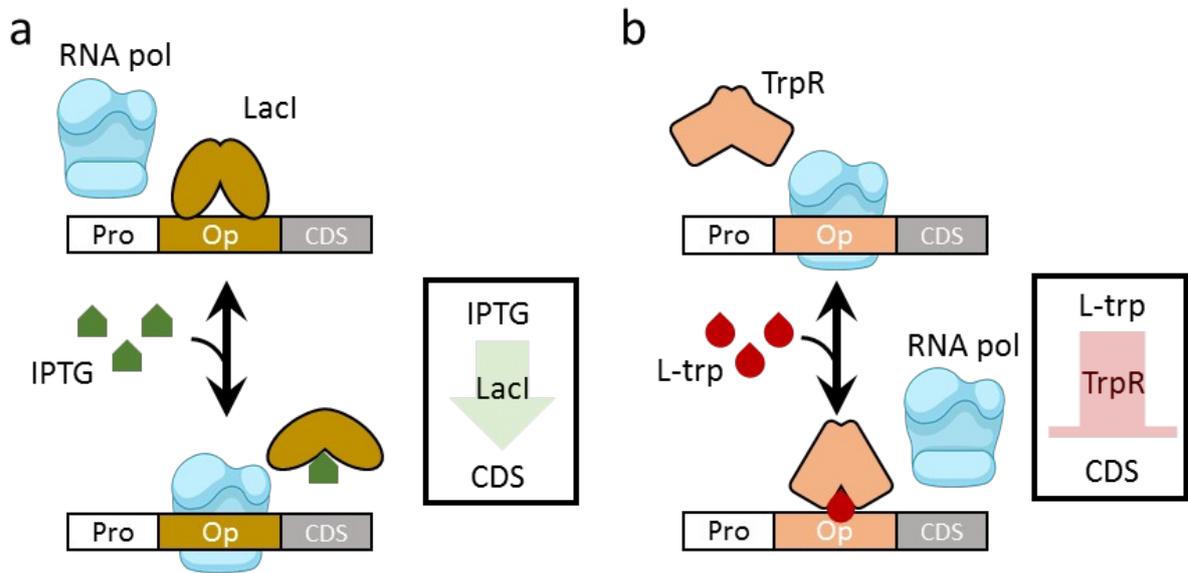


**Figure 29 Lac operon map.** Annotations according to SBOLVisual 1.0.0 standard glyphs<sup>99</sup>. The operon contains first the lacI open reading frame (ORF), followed by a polycistronic transcript of lacZ, lacY and lacA genes, responsible for lactose breakdown, import and beta-D-galactoside acetylation.

Most studies, concerning Lacl made use of Isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) – an allolactose analogue, that is not broken down by beta-gal yet binds Lacl to induce conformational changes<sup>100</sup>. We will not dwell much on intricate detail but would like to briefly mention that several variants of the operator sequence exist with varying affinity for Lacl. These are hypothesised to enable formation of complex structures: the dimer of dimers looping DNA, to make the promoter region inaccessible<sup>101</sup>. This is also hypothesised to enable further cooperativity – a situation, where additional IPTG leads to multiplicative, rather than linear response explained through Lacl monomers responding in a coordinated fashion, rather than independently.

Tryptophan repressor (TrpR) is another transcription factor found in *E. coli* and responsible for core metabolism. The dimer protein regulates tryptophan (L-trp) synthesis through binding upstream of the attenuator mentioned above. As L-trp is usually abundant in the cell, TrpR switches off expression of the operon upon L-trp binding. When L-trp is depleted and is no longer available to bind TrpR, TrpR changes conformation and becomes unable to bind the operator sequence. Therefore, expression of L-trp synthesis genes is resumed and L-trp production ensues. Like with Lacl, extensive studies were carried out to discover both TrpR and operators site variants<sup>102</sup>. We will forego most details but will mention that to the extent of our knowledge TrpR does not form a tetramer structure, unlike Lacl. A thermophilic homologue TrpY, has been reported exhibiting such behaviour<sup>103–106</sup>.

Together, TrpR and Lacl represent transcription factors of opposite logic of action. That is, in one case, the ligand leads to gene expression and in other repression (**Figure 30**). In this work we will use both in the same circuit to demonstrate multiple small molecule input processing. We will also use TrpR to detect L-trp, the product of TrpB we have mentioned before. In particular, we seek to detect TrpB activity and through use of a molecular program to link it to TrpB gene replication.



**Figure 30 Summary of LacI and TrpR modes of action.** dsDNA fragments are shown as rectangular blocks containing promoter (Pro), operator site (Op) and protein coding sequence (CDS). (a) LacI dimer is bound, until IPTG is added. Once this occurs, LacI changes conformation and can no longer bind. This allows RNA polymerase to progress through the operator site and transcribe the CDS. In short, IPTG induces CDS expression. (b) TrpR dimer is unbound, allowing RNA pol to proceed freely. Once L-trp is added and bound to TrpR, its conformation changes, allowing it to bind DNA. This prevents RNA pol from transcribing CDS. In short, L-trp inhibits CDS expression.

## 0.3 Protein engineering

Protein structure can be intentionally altered to affect its functional properties. This can be done to improve stability under different reaction conditions or to alter specificity and/or catalytic rate if applicable. Broadly, two complementary approaches exist (aside from searching for a new candidate protein in nature)<sup>107</sup>. Rational design, as the term would suggest, relies on the detailed understanding of the functional correlation to the structure to introduce changes. Directed evolution, in turn, similar to natural evolution, relies on introduction of variability to the parent gene with successive screening to discover variants with desired traits. Unlike nature, the desired traits are defined by the experimenter.

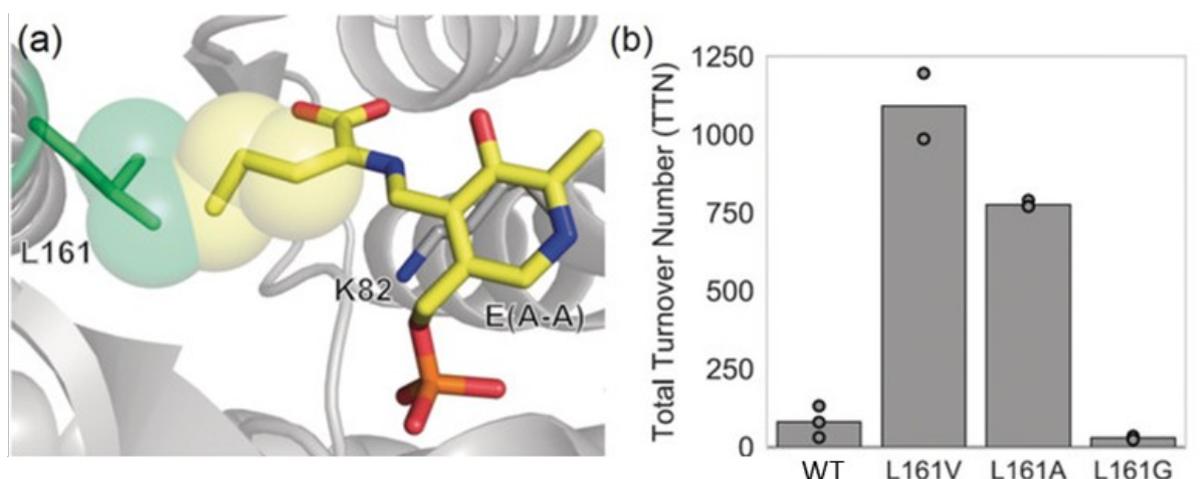
In this work, we aimed to improve an existing directed evolution method – PEN CSR. We will discuss it in Chapter 2, dedicated to it. In particular we sought to make it suitable for small molecule producing enzymes, whilst preserving the general pipeline. However, we will first introduce the rational design methods as they do present an alternative, we would like to inform the reader about. We will then follow with an introduction to general directed evolution concepts and continue the discussion in Chapter 2.

### 0.3.1 Rational design

Rational design approach exploits different insights accumulated about proteins to guide future designs. These can be based on three-dimensional structural information or primary structure, where the former is unavailable. If the protein structure is well-resolved, predominantly via X-ray crystallography, although NMR and cryo-EM have also made great strides, this can be used to speculate about the catalytic mechanisms and current limitations.

In case of TrpB, for example, Christina Boville's group worked on improving a variant to accept  $\beta$ -branched Ser analogues with longer alkyl chains ( $\beta$ -ethyl- and  $\beta$ -propylSer)<sup>108</sup>. A structure was previously obtained and deposited to the Protein Data Bank (PDB) – public biological structures repository (PDB: 5VM5). Via modelling  $\beta$ -EtSer bound to PLP (the co-factor we introduced before) in the active site, they noticed that the conformation of the latter results in a steric clash (**Figure 31a**). The clash was alleviated through substitution of long leucine side chain with valine, alanine and glycine, that possess shorter side-chains. The total turnover number (TTN) increased by 14 and 10-fold for L161V and L161A though decreased 2-fold for L161G (**Figure 31b**). L161A, as the smaller successful side chain variant, was then taken forward for future studies we will mention below.

Such insights are not always successful as an effect of a mutation can be challenging to predict. This is exemplified by the L161G variant, inferiority of which has not been rationalised by the authors of the report. As we mentioned, proteins are rather dynamic and not all isoforms are usually captured in static structures based on immobilised protein imaging. Computational models can account for this but can also be far from perfect as we will discuss. Rational design can be guided by approaches relying solely on the protein sequences or alternatively those attempting to infer the corresponding three-dimensional structures first.



**Figure 31 Rational engineering of TrpB example.** Adapted from Boville et al.<sup>108</sup>. (a) PLP-bound β-ethyl-serine (yellow) clashes with L161 of TrpB (green). Relevant van der Waals radii are shown as spheres (b) β-EtTrp production by the starting variant of TrpB (WT) and those with L161V, L161A, or L161G mutations.

### 0.3.1.1 Sequence-based rational design

If structural information is unavailable, sequence analysis can sometimes aid design. Amino-acids strongly conserved between species usually play an important role and therefore can be targeted for exploration. These are identified through multiple sequence alignment (MSA) methods, with Clustal W, MAFFT, K-align as well as MUSCLE and T-coffee being among those, commonly known. An example of use of such tools includes consensus sequence design approach. The most commonly occurring amino acids at the conserved positions are combined in hope of a beneficial effect. This was exemplified by Boris Steipe and colleagues. 10 consensus-guided mutations were independently made on the murine antibody McPC603, with 6 yielding increased stability, 3 being neutral and 1 leading to destabilisation in line with their initial predictions<sup>109</sup>. Multiple stabilisation works followed aimed at enzymes also concerning thermal stability<sup>110–112</sup>. Resistance to low pH and peptidases as well as an increase in specific activity has also been achieved, although through a combination of methods<sup>113</sup>. An interesting variation involved the separate discovery of desirable catalytic and stereospecific features of existing enzymes. Those were then combined *in silico* and searched for in the existing databases to discover novel enzymes possessing both features<sup>114</sup>.

Moving on from single mutation pooling, a phenomenon of coevolution has been observed. It highlights the fact that the hypothesis of independence and additivity of mutational effect is overly simplistic. Also known as correlated mutation, covariation or co-substitution, it describes a situation where the residues seem to mutate in ensemble. The concept was presented by Thomson as “reciprocal evolutionary change in interacting species” covering the organismal level<sup>115</sup>. When considering “species” of individual amino acids within a protein, reciprocity is still the key: mutation at one locus changes evolutionary pressure at another and so forth. This can be to retain overall structure or function, despite the substitutions – subsequent mutations compensate for the first, allowing complex fitness landscape transitions. These amino acids can be distal in the primary structure but are frequently close in the tertiary or quaternary. Hence, observation of such processes sheds light on structurally important groups of residues throughout the protein. An array of

computation tools exists employing various approaches for the analysis and has been reviewed<sup>116</sup>, whilst the bird's-eye view of the approach is similar to that mentioned above in the context of MSA. We would only note that comparisons of the methods become available from time to time<sup>117</sup> and integrated online suites of tools have been created<sup>118</sup>.

A striking example of the power of such approach is work by Socolish *et al.*<sup>119</sup>. They have managed to define some artificial protein sequences based only on MSA to generate folds that replicate thermodynamic properties of the natural ones. Coupled conservation had to be assumed as site-independent model gave rise to less successful folds, even though in both cases the sequences generated were equally divergent from the natural WW domains. Thus, showing that a statistical energy function capturing coevolution is essential and sufficient to replicate native structures. This was done on the WW domain – a small, independently folding sequence binding proline-containing peptides and possessing conserved tryptophan residues that give rise to the name. Despite the size of the domain, it is remarkable that even the structure of one of the artificial proteins closely agrees with a WW fold. Therefore, sometimes, protein folding problem can be overcome.

A more applied example concerns aspartokinase from *Corynebacterium glutamicum* (CgAK), which is useful for production of methionine, lysine, and threonine but is allosterically inhibited by the products. That is, the activity is impeded by product binding at a distal regulatory site. Chen *et al.* have efficiently used coevolutionary analysis to define the interaction network of the regulatory residues. They identified most of the previously located sites, which responded to random mutagenesis, as well as a number of novel ones. Through replacing amino acids with those least frequently found or with differing chemical properties, inhibition could be disrupted. Overall, they produced 14 mutants, partially or fully desensitised to product inhibition<sup>120</sup>. Apart from the residues immediately involved in inhibitor binding, which could be inferred through structural studies, others have been found to have an effect. Those cannot be predicted or easily explained through static crystal structures and are most likely involved in the signal transduction across the protein<sup>121</sup>. This, once again, highlights the power of the sequence-based coevolutionary methods.

### 0.3.1.2 Structure inference-based methods

Despite the successes of sequence-driven methods, three-dimensional considerations are of use as they intimately intertwine with function. Structure determination involving pure protein production, subsequent preparation and equipment-dependent studies is much more laborious than nowadays sequencing. Therefore, computational methods are used to project trends from structures obtained onto less studied proteins. The structural features of a protein are usually more conserved than the sequence making this possible. The methods can be broadly categorised into: *Ab initio* methods, where the structure is determined from scratch; fragment-based methods, where particular known motifs are identified; homology modelling, where structures, similar in amino acid sequence are used as a template; and protein threading: approach, somewhat similar to homology but relying on potential fold similarity, rather than primary sequence.

These go greatly beyond the scope and therefore we invite the reader to look in the Appendix A for a brief overview, if desired.

### 0.3.2 Directed evolution

Mutations occur during DNA replication and propagation, frequently influencing the underlying protein sequences. The changes can be beneficial – creating novel functions or improving existing ones for the current environment—or detrimental to the host. The idea popularised by Charles Darwin is that over millions of years nature has converged to particular sequences, well-optimised to their settings to ultimately facilitate procreation, whilst getting rid of the less efficient ones. Replication of this phenomenon on the molecular level can provide us with insights about natural evolutionary processes. It can also be an immediate way of creating proteins with desired properties by replacing nature in guiding the process. This gives rise to the term “directed evolution” (DE).

Akin to natural evolution the process relies on two iterative steps: creation of diversity and screening or selection of the variants according to the desired criteria. Concerning the former, it is thought that nature might have explored only a small subset of all possible sequences: a protein can be stuck in a local maximum of evolutionary fitness (suitability for gene reproduction). Also, if not enough selection pressure is applied, sometimes optimization can arrive at sufficient yet not the very best of the solutions. Diversity is achieved through creation of a genetic library. This can be generated from the initial protein sequence through asexual and sexual methods. The former consists of a multitude of random and focused mutagenesis methods. The latter of homologous or non-homologous recombination.

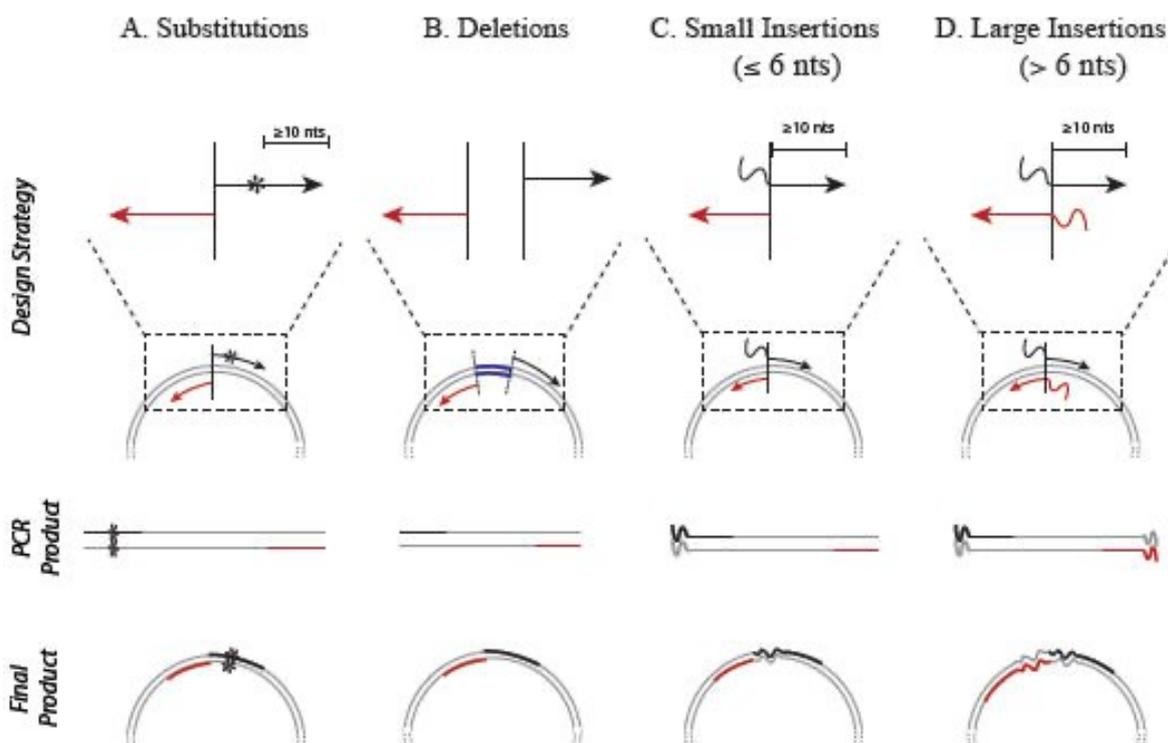
Briefly, asexual methods start with a single initial variant. Random mutagenesis requires no insight about protein structure. Mutations can be introduced chemically, through mutator strains, error prone PCR/RCA among others<sup>122</sup>. As PCR (and sometimes RCA) is typically part of the workflow anyway, it is a tempting option. Low fidelity polymerase is used with reaction conditions intentionally worsened. The method is simple, however, as the gene has to be ligated onto a backbone, this limits the throughput. Also, mutation occurrence is usually biased, for example, towards transitions.

DNA can be mutagenized *in vivo*. This can be due to physiochemical treatment or inherently low fidelity of DNA replication of the host (through disabling DNA repair machinery). This approach also allows straightforward continuous mutation of the DNA with complication being host genome suffering equivalent mutational load. The strain can therefore get progressively sick. This can be overcome through plasmid extraction and re-cloning but a limit on the mutational rate is still imposed. Complex protein-fusion/orthogonal systems are developed to specifically target the gene of interest *in vivo*, whilst sparing the host. One such example is the OrthoRep system developed by Chang Liu’s group, where targeted plasmid mutates  $10^5$  times faster thanks to target plasmid-specific DNA polymerase<sup>123</sup>.

Even for a small protein of 100 amino acids,  $20^{100}$  or about  $10^{130}$  combinations are possible, making the space difficult to explore. This is further complicated by the redundancy in the genetic code. Therefore, mutagenesis can be focused on regions of interest, if such regions can be identified. Site saturation mutagenesis (SSM) can be exemplified by site directed mutagenesis, commercial kits for which are widely available<sup>124</sup>.

The method can be summarised as relying on primers to specifically incorporate mutations during PCR. These are used on a plasmid and land at the mutation site pointing outward. In case of point mutations, the mismatch is encoded on the primers. In case of insertion, additional sequence is encoded on primer 5’-end, whereas deletion is achieved

through primers not encompassing the full sequence (**Figure 32**). The mutations are incorporated through PCR and biologically-derived non-mutated template is degraded by DpnI endonuclease enzyme, specifically degrading biological DNA that is methylated. Linear DNA is then ligated and transformed into bacteria. An array of methods follows a similar pipeline of targeting small sections of DNA with mutagenesis<sup>125</sup>.



**Figure 32 Site directed mutagenesis examples.** As found on <sup>126</sup>. Top to bottom: Primer annealing to original template, PCR product and final plasmid recovered from the host. In cases of base substitution (A), deletion (B), Insertions (C-D).

Sexual methods aim to harvest potential for recombination: combination of properties from parental genes, similar to nature. Like in nature, this allows accumulation of positive traits. A backcross with parent is also possible as an additional tool to eliminate deleterious mutations from a mixture. In this sense, homologous recombination is more intuitive and somewhat similar to consensus sequence design approach mentioned before but in a tube. An array of methods exists both *in vitro* and *in vivo*. In this work we will briefly mention our exploration of one. Non-homologous shuffling in turn allows for exploration of structurally similar but sequence-differentiated (a common instance as we mentioned before) chimeras as well as domain reorganisation within large proteins.

DE library generation methods appear to offer endless potential for design landscape exploration. However, this is limited by our ability to analyse the variants produced. For derivation of binding properties, affinity-based approaches can be used. The target is immobilised and only successful variants bind to it, whilst the rest are washed away. The proteins are linked to DNA through “display” methods, where some part of expression machinery links NA code to the protein displayed on it. Most commonly known is the phage display: protein variants are expressed on the phage surface<sup>127</sup>. Successful binding then

insures gene retrieval from the host. Similarly, cell-surface display, has been demonstrated for a variety of cells<sup>128</sup>. Likewise, cell free methods include ribosome display, mRNA display, DNA display and most interestingly, *in vitro* compartmentalisation that we will discuss in Chapter 2.

In case of enzymes that exhibit catalytic activity but not binding, the approaches above are not suitable. Some *in vivo* selections rely on linking enzymatic activity to survival and reproduction. The simplest case is use of auxotrophic strains and nutrient omission to only allow nutrient producers to grow. Metabolic networks can be altered to make a seemingly metabolism-irrelevant enzyme important<sup>129</sup>. However, some enzymatic activities might be orthogonal or detrimental to the cells in which case the screening is more complicated. In this case protein variants have to first be produced and then individually evaluated.

### 0.3.2.1 A case of Tryptophan synthase subunit B directed evolution

TrpB, that we mentioned before, has recently attracted a lot of interest as the unit is highly promiscuous: the catalytic site can accommodate a variety of indole analogues as well as  $\beta$ -branched L-ser analogues, giving rise to a great repertoire of products, otherwise difficult to synthesise. However, as mentioned above, activity of TrpB is reliant on activation by TrpA, so even if reactants are fed directly, TrpB displays little activity on its own (~5%). Therefore, standalone units were engineered. In case of *Pyrococcus furiosus* (*PfTrpB<sup>WT</sup>*), a single mutation T292S reportedly increased standalone catalytic efficiency by 20-fold (*PfTrpB<sup>T292S</sup>*) bringing it above that of the wild-type complex<sup>130</sup>. This was done through a DE approach: error-prone PCR library was transformed into *Escherichia coli* and 528 clones were manually screened. This was somewhat facilitated by the fact that the protein is thermostable and therefore can be extracted by heat-shock and immediately screened. However, performing this in a 96-well-plate format still posed some amount of work. In particular, it involved picking individual bacterial colonies, culturing them separately to sufficient densities, induction of protein expression, subsequent thermal lysis and plate reader absorbance-based enzyme kinetics assay.

Further recombination of 12 variants and screen of 1408 clones identified a 5-mutation variant yielding further 2.6-fold improvement. Another round of mutagenesis with 1144 clones screened finally brought a six-mutation variant with yet another 1.7-fold improvement on top, that is 83-fold overall. Although the achievement is remarkable, the number of clones screened is large. Homology-based approach was then used to transfer some of the mutations onto protein relatives. Interesting to us, *Escherichia coli* homologue was improved by two-fold through 2 point mutations yielding *EcTrpB<sup>M149T N171D</sup>*<sup>131</sup>. These, however, were taken from another of the 528 clones from the original *PfTrpB* screen as T292S was already present in *EcTrpB*. The repertoire of the methods just described allowed Arnold and colleagues to greatly enhance the rate of production of an impressive range of analogues in a variety of settings<sup>73,131–133</sup>, although they have most recently opted to explore OrthoRep due to the throughput potential<sup>134</sup>. In this work we will propose an alternative approach.

## 0.4 Molecular programming

Finally we will present here a brief introduction to molecular programs. This thesis uses molecular programs that are both inspired by natural regulatory motifs and aim to mimic the

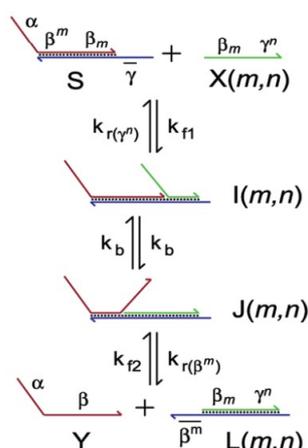
---

natural process of selection. They are therefore at the foundation of our work both as a platform for natural small molecule sensing reproduction and that for directed evolution workflow.

Taking inspiration from natural regulatory motifs, such as the ones described above, artificial circuits have been proposed. Molecular programming broadly falls into the field of synthetic biology, that aims to verify system understanding by building, a general approach attributed to Richard Feynman. Systems constructed can also be of practical use. To differentiate from frequently broad and ambiguous definitions of synthetic biology, we can consider molecular programming to restrict itself to *in vitro* circuits that are constructed bottom up<sup>135</sup>. Ultimately, we shall consider a molecular program system as a molecular machine capable of taking various inputs and responding with outputs, having performed some pre-defined computations in-between.

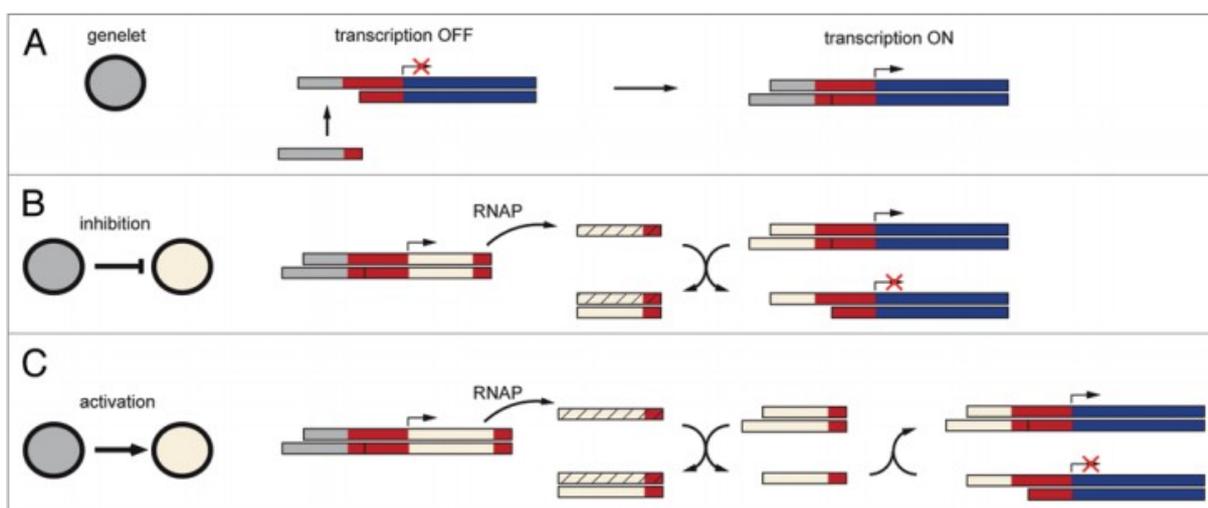
Such systems were shown to emulate, *in vitro*, a range of phenomena, starting from electronic circuits<sup>136,137</sup>, cellular regulation networks<sup>138,139</sup>, ecological webs<sup>140,141</sup>, and even neural architectures<sup>142</sup>. The modularity and chemical simplicity of the approaches usually allow rational and scalable circuit assembly, culminating in information processing systems whose performance is orchestrated by hundreds of different molecular components<sup>142</sup>. It is remarkable that just short strands of DNA can be sufficient for complex behaviors. Below, we will consider first an example of a DNA only-based circuit, and then move onto those including enzymes.

As first remarked by Ned Seeman<sup>143</sup>, due to predictable Watson-Crick based-pairing DNA immediately offers itself as a building material. An explosion of work followed paving the way to the DNA origami field. This verified that DNA strand interactions are understood sufficiently well to build large intricate structures. In parallel to this definition of self-assembly, DNA was proposed as a computational tool. One of simplest, yet versatile examples is the toehold-based DNA strand displacement circuits<sup>144</sup>. As affinities of oligos can be easily predicted, a system can be designed, where an oligo of greater complementarity will gradually displace that of lesser in a double-stranded complex with a given template. The oligo introduced serves as an input, whilst that ejected – as an output (**Figure 33**). The latter can in turn diffuse to some other complex or several thus continuing an execution of the program.



**Figure 33 A toehold exchange reaction and its mechanism.** From<sup>145</sup>. Complex S is disrupted due to arrival of invading strand X at  $\gamma^n$ . Intermediate I is formed and as the branch migrates converts to intermediate J. At this point Y spontaneously dissociates due to insufficient binding.

However, it is obvious that in such cases of domino-fall-like programs the reaction can only cascade once. Utilised strands have to be degraded and new – synthesised to maintain dynamic response of the program to the changing inputs fed in. Akin to nature, enzymes present an interesting option as they are naturally posed to synthesise and degrade NAs. An example commonly used is the transcription-based system developed by Jongmin Kim and Erik Winfree<sup>146,147</sup>. As summarised in a recent review<sup>148</sup>, the circuits rely on “genelets” – partially double stranded DNA, producing some output only when the upstream T7 promoter is complemented with a short ssDNA input. Inhibition is possible by balancing affinity of the activating input towards the genelet and an inhibiting input. Although RNA outputs cannot directly affect the genelets as DNA-RNA hybrids are inefficient in transcription, they can bind ssDNA inputs (**Figure 34**). Given sufficient amount of NTPs to serve as fuel, as well as T7 RNA polymerase and RNase H (RNA-degrading enzyme), the reaction can stay responsive and out of equilibrium for a long time and exhibit such complex behaviours as oscillations.



**Figure 34 Genelets mode of action.** Adapted from<sup>148</sup>. On the left, the circuit overview is shown, on the right, the mechanism in more detail. (A) The genelet is partially double stranded with T7 promoter (red) being incomplete. Upon binding of activating DNA (gray/red) transcription can start. (B) Activating strand can be removed if an inhibitory RNA molecule is produced and complexes the activating DNA. (C) Activation can be achieved through double inhibition: compared to (B), another RNA is produced to bind the first inhibitor.

#### 0.4.1 PEN-DNA toolbox

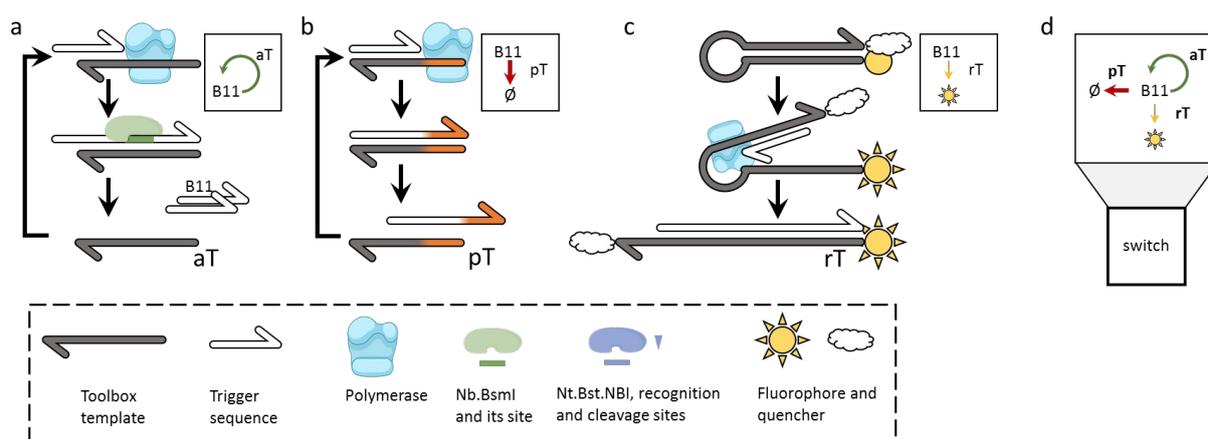
A particular ensemble of molecular programming tools, called PEN-DNA toolbox is gradually expanding and appears to be promising<sup>149</sup>. It has already allowed the rational building of a variety of molecular systems, including batch oscillators<sup>150</sup>, programmable pattern formation<sup>151</sup>, or ultrasensitive multi-input sensors<sup>152</sup>. The PEN-DNA employs Polymerase, Exonuclease and Nickase (PEN) enzymes as the chemical machinery. Short protected single-stranded oligos (templates) with an input-output structure provide instructions. Data flows are implemented via the endogenous production, exchange and decay of smaller strands of DNA called signals. The sequence of these signal strands is designed in such a way that i) the  $T_m$  of a single hybridized signal is similar to the reaction temperature, yielding dynamic binding and unbinding to templates and ii) it carries recognition sequence site enabling productive nicking/polymerization cycles on the template.

Modularity and biocompatibility of the system has prompted us to consider it for use in our work.

Here, we will build on three essential modules from the PEN-DNA toolbox: the first one is the autocatalytic module (aT, for autocatalytic template) (**Figure 35A**), whose encoding oligo consists of a dual repeat of the complement of an arbitrary “signal” sequence. When a cognate trigger binds to an aT, it is extended by the polymerase, the resulting full duplex is nicked, releasing two copies of the trigger. aTs thus provide positive feedback loops and exponential amplification of signals. However, as noted in a number of reports<sup>153,154</sup>, this motif is intrinsically sensitive to spurious initiation.

The second module, a pseudotemplate (pT) provides a solution to this instability issue. pT (**Figure 35B**) also has a trigger binding site, which is followed by a short poly-T tail (in this paper 2 bases). Additionally, the binding energy of the trigger for pT versus aT is biased in favor of pT by removing a few bases to the 3' extremity of the aT. At low concentration, triggers thus bind preferentially to pT, are extended by a short tail, which deactivates them. Above a certain concentration threshold, this sink gets saturated and exponential amplification ensues. The combination aT/pT thus creates a bistable switch that can be adapted to sensitive detection of a variety of inputs<sup>152,155</sup>.

The dynamics of these systems can be observed with non-specific dsDNA fluorescent dyes or by a variety of sequence-specific fluorescent reporters based on DNA-dye conjugates. Here we will use a trigger-specific fluorescent probe, we called reporter template (rT) (**Figure 35C**). It consists of a hairpin-structure DNA with the two ends carrying a quencher and a fluorescent dye. Upon binding and extension of the trigger the hairpin structure is disrupted, increasing the separation between the two moieties and giving rise to fluorescence. Since the trigger is short, this form is relatively unstable, but it is captured by the polymerase, which extends the trigger along the probe, thus making the unquenching reaction essentially irreversible. We shall refer to this 3-module assembly (aT, pT, rT) as a “switch” throughout the text (**Figure 35D**). A particular switch we developed will serve as a foundation for our work.



**Figure 35 PEN-DNA toolbox elements.** Polymerase, Exonuclease and Nickases (PEN) are used to drive a series of modules. (A) The autocatalytic module where ssDNA signal B11 binds the dual-repeat template aT, is elongated and nicked, thus releasing two identical triggers. (B) The pseudo-template module where the signal binds pT, is extended with a TT 3'-end, and released. The TT-tailed signal cannot initiate an aT cycle and will eventually be degraded. The pT reaction is faster than the aT one, but has limited capacity, leading to an adjustable threshold effect (C) Reporting module where the signal is extended over the hairpin structure, therefore dequenching a fluorophore, giving rise to fluorescence. (D) The combination of aT, pT and rT provides a bistable amplification switch, whose state is controlled by the concentration of B11 signal.



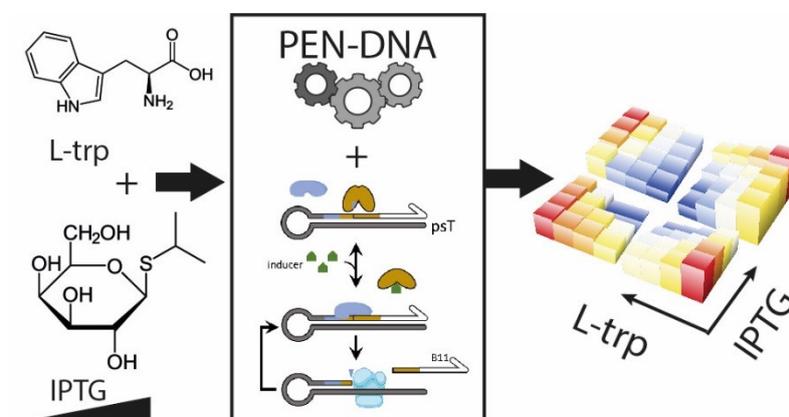
# Chapter 1: Sensing small molecules as multiple molecular program inputs.

## 1.1 Aims

In this chapter we aim to introduce the method we propose to interface small molecules with molecular programmes. We will outline our motivation based on discussion of current implementations that despite some promising results did not fully tap into the potential of the computational layer of the molecular programs. We will follow with the inspiration of our design strategy, built on miRNA detection tools previously developed by our group within PEN framework.

The original temperature constraints of the existing PEN systems, which were not suitable for our mesophilic transcription factor work will be presented together with our approach to overcome them. In particular, by creating our own system suitable for lower temperatures.

Finally, we will present the design strategy for our sensing modules, demonstrate their functionality and present how this architecture can be coupled with DNA-level information processing for two-input analogue adder implementation for IPTG and L-trp concentration inputs (**Figure 36**).

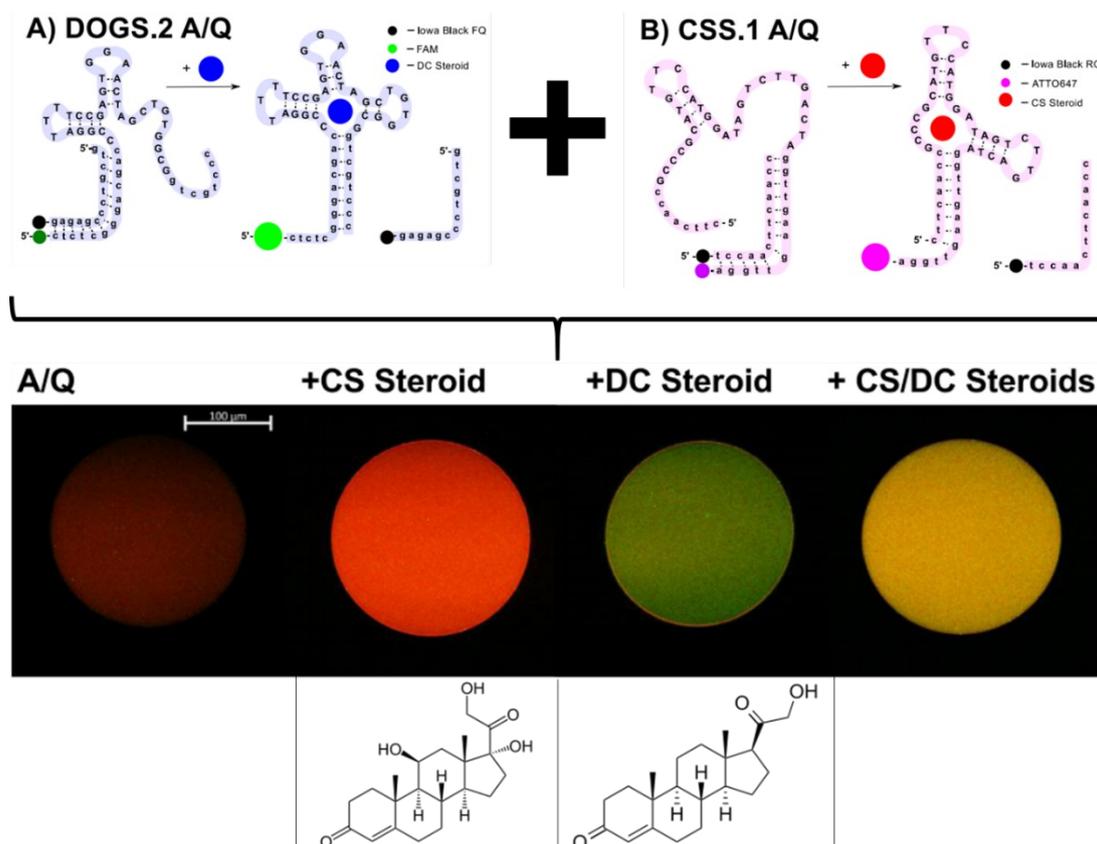


**Figure 36 Graphical abstract of chapter 1.** Two small molecule inputs (L-trp and IPTG) are processed in various ways by a PEN-DNA circuit combined with the sensing modules to provide combined outputs.

## 1.2 Motivation

One critical limit of most *in vitro* DNA-based approaches, such as the ones we mention in (0.4) is that they only accept inputs presented as nucleic acid species. In contrast, natural networks, such as the ones we have mentioned, as well as many *in vivo* synthetic circuits, accept a variety of chemicals, including small molecules, as inputs. The *lac* operon, for instance, can be viewed as a circuit that optimizes the metabolic response of a bacteria on a 2-dimensional glucose/lactose concentration space<sup>94</sup>. Synthetic biology also builds on the rich catalogue of regulatory elements, such as allosteric regulators or riboswitches, to engineer a variety of circuits responding to small-molecule inputs<sup>156,157</sup>.

Following the biological inspiration, early studies explored two strategies, a direct or an indirect one, to connect nucleic acid reactions networks and small molecule chemistry. The direct strategy leverages DNA or RNA-based aptamers<sup>158,159</sup>. A simple example of such approach was recently reported by Lakin and colleagues. It involves the use of two co-localized aptamers to respond to deoxycorticosterone (DC) and cortisol (CS)<sup>160</sup>. They demonstrated one-pot specific detection of two very similar molecules to produce a combined colorimetric output (Figure 37).



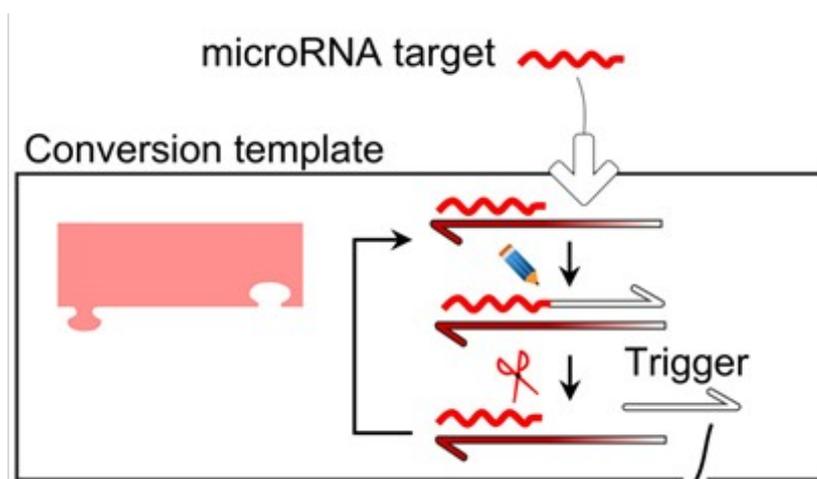
**Figure 37 Aptamers used together to detect deoxycorticosterone (DC) and cortisol (CS).** Adapted from <sup>160</sup>, please refer to the original work for details. Above: schematics of the two aptamers used together as well as their mode of action. Middle: illustration of the implementation. Left to right, negative control, cortisol (CS), deoxycorticosterone (DC) and a mixture of both are detected by droplets containing identical assembly of the two aptamers above. Bottom: structures of CS and DC. Although the authors did not explicitly specify whether they used 11-Deoxycorticosterone or 21-deoxycorticosterone, we assumed it was the former based on the reference they provided, where only the former is discussed as detected by the aptamer from this work<sup>161</sup>.

A large variety of well-designed aptamers is available. Ricci *et al.* demonstrated that their binding affinity can be finely and controllably tailored through a series of rational methods<sup>162</sup>. Ability to shift the response curve by simply shortening the aptamer or adding an allosteric inhibitor strand is of great advantage. However, this approach is limited by how low the binding affinity of an aptamer to a ligand can be. The aptamer affinity for its ligand can also be prone to decrease in face of modifications to bestow additional functionality. In case of Ricci's group work, the parent aptamer has  $K_d$  to cocaine of 4  $\mu\text{M}$ , whilst addition of fluorescent labels changed it to 824  $\mu\text{M}$ . Similarly, L-trp aptamer has been reported with  $K_d=2$   $\mu\text{M}$ , however, only under conditions of pH 6–7.5, a high concentration of  $\text{Na}^+$  ( $\geq 500$  mM) and 2–5 mM  $\text{Mg}^{2+}$ . Alternative conditions, such as increase in pH or decrease in salt strongly affected the performance<sup>163</sup>. Subsequent in-buffer sensors showed a detection limit of 6.97  $\mu\text{M}$ . As we aimed to integrate sensing into PEN with the buffer at pH 8.9 with 10 mM NaCl, we were anxious about compatibility of the approach. Nonetheless we would note that it was successful in other settings of simple sensors<sup>164,165</sup>.

In contrast, the indirect approach uses an intermediate molecule to bridge nucleic and small molecule chemistries. A prominent example is provided by some natural or engineered<sup>166</sup> transcription factors (TFs) we described in (0.2.6.2.2). This allowed the building of simple *in vitro* sensors of remarkable sensitivity<sup>167,168</sup>. For example, report of Zhao<sup>167</sup> and colleagues demonstrates interference of the tryptophan repressor (TrpR) with rolling circle amplification (RCA) process, presumably competing with DNA polymerase for DNA binding. The authors claimed that the interference with the polymerase's ability to bind was the mechanism of action and this was dependent on the quantity of L-trp added: the more L-trp is added, the greater is the capacity of TrpR to bind. The detection limit established was impressive 0.77  $\mu\text{M}$  (S/N=3). In our opinion, it is in part owed to optimisation of the reaction assembly: 5pM circular template was mixed with 2.5  $\mu\text{M}$  TrpR - over  $10^6$  excess allowing to compensate for reported dissociation constants ( $K_d$ ) of the TF being around 2.6 nM for the operator site at 25  $\mu\text{M}$  for L-trp<sup>169</sup>. This approach was not taken further to the computational level or generalised.

### 1.3 Design strategy

We hypothesised that interference, similar to that, reported by Zhao *et al.*, is possible with DNA polymerases and nickases used in PEN-DNA toolbox. This would create an interface through which molecular programs could "sense" the presence or absence of a given small biomolecule. Due to the existing framework of a bistable switch-based detector<sup>155</sup> we decided to reuse the architecture. A sensing module – the conversion template (cT) – has previously been proposed to sense medically-relevant miRNA and convert the signal into DNA PEN-acceptable input. This template allows binding of signal, in this case, miRNA to the 3'-end. miRNA is then extended according to the template coding for a trigger sequence, which is nicked and released. The extended miRNA remains bound thus, insuring linear production of the trigger (**Figure 38**).



**Figure 38 Conversion template for miRNA detection in PEN.** Adapted from [155]. microRNA binds to CT (gradient-coloured). Polymerase (pencil pictogram) extends miRNA to add trigger sequence. Nickase (scissors pictogram) introduces a ss break, releasing the trigger. miRNA remains bound and the cycle repeats.

### 1.3.1 Circuit design and optimization

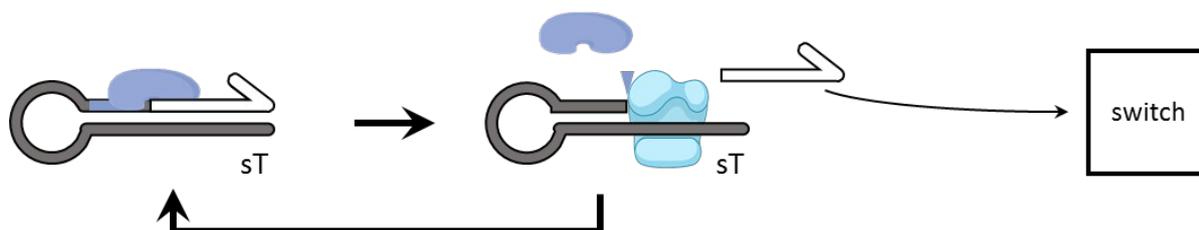
The miRNA detection strategy just presented was used with a particular bistable switch system<sup>155</sup> (concept introduced in 0.4.1). It consists of several oligonucleotides named after their function with “ $\alpha$ ” added to represent association with the particular switch “ $\alpha$ ” (**Table 1**). As this circuit was used at 50 °C, we shall refer to it as “switch<sub>50</sub>”. We considered using the said circuit.

Other circuits previously reported were also typically run at a temperature ranging from 45 to 50 °C, which is not appropriate if one wishes, for example, to use them to sense protein components coming from a mesophilic organism or use transcription factors from said organism. The first step was therefore to adapt the working temperature of the switch down to 37 °C.

Initially, we considered re-using switch<sub>50</sub> at 37 °C as it was already available in the lab. We prepared the reaction assembly as previously reported and can be found in the methods to be studied between 50 °C – the original temperature, and 40 °C – the lowest available in the “gradient” setting in the machine used. We then prepared two sets of samples: positive controls, containing what we shall call a source template (sT) at 0.5 nM and negative controls lacking it.

The source template is a component we will use, as the name suggests, as a constant source for linear production of the trigger sequence. It consists of a hairpin structure with a trigger sequence on the 5′-end and a nicking site just before. Similar to the miRNA sensing template, this allows cycles nicking, trigger dissociation and self-primed extension restoring the hairpin (**Figure 39**).

This activity allows addition of small amounts of the template that result in a gradual build-up of the trigger sequence and intentional start of the switch as opposed to sporadic self-start we mentioned in the introduction. This can be useful, for example, to simulate circuit response to a gradual build-up of an input. We will soon show examples of detailed designs and their uses, whilst here, we used a sequence we will discuss in greater detail in 3.3.4.2.



**Figure 39 Source template (sT) concept.** The sT DNA hairpin is nicked by Nt.BstNBI nickase (blue), trigger sequence dissociates (white) and can serve as input for the switch. At the same time, polymerase (light blue) binds partially doublestranded DNA and extends it to recreate the original hairpin. The cycle is repeated. Blue triangle represents the nicking site, whilst blue section of the template corresponds to the nickase recognition domain.

When we ran the experiment, we observed that the ability of the switch to start thanks to  $sT\alpha$  appeared to decay quickly as the temperature decreased (**Figure 40**). At the same time, for temperature of around 42 °C and below, a non-specific start of the circuit (i.e., start in the absence of  $sT\alpha$ ) appears and becomes dominant.

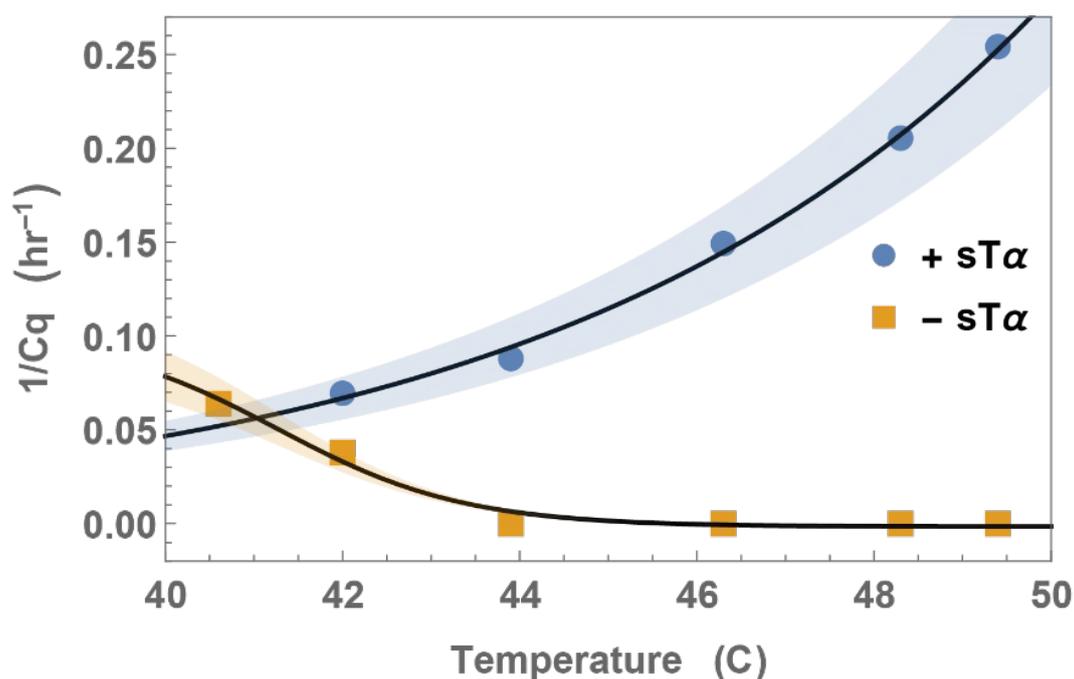
We concluded that at 37 °C the specific and non-specific start times would be bound to overlap. The oligonucleotide components of the switch were designed to have a particular melting temperature. Therefore, going outside of this regime to lower temperatures makes trigger dissociation – the process responsible for the dynamic behaviour of the system and information transfer, less probable. Coupled with slowing down of thermophilic enzyme activity with optimal temperature ranging from 55 °C to 70 °C, this explains the decrease in the starting rate of the positive controls.

The same phenomenon causes the self-start in some negative controls. The pT that acts as a threshold for the switch start becomes less efficient at lower temperatures as deactivated triggers are less likely to dissociate. Therefore, they saturate the sink and overflow towards the aT starting the switch.

#### Oligos used in switch<sub>50</sub>.

Name	Sequence
$\alpha$ (trigger)	CATTCTGGACTG
aT $\alpha$	C*A*G*T*CCAGAATG - CAGTCCAGAA
pT $\alpha$	T*T*T*TT-CAGTCCAGAATG
rT $\alpha$	Atto633*A*T*TCTGAATGCAGTCCAGAAT BHQ2
sT $\alpha$	TG-CAGTCCAGAA-CCTATCAATGATA-GACTC-TGCAAT TTTTATTGCAGAGTCTATCATTGATAGGTTCT

**Table 1. Oligos used in the switch<sub>50</sub>.** As in <sup>155</sup>. Stars represent phosphorothioate modification. Fluorophore and quencher modification positions are shown for rT. Dashes are shown for ease of interpretation.

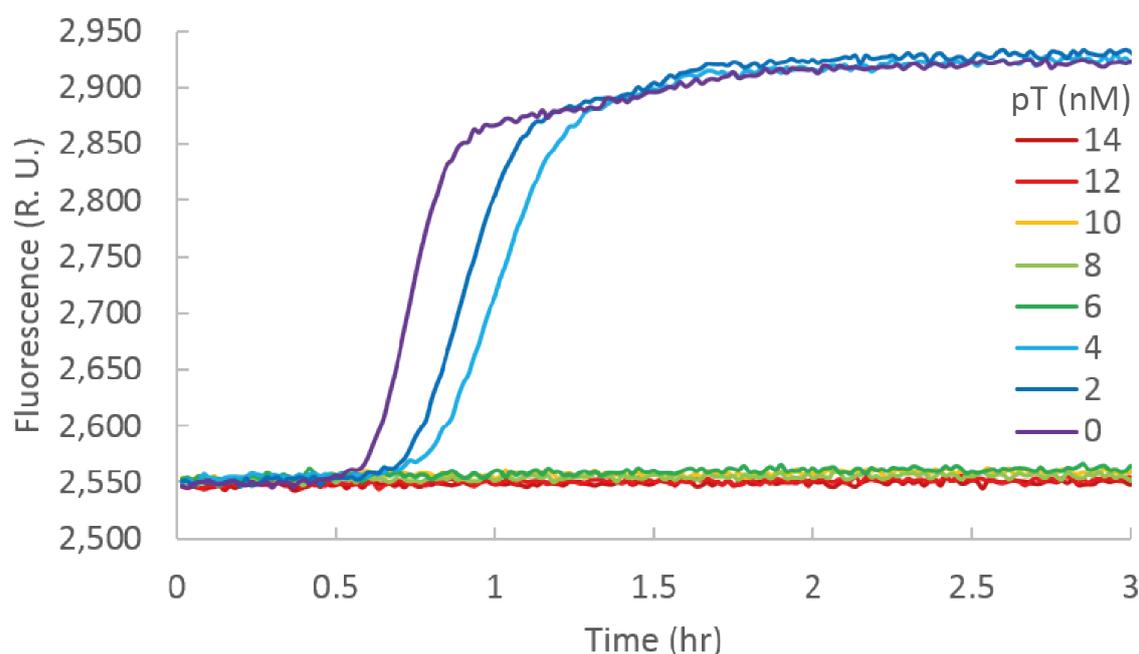


**Figure 40** Switch<sub>50</sub> performance of a range of temperatures. The switch was tested with (circles) or without (squares) source template at various temperatures. The expected behaviour (fast switching in presence of sT and no switching in the absence of sT) is observed down to 42°C, although the discriminative power seems to deteriorate. Below 42°C, no discrimination was observed for Switch<sub>50</sub>.

Since switch<sub>50</sub> was not suitable for us we decided to design our own. In particular, we found that shortening the signal sequence and decreasing its melting temperature to around 26 °C could still provide reasonable amplification kinetics. We then optimized the concentration of the switch components to achieve robust bistability, yet reasonable threshold of activation. That is, the switch remained off in absence of an input yet turned on, when expected input was provided.

The new switch sequences are reported in (4.2.3). The trigger sequence consisted of 11 bp and was thus termed “B11”. We shall occasionally refer to it like so, interchangeably with “the trigger”. We next had to find out the correct balance between aT and pT. Insufficient pT would result in overly weak thresholding and undesired self-starting due to the spurious initiation. Excess pT would create an unnecessary drop in sensitivity as stronger signal would be required to start the switch.

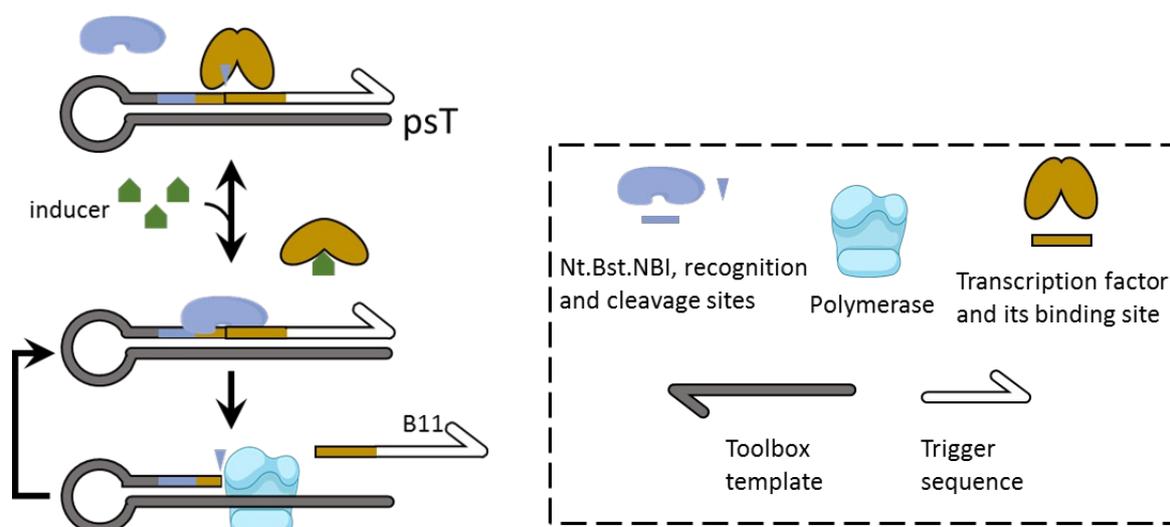
For 50 nM aT and 10 nM rT conventionally used, we determined 6 nM of pT to be minimally sufficient for bistability (**Figure 41**). We chose to proceed with 7 nM of pT to increase robustness of future experiments.



**Figure 41 Fluorescence time-trace of the switch with a range of pT concentrations.** The 37 °C reaction was assembled as reported bar only containing aT and rT oligos. Samples with 0, 2 and 4 nM pT start sequentially with little delay whilst 6-14 nM pT samples remain off. The reaction was observed for 15.65 (hr) with no significant further change in fluorescence.

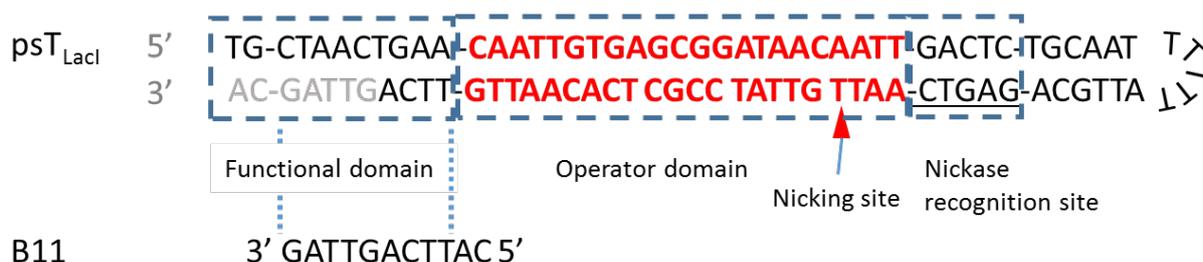
### 1.3.2 Sensing module concept

We combined the linear trigger output, as in case of miRNA detection described, with the TF interference. We called the resulting module the protein sensing template (psT) (**Figure 42**). A psT is a hairpin-shaped structure containing in the stem a nickase site followed by two variable elements: a protein binding site (operator) and an output-encoding sequence. For the nickase Nt.BstNBI, the cleavage site is located 4 bp downstream of the recognition site, so it can be designed to overlap with the operator site without sequence constraints (see 1.4.3 for details). In absence of the cognate operator-binding protein, this template acts simply as a constant source of its output sequence, through self-primed extension-and-nicking cycles. Here the output sequence is the switch's signal sequence, so a sufficient concentration of free psT will eventually turn the switch on. When the DNA-binding protein attaches to the psT, it blocks access for the nickase because of steric hindrance. In that configuration, the output production rate is expected to drop, and the switch start-up time to be delayed or abolished.



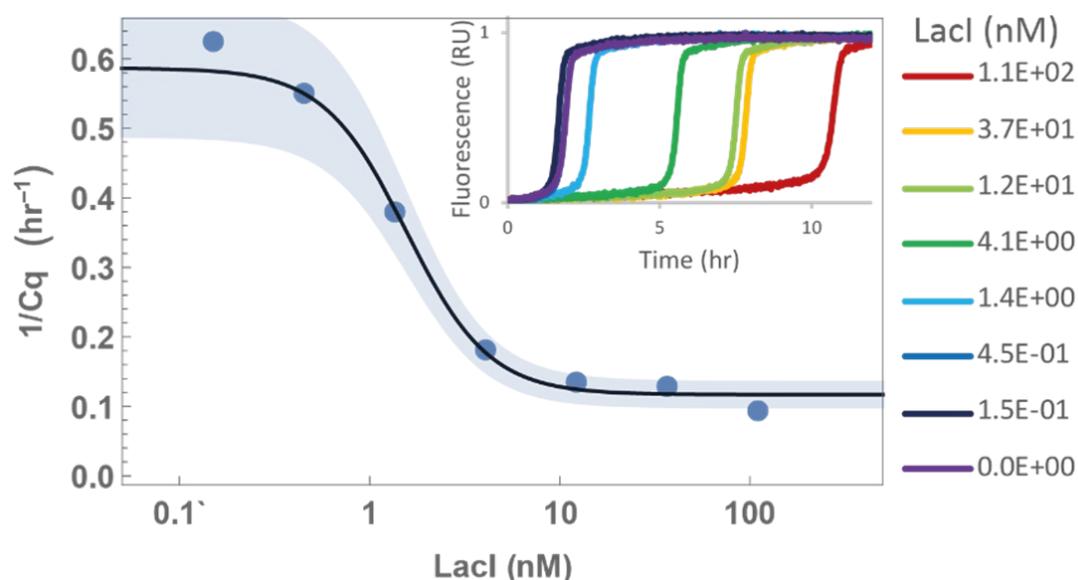
**Figure 42 Schematic of the protein sensing template (psT).** TF occludes the Nt.BstNBI nicking site on psT blocking its ability to produce signal via sequential nicking and polymerization cycles. Inducer addition results in TF unbinding and signal production.

To validate this strategy, we expressed and purified *Escherichia coli* LacI. LacI dimers bind DNA containing one of the three naturally occurring operator sequences with an affinity ranging from 1 pM to 100 nM<sup>96,170–172</sup>. In parallel, we designed the corresponding psT, with the B11 signal chosen as the output and LacI O<sup>1</sup> sequence – the operator sequence with the lowest K<sub>d</sub> (**Figure 43**). The B11 output sequence was modified in a “-2+2” fashion: two nucleotides were removed from 5′-end and two added on the 3′-end to increase aT complementarity, whilst decreasing that to pT. This makes psT output more potent in starting the switch. The remainder of the detailed designs used in this chapter is similar and can be found in **1.4.3**.



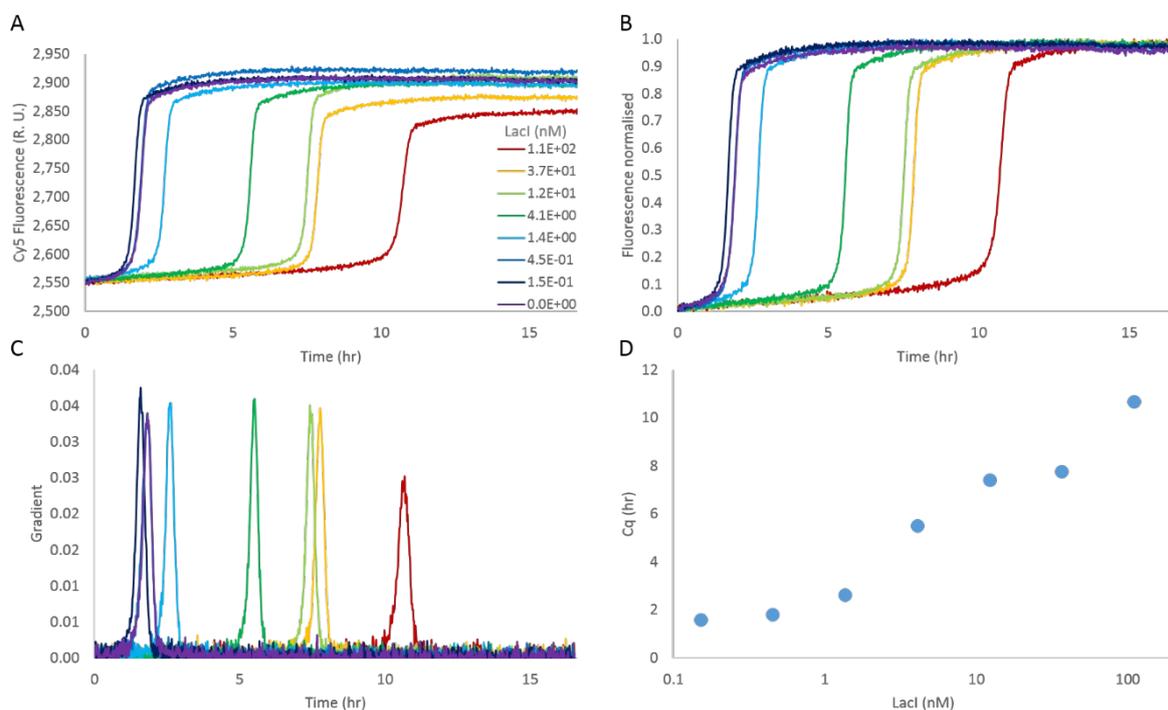
**Figure 43 psT<sub>LacI</sub> detailed design.** Here, going from right to left along the hairpin structure: polyT region and 6 bp, providing hairpin backbone; Nt.BstNBI recognition site (underlined); operator domain for TF binding (red), within which Nt.BstNBI nicking site is indicated (red triangle); functional region consisting of modified B11 signal sequence (shown standalone as well). The part of the sequence that was not chemically synthesised, but filled in by DNA polymerase upon reaction initiation, is shown in gray. Dashes are used for ease of interpretation.

We combined psT<sub>LacI</sub> with the switch and tested a range of LacI concentrations. As expected, we observed a direct effect on the switch response, with more TF leading to a delayed initiation (**Figure 44**). The repressive effect seems to saturate close to 4 nM of LacI dimer, consistent with the concentration of psT (5 nM), and the reported tight affinity of the TF for its operator<sup>173</sup> (~1 pM).



**Figure 44**  $psT_{Lacl}$  response to Lacl. Time-response to a range of Lacl dimer added to the switch with  $psT_{Lacl}$ . Inset – normalized fluorescent time traces.

For standardized quantitative analysis, we define the time of switching ( $Cq$ ) as the time point where the gradient of  $rT$  fluorescent trace is the steepest. Since in some conditions (strongly repressed  $psT$ ) the switch does not start at all in the experimental timescale, we find it useful to use  $1/Cq$  values. This allows to represent non-starters as 0 and to assign the highest values to the fastest reactions. **Figure 45** shows all of the steps of arriving at **Figure 44**: raw fluorescence traces are curated and normalized between 1 and 0. Differences between 2 points 10 minutes apart are then calculated (last 10 minutes are excluded) and the lower time-point corresponding to maximum difference value is extracted. The data are manually curated to identify flat traces with no distinctive switch-driven fluorescence increase<sup>174</sup> which are assigned  $1/Cq=0$ .

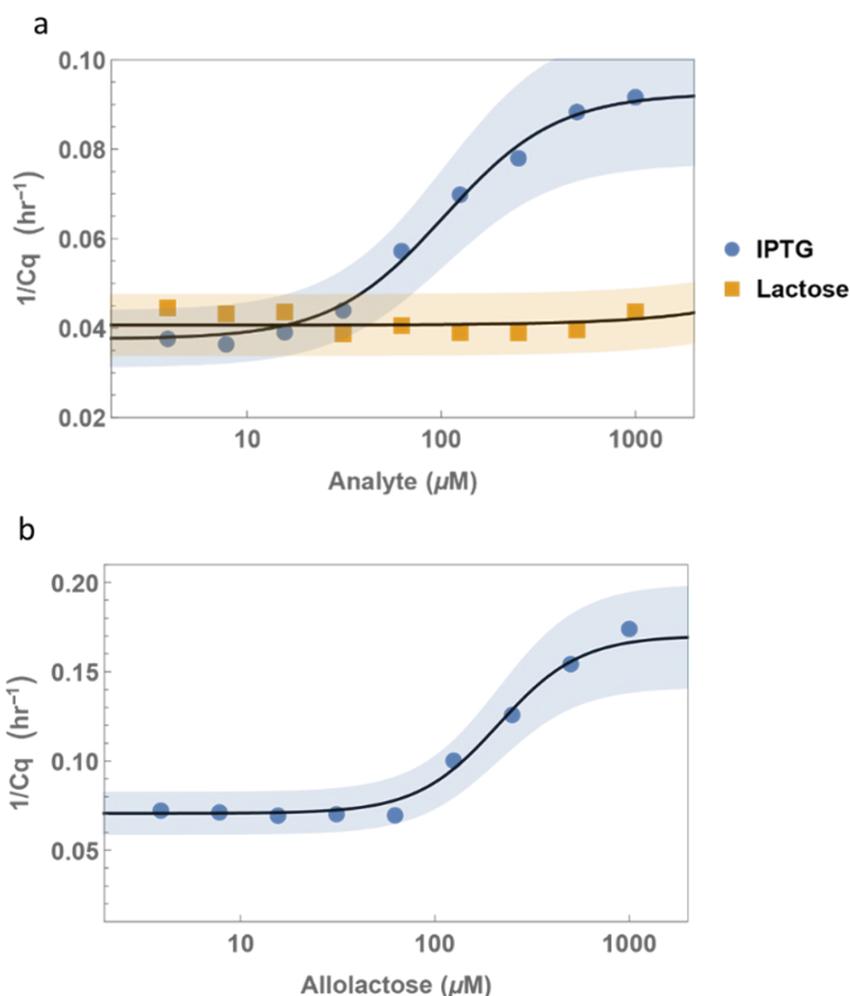


**Figure 45 Data processing workflow for Figure 44.** (A) Raw fluorescence reading for reaction assemblies containing varying concentrations of Lacl dimer (in nM). (B) Same data as in A, but normalised between 1 and 0. (C) Derivatives of the curves in B. (D) Highest derivative time points for each Lacl concentration.

## 1.4 Small molecule sensing using positive and negative TF logics

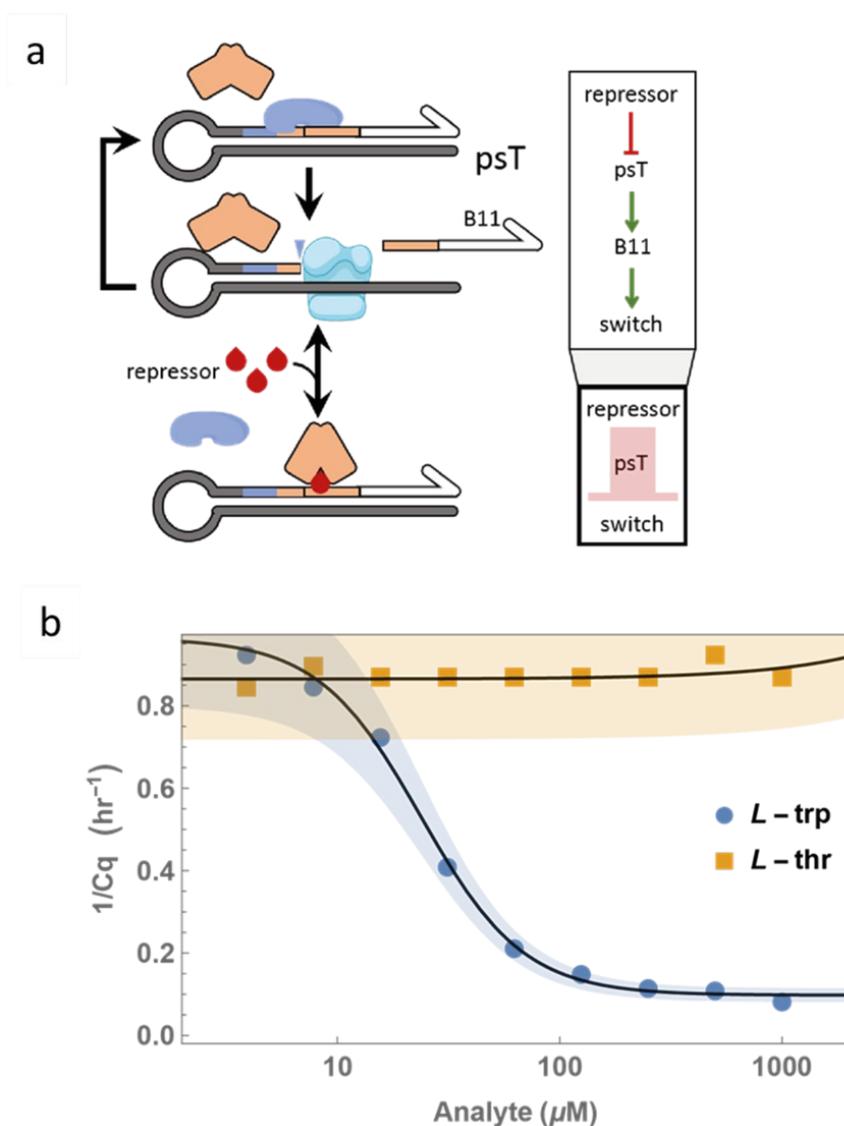
### 1.4.1 Direct signal relay

The classic inducer isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) we mentioned before can decrease Lacl affinity to DNA by over 4 orders of magnitude<sup>173</sup>. We therefore tested whether it has an effect on the circuit. A range of IPTG concentration going from 0 to 1000  $\mu$ M was added to the DNA circuit described above. We observed indeed that the DNA system now responds to IPTG (**Figure 46a**). The behavior is reminiscent of the typical sigmoidal bacterial induction curve for Lacl: more IPTG led to smaller Cq, that is, higher switch activity. The apparent  $K_d$ , close to 100  $\mu$ M, appears reasonable, provided the reported one is around 23  $\mu$ M<sup>173</sup> at pH 9.2, close to our pH 8.9. As IPTG is an analogue of the naturally occurring allolactose, the true effector of this system, we achieved similar results with the latter (**Figure 46b**). As a control, we used the same concentration range of lactose, also a close chemical analogue – a metabolic precursor of allolactose with no significant allosteric effect on Lacl. As expected, the switch circuit did not respond at all to lactose (**Figure 46a**). These results thus suggest that the specificity of the natural allosteric transducer has been maintained upon embedding into an artificial *in vitro* circuit.



**Figure 46**  $psT_{LacI}$ -LacI-based analyte detection. (a) Time-response to a range of IPTG (circles) and lactose (squares) added to the switch. (b) Time response to a range of allolactose.

While IPTG, or allolactose disrupt the LacI-DNA complex, other TFs may follow the opposite logic: an example from before, TrpR, only binds its operator sequence in presence of L-trp<sup>175,176</sup>. To test whether this class of sensing proteins could be used with DNA-based circuits, we purified *E. coli* TrpR and designed the corresponding  $psT$  presenting the TrpR operator site and the same B11 signal as output sequence (**Figure 47a**). When the reaction was assembled as before, but with TrpR and  $psT_{trpR}$  as input module, we observed that the switch indeed responded to the presence of L-trp. However, the response was inverted compared to the previous case, with more L-trp leading to larger Cq. Here as well, we could confirm the specificity of the sensing circuit using another amino acid as analogue: a concentration of L-threonine (L-thr) as high as 1 mM had no detectable effect on the switch. The Hill curve fitting on the L-trp range yielded a  $K_d$  value of 24  $\mu M$ , again consistent with the reported value of  $\sim 20 \mu M$ <sup>92,177</sup> (**Figure 47b**).



**Figure 47 Design and implementation of  $psT_{TrpR}$ .** (a) TF blocks the Nt.BstNBI site on psT upon repressor addition and stops B11 production via sequential nicking and polymerisation cycles. (b) Time-response to a range of L-trp (circles) and L-thr (squares) added to the switch with TrpR and  $psT_{TrpR}$  mixture.

### 1.4.2 Signal inversion

In practical applications, being dependent on the TF internal allosteric logic, i.e. its positive or negative response to the molecular stimulus, may seem inconvenient. Fortunately, the PEN toolbox offers straightforward mechanisms for logical inversion<sup>178,179</sup>. Here we decided to replace the B11 output of the psT with the complementary sequence – antiB11 (aB11). aB11 would therefore bind B11 in the circuit, preventing it from participating in the autocatalytic look and therefore preventing the start of the switch (see part of **Figure 48a**).

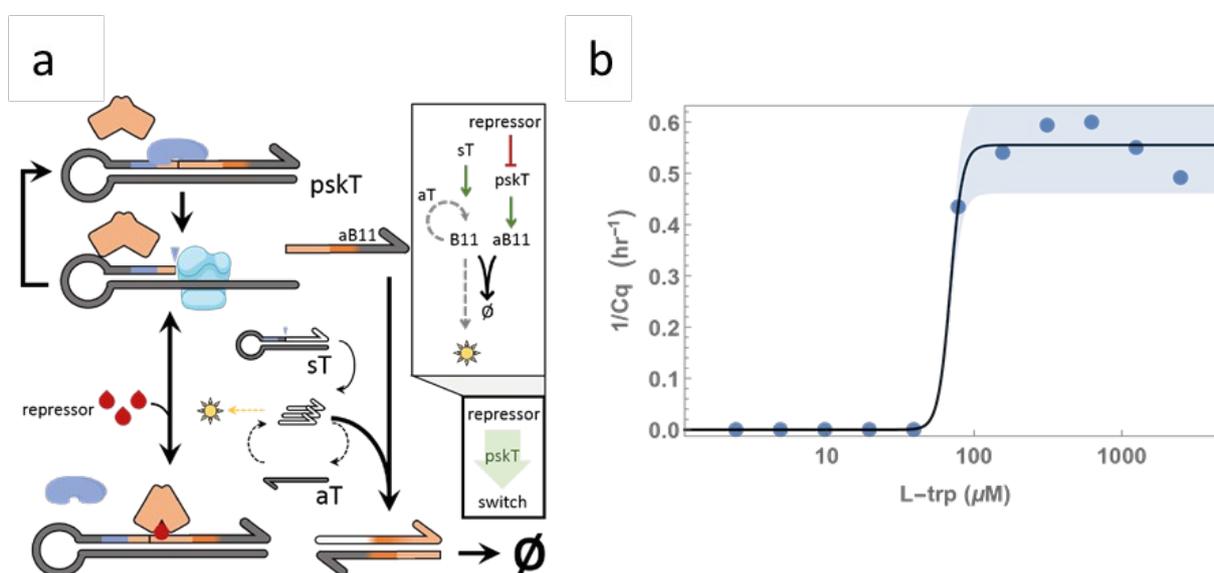
For clarity we named these repressing input modules protein sensing *killer* Template (pskT), in line with recently described killer Templates (kT)<sup>152</sup>. Use of such module on its own presents some difficulty: the bistable switches are designed to be OFF, unless activated. Hence, additional inhibition in form of an antitrigger sequence will not be apparent from the fluorescent trace.



variant will produce quantity of L-trp, sufficient to activate the switch faster. This, in turn, can be translated to gene replication.

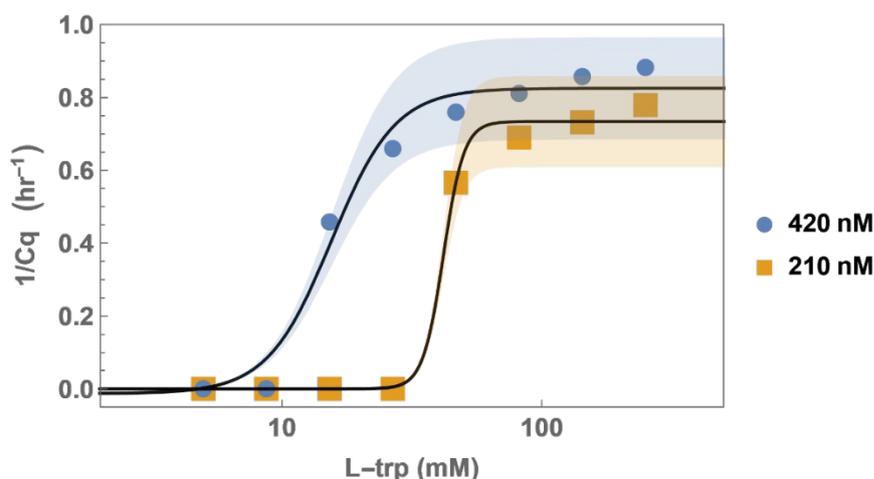
The  $K_d$  is  $69 \mu\text{M}$ , close to  $24 \mu\text{M}$  for the  $\text{psT}_{\text{TrpR}}$ , echoing the shift in  $K_d$  for the LacI system. Interestingly however, here the competitive mechanism implemented by this additional inversion layer resulted in a much sharper transition than the corresponding direct design. We observed a Hill coefficient shifting from 1.9 to 10.5. This observation is reminiscent of similar ultrasensitive behavior in system based on inhibitor titration<sup>180,181</sup>. Briefly, the authors show that such state is achieved when there is a finite sink strongly binding the input. Once saturated, the overflowing inducer suddenly impacts the circuit.

We believe this example shows that the DNA circuit can also be used to reshape the functional response of the allosteric element.



**Figure 49 Design and implementation of  $\text{pskT}_{\text{TrpR}}$ .** (a) aB11 is produced from pskT via sequential nicking and polymerisation cycles and the switch is initially deactivated. Repressor addition results in TF binding and aB11 production disruption leading to switch activation. (b) Time-response to a range of L-trp added to the switch with  $\text{pskT}_{\text{TrpR}}$  and TrpR mixture.

We did not investigate systematically the impact of TF concentration on the circuit response. However, we expect it to affect the apparent  $K_d$ . In **Figure 50**, we compare the response to L-trp for two TrpR concentrations, in the case of the inverted design with pskT (positive response). The 2-fold dilution notably affects apparent  $K_d$  which increases, as expected, but also the response becomes sharper. A further dilution to 42 nM resulted in loss of discrimination of the circuit as pskT repression became insufficient to overcome inhibition for all L-trp concentrations (not shown). This tuning opens up the possibility to adjust the sensing range and dynamics, depending on the application.



**Figure 50** Effect of TrpR concentration on apparent  $K_d$  of the L-trp pskT-based sensor. Final TrpR concentration of 420 nM (circles), 210 nM (squares). Conditions as in Figure 2F, apart from TrpR concentration.

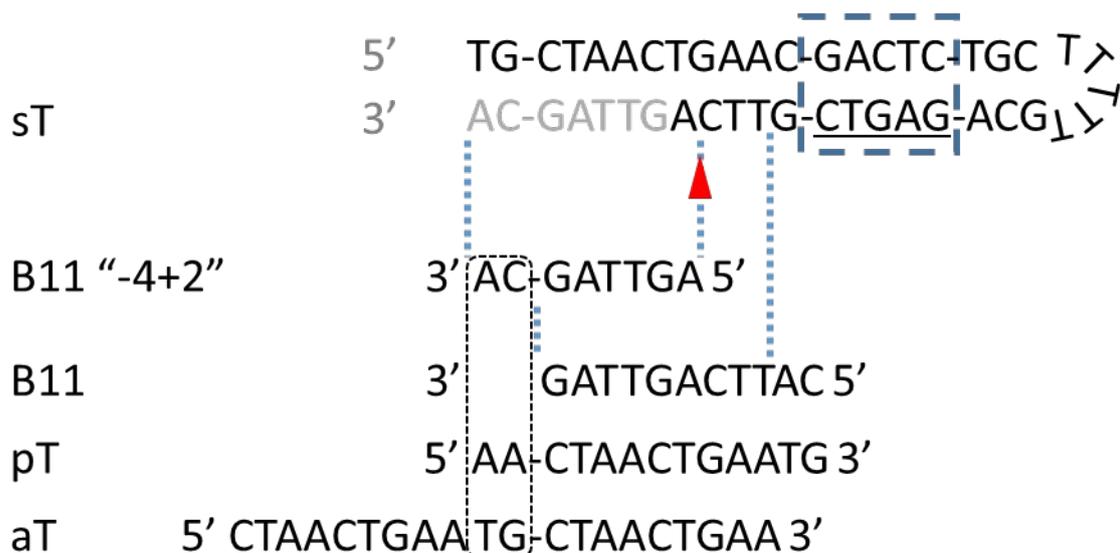
### 1.4.3 Design details

Here, we provide more information behind our design process. We started with the design of the Source Template (sT), a hairpin structure, containing an Nt.BstNBI site followed by the B11 signal sequence. Starting from the fully elongated form, sequential nicking, B11 dissociation and polymerase extension result in constant production of B11, proportional to sT concentration. To increase circuit start efficiency, we added 2 bp at the 3'-end, to match aT but not pT. This allowed to bypass the thresholding module and increase affinity for aT. We also shortened B11 sequence on the 5'-end to compensate for 3' extension and ease dissociation from the source template (**Figure 51**).

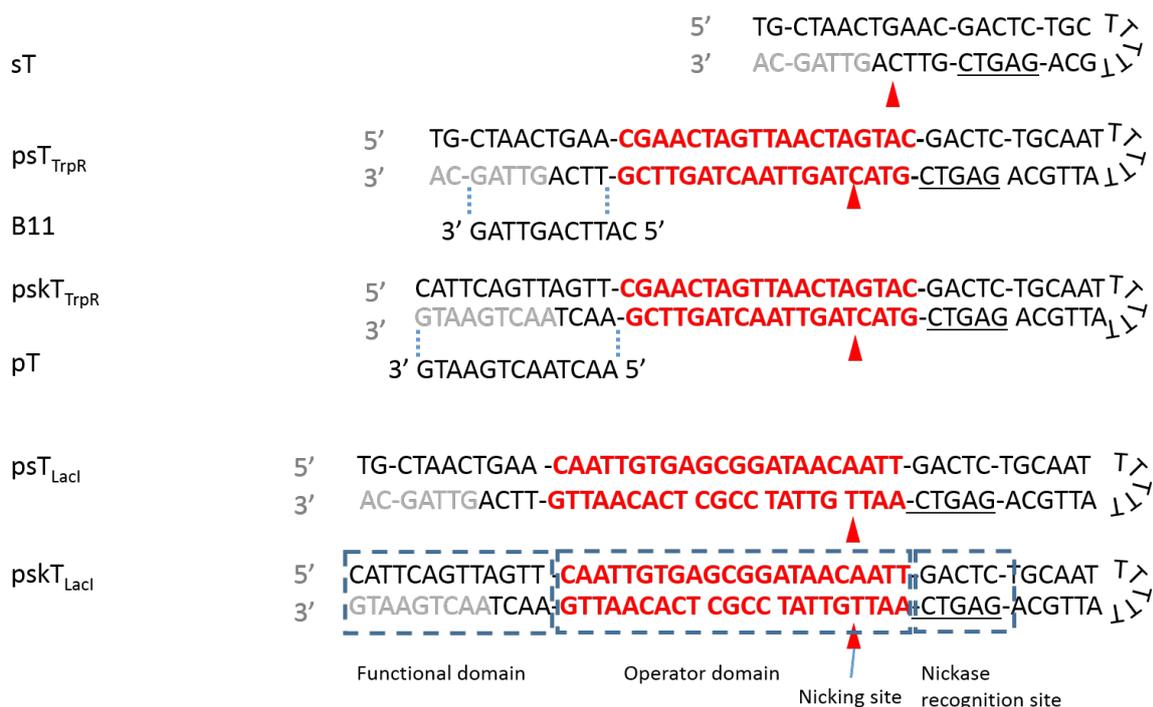
We propagated the strategy above to Protein Sensing Templates (psTs). The same hairpin structure was taken and TrpR or LacI operator site was inserted between the Nt.BstNBI site and B11 output signal (**Figure 52**). When operator site is not occupied constant signal, production occurs like in case of the sT. However, this means that the output sequence is longer as it is now catenated to a part of the operator sequence. Although this makes spontaneous dissociation unlikely (post nicking), this issue is alleviated by the fact that Bst polymerase Large Fragment used to drive the circuit has strand displacing activity.

One can observe that the LacI operator site consists of 22 bp and is 3 bp longer than TrpR site. We think this could have contributed towards worse performance of LacI-based templates, by reducing the rate of output production of the corresponding psT. This constraint was in part mitigated by adjusting the concentrations of the templates.

Protein sensing killer Templates (pskTs) were based on corresponding psT designs with replacement of output from B11 signal to antiB11 (aB11) sequence, similar to the pT sequence. As before, the resulting output contains part of the operator site as a tail.



**Figure 51 Detailed design of the source template (sT).** sT is shown along with its product B11 “-4+2”. B11, pT and aT are shown for reference. Blue dashed box represents Nt.BstNBI recognition sequence, whilst the red triangle – the nicking site. Black finely dashed line box highlights the “+2” nucleotides, complementary to the aT, but not the pT. Light blue dashed lines are shown for ease of interpretation.



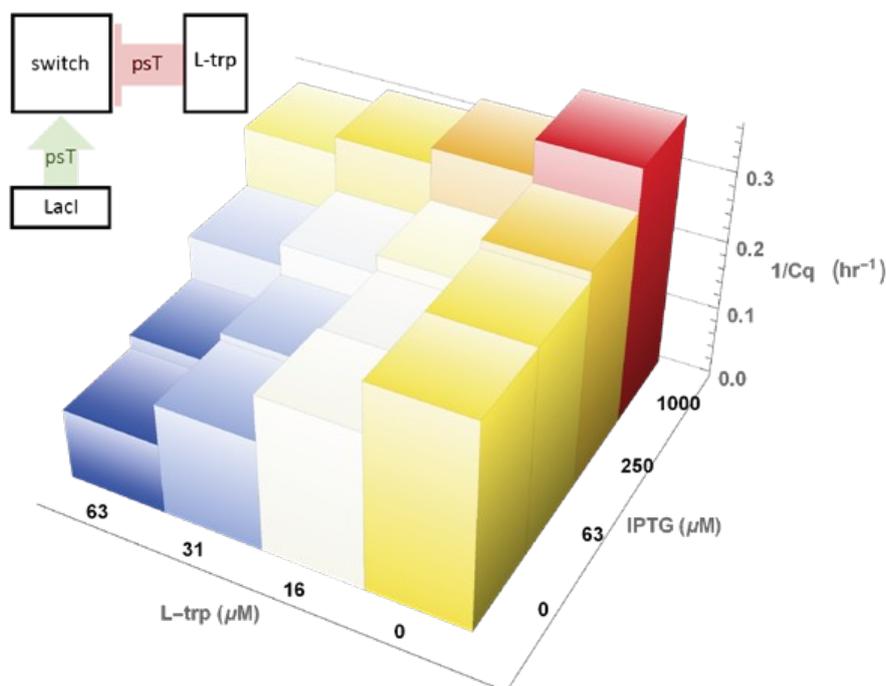
**Figure 52 In detail representation of the sensing modules used in this work.** Here, going from right to left along a hairpin structure: polyT region and 6 bp, providing hairpin backbone; Nt.BstNBI recognition site (underlined); operator domain for TF binding (red), within which Nt.BstNBI nicking site is indicated (red triangle); functional region consisting of pT sequence or modified B11 signal sequence (shown standalone as well). The part of the sequence that was not chemically synthesised, but filled in by DNA polymerase upon reaction initiation, is shown in gray. Dashes are used for ease of interpretation.

## 1.5 Multiple input implementation

Transducing small molecule concentration into DNA reactions as above in principle allows us to deploy the full toolbox of molecular information processing reported before, including logic functions, pattern recognition, signal amplification, etc, to the sensing of metabolites. To demonstrate this idea, we proceeded to create a two-input molecular classifier, based on IPTG and L-trp.

We combined both  $psT_{LacI}$  and  $psT_{TrpR}$  at 2.5 nM and 0.13 nM accordingly in a circuit, otherwise identical to the one, used for the individual sensors. We then added combinations of concentrations of both analytes (**Figure 53**).

We observe, that IPTG increases the starting rate of the circuit, whilst L-trp diminishes it. This is in line with their individual action through corresponding  $psTs$ . For the steps of concentrations chosen, the response is equivalent in both IPTG and L-trp dimensions. We therefore claim we have successfully build an analogue adder.

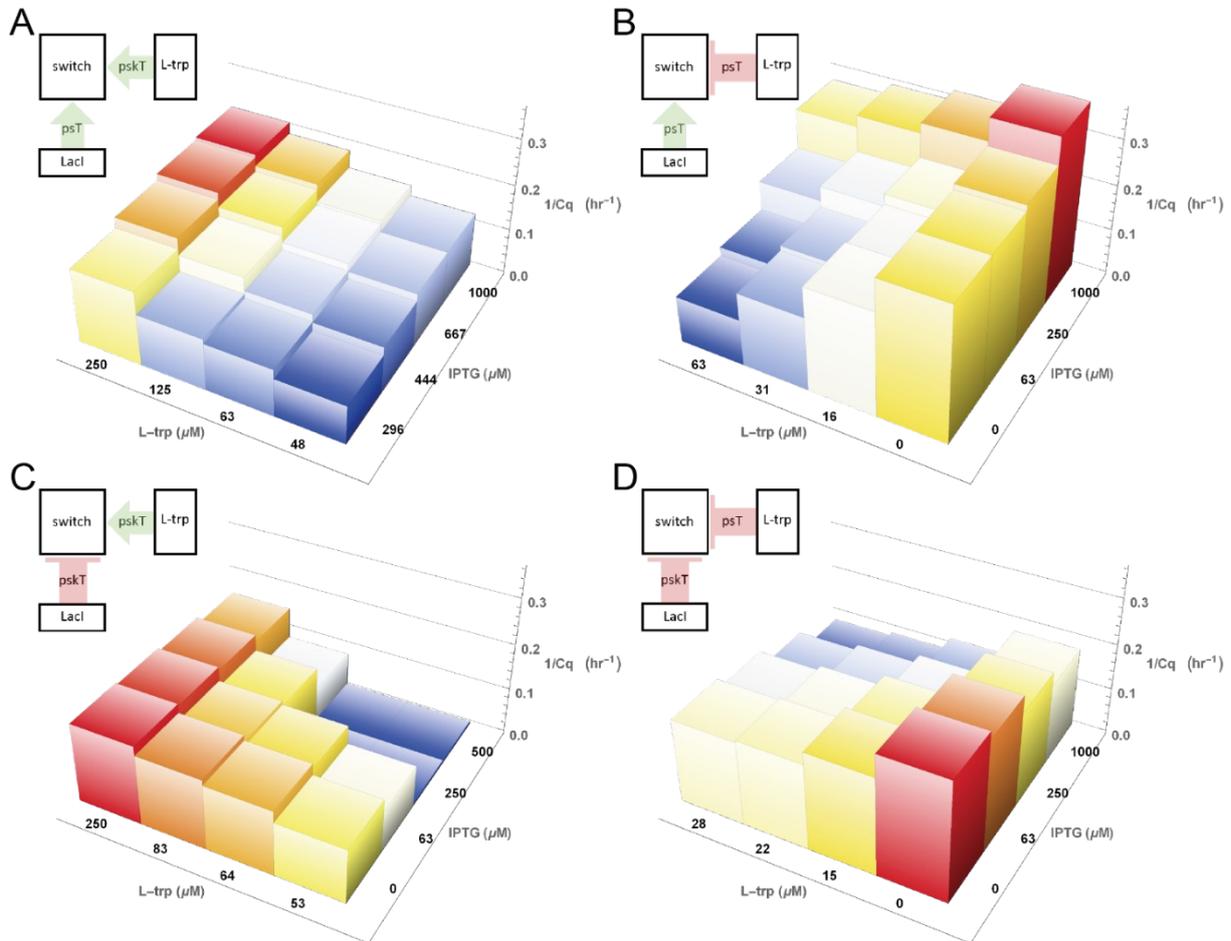


**Figure 53**  $psT_{LacI}$  and  $psT_{TrpR}$  both driving the circuit. Inducer concentrations are shown along the base of the plot against  $1/Cq$ . Colour highlights the relative magnitude in the  $1/Cq$  axis between the samples.

However, the logical inversion mechanism reported above allowed, us to design four possible 2D systems (including the one just shown). We therefore went on to construct them all (**Figure 54**).

We combined two input channels into an integrated response, whatever the logic, through adequate matching of the time scales and reaction strength. Fortunately, in DNA-based chemical reaction networks these parameters are controlled by the reaction kinetics, and ultimately by the encoding DNA concentrations. We could therefore balance the systems, by adjusting  $psT/pskT$  concentrations to change the inputs strength, and  $pT$  to adapt switch sensitivity. Only given these adjustments, we obtained a series of plots showing rotational symmetry. This tunability is both useful for rapid rewiring of a circuit and for adaptation to

different transcription factors. In the plots shown in **Figure 54**, we tuned the system to obtain an approximate weighted log-sum on the inputs, where the weight can be positive or negative, but other responses are possible as we observed during optimization. The weight and shape of the contribution of a particular input can be varied, which can be used to prioritize one analyte over another or binarize the response as we will show below in 1.5.1.

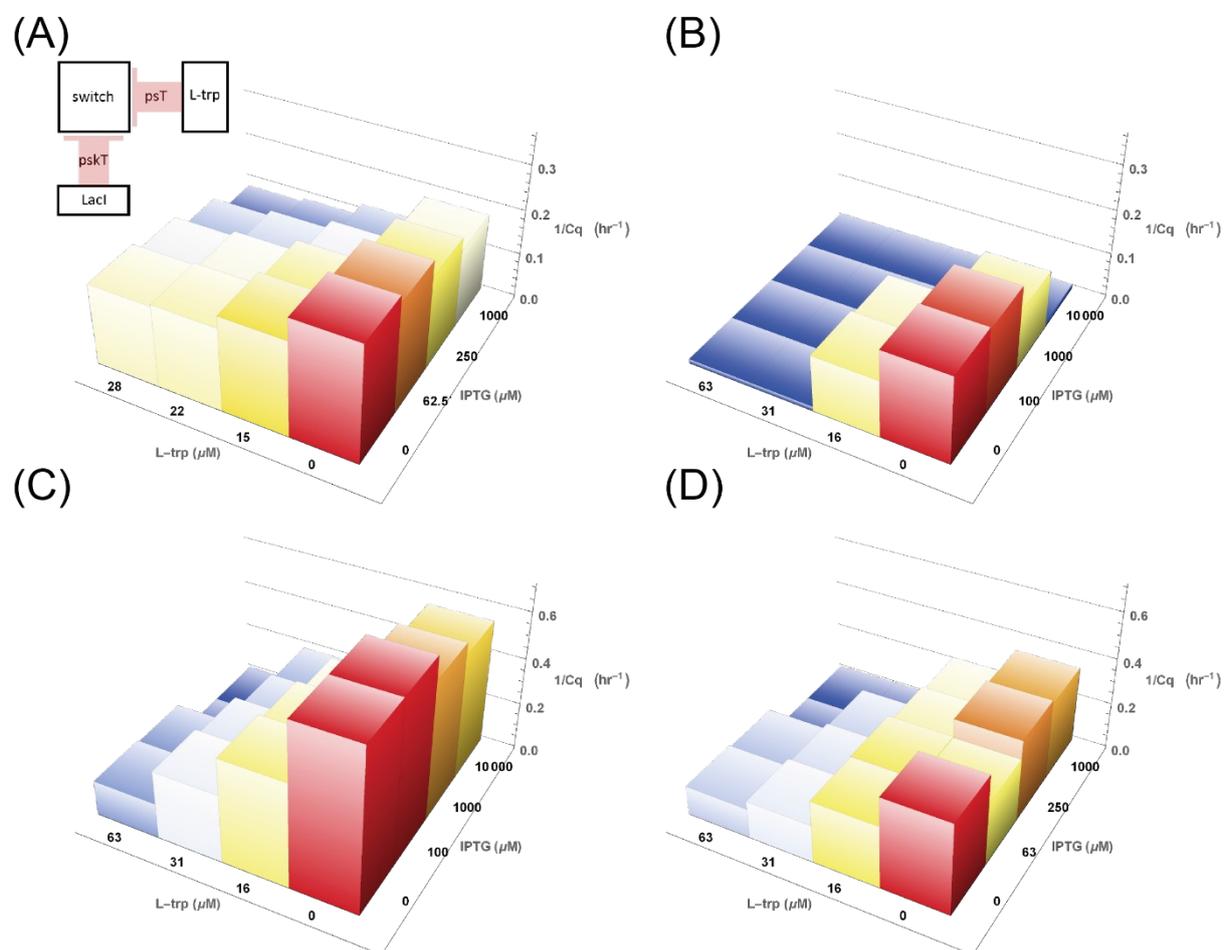


**Figure 54** Combinations of protein sensing (killer) templates with ranges of concentrations of two inputs. LacI with TrpR are present in all samples. Circuit schematics are shown above the figures. Colour range represents relative  $1/Cq$  variation within individual plots. (A)  $\text{pskT}_{\text{TrpR}}$  and  $\text{psT}_{\text{LacI}}$  for L-trp inverted sensing, IPTG direct sensing. (B)  $\text{psT}_{\text{TrpR}}$  and  $\text{psT}_{\text{LacI}}$  for L-trp direct sensing, IPTG direct sensing. (as before) (C)  $\text{pskT}_{\text{TrpR}}$  and  $\text{pskT}_{\text{LacI}}$  for L-trp direct sensing, IPTG inverted sensing. (D)  $\text{psT}_{\text{TrpR}}$  and  $\text{pskT}_{\text{LacI}}$  for L-trp inverted sensing, IPTG inverted sensing.

### 1.5.1 Multiple input processing variations

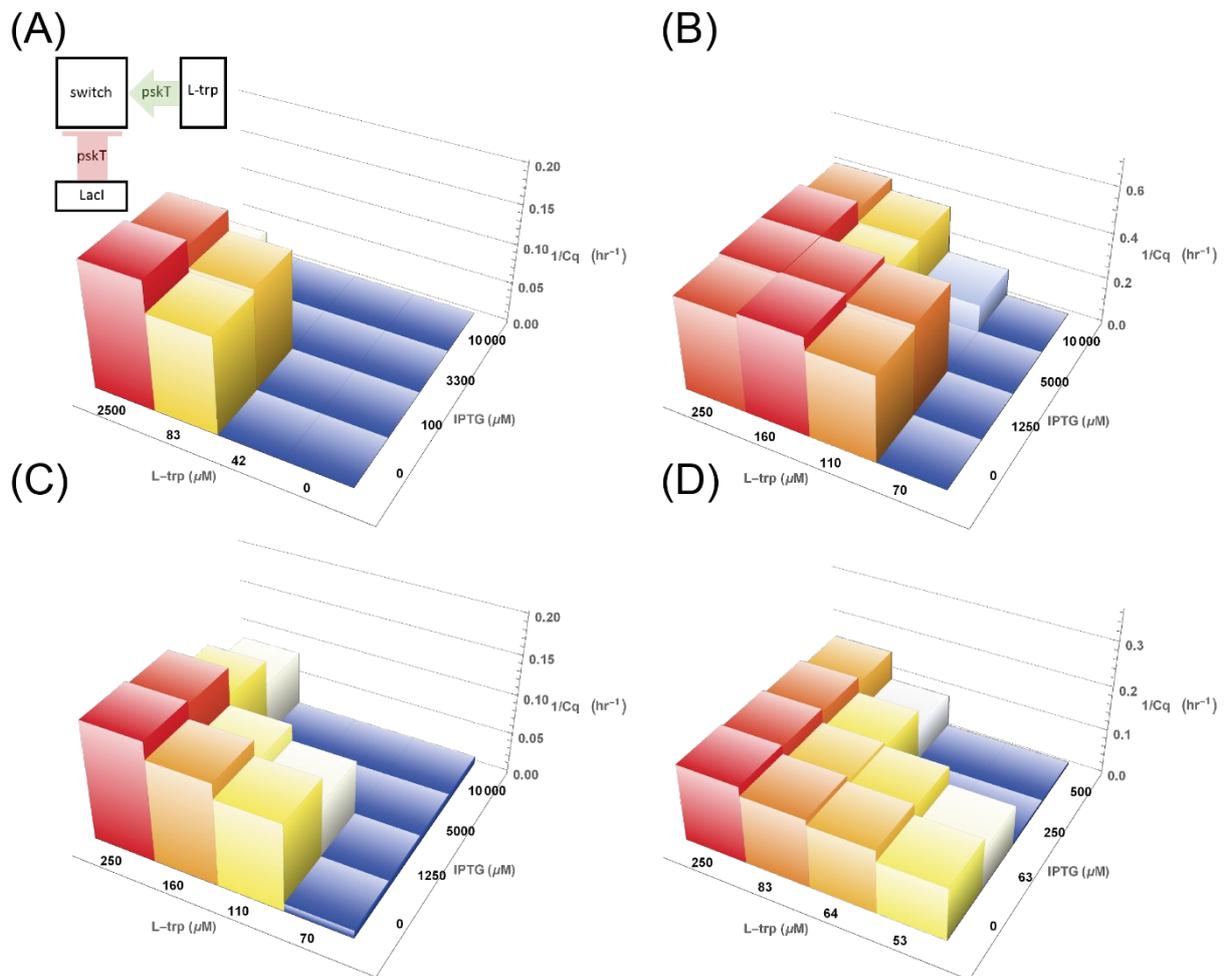
By adjusting the experimental conditions, we found it possible to obtain various shapes for the response of a system to its two inputs. For example, compared to the system shown in **Figure 54D**, reproduced here as **Figure 55A**, increasing pT concentration from 1nM to 4nM, setting a higher threshold, as well as decreasing  $psT_{TrpR}$  from 62.5 pM to 12.5pM, sets a much sharper cutoff in both dimensions (**Figure 55B**).

In **Figure 55 C,D** conditions were similar to **Figure 55B**, and only sensing module concentration was varied:  $psT_{TrpR}=125$  pM,  $pskT_{LacI}=7.5$  nM and  $psT_{TrpR}=62.5$  pM,  $pskT_{LacI}=10$  nM respectively. The resulting slow system in **D** becomes almost insensitive to IPTG input, whatever the L-trp concentration. The system in **C** applies a larger weight to L-trp, and a smaller to IPTG, compared to the original design.



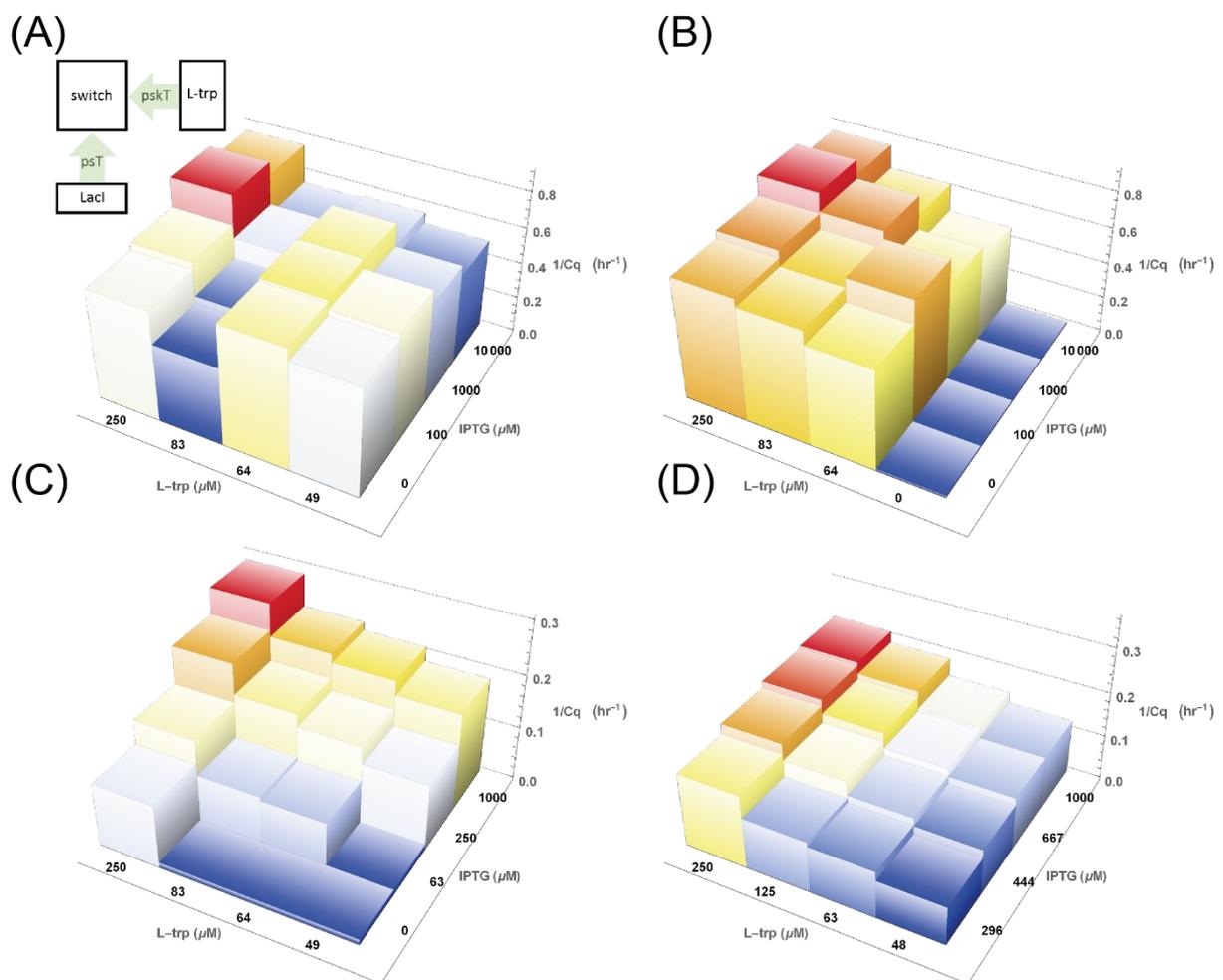
**Figure 55** Alternative plots for **Figure 54D** with altered input balance. (A) Original plot:  $pT=1$  nM,  $psT=0.063$  nM,  $pskT_{LacI}=40$  nM. (B) As in A bar  $pT=4$ nM,  $psT_{TrpR}=12.5$  pM,  $pskT_{LacI}=10$  nM. (C) As in B bar  $psT_{TrpR}=125$  pM,  $pskT_{LacI}=7.5$  nM (D) As in B bar  $psT_{TrpR}=62.5$  pM,  $pskT_{LacI}=10$  nM.

pskTs tend to provide sharper transitions implementing binary classification (**Figure 56A**). The system is very sensitive as doubling  $sT$  concentration (from 8 to 16 pM) has a large impact on the classification limits (**Figure 56B**). To obtain the smooth log-linear summing, we used 8 pM  $sT$  but rebalanced  $pskT_{TrpR}$ ,  $pskT_{LacI}$ , from 5 nM of each to 7.5 nM of  $pskT_{LacI}$  and 2.5 nM of  $pskT_{TrpR}$ . This corrects the bias towards stronger TrpR-based element and brings more samples into the window of observation as the total amount of pskT leads to less repression (taking into account that LacI-based elements are less efficient) (**Figure 56C**). Finally, we made a fine adjustment to  $pskT_{TrpR}=1.5nM$ ,  $pskT_{LacI}=7.5nM$ , to further rebalance and speed up the switch to produce the final plot reported (**Figure 56D**).



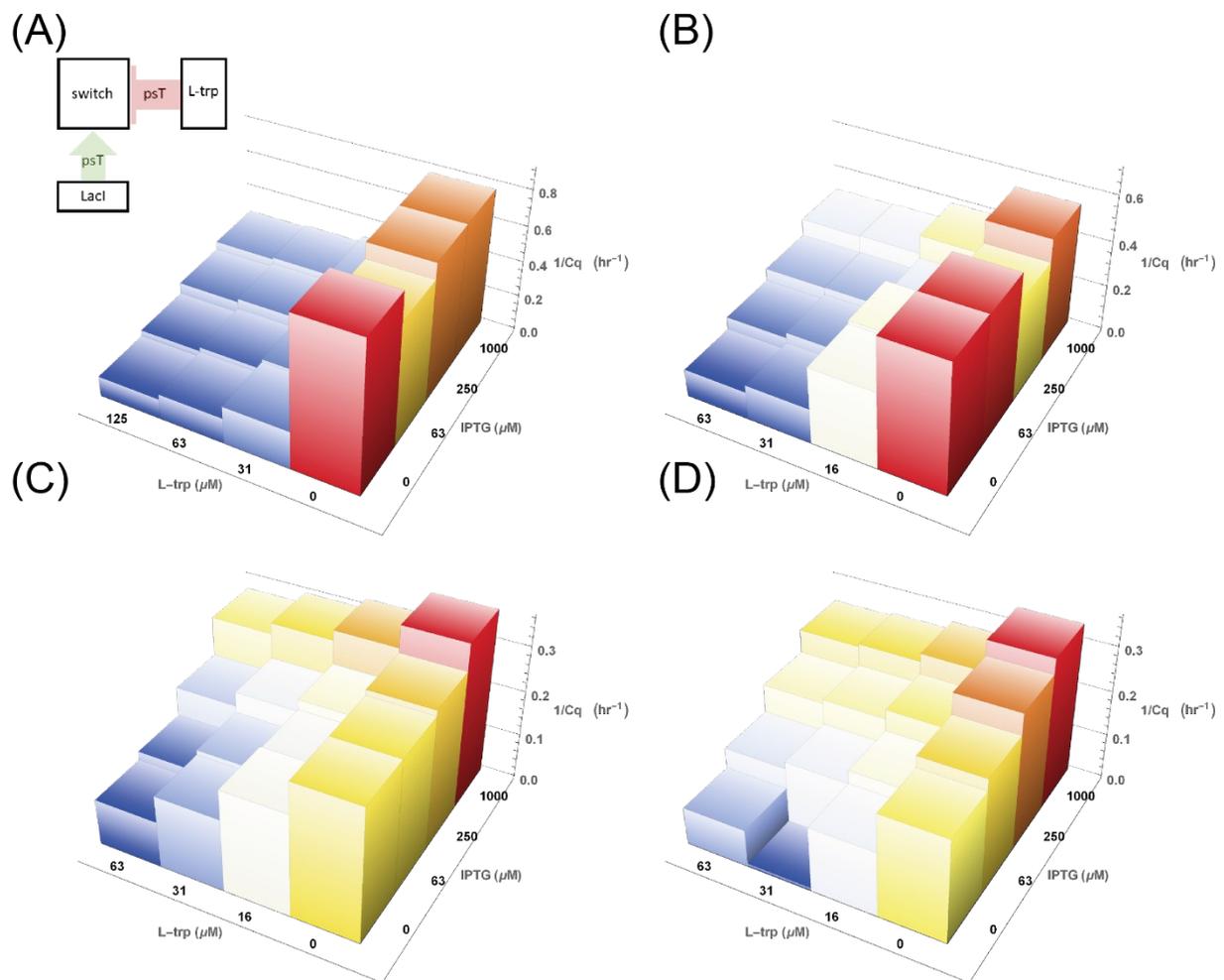
**Figure 56** Alternative shapes for the positive/negative circuit with altered input balance. (A) As in **D** bar  $pskT_{TrpR}=5$  nM,  $pskT_{LacI}=5$  nM. (B) As in **A** bar  $sT=16$  pM. (C) As in **A** bar  $pskT_{TrpR}=2.5$  nM,  $pskT_{LacI}=7.5$  nM. (D) Original plot:  $pT=4$  nM,  $pskT_{TrpR}=1.5$  nM,  $pskT_{LacI}=7.5$  nM,  $sT=8$  pM.

The positive/positive design shown in **Figure 54A** combines psT and pskT. This can present a challenge for balancing the circuit as pskTs tend to have less impact on the circuit. On the one hand, the circuit has to start in reasonable time and respond to the psT TF binding. On the other, the pskT should be able to compete and have sufficient delay effect, that TF-related relief would be resolved. **Figure 57A** shows a system starting too fast for input discrimination to occur. Decreasing the amount of psT<sub>LacI</sub> ten-fold from 2.5 nM slows down the system but makes it little sensitive to IPTG (**Figure 57B**). Another option to enter the dynamic range is to increase pT (from 4 nM to 7 nM in **Figure 57C**). This provides a reasonable window of observation, which we narrowed down to arrive to the final plot reported (**Figure 57D**).



**Figure 57** Alternative plots for the positive/positive circuit with altered input balance. (A) As in D bar,  $pT=4nM$   $pskT_{TRP}=10$  nM,  $psT_{LacI}=2.5$  nM. (B) As in A bar  $psT_{LacI}=0.25$  nM. (C) As in A bar  $pT=7nM$ . (D) Original plot:  $pT=7nM$ ,  $pskT_{TRP}=10nM$ ,  $psT_{LacI}=10$  nM.

Finally, the psT-only (IPTG-positive, L-trp-negative) design was the easiest to optimise. This is probably due to more robust performance of the psTs in general, coupled with lack of need for inter-species (psT/pskT) balancing. Therefore, we did not explore much beyond reported in **Figure 54B**, reproduced in **Figure 58C**. However, our first attempt used 1 nM of psT<sub>TrpR</sub> and 2.5 nM of psT<sub>LacI</sub> and was not in line with our expectation. Unless fully repressed, psT<sub>TrpR</sub> pushed the circuit into “always on” mode, hiding potential effects of the LacI inputs (**Figure 58A**). Decrease to 0.5 nM of psT<sub>TrpR</sub> and doubling psT<sub>LacI</sub> to 5 nM, as well as shift in L-trp range showed some effect (**Figure 58B**). Further halving of psT<sub>TrpR</sub> concentration allowed us to arrive at the plot reported (**Figure 58C**). We also tried halving psT<sub>LacI</sub> concentration to the original in hope of better IPTG resolution but the effect was marginal and the experiment yielded an outlier (**Figure 58D**).



**Figure 58 Alternative plots for the negative/positive circuit with altered input balance.** (A) As in C bar, psT<sub>TrpR</sub>=1 nM, psT<sub>LacI</sub>=2.5 nM. (B) As in C bar psT<sub>TrpR</sub>=0.5 nM. (C) Original plot: psT<sub>TrpR</sub>=0.25 nM, psT<sub>LacI</sub>=5 nM. (D) As in C bar, psT<sub>LacI</sub>=2.5 nM.

## 1.6 Summary

In this chapter we outlined the existing state of the art relating to small molecule sensing in molecular programs. We presented the better explored aptamer approach as well as the intrinsic sensitivity limitations. This was complemented by transcription factor -based sensor, which although achieved promising results were not integrated with downstream computation.

We moved to present the existing approach for ultrasensitive detection of micro RNA with the PEN toolbox whose architecture we intended to reuse. This was followed with the discussion of bistable switch parameters, namely the gap between specific (input-dependent) and non-specific (spontaneous) starts. This is determined on the one hand by the capacity of the aT and related assembly to start the switch, determining the fastest time to achieve the ON state. On the other hand, this depends on the capacity of the pT sink to prevent self-start of the system. Through use of a source template (sT) – a constant linear source of input sequence (trigger), we demonstrated that both specific and non-specific starts are temperature-dependent. They converge with decrease in temperature, needed for transcription factor use, in case of the switch reported for the micro RNA detection.

This led us to design our own switches, optimized to work at 37 °C through oligonucleotide sequence adjustment. We demonstrated how we varied the pT concentration to achieve bistability. This was complemented by the introduction of our protein sensing template (psT) concept. We demonstrated the concept of the mechanism of action dependent on a transcription factor, such as LacI. A particular LacI-specific design was shown as well as an experiment demonstrating that increasing concentrations of LacI interfere with psT-driven start of the circuit. The concentration of LacI required was in line with expectation. We also introduced the workflow of time-fluorescence trace data processing used throughout the text to produce easily interpretable  $1/C_q$  values.

Specific detection of a range of IPTG, the LacI ligand was demonstrated confirming our hypothesis that it will affect the extend of the transcription factor interference. We followed by successfully porting the approach to TrpR – a transcription factor with logic of the response inverted to that of LacI. A concept of protein sensing killer templates (pskTs) was introduced and successfully implemented to invert native response of the transcription factors used before.

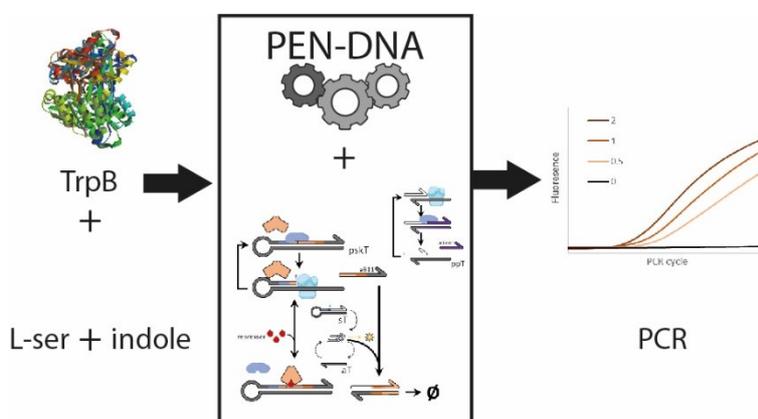
Finally, we demonstrated two small molecule inputs processing through an analogue adder. The DNA processing layer was used to implement the four combinations possible, when each input is relayed either directly on in inverted fashion. We also observed that the shape of the response can be changed whilst the relative weight of the inputs in summation can be shifted.



## Chapter 2 Towards evolving small-molecule enzymes *in emusio*.

### 2.1 Aims

In this chapter, we aim to show that the sensing method demonstrated before can be coupled with a biologically relevant chemical output rather than fluorescence. In particular, we use integration into directed evolution method, previously developed within our group: *in vitro* PEN-CSR as an example. We will hence start by briefly presenting the method and the considerations. We will follow by demonstration of one-pot L-trp synthesis and detection, followed by incorporation of gene-amplification output via PCR (Figure 59). The latter will include some new considerations we discovered. We will conclude by preliminary results, suggesting our circuit could be used in the microfluidic setting of PEN-CSR.



**Figure 59 Graphical abstract of chapter 2.** TrpB and its reactants are added to a molecular program. The activity is detected and translated into PCR gene amplification.

## 2.2 Introduction

In this chapter we aimed to expand on the directed evolution pipeline development carried out in parallel with us by Rémi Sieskind of our group. His work was recently presented in his thesis <sup>182</sup>. Therefore, we would like to outline the concept of compartmentalized self-replication followed by the particular details of the method developed in our group (the *in vitro* PEN-CSR) and the limitations of the methods we sought to overcome.

### 2.2.1 Compartmentalised self-replication

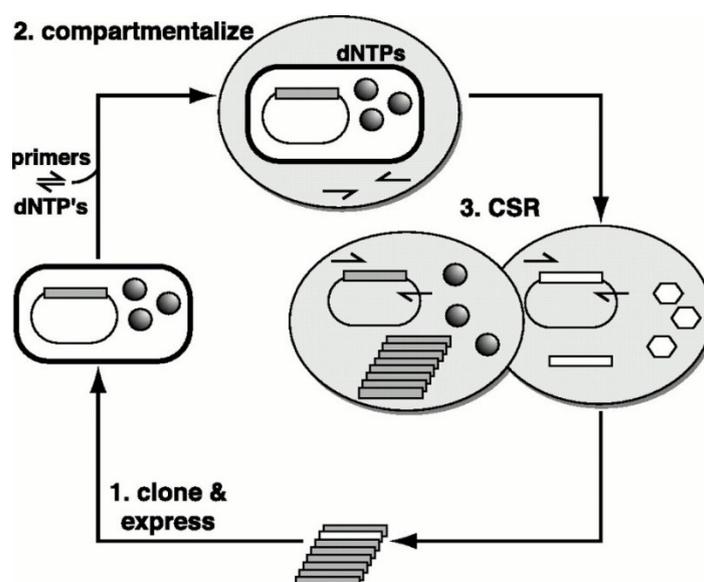
Compartmentalised self-replication (CSR) is a concept pioneered by Philipp Holliger <sup>183</sup>. It is a directed evolution method with a powerful selection tool compatible with thermostable DNA polymerases. It works as follows (**Figure 60**):

1. A mutagenic library of PCR-compatible DNA polymerase constitutively expressed is transformed into *E. coli*.
2. The bacteria are distributed into microdroplets, with buffer, dNTPs and primers suitable for amplification of the polymerase gene transformed.
3. Microdroplets are subjected to PCR thermocycling. Cells burst, releasing the DNA polymerase expressed and the corresponding plasmid, acting as a template for the PCR.
4. Droplets are disrupted and PCR products are pooled for evaluation by sequencing or another round of selection. The DNA library is enriched in sequences, coding for the more successful enzymes.

This is a powerful approach that proved invaluable for development of better variants of Taq polymerase. It allows an assay of individual mutant's enzymatic activity through compartmentalisation. At the same time, the throughput of the selection is not limited by a particular scale as the library is examined in parallel rather than sequentially. This is due to the concept of self-selection: each droplet in itself both assays the fitness of the particular variant and replicates the coding gene in response to the fitness. This process is governed by a selection function that can vary.

It is also important to highlight that this is possible due to the maintenance of genotype-phenotype linkage. A particular freshly transformed bacterium carries both the unique gene and the corresponding enzyme it conveniently expresses within. Once compartmentalised, even after lysis these remain associated and separated from other DNA-protein variant pairs during the selection stage.

Despite the breakthrough aspects of the method, it is evidently limited in its scope to the DNA polymerases that can be expressed in *Escherichia coli*. Therefore, our group aimed to develop PEN-CSR, an alternative, suitable for a broader scope of enzymes.

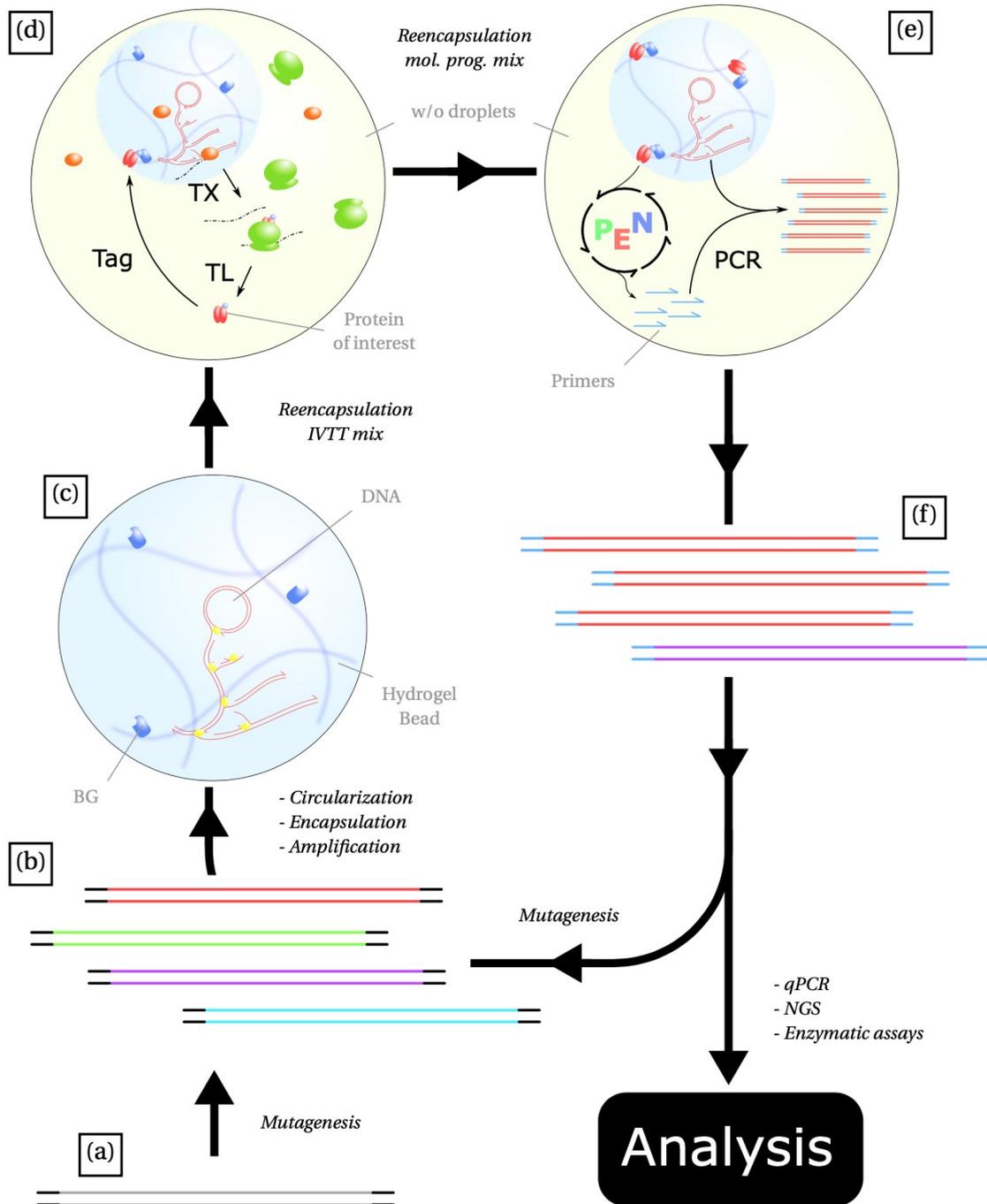


**Figure 60 Original CSR workflow.** As in<sup>183</sup>. (1) Thermostable polymerase library is generated and transformed into *E. coli*. (2) Bacteria are encapsulated with PCR reagents bar the polymerase, with primers flanking the polymerase gene within bacteria. (3) During PCR cycling cells burst, releasing the polymerase variants expressed. Better variants are able to replicated their own genes better, therefore enriching the overall library.

## 2.2.2 *In vitro* PEN Compartmentalized Self-Replication

*In vitro* PEN Compartmentalized Self-Replication (PEN-CSR) is a method, recently presented by Rémi Sieskind<sup>182</sup> of our group. It is a development from a *semi-in vitro* PEN-CSR method developed by Adèle Dramé-Maigné<sup>184</sup>, in turn inspired by the original CSR. Dramé-Maigné sought to broaden the scope of the enzymes that can be assayed by CSR by interfacing enzymatic activity with the PCR output through primer production controlled by a molecular program at the selection stage. Otherwise, the workflow was similar to the original CSR. However, when working with a nickase as a proof of principle target, she ran into problems related to the toxicity of the enzyme expressed towards the host. Presumably, this was due to the host DNA damage.

This presents an example of the directed evolution target toxicity problem we alluded to in 0.3.2. Furthermore, use of living hosts poses other problems. It may introduce an undetectable selection bias, for example, associated with how well a protein variant is expressed in the host or how well such host grows. In context of the nickases, one might easily imagine that in a freshly transformed bacterial population, those clones that are not burdened by an efficient nickase damaging their genome will be at an advantage. Therefore, they will rise in numbers at the pre-selection stage, counteracting the following selection. Another aspect is the environment one wishes to optimise the protein for: it may be incompatible with host components. In the case of Dramé-Maigné, bacteria were observed to release native exonucleases that partially disrupted the molecular program performing selection.



**Figure 61 *In vitro* PEN-CSR scheme.** As reported and described by Sieskind<sup>182</sup>. (a) The initial gene is extracted from its host organism or ordered from a gene synthesis company. (b) The gene is mutated by error-prone PCR (epPCR) to form a library. (c) Each mutant is then circularized and encapsulated with the precursors of an hydrogel and the reagents necessary for a Rolling-Circle Amplification (RCA). The droplets are incubated so that the gel can form and the RCA can occur. (d) After emulsion break and washes, the newly formed beads containing the amplified DNA is re-encapsulated with an IVTT mixture. The SNAP-tagged mutant proteins are expressed and then bind the benzylguanine (BG) tag substrate attached to the beads. (e) After emulsion break and washes, the beads displaying the proteins are then re-encapsulated in the molecular program mixture that sense the activity and convert it into an amount of primers. A PCR step follows, the efficacy of which depends on the amount of primers produced. (f) After emulsion break, the final library, enriched in the best mutants, is retrieved.

Therefore monumental work has been carried out by Sieskind to bring together an *in vitro* microfluidics pipeline for directed evolution of proteins<sup>82</sup>. As summarised by him in **Figure 61**, it starts with library creation through error prone PCR we mentioned in the introduction (0.3.2). The output is a library with more successful genes enriched.

For display, the genes on an expression plasmid (see 2.2.3.1) are first encapsulated in microemulsion, aiming to contain 1 variant per droplet or less. As single copy of a gene is insufficient for reasonable expression, the plasmid is amplified through rolling circle amplification (as in **Figure 8**). In parallel, the concatenated product is immobilised throughout a hydrogel simultaneously formed. We may refer to the hydrogel beads as “THA beads” as they are made out of thiolated hyaluronic acid. DNA, now amplified and on a solid support can be re-encapsulated in *in vitro* transcription translation mixture (IVTT) (see 2.2.4) that allows protein expression. The protein is likewise immobilised on the bead, via a use of a protein tag (see 2.2.3.1.1). The result is a solid support with both a particular gene and the corresponding protein variant both attached. The genotype and the protein, capable of representing a phenotype are therefore linked, similar to living organisms. The bead is then finally re-encapsulated again into a PEN-based molecular program capable of selection.

For selection, desired protein activity is taken as a positive input, whilst undesired can be penalised (as shown in 1.5). Some form of computation is carried out and primers are produced as an output (see 2.2.5). At a chosen timepoint, determined through fluorescence monitoring, the reaction temperature regime is switched to PCR cycling. At this point all conditions for PCR are satisfied, with primers being the limiting factor. Therefore genes in emulsion droplets are differentially amplified, dependent on primer availability. As just described, this is dependent on protein activity, in turn linked to the particular gene sequence coding for it. Therefore, desired genes are amplified, whilst undesired are not, according to the selection criteria. The resulting library can then be examined and/or taken forward for another round of the process.

Whilst promising, so far this *in vitro* process has been applied to polymerases and nickases, that are naturally present in the PEN toolbox. The reader can easily envisage that a more processive or higher fidelity polymerase will be more efficient at running the molecular program at a desired temperature. This will in turn lead to earlier switching and primer generation. If thermostable it will also contribute to the PCR.

The nickases, being an integral part of the circuit, likewise allow to select for cleavage rate. Similarly, environmental conditions can be changed for the corresponding adaptation. Most interesting, however, is the potential to alter the specificity for different recognition sequences.

However, a large number of enzymes does not belong to either of the classes of enzymes and produce reasonable small molecule chemicals instead. The case we presented throughout this thesis is the TrpB. We therefore thought that incorporation of our L-trp sensing interface into *in vitro* PEN-CSR would serve as a proof-of-principle, that such TF-based interfaces can in general be deployed.

This adjustment would only concern the selection stage of the process, whilst the remainder of the *in vitro* PEN-CSR, notable the display step would stay unchanged. We therefore specifically focused on immobilising the enzyme on the beads and the subsequent selection step as declared in (2.1).

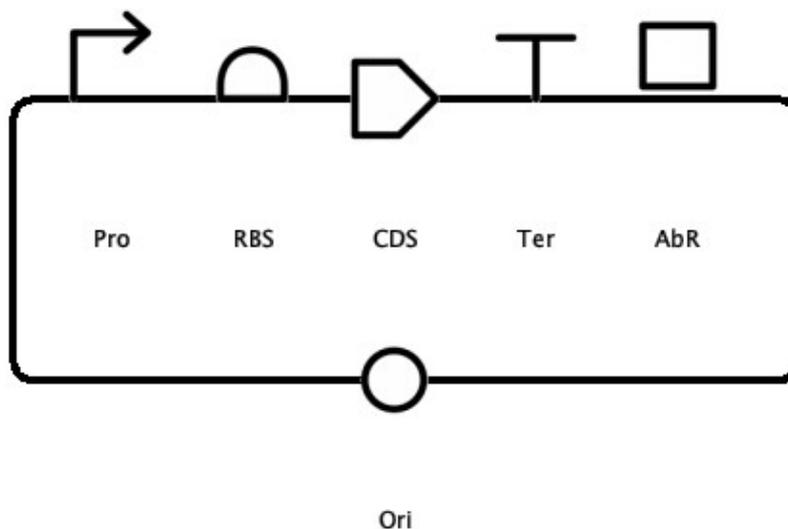
### 2.2.3 *In vitro* PEN CSR components

Before we proceed to describing the experiments, we will introduce and explain some of the molecular components that have been used in this work and alluded to above. In particular we will introduce: the concept of plasmids, that are usually used for protein production, including protein tags, used for their immobilisation; *in vitro* transcription translation (IVTT) system used for protein generation within *in vitro* PEN-CSR; and primer producing templates, responsible for primer generation within the molecular program, hence facilitating PCR.

#### 2.2.3.1 Plasmids

Plasmids are small (typically 1 – 200 kbp) dsDNA molecules found in bacteria<sup>185</sup>. They are capable of replication and carry genetic information yet are independent of bacterial chromosomes. An artificially constructed plasmid of academic use typically consists of: an origin of replication (ori), driving autonomous replication; antibiotic resistance cassette (AbR), enabling resistance to a marker antibiotic for selection purposes; as well as some cassette of interest, flanked by functional regions<sup>186,187</sup>. For example, if one wishes to produce a protein, the optimised coding sequence (CDS) will be preceded by: a promoter, driving transcription to produce the mRNA<sup>50</sup> and a ribosomal binding site (RBS), driving translation<sup>188</sup>. The CDS can be followed by sequences coding for potentially useful protein tags, if desired, followed by a stop codon, terminating translation<sup>189</sup> and transcriptional terminators<sup>190</sup> (**Figure 62**).

The protein of interest can be expressed constitutively. However, both natural and artificial systems usually control rate of protein synthesis by regulating both transcription and translation, as we have mentioned (0.2.6)<sup>191,192</sup>. We would note that we used such approach during protein production. It allows bacterial cells to grow to a desired density before introducing the metabolic load of protein overexpression. Once protein production is induced, bacteria tend to divert a lot of resources to the task leading to growth slowing down<sup>193</sup>. Interestingly, one of the systems we used was regulated by LacI and IPTG, highlighting their ubiquity.



**Figure 62 Generic plasmid map.** Annotations according to SBOLVisual 1.0.0 standard glyphs. The plasmid contains an origin of replication (ori), antibiotic resistance cassette (AbR), promoter (Pro), ribosomal binding site (RBS), protein coding sequence (CDS), transcriptional terminator (Ter).

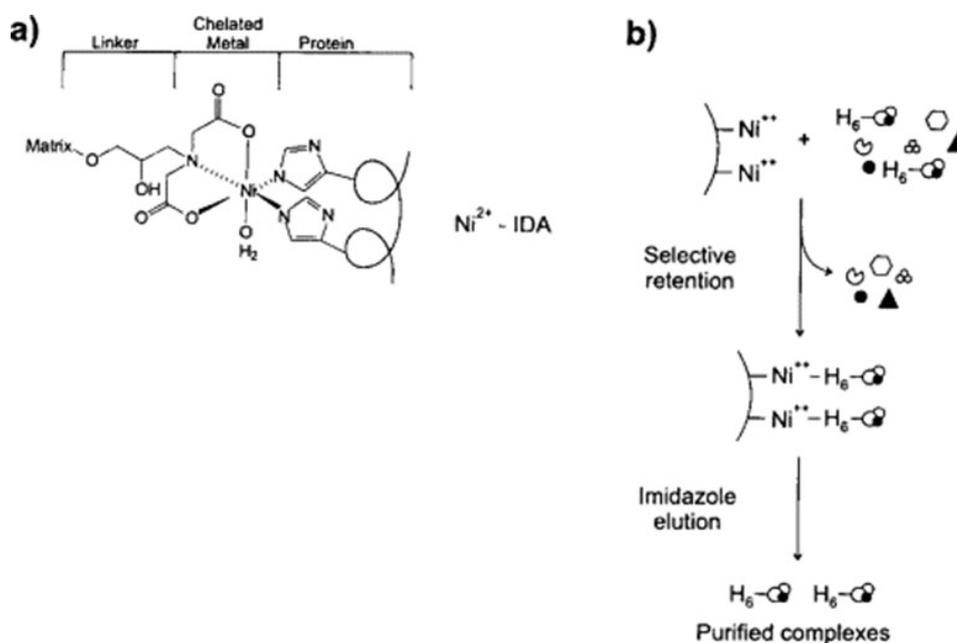
In this work plasmids were used in two contexts. First, they were transformed into *E. coli* to purify large quantities of proteins necessary in bulk (or the plasmid itself). This was convenient, as bacteria rapidly multiply, thus multiplying the internal plasmid/amount of protein collectively expressed. Second and more important, they served as mock templates in the relevant *in vitro* PEN-CSR steps. In particular, protein expression in the IVTT and PCR amplification.

### 2.2.3.1.1 Protein Tags

Protein tags are functional sequences typically added in a modular manner at the beginning or the end of the main protein coding sequence. On the DNA-level they are usually carried by a premade vector like the one we have just described, hence targeted addition of the coding sequence of interest into a plasmid results in a combination of two sequences within the same transcription and translation frames.

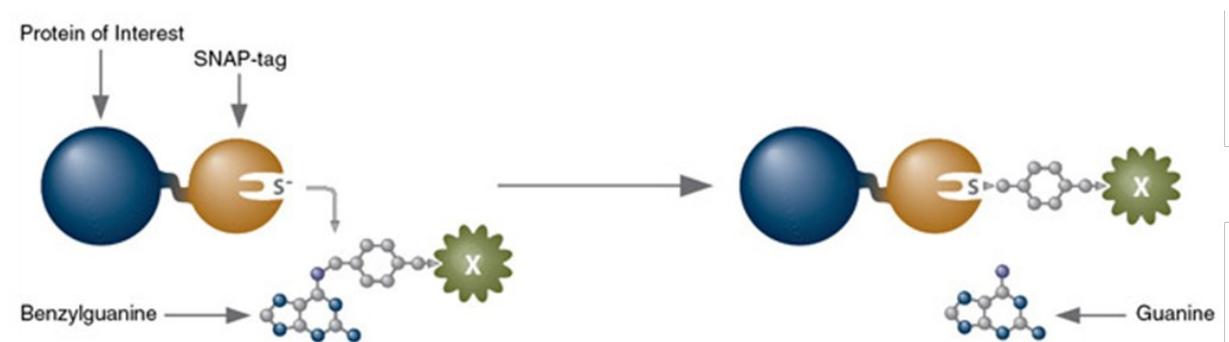
On the protein level, these are engineered to have predetermined functions expected to be maintained in fusion with the protein of interest. Of interest to us are two such tags the His-tag and the SNAP tag, both used for protein immobilisation, typically, in protein purification context.

The His-tag was added to all proteins used in this work for purification on nickel gravity flow columns. Briefly, 6 histidine residues at the end of the protein sequence coordinate the divalent metal ions, immobilised on the column matrix (**Figure 63a**). Thus, the protein is retained, whilst the rest of the cell lysate flows through. The protein is then released by adding a competitor molecule – imidazole. Hence, pure protein is obtained (**Figure 63b**). This tag consists of just 6 residues and is routinely used so we had little reservations about it.



**Figure 63** The His-tag purification process. Adapted from <sup>194</sup>. (a)  $\text{Ni}^{2+}$  adsorption centre interacts with histidine residues on the protein. (b) Schematic of the purification. Protein mixture is run along the matrix with tagged proteins being retained. Upon imidazole addition, the protein of interest is released.

SNAP-tag, based on the O<sup>6</sup>-alkylguanine-DNA alkyltransferase (hAGT)<sup>195</sup>, on the other hand, covalently binds to benzylguanine substrate, when intact (**Figure 64**). This is useful to ensure robust protein immobilisation on the hydrogel beads used in PEN-CSR as the substrate is easy to incorporate into the matrix. However, due to its mammalian origins it consists of 182 residues – a considerable amount. This can disrupt the folding of the main protein during translation. Similarly, SNAP tag can sometimes interfere with catalytic activity, as observed by Sieskind. As demonstrated below, this was taken into consideration.



**Figure 64 SNAP-tag mode of action schematic.** Adapted from<sup>96</sup>. SNAP-tag reacts with benzylguanine to covalently attach to the bead (X) and release guanine. Protein of interest is therefore also immobilised.

## 2.2.4 *In vitro* transcription translation IVTT

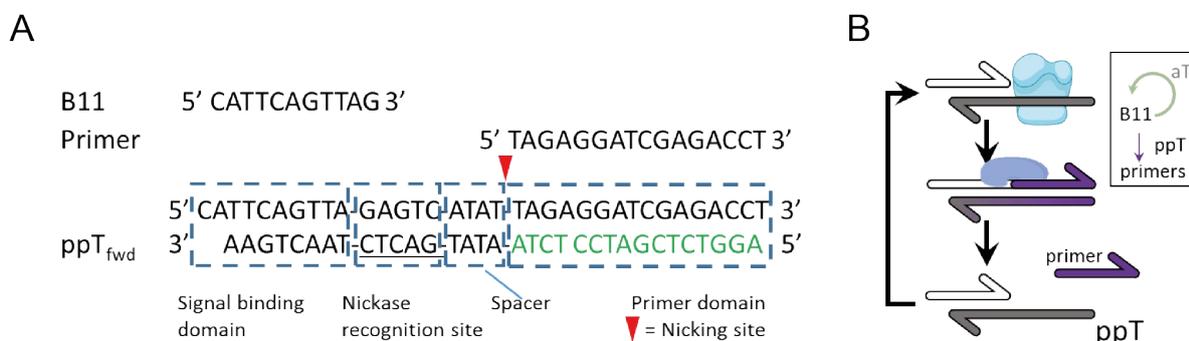
*In vitro* transcription translation (IVTT), as the name suggests is an approach, allowing protein expression outside of a living organism (a host). Several mixtures were proposed to simulate different environments. We used the *E. coli* derived one, optimised in our lab by Sieskind and Caschera and reported in the methods. As mentioned, the main motivation of avoiding living hosts are the hurdles associated with them: the protein evolved may be toxic to the host impeding expression or imposing undesired selection penalties. Also, evolution of the protein for conditions not suitable to the host is impossible.

## 2.2.5 Primer producing templates

In this chapter we aim to have the output of the PEN-DNA circuit, instead of producing a fluorescent signal, to enable the amplification of a full-length protein-encoding gene. The goal is to obtain an enzymatic activity-to-genetic amplification circuit that can be used to build autonomous selection networks for micro-compartmentalized directed evolution protocols. We thus complemented the reporter template by two primer-producing templates (ppT). These were first used by Dramé-Maigné<sup>184</sup> in a PEN-CSR process. Ours will use the signal B11 as input, and will produce two primers targeting the gene of interest. After isothermal incubation, primers are then produced according to the timing of ON-switching of the PEN circuit (with earlier switching leading to more primer accumulation), and can be used during a subsequent PCR. The primer-producing templates are designed as follows: 3' to 5', it contains the B11 binding site, a Nt.BstNBI recognition/nicking site with a 4 base pair spacer, and then the primer encoding sequence (**Figure 65A**).

Following switch activation, B11 signal binds on ppTs and is extended. Cycle of nicking, product dissociation and polymerase extension will ensue, like previously described for pST.

The result is a linear production and accumulation of primers, conditional to switch activation (**Figure 65B**). This process is called isothermal primer amplification or IPA.



**Figure 65 Primer producing template (ppT) design.** (A) On the bottom: single-stranded ppT<sub>fwd</sub> is shown with B11 signal bound and extended. Going left to right it contains a signal binding domain, nickase recognition site, a spacer and primer-coding domain. (B) Schematic summary of primer production. Bound signal is extended, nicked and primer is released in cycles.

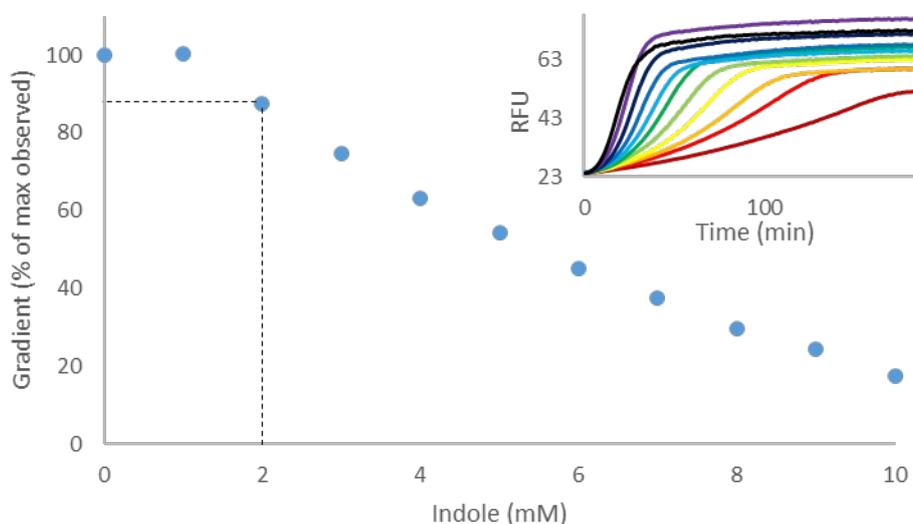
## 2.3 Results

### 2.3.1 Combining the sensor with analyte synthesis

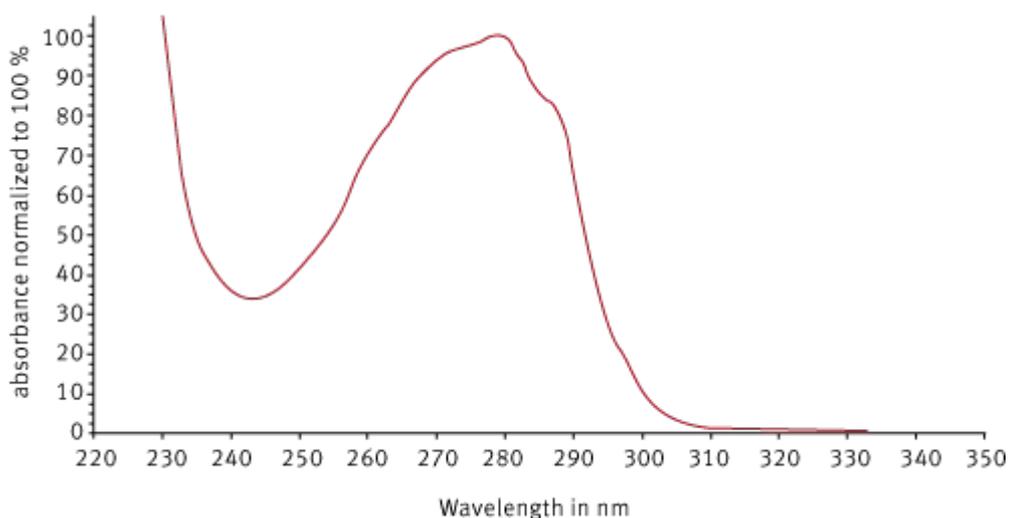
We took forward our insights from the previous chapter in hope of using a plug and play approach of simply adding an enzyme of interest – tryptophan synthase, as well as the reactants and expecting the detection to not be disrupted. However, we have quickly discovered some steps requiring consideration.

We took the assembly as before, but decreased pT to 5 nM combined with 5 nM pskT – prototype concentrations we were exploring. More importantly, we added 5 nM of sT – a rather large quantity to stimulate instant start of the switch.

We observe that indole – the substrate of TrpB, even in relatively small concentrations appeared to be disruptive for the molecular program (**Figure 66**). Unsure whether the effect was due to the solvent – dimethyl sulfoxide (DMSO) – or indole itself, we sought to minimise both. We have switched to using 2 M DMSO stock of indole, which was then diluted in water to the concentration of 0.0225 M. This is what we determined to be the solubility limit at room temperature. (Heating to 70 degrees though strongly helps with quickly re-dissolving the stock, when taken from a freezer.) We moved to using indole in concentrations below 2 mM, when with molecular program. The figure below demonstrates only 10 % loss in gradient slope at this concentration, yet if at least 50% of substrate is converted into L-trp, this should give us sufficient concentration (1mM) for reasonable detection.



**Figure 66 Effect of indole on miR-based reactions.** Here identical pskT/sT-based reaction is run with the sole difference of different amount of indole added in mM amounts as indicated in the legend.



**Figure 67 Characteristic tryptophan absorbance spectrum.** This spectrum is reported by the manufacturer to be produced on a machine of the same series as the one we were using (CLARIOstar)<sup>97</sup>.

## 2.3.2 Detecting *in situ* synthesis

### 2.3.2.1 Detection with conventional methods

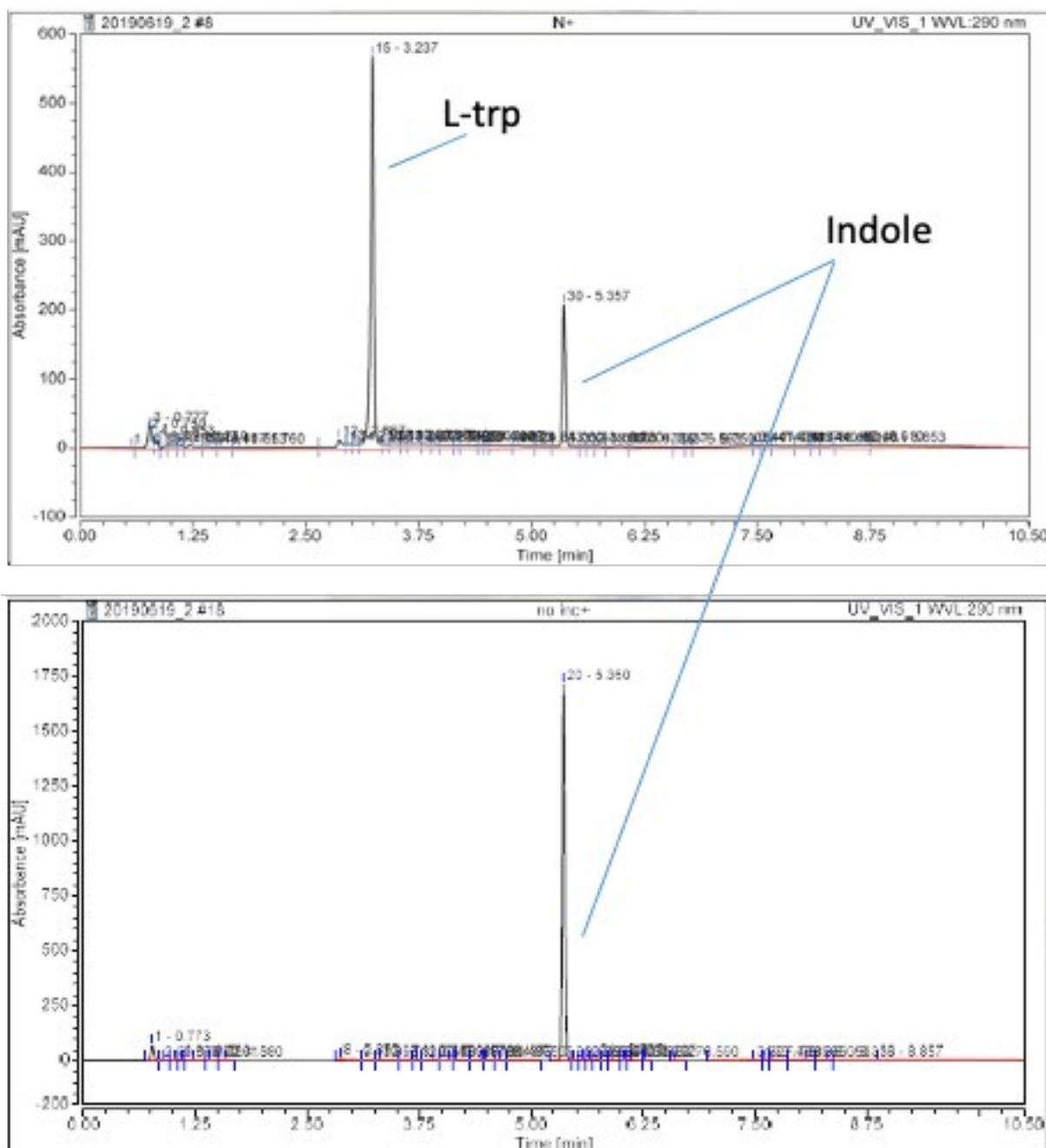
With great participation of Camille Lescanne during her internship, we have optimised the program and verified that TrpB can indeed synthesise L-trp in our buffer, provided the reactants and the co-factor (PLP) were added. We first attempted to perform the verification through absorbance tracking as described by the Arnold group<sup>131</sup>. Surprisingly, this approach failed for us. The method relies on characteristic absorbance of L-trp at 290 nm and lack of notable absorbance in water beyond 310 nm<sup>197</sup> (**Figure 67**). We therefore expected to observe

an increase only at the specified wavelength providing us with a preliminary test. However, this approach was not successful. Briefly, whenever we observed absorbance rise, it did so beyond 310 nm similarly to 290 nm wavelength and therefore we were unsure whether L-trp appearance was the cause. We present more details in Appendix B.

We therefore chose to rely on other analytical methods. Thin layer chromatography (TLC) was instrumental for proof-of-principle runs. TLC plates were run in a mobile phase consisting of 2 volumes of methanol for 1 volume of deionized water. 10mM L-trp and 3mM indole were spotted on the TLC plate for reference along with samples of interest. This allowed us to qualitatively assess, whether indole in a reaction of interest remained or was converted into L-trp with a detection limit approximated by us to be about 0.5 mM. We therefore used the method to verify that L-trp synthesis was possible in miR, setting an expectation that we could therefore detect *in situ* synthesis with our sensor. TLC was, however, difficult to record, which is why we are not reporting the results.

To verify synthesis in miR buffer with an alternative method, HPLC was carried out. This was done on the facilities and with great help of Mrs. Emmanuelle Demey from the SMBP Lab of ESPCI. For example, 2.4 mM indole, 3 mM L-serine, 40  $\mu$ M PLP and 5.6  $\mu$ M EcB were incubated overnight in 1x miR buffer and showed a dominant peak, corresponding in time to that obtained when pure L-trp is run on the column (not shown). A smaller peak is observed, corresponding similarly to indole (**Figure 68**). In contrast, a freshly prepared mixture only showed the indole peak.

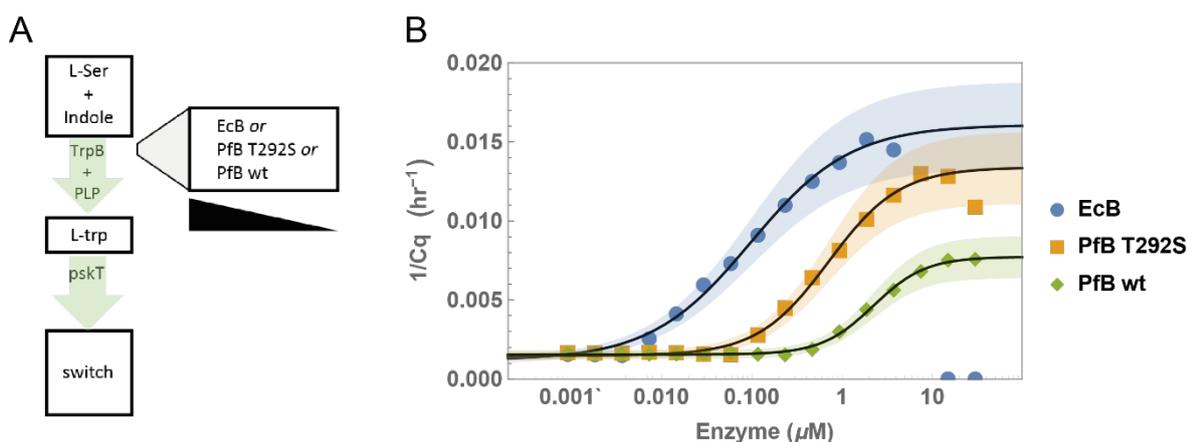
Based on results from both TLC and HPLC we concluded that our proposed L-tryptophan synthesis mixtures indeed produced the desired product, which would then cause the molecular program start.



**Figure 68** Sample chromatograms of filtered reaction mixtures. Above: incubated at 37°C overnight. Below: freshly prepared. Conditions: 2.4mM indole, 3mM L-serine, 5.6uM EcB. 3.23 min peak corresponds to L-Trp and 5.36 minutes corresponds to indole. These were obtained with 2.1 x 100 mm C-18 silica column with acetonitrile/water: 0% acetonitrile for 1 min, 0% to 100% over 5 min, 100% for 4 min and then a wash with 100% water for 10 minutes before the following sample. The enzyme was filtered out using 3 kDa filters prior to loading.

### 2.3.2.2 Detection with a molecular program

We proceeded with an attempt to detect synthesis within the molecular program assembly. This is how it would have to happen in emulsion. We moved forward to exploring detection ranges for TrpB synthase variants (mentioned in 0.3.2.1) added to our pskT<sub>TRP</sub>-based sensor. (**Figure 69**). Our usual miR assembly was used with 10 nM pskT<sub>TRP</sub>, 2 nM sT, 10 mM L-serine, 80 μM PLP, and 5 mM indole. We then added various concentration of TrpB variants we have previously mentioned.



**Figure 69 Sensing of L-trp production *in situ*.** (A) L-serine and indole are converted to L-trp by one of TrpB variants, a metabolic activity that is sensed *in situ* by the switch. (B) Switch response to L-trp production by various concentrations of EcB (circles), PfB T292S (squares), PfB wt (diamonds).

Later start times appear to correspond to lower enzyme concentrations. The enzymes align themselves in an expected pattern: EcB, then PfB T292S, then PfB being the slowest.

This appears to be in line with the reported activity: EcB is at advantage due to its mesophilic nature, improved version of PfB comes second, followed by native PfB. From here, we also establish a detection limit of about 7 nM for EcB, which could be a reasonable amount to produce in an IVTT reaction – micromolar concentrations of GFP have been produced in IVTT<sup>198</sup>.

The use of different variants also simulates different variants that could be present in a potential library being selected. To illustrate the idea let us consider a situation where we would be carrying out selection in a library containing EcB and PfB T292S with concentration of either protein in a droplet being, for example, 0.1 μM. Now, there exists a window of about 3 hours in this particular experiment, where we are only observing switch start for the EcB variant. This could be translated into primer production time, which in turn will contribute to amplification of the gene, coding for a variant that is faster at L-trp production. The gene will therefore be enriched in the total library after selection.

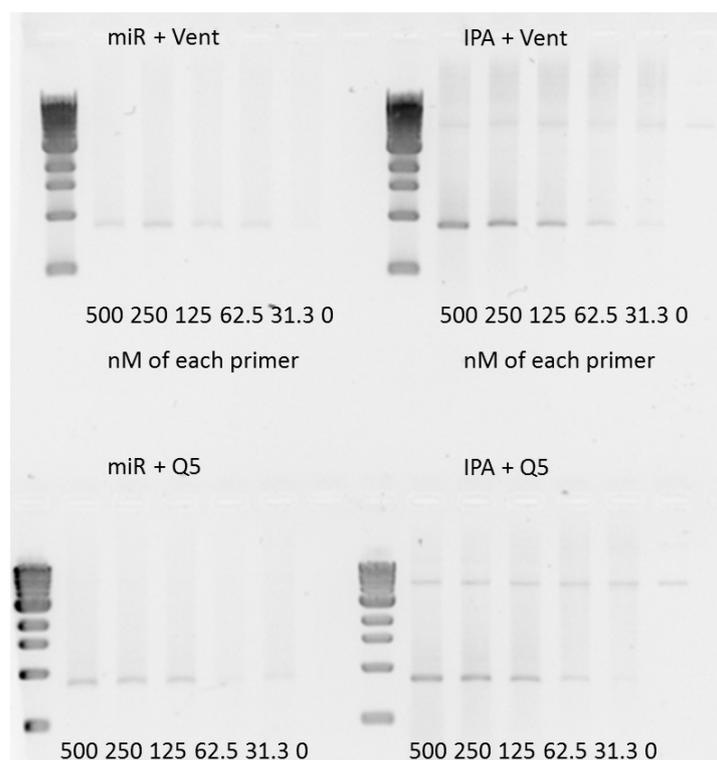
### 2.3.3 Optimising the sensor to work with IPA

Given successful sensing results yielded through considerable reaction assembly optimisation we tried adding primer producing templates (ppT) and Vent or Q5 polymerases, known not to contribute to the sensing, to the miR-based reaction in hope of PCR. This simplistic approach was not fully successful (not shown). As previously reported by Adèle Dramé-Maigné and Rémi Sieskind, the miR buffer is optimised for nickase activity and the molecular program, however, too high magnesium concentration becomes creates an issue for PCR-reactions.

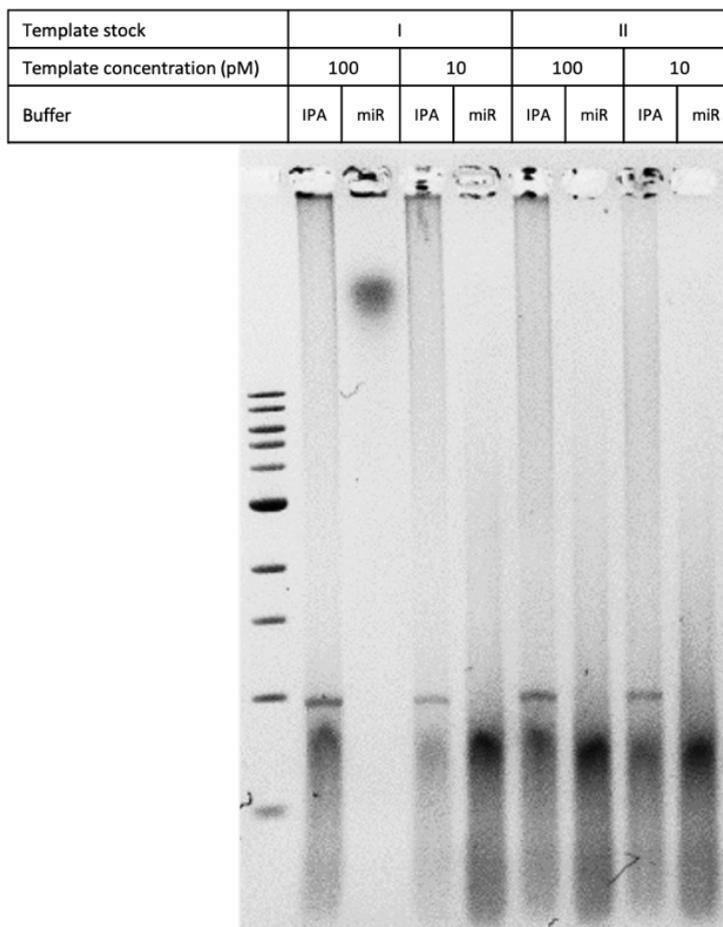
We confirmed this through a series of two experiments. In both we compared miR buffer along with the IPA buffer (see compositions in materials and methods). The latter was proposed by Rémi Sieskind and has demonstrated greater PCR potential in his work.

At first, we used high PCR template concentrations (1 nM) (**Figure 70**). In this context the IPA buffer shows a slightly more robust amplification. We disregarded this at first as it appeared to be outweighed by the problem of re-optimising the sensor for alternative conditions as well as unexplained smearing and a secondary band.

However, we then thought that 1 nM of template might not be an achievable concentration in emulsion and testing should focus on lower concentrations. IPA was at a clear advantage (**Figure 71**). We therefore have taken it the IPA buffer forward.



**Figure 70** Agarose gel of PCR products in miR and IPA buffers with Vent and Q5 polymerases. Buffer and polymerase as shown on top of a set of samples, final primer concentration shown along each set in nM. The dominant band corresponds to the desired product from 1 nM template.



**Figure 71** Agarose gel with PCR products of Vent polymerase in miR and IPA buffers with two template concentrations. Top row number shows the stock of template used, quantified as being of the same concentration. This is followed by supposed template concentration. Bottom row designates the buffer.

pR070 plasmid provided by Rémi Sieskind, containing sfGFP (see 4.3.3) was used as template in both experiments with expected product size of 900 bp. Primers used were iR337 and iR338, also reported in the methods and equivalent in sequence to the primer producing template output. 30 cycles of: (95 °C, 30s; 61 °C 15 s; 72 °C 60 s) done, along with 5 minute 95 °C heating at the start and final extension at 72 °C for 5 minutes.

### 2.3.3.1 Bulk enzyme activity selection

After some optimisation of the sensor to work in IPA buffer with suitable resolving capacity we have proceeded with an attempt to reproduce the detection of a range of enzyme concentrations followed by PCR.

The IPA assembly (see methods) was complemented with 24 nM of pT, 7 pM sT, 10 nM pskT<sub>TrpR</sub>, 5 mM L-serine, 80 µM PLP, 2.5 mM indole, 10 nM of each primer producing template and 120 pM of pR070 template. PCR cycling was carried out as before.

Our first attempt looked at a broad range of both EcB and Pfb T292S (from now on, just “Pfb” for this chapter). Although we did not observe a clear trend in start time differentiation between different enzyme concentrations (**Figure 72a**, **Figure 73a**), we still proceeded with real time following of PCR cycling to yield some encouraging results (**Figure 72b**, **Figure 73b**).

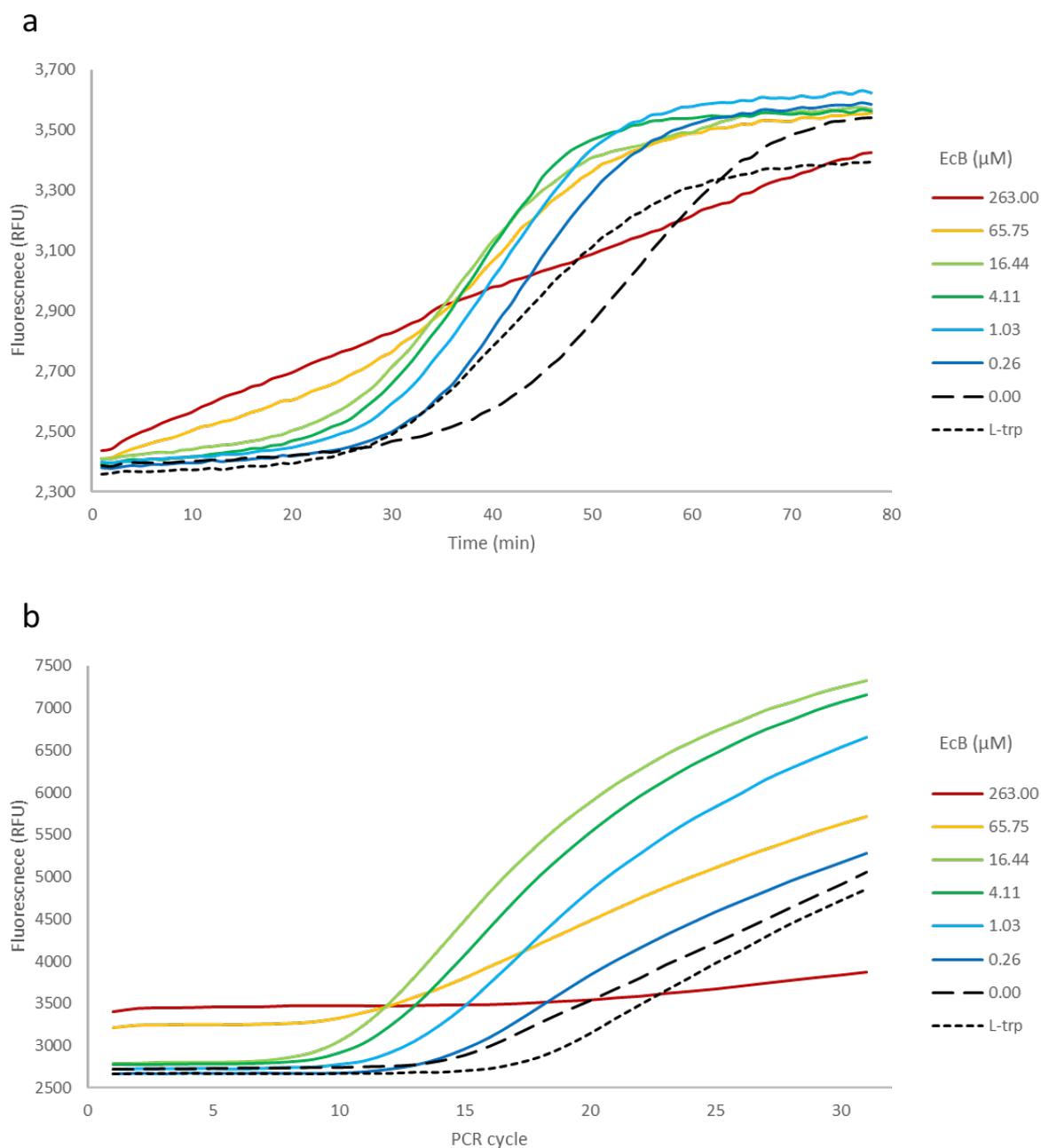
When considering EcB, it seems that large excess is detrimental to both molecular program and PCR, however concentrations below 16 µM start amplification in the correct order for both PCR and the molecular program run (**Figure 72**). It is surprising that minimal difference in start time of the molecular program results in strong differentiation of PCR traces. We suspect this might be both due to rapid primer production. Also, the production could be starting before the observed switching due to ppT activation before that of the switch: the ppTs used in this experiment share the same sequence, complimentary to the trigger as the aT. They therefore have the same probability of being activated as the aT (this is discussed in 3.3.7.1).

We can also note, whilst the negative control is somewhat in line with expectations (although it should not start with a perfect sensor), the positive control is behind the protein samples. We are not sure why, but the addition of protein storage buffer, compared to aqueous L-trp could be having an impact on the switch. For future reference, it means that only strict like to like comparisons are valid, when it comes to sensitive systems, such as the one considered here.

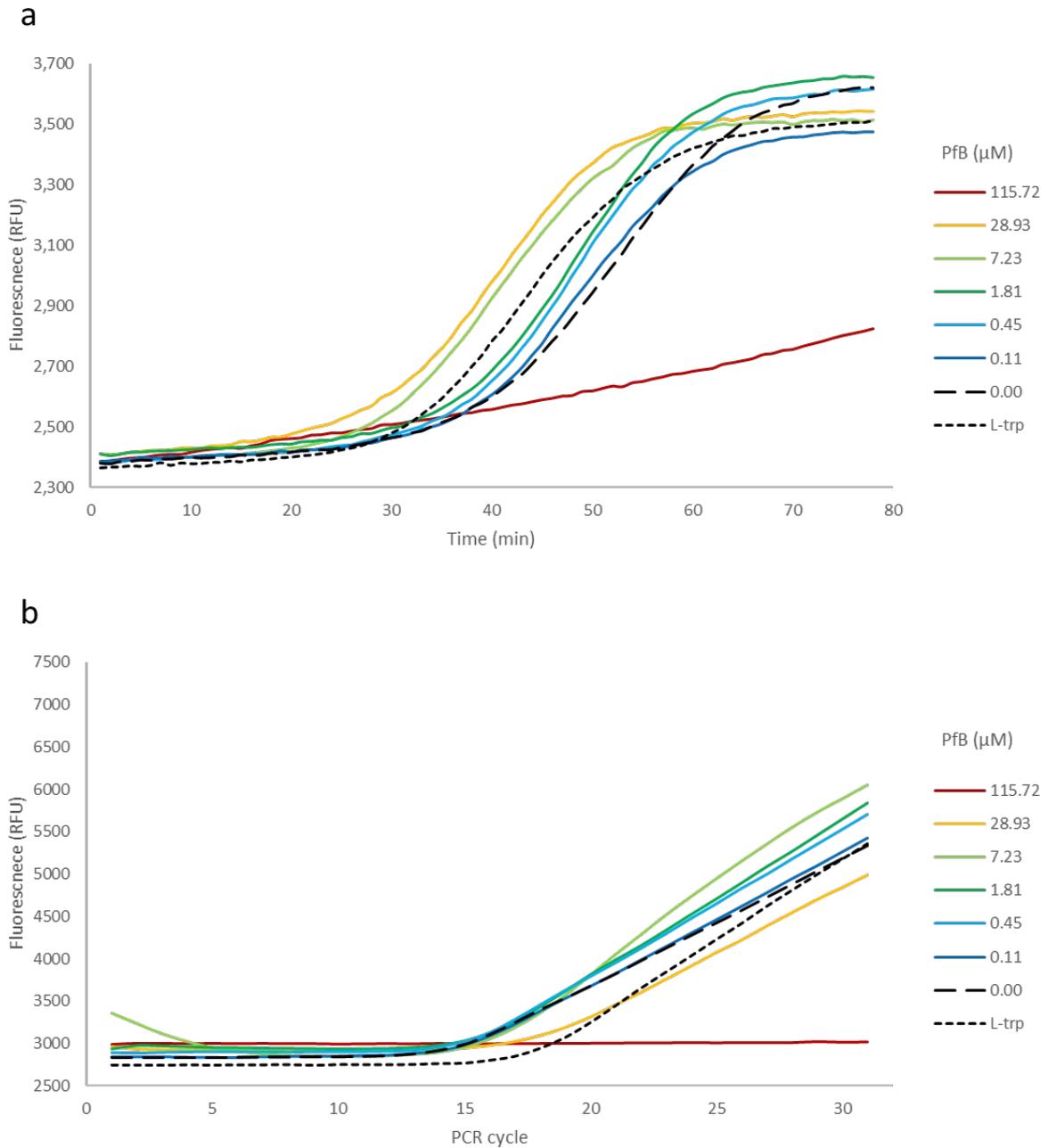
There is no PCR differentiation of Pfb, despite molecular program detection comparable to EcB in trends (**Figure 73**). The investigation will be described later in (2.3.3.2).

Gel electrophoresis revealed apparent lack of the desired product. The size expected was around 900 bp in this case for a mock de-GFP fragment flanked by standard primer landing sites (**Figure 74**).

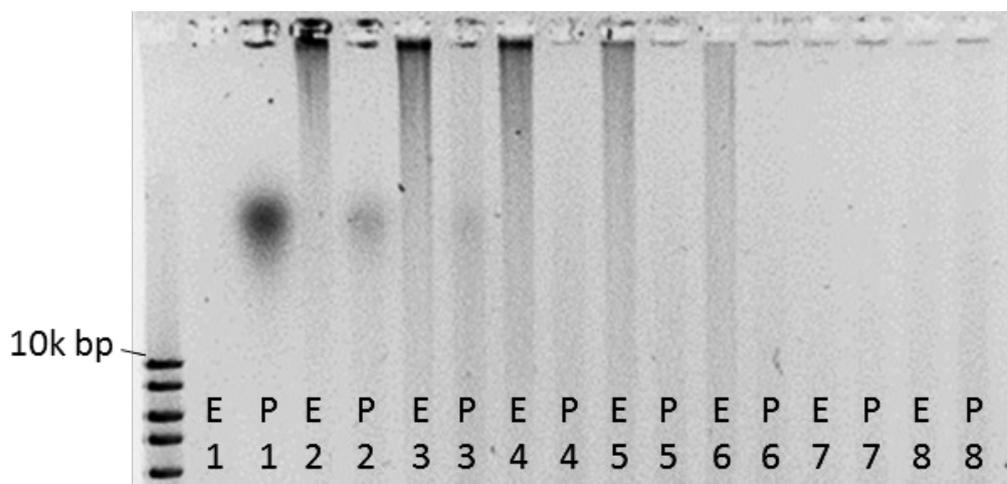
Higher enzyme concentrations were apparent on the gel in different forms. We speculate that this could be some form of product concatenation or aggregation, possible together with the protein added. The DNA could in principle be retrieved by nested PCR in case of EcB as it showed a reasonable PCR trace in this experiment (not shown).



**Figure 72** Detection and amplification of various concentrations of EcB in IPA. (a) Fluorescent traces of molecular program for two enzymes. (b): Real time fluorescent traces of PCR amplification of the samples above. Coloured legend represents concentration of the enzyme in  $\mu\text{M}$  shared for two plots. Positive control, containing 1 mM L-trp is shown as short dashed line.



**Figure 73** Detection and amplification of various concentrations of PfB in IPA. (a) Fluorescent traces of molecular program for two enzymes. (b) Real time fluorescent traces of PCR amplification of the samples above. Coloured legend represents concentration of the enzyme in  $\mu\text{M}$  shared for two plots. Positive control, containing 1 mM L-trp is shown as short dashed line.

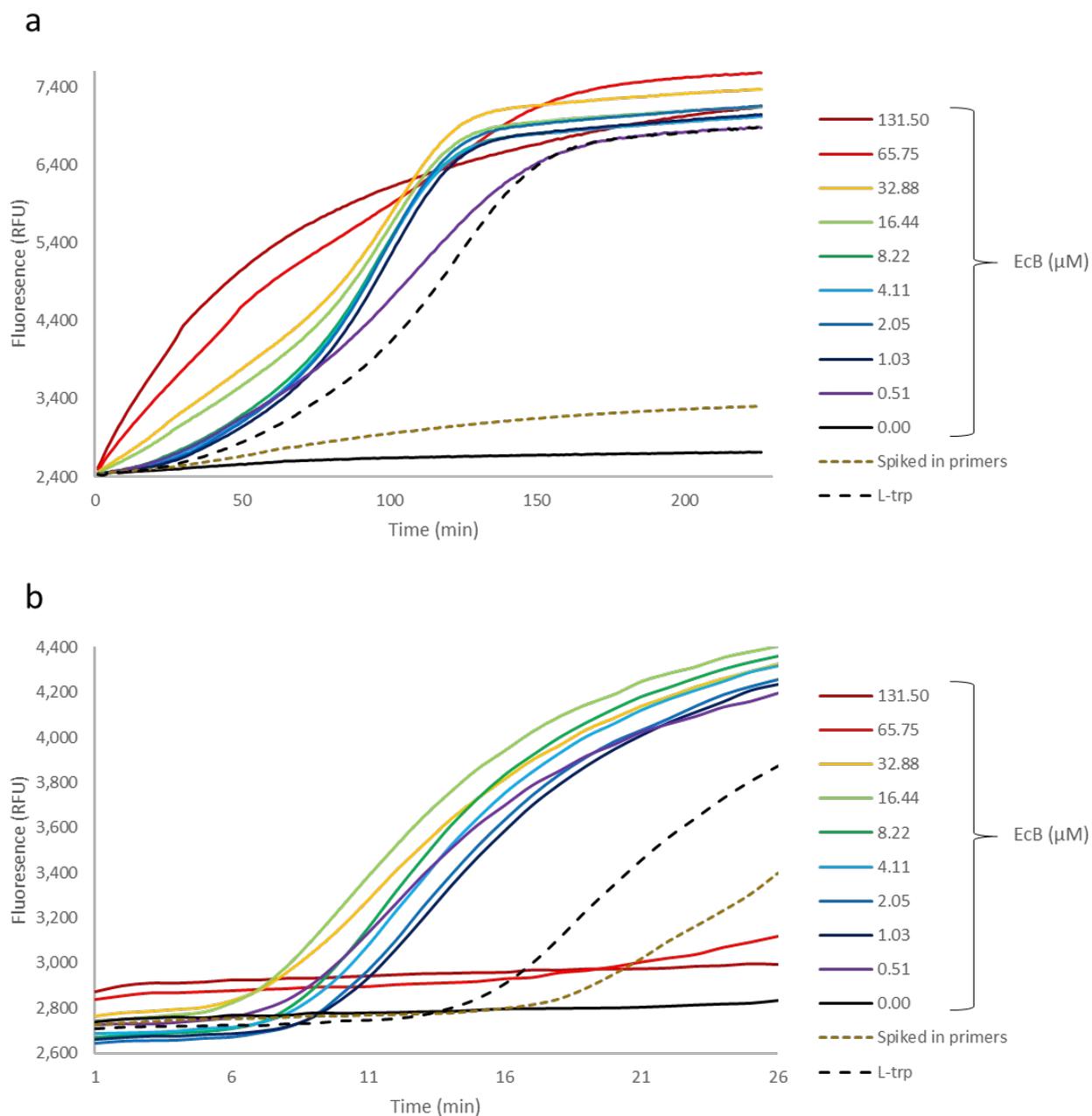


**Figure 74** Agarose gel of PCR products retrieved from reactions shown in **Figure 72**, **Figure 73**. Samples are loaded in an alternating pattern in the same order: **Figure 72** is E for EcB sample, **Figure 73** is for P – PfB sample.

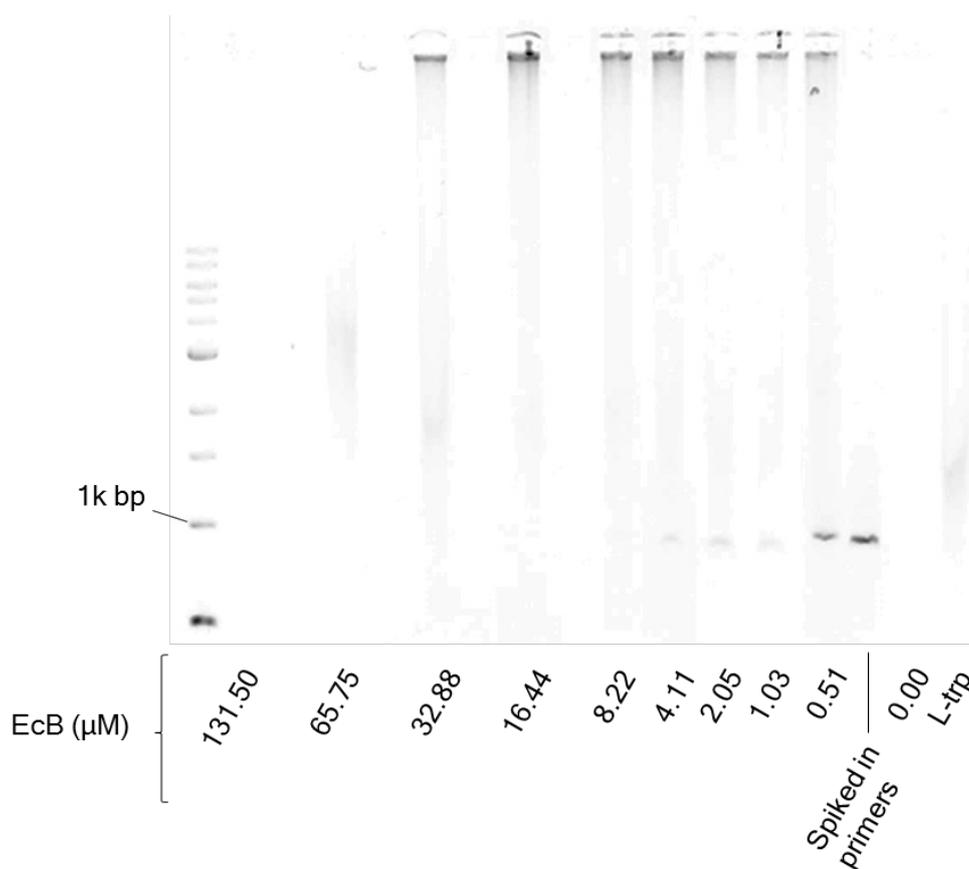
We have repeated the experiment with EcB with a finer range of concentrations and 60 pM of template (**Figure 75**). The results observed confirm: 1. Excess of enzyme added indeed disrupts the circuit, producing an atypical fluorescence curve. Switch-like behaviour is lost and is replaced by a non-sigmoid curve, saturating when all rT is opened. 2. At lower EcB concentrations switching occurs in a non-resolved manner with PCR traces corresponding.

We also observe earlier start of the PCR for protein-containing sample than positive controls containing primers, manually added. It could be that *in situ* produced primers exceed those added manually here (500 nM).

We analysed the PCR products on an agarose gel (**Figure 76**). Compared to previous results (**Figure 74**), here the expected product of 900bp is apparent. Whilst recommended amount of primers gives the sharpest band, similar ones are visible for the 4 lowest concentrations of the enzyme. Moreover, band intensity increases, as the enzyme concentration drops from 4.11 to 0.51  $\mu\text{M}$ . At higher concentrations and for 1 mM L-trp sample that provided a qPCR trace only diffuse smears are visible in the middle of the lanes as well as extending from the top. We are not sure of the cause but speculate this could be once again related to concatenation caused by primer overload. Where overall protein concentration is excessive, no product is visible in qPCR trace and on the gel.



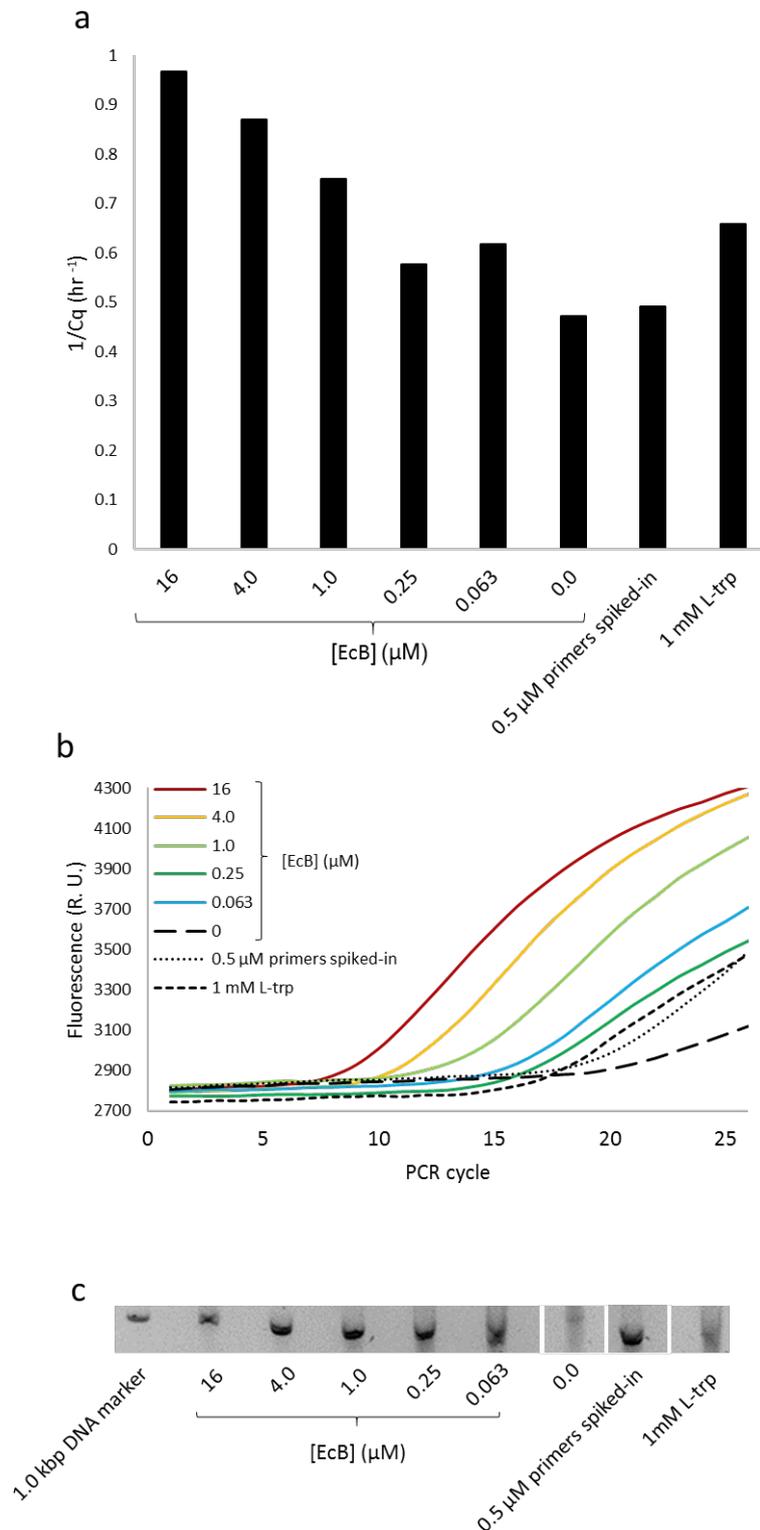
**Figure 75** Detection and amplification of a fine range of concentrations of EcB in IPA. (a) Fluorescent traces of molecular program for two enzymes. (b): Real time fluorescent traces of PCR amplification of the samples above. Coloured legend represents concentration of the enzyme in  $\mu\text{M}$  shared for two plots. Positive control, containing 1 mM L-trp is shown as short dashed line.



**Figure 76** Agarose gel of PCR products retrieved from reaction shown in Figure 75. Samples are loaded as indicated at the bottom. Some were rearranged digitally for ease of interpretation. “Spiked in primers” represents a control with 0.5  $\mu\text{M}$  primers spiked-in before PCR and “L-trp” – 1 mM of L-trp added instead of the enzyme.

We adjusted our range of EcB to confirm our results (**Figure 77a-b**) but also achieve more satisfactory bands on the gel as seen in **Figure 77c**: all of the fragments match the expected size of 900 bp, the high molecular weight smearing effect diminishes with decrease in enzyme concentration and qualitatively the intensity of the band increases. Higher amount of 900 bp DNA could explain a slight downward shift on the gel. As expected, primers added at the end of the run to a negative control result in a sharp band. The negative control without primers added also presents a band. This can be explained by the fact that the switch was also activated, although late, some primers could have been produced.

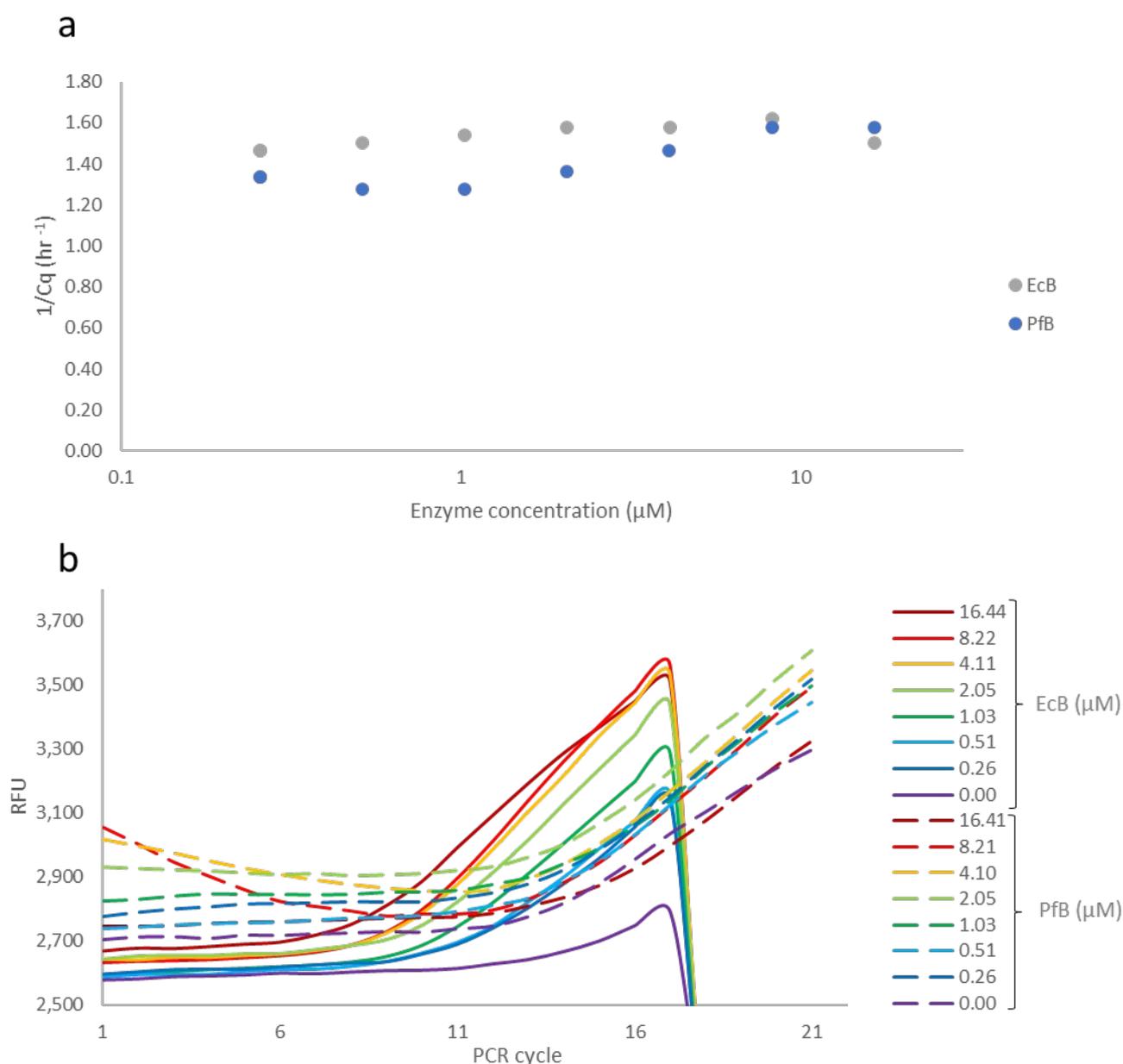
We believe our results demonstrate that PEN environment is suitable for enzyme activity selection. Previously used molecular programs can be adapted to be compatible with the enzymatic activity. This activity is sensed, processed and translated into production of primers. Under appropriate conditions, these in turn, result in gene amplification, representative of the efficiency of the enzyme at producing the small molecule detected by the circuit. Hence, here we have constructed a circuit suitable for selection, that could be incorporated into *in vitro* PEN-CSR.



**Figure 77 EcB activity detection and gene amplification response.** Samples consist of various concentrations of EcB, control with 0.5 μM primers spiked-in before PCR and 1 mM of L-trp added instead of the enzyme. **(a)** 1/Cq extracted during molecular program run. **(b)** real-time PCR amplification trace of PCR launched 40 min after first complete start in **a**. **(c)** Section of agarose gel of the PCR products retrieved from reactions shown in **a-b** along with a 1.0 kbp DNA ladder. "0.0" and "0.5 μM primers spiked-in" lanes were switched around horizontally through graphical editing for ease on interpretation. All samples were run on the same gel.

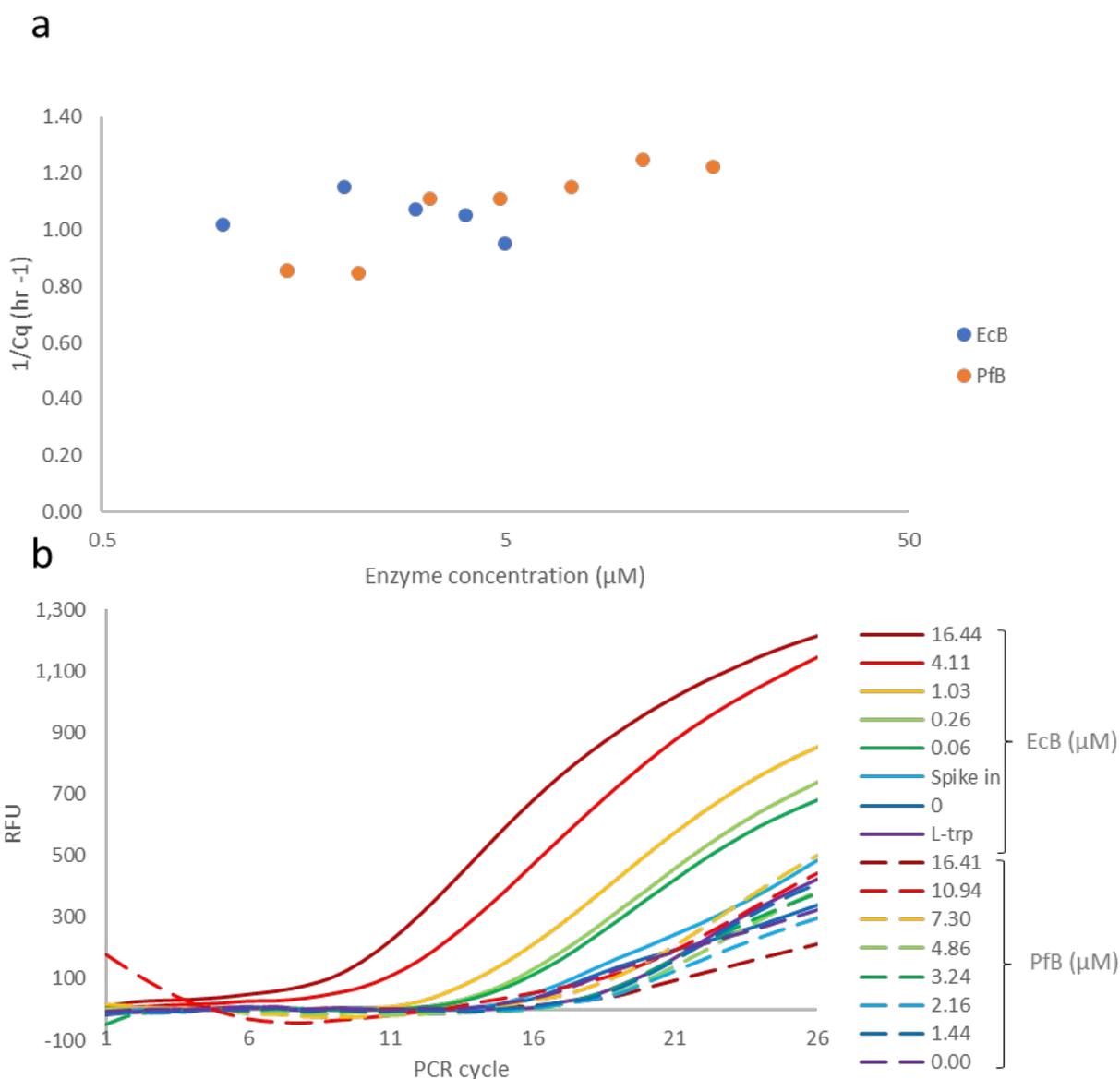
### 2.3.3.2 Tryptophan synthase from *Pyrococcus furiosus* strongly binds DNA

Over a series of preliminary experiments, we have noticed that despite side by side overlapping performance of PfB and EcB regarding catalytic activity, DNA amplification consistently failed for the former. For example, **Figure 78a** shows seemingly similar switching for both enzymes with potential acceleration at higher concentrations. Although the highest concentration could be considered toxic. However, when moving to PCR cycling (**Figure 78b**), EcB range appears to be resolved in curve steepness and starts the amplification earlier. PfB start is indiscriminatory, later and leads to non-specific DNA amplification.



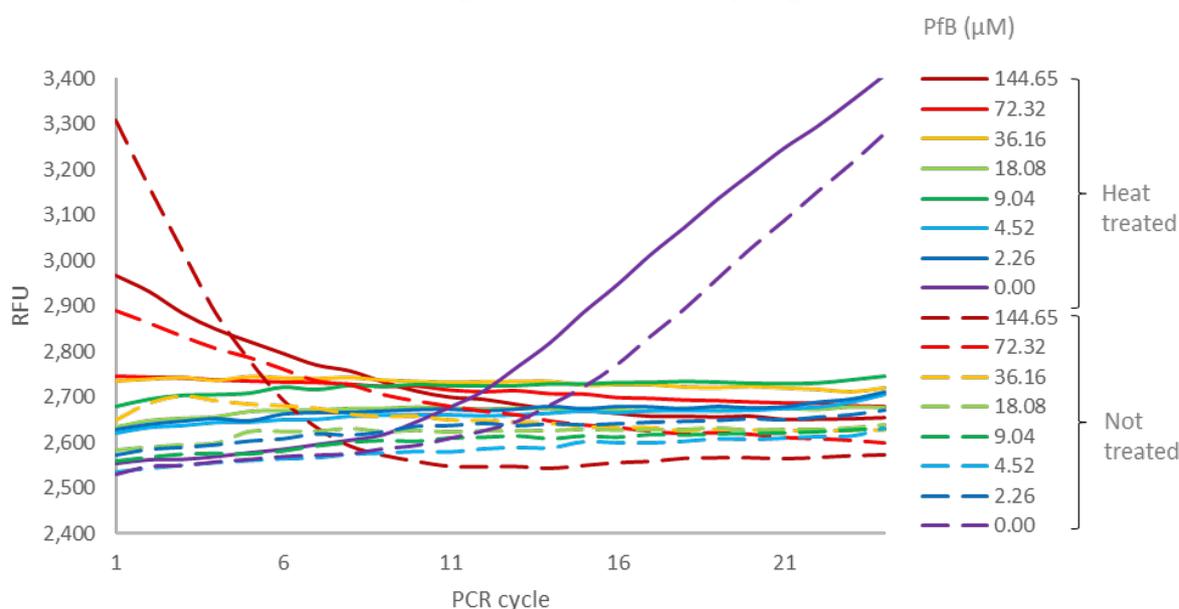
**Figure 78 PEN enzyme activity detection and PCR amplification.** EcB and PfB variants were assayed in IPA buffer with enzyme concentrations shown at 8 nM pT in IPA. **(a)** Switch activation times for the two enzymes. **(b)** PCR fluorescent traces of samples shown in **a**.

We tried to improve the resolution by increasing the amount of pT from 8 nM to 15 nM. This was expected to slow down the switch but improve the resolution. Surprisingly, there was little effect on the switching times (**Figure 79a**). However, PCR of EcB was better resolved. Interestingly, it performed much better than the samples with spiked-in primers or 1 mM of L-trp, presumably, due to overproduction of both in the enzyme-containing mixes (**Figure 79a**). Nonetheless, PfB amplification still failed. This lead us to suspect that it somehow interferes with the PCR step.



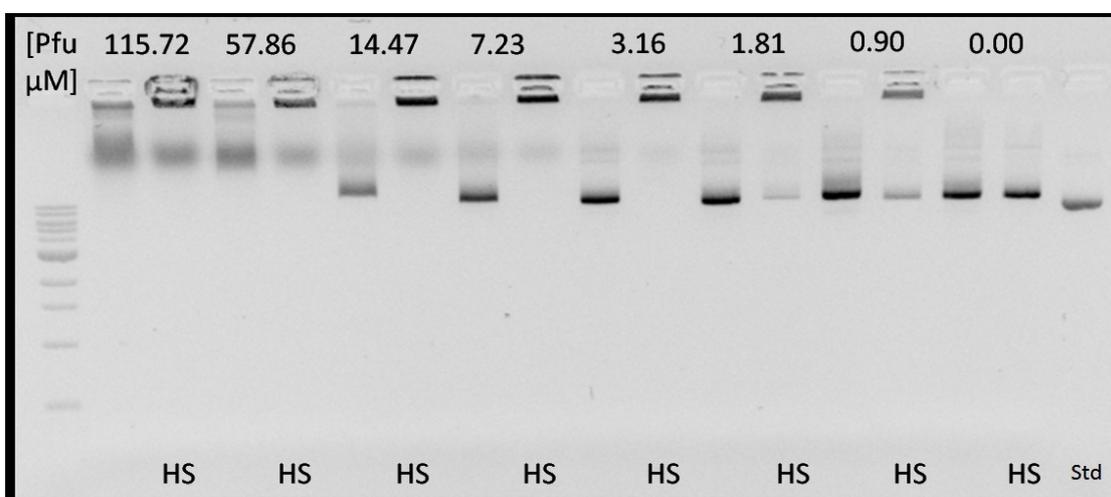
**Figure 79 Improved PEN enzyme activity detection and PCR amplification.** EcB and PfB variants were assayed in IPA buffer with enzyme concentrations shown at 15 nM pT in IPA. (a) Switch activation times for the two enzymes. (b) PCR fluorescent traces of samples shown in a. Controls with 0.5  $\mu\text{M}$  primers spiked-in before PCR and 1 mM of L-trp added instead of the enzyme is added.

We have therefore tried to run standalone PCR amplification as per commercial protocol with various concentrations of PfB added. We expanded the range of concentrations to check if this could have some effect. This was done either with a 10-minute pre-incubation at 95 °C in hope of denaturing PfB or without. Irrespective of the treatment or the concentrations, all samples containing PfB failed to amplify (**Figure 80**).



**Figure 80** PCR amplification with various concentrations of PfB. Concentrations show in legends with solid line samples incubated at 95°C for 10 minutes prior to PCR.

The unusual effect of PfB was investigated. It seems that when heated to 95 degrees for as little as 1 minute, it irreversibly aggregates with DNA, resulting in in-well blockage in agarose gels. Some binding is also occurring without heating (**Figure 81**). At 1.8  $\mu$ M of PfB there appears to be some free DNA after a heat shock, leading us to believe that in case of our 5.3 kbp plasmid added at a concentration of 1.8 nM, there exists roughly a 1000:1 ratio of DNA capture.

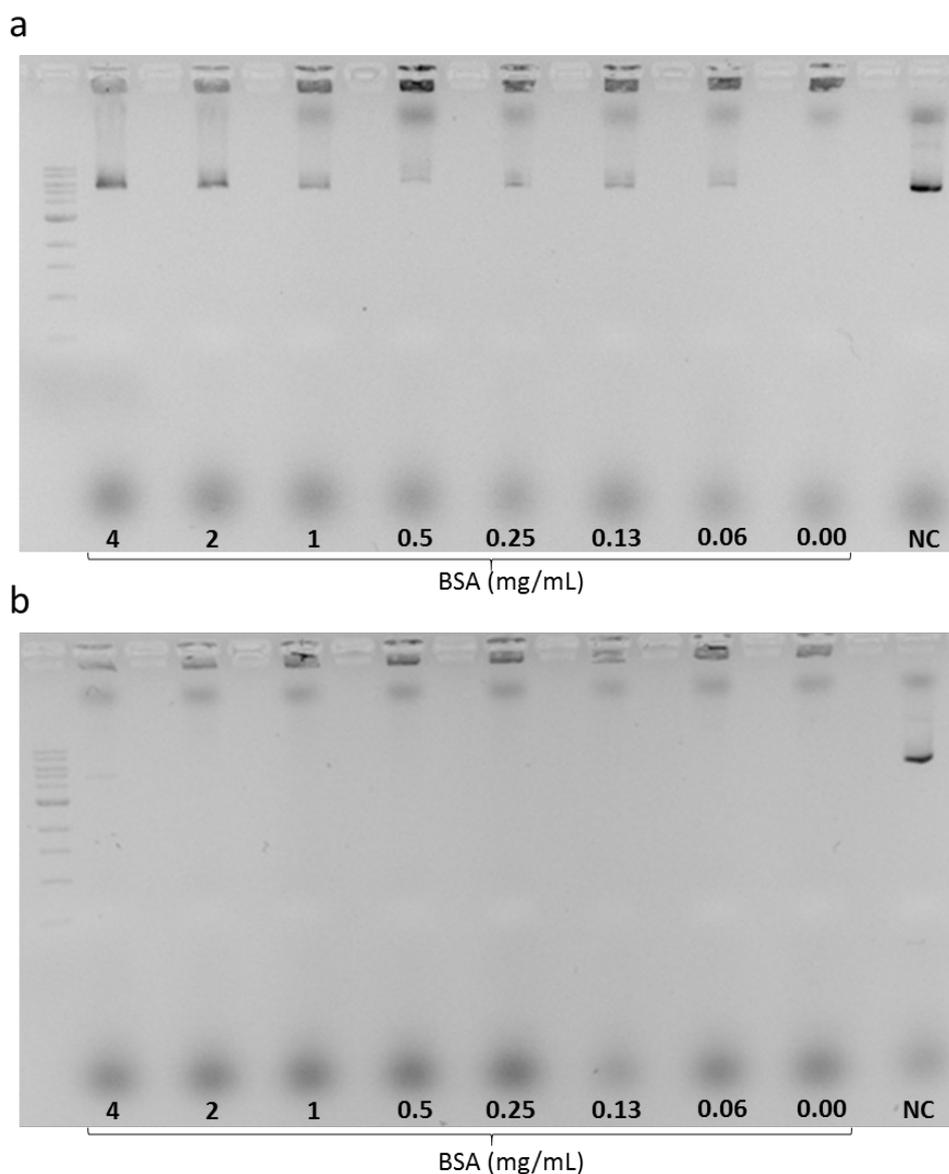


**Figure 81** Agarose gel electrophoresis showing DNA capture by PfB. Concentrations show in legend. The top band is DNA. The bottom band is free DNA. The plasmid, loaded directly without incubation in the IPA buffer.

We speculate when heated, the protein becomes more labile and residues with DNA-binding potential are somehow exposed, leading to DNA-protein aggregation.

We attempted to alleviate the issue through a series of potential quick fixes: use of yeast RNA decoy and adding BSA. We hoped that these would competitively bind the PfB and protect DNA. However, going as high as 0.25g/mL of yeast RNA showed no improvement during the electrophoretic mobility shift assay (EMSA), conducted as in (**Figure 81**) (not shown).

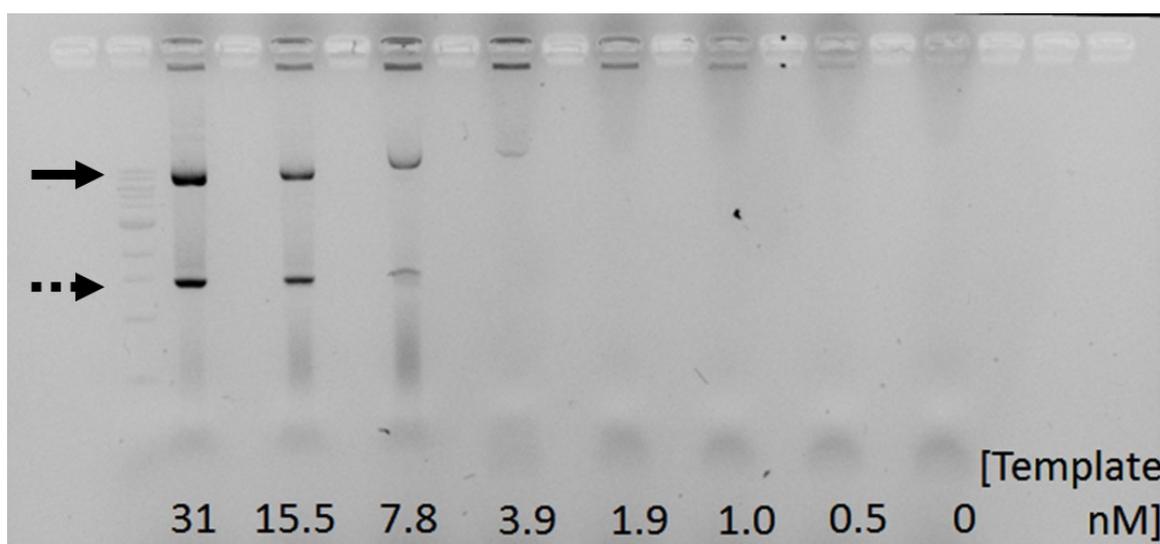
BSA showed some promise at EMSA stage when going as high as 4 mg/mL (**Figure 82a**). However, this did not help with PCR (**Figure 82b**). We speculate this might be due to exceptional stability of PfB, compared to BSA.



**Figure 82 BSA as DNA competitor for PfB binding.** Range of BSA concentrations as shown added to the reactions. NC represents bare plasmid run on the gel. (a) Template DNA incubated in IPA buffer with PfB and BSA at 95 for 5 minutes and run on the gel. (b) PCR reactions run as per manufacturer's instructions (NEB, Vent exo-, with primers and template as before) with extra BSA added.

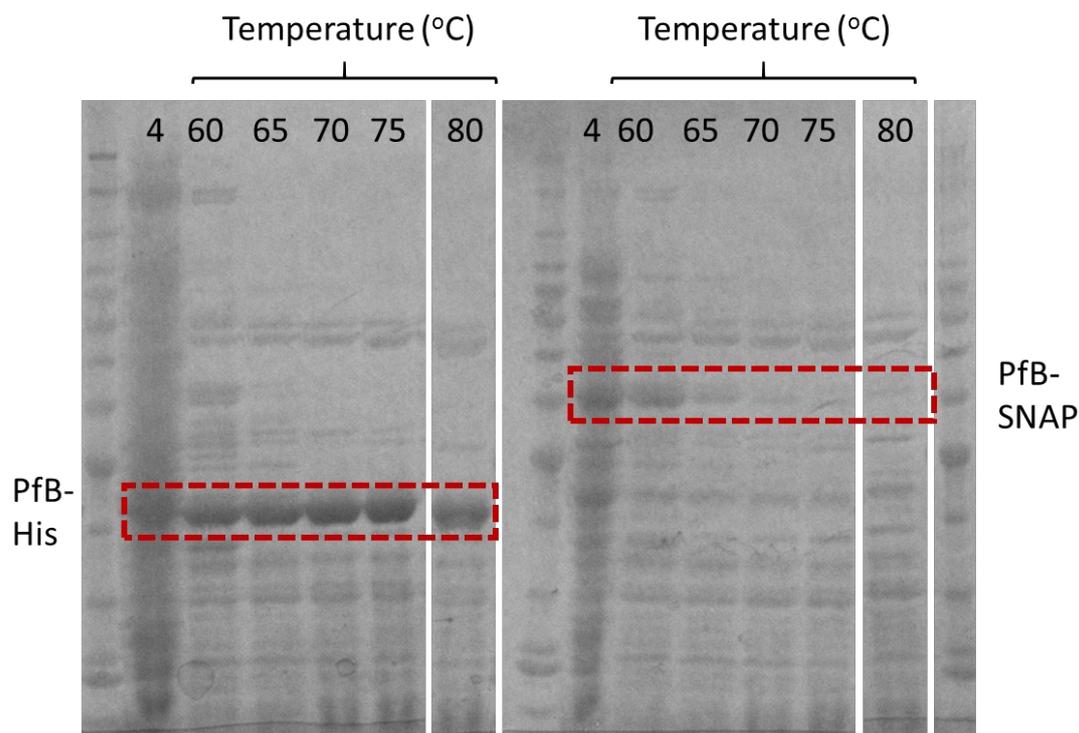
In line with our EMSA prediction we managed to obtain a PCR product, when template was close to or above 1:1000 ratio to the PfB (**Figure 83**). That is, for 16  $\mu\text{M}$  of PfB in this experiment, at least 7.8 nM of template was needed to observe a PCR product. The latter would appear to present lesser quantity of DNA than the template.

In principle, we did demonstrate detection of PfB activity at concentrations as low as 0.1  $\mu\text{M}$ , setting a requirement for only hundreds of pM template DNA that could be achievable. However, we were anxious about potential uncertainties and decided to abandon our exploration of the PfB.



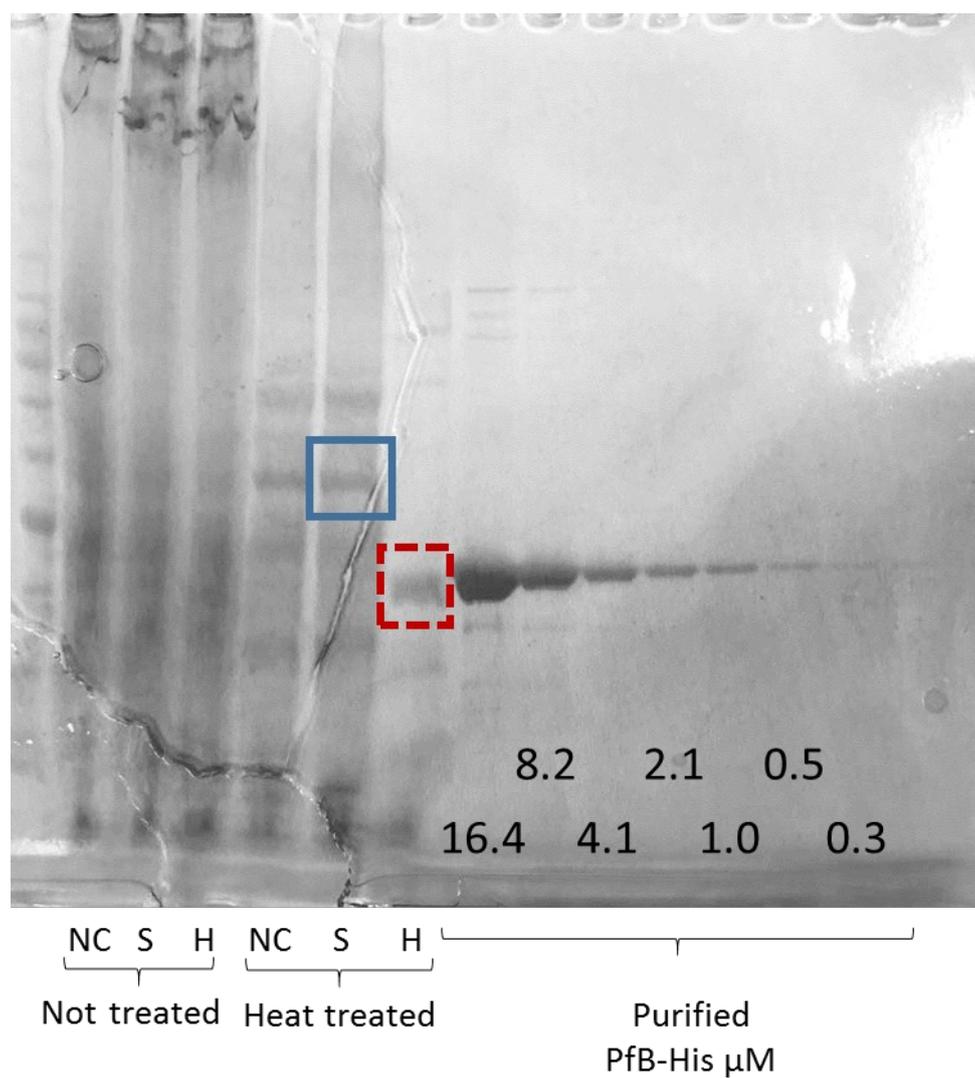
**Figure 83 Range of template for PCR with PfB.** Plasmid template is shown with a solid arrow, whilst PCR-product with a dashed one. Template concentration shown along the bottom, PfB concentration was at 16  $\mu\text{M}$ .

Thermostability of PfB does not only make it tolerant to mutational load but also seemed attractive to us due to potential convenience of purification. For proof-of-principle experiments, the protein could be purified in bulk via precipitating *E. coli* proteins through heating. To our disappointment, although this is the case for PfB, labelled with just the His<sub>6</sub>-tag, SNAP tag disrupts stability. Therefore, the 64 kDa protein quickly vanishes on the SDS-PAGE gel as temperature of treatments increases, unlike 45 kDa His<sub>6</sub>-tagged version that persists (**Figure 84**).



**Figure 84 Heat shock purification of Pfb protein.** Pfb-His and Pfb-SNAP fusions are labelled with red dashed rectangles. Cell lysate in 50 mM potassium phosphate buffer pH 8.0 with 200  $\mu$ M PLP was subjected to temperature treatment indicated (60-80  $^{\circ}$ C) for 20 minutes, then spun down at 16,000 G for 20 minutes at  $4^{\circ}$ C and run on the gel.

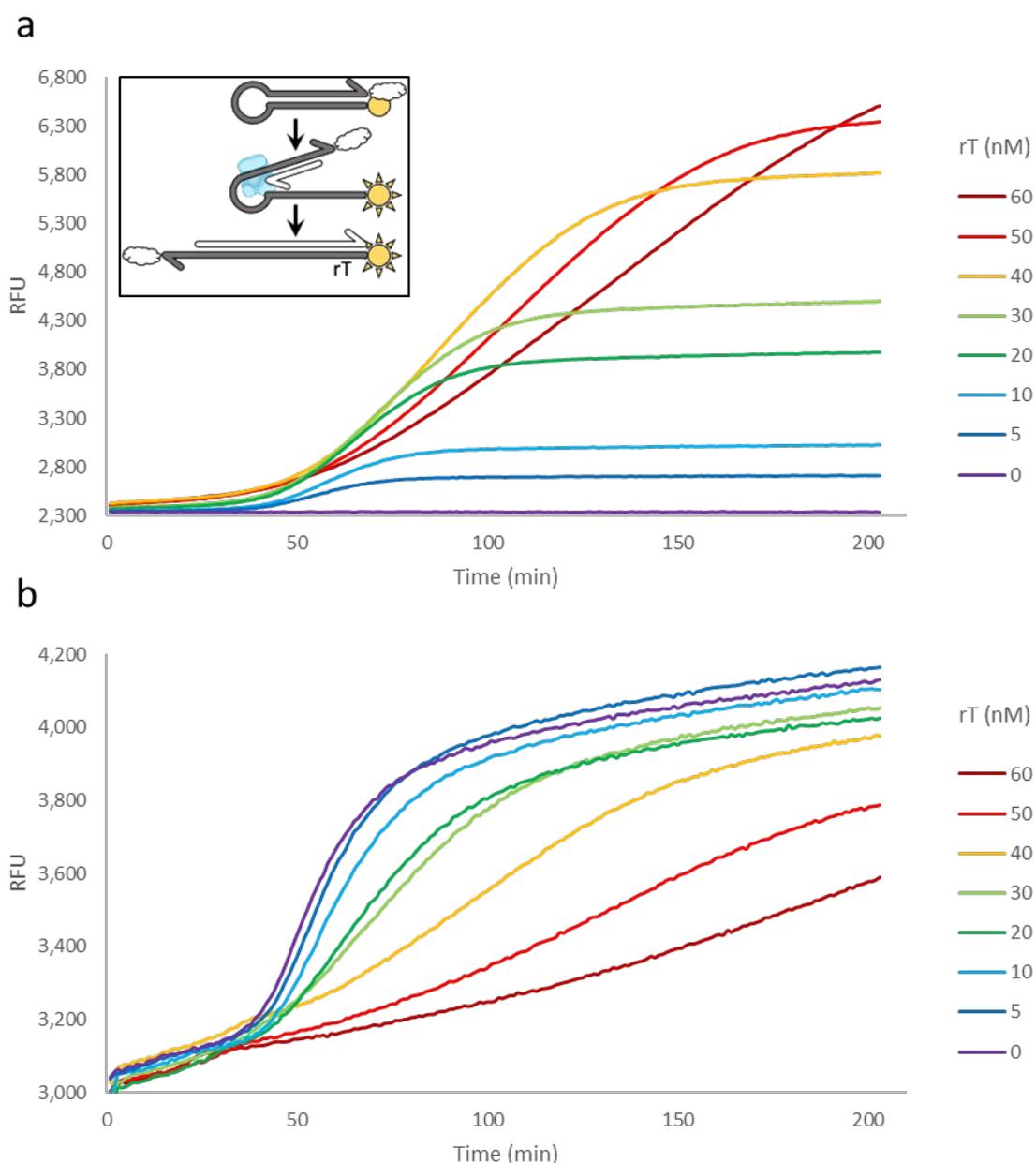
Although the inability to use the SNAP-tag makes such process prohibitive some of our experiments, this method of purification can be useful in other circumstances. This is why we decided to briefly see if IVTT could be coupled with heat-shock for small-scale rapid pure production. Such approach is tempting, as a typical purification protocol requires cell preculture, up to a day of induction and another for the subsequent purification. Here we simply added water, Pfb-His or Pfb-SNAP coding plasmids to IVTT mixes and waited for 2 hours. Each sample was split with 10  $\mu$ L either treated at 80 $^{\circ}$ C for 20 minutes or not, followed by clarification through centrifugation. SDS-PAGE results appear to us to present sufficient quantity and quality of Pfb for potential on-bead immobilisation and detection, for example, if such scenario were the case (**Figure 85**). Though it is difficult to evaluate from the gel, Pfb-His band seems to be more apparent than 0.3  $\mu$ M of protein, which is way above our detection limit of around 0.1  $\mu$ M according to **Figure 69**. Pfb-SNAP lysate did not clarify as well and the negative control has a comparable band at 64 kDa, making results difficult to interpret. We would note, that SNAP-tag in itself also provides a mean for purification.



**Figure 85 IVTT expression and heat-shock purification of PfB.** Negative control (NC), PfB-His (H) and PfB-SNAP (S), heat treated at 80 °C for 20 minutes or not (as indicated) along with a range of PfB-His (16-0  $\mu\text{M}$ ) run on SDS-PAGE. PfB-His band is indicated with red dashed box, PfB-SNAP with blue solid box.

### 2.3.3.3 *rT* hinders IPA PEN-DNA circuits

For a large number of experiments, we kept *rT* as 50 nM as this concentration was a once commonly used and allowed easy future detection in a microfluidic setting. We have therefore used 50 nM in all of the IPA-related experiments reported so far. *rT* saturates linearly, and, unlike *pT* does not regenerate. Therefore, we were rather surprised to find, during a retrospective investigation, how much of an effect it had on our circuit. In particular, we ran IPA-PCR reactions as before, but excluded *pskT* to ensure instant start. Whilst probe-specific fluorescence indicated similar start times, yet longer times to fully saturate *rT* and correspondingly higher fluorescence (Figure 86a), Eva Green fluorescence, driven mostly by *aT* saturation, indicated much sloppier switching of the circuit (Figure 86b).

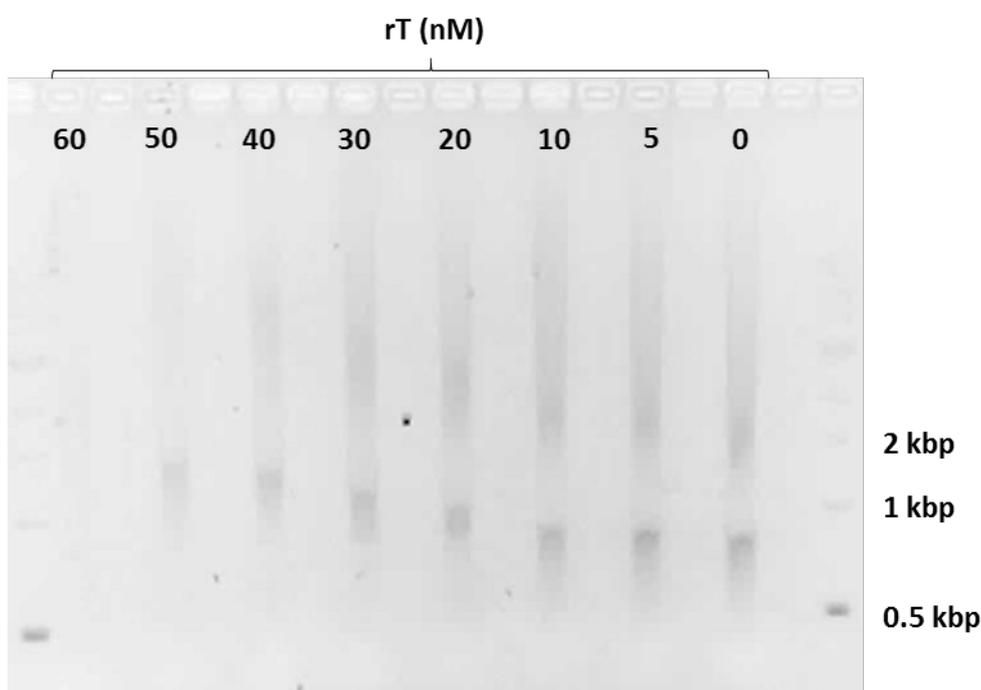


**Figure 86 Range of *rT* in IPA reaction.** Concentration of *rT* varied as indicated in the legend. (a) Cy5 fluorescent trace. Inset: *rT* mode of action. (b) Non-specific FAM fluorescent trace.

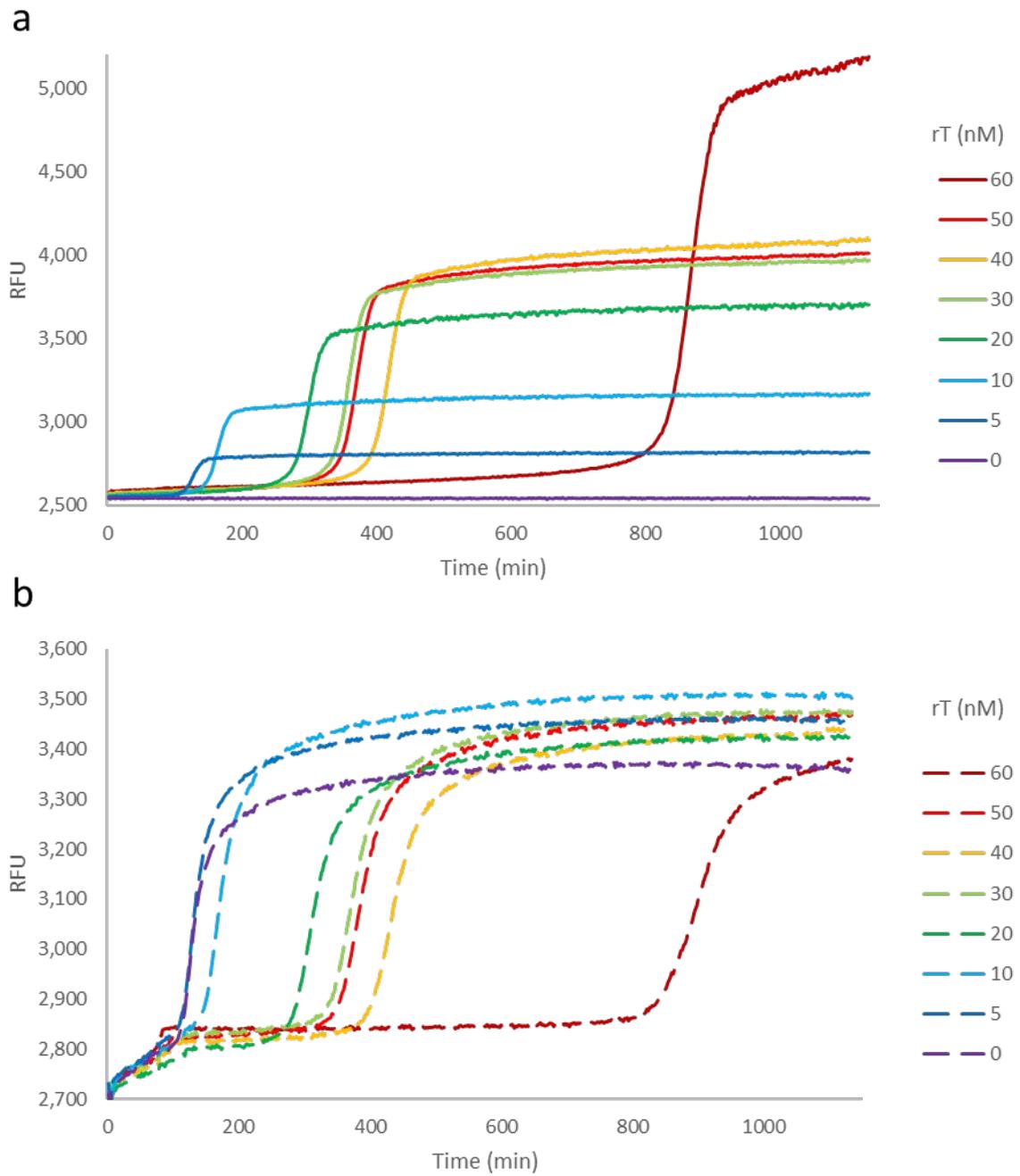
Following PCR cycling and gel electrophoresis analysis, we have made two interesting observations, related to the phenomenon above. The first one concerns an apparent PCR product drift on the agarose gel (**Figure 87**). We are not exactly sure why this is the case, but although not shown here, we have observed similar inconsistencies in apparent product size in other experiments. However, unlike other experiments, where no pattern could be established, here, it could be related to the rT concentration.

The second one relates to a similar verification for our miR buffer-based reactions (such as some of the exemplary ones, reported in (0)). We see that such assembly was more robust (**Figure 88**). Higher rT quantities reached higher fluorescence levels. This is reasonable, given that a higher total amount of fluorophore is present in the system.

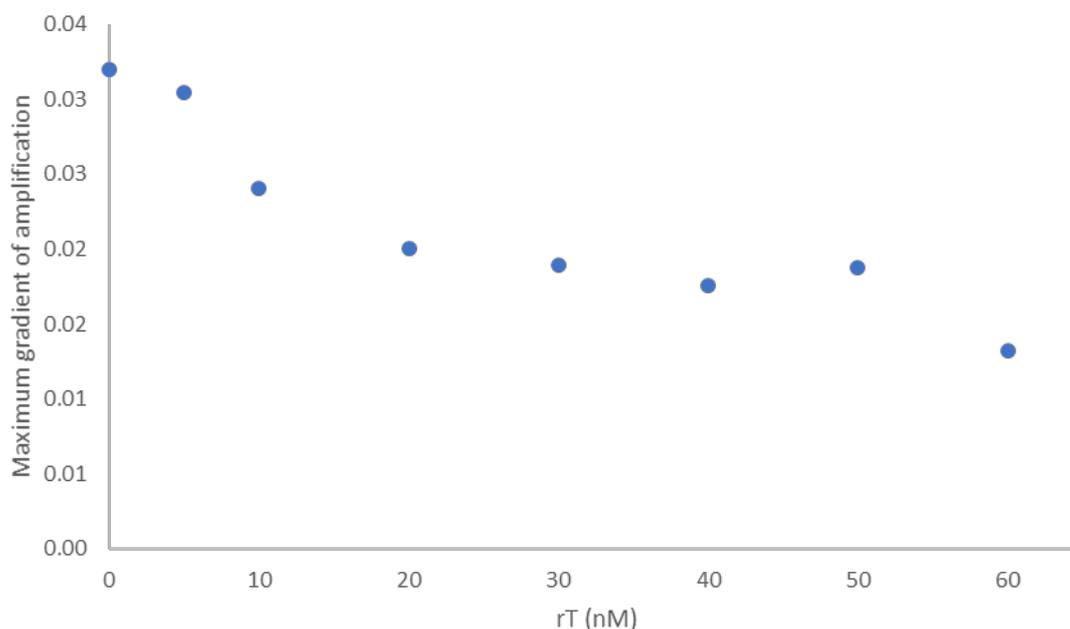
In either Cy5 or FAM channel the gradients appeared to be unchanged, despite a shift in start time. This could be due to more robust bistability in the miR buffer. A closer examination of the data, however, also demonstrates a slight gradient decrease (**Figure 89**). We took this into account for future work.



**Figure 87** Range of rT in an IPA reaction. Reactions were run in otherwise identical mixtures as reported in **Figure 86**.



**Figure 88 Range of rT in a miR reaction.** Concentration of rT varied as indicated in the legend. (a) Cy5 fluorescent trace. (b) Non-specific FAM fluorescent trace.



**Figure 89** Effect of rT on switching gradient in miR. Maximum gradients of normalised points 10 minutes apart are plotted for the traces shown in **Figure 88b**.

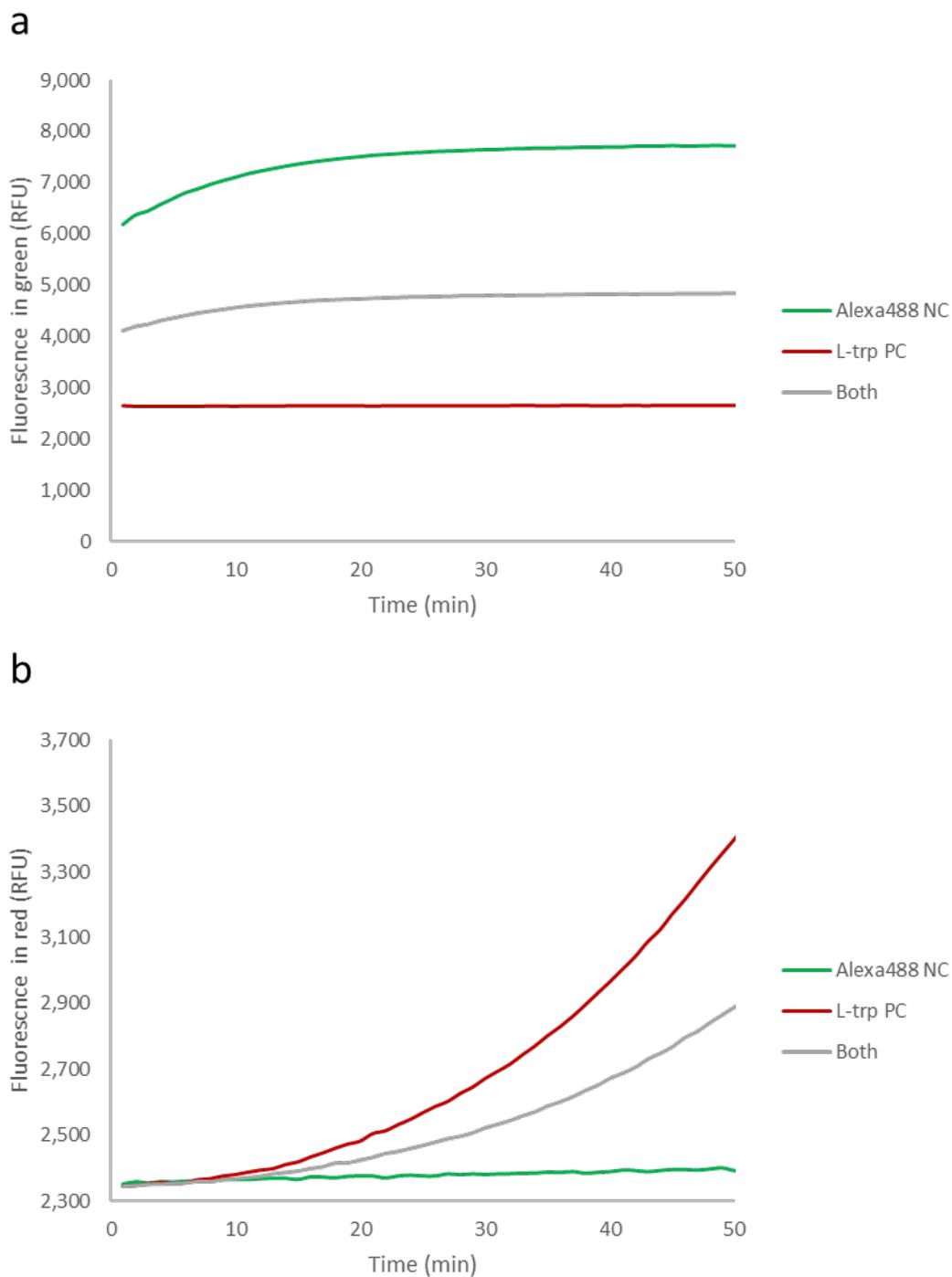
#### 2.3.3.4 *L-trp* synthesis is suitable for in emulsion detection

During previous PEN-CSR work in emulsion compartmentalisation has been verified to be compatible with all of the components we plan to re-use. However, small-molecule substrates and products are new. In particular, we were curious about indole, known to be easily soluble in oil and L-trp, as it did not dissolve too readily in water, hence could migrate through the oil phase.

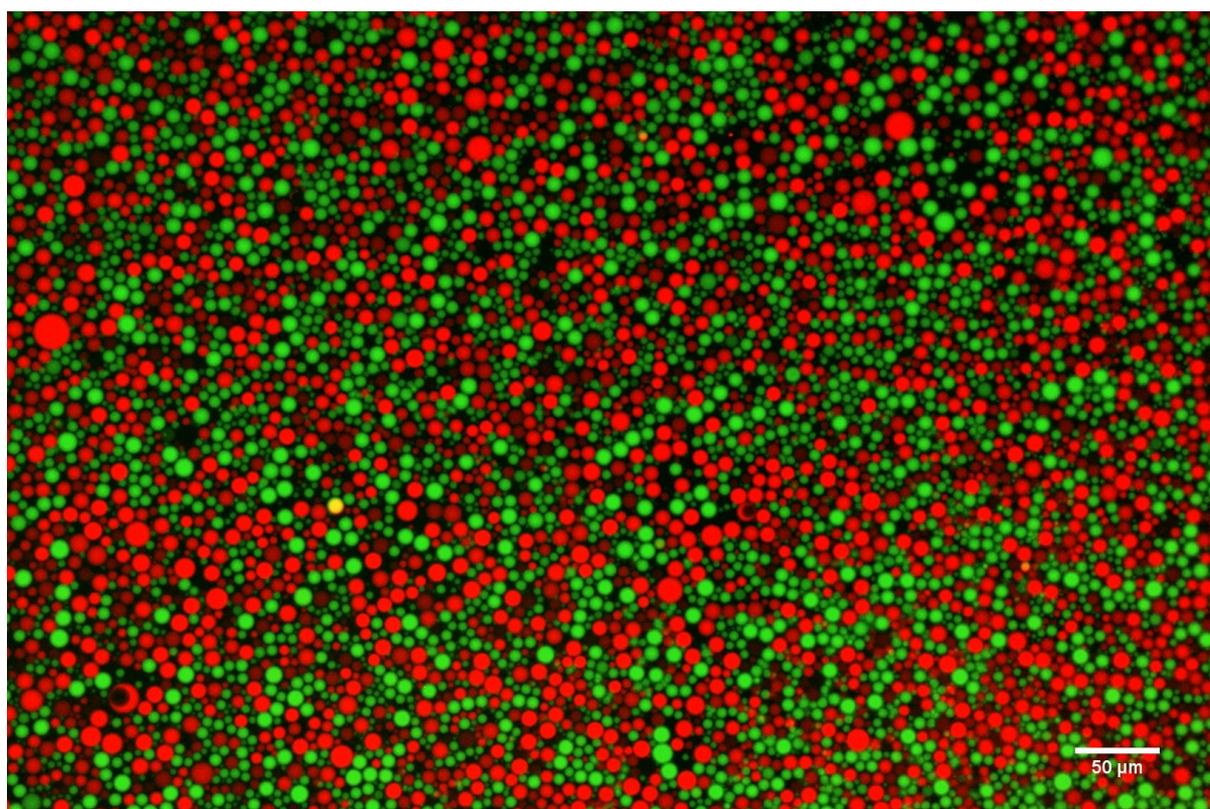
Concerning L-trp we carried out a simple proof of principle. One molecular program master mix was split into two samples. Either Alexa 488 dye was added, fluorescing in green, or 1 mM L-trp, known to start the circuit and therefore generate rT probe fluorescence in Cy5 channel that is red. We then emulsified the components through a simple bead-beating method – shaking the oil-reaction mixture with a metal bead in the tube – and observed green and red fluorescence for emulsions both separated and incubated together.

Bulk observation was in line with expectations: 3 stable levels of green fluorescence represented different amounts of Alexa488 (**Figure 90a**). Cy5 observation allowed us to determine the starting of the positive control to observe the emulsion under the microscope (**Figure 90b**).

To our relief, microscope examination indicated lack of cross-talk between the droplets. They were either green, corresponding to negative control Alexa488, or red, corresponding to only non-negative control samples (positive control being the only alternative). The brightness corresponded to non-mixed samples (not shown). Only one yellow droplet was observed which we consider an artefact caused by droplet fusion, rather than L-trp dissipation (**Figure 91**).



**Figure 90 L-trp behaviour in emulsion.** Green (a) and red (b) fluorescence traces of three emulsion samples incubated at 37°C. “Alexa488 NC” contains the corresponding dye, “L-trp PC” – 1 mM L-trp. “Both” contains the two emulsions mixed in equal proportion.



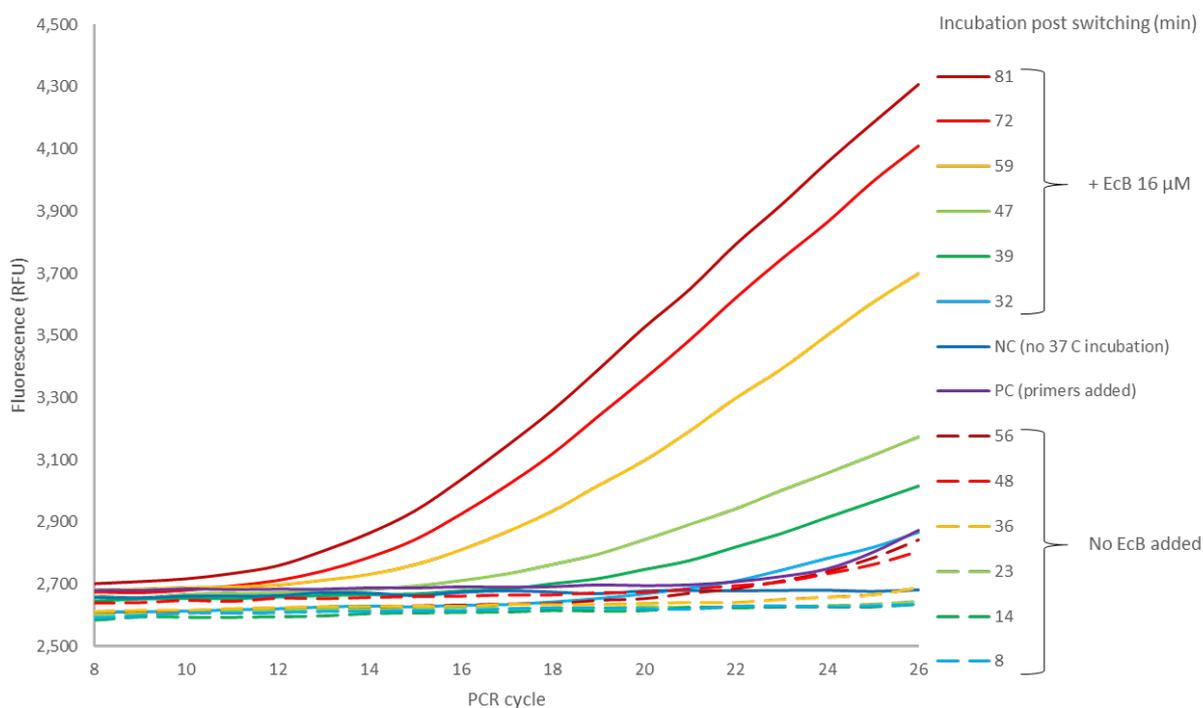
**Figure 91 Co-incubated L-trp and no-L-trp emulsions micrograph.** Cy5 fluorescence is observed in red for droplets, where the switch has started. Green fluorescence is observed in Alexa488-containing negative control droplets. All observable droplets bar one are either green or red but not both.

We have likewise assessed, whether indole dissolved in water can migrate into the fluorinated oil, when the two are vortexed and incubated together. Through absorbance measurement we determined that the concentration of indole in both solvents seemed to roughly equalise between the two (not shown). This was of little concern, as supplying fluorinated oil with indole dissolved during emulsion incubation is possible and can potentially increase the amount of indole available during L-trp synthesis.

#### 2.3.3.5 Primer production time affects the quantity of the PCR product

Whilst this may seem trivial we felt it was important to verify that at least for some time range post-switch start indeed correlates with the amount of product amplified. It is indeed the case, at least for short incubation times in early miR-based experiments we performed (**Figure 92**).

It occurred to us to cause switching through *in situ* synthesis of EcB. Therefore we had a group of earlier starters and slowed starters (lacking EcB but self-starting), differing in start times by about 25 minutes that were removed in pairs every 10 minutes. It is therefore interesting to note, that for comparable post-switching times, EcB-containing samples amplify better. As both sets of controls: with or without EcB (the latter not shown) amplified identically, we can speculate that some reaction occurs during the sensing stage. We can only speculate, that some side-effect of L-trp synthesis makes PCR conditions more favourable.

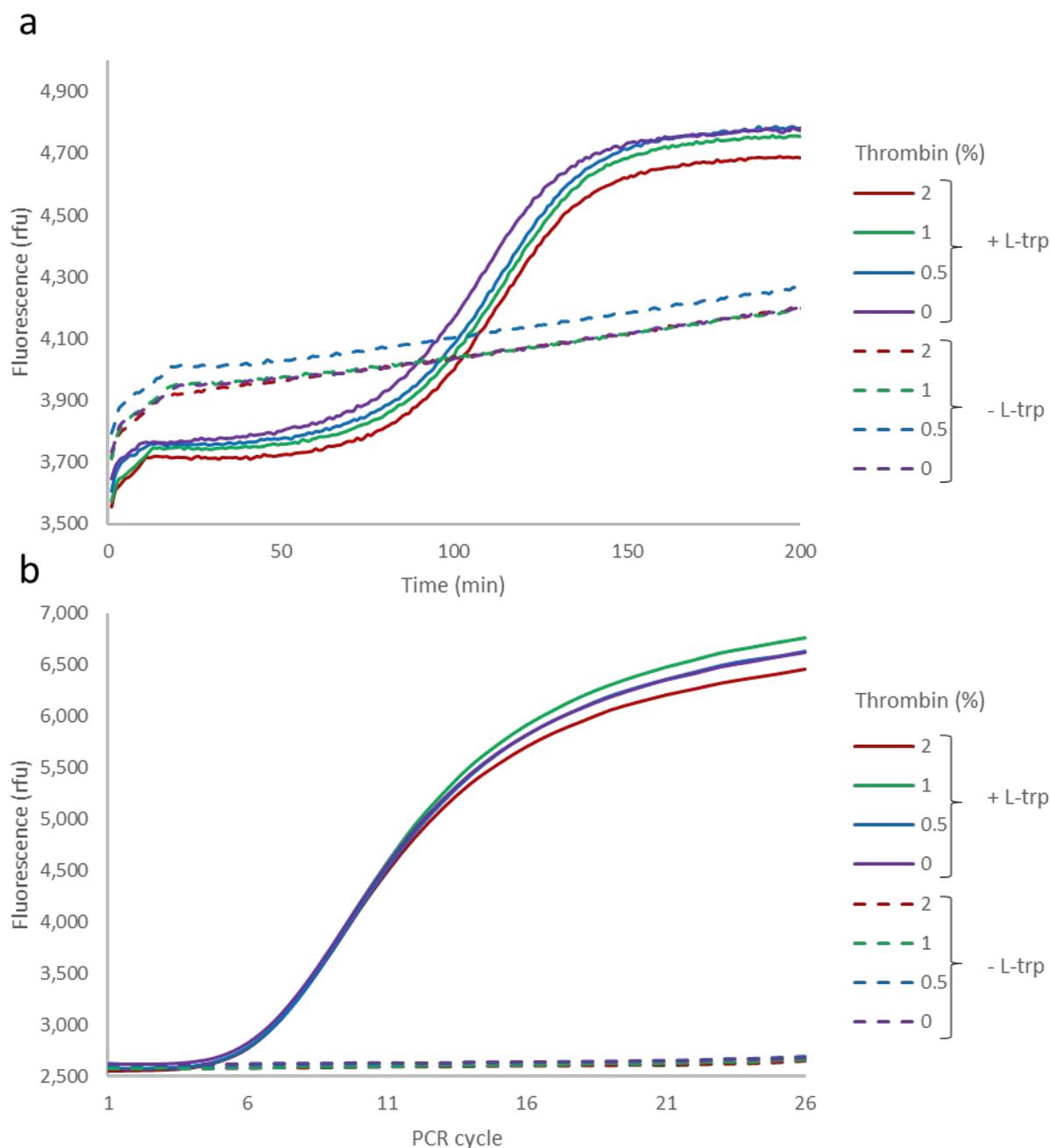


**Figure 92 Impact of primer production time on PCR.** Samples were incubated post-switching indicated amount of time and removed in pairs (matching colour), consisting of a sample containing 16  $\mu\text{M}$  of EcB or not. Negative control (NC) consisted of the same master mix containing 16  $\mu\text{M}$  of EcB or not (the latter not shown as appeared identical) but kept on ice until PCR cycled. Positive control (PC) was identical to NC bar recommended amount of primers was added just before the PCR. Here, the switching time was determined by reaching an arbitrary fluorescence threshold, as not all samples had time to produce a full sigmoidal trace.

### 2.3.4 On-bead enzyme activity selection

Our final goal was to incorporate our TrpB sensing into the workflow of Rémi Sieskind. To this end we verified, whether C-terminally SNAP-tagged TrpB retained its activity in bulk and on THA-beads.

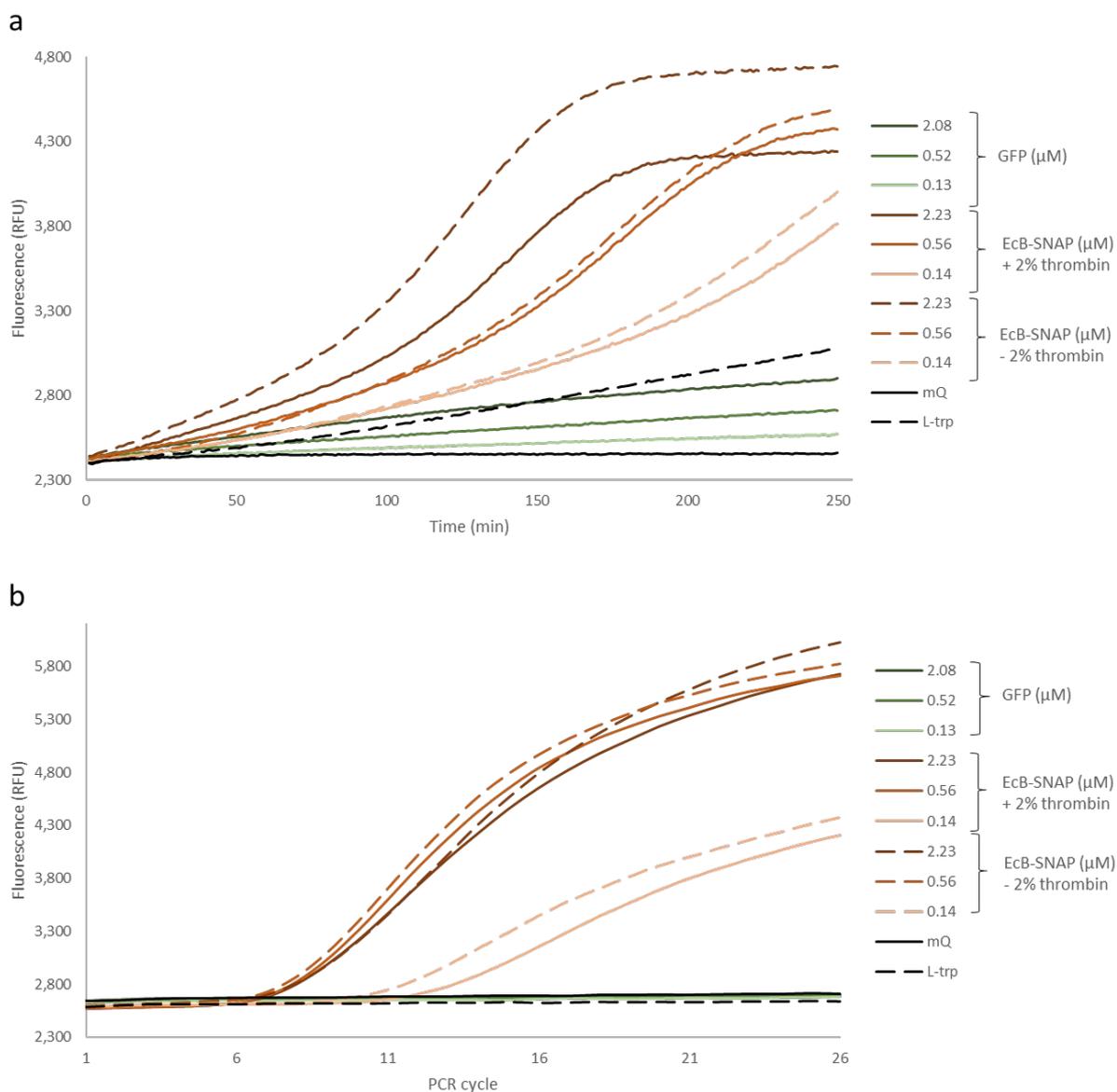
We first tested, whether thrombin protease, used by Sieskind in his workflow affected the circuit. It is added by Sieskind to cleave off the enzyme of interest from SNAP by targeting an appropriate specific site incorporated between SNAP and the enzyme to ensure the latter is not hindered by the former. We made a range of 0.02-0 U/ $\mu\text{L}$  of thrombin with and without tryptophan to see if it had any disruptive effect on the circuit. 0.02 commercial units per  $\mu\text{L}$  or 2% was the concentration successfully used by Sieskind. From this experiment, we observe that thrombin protease has no effect across a range of concentrations on the detection of L-trp by the switch. It also does not appear to affect the subsequent isothermal primer production nor the PCR amplification stage (**Figure 93**).



**Figure 93 Range of L-thrombin in IPA reaction.** 2-0% of thrombin added to samples containing 1 mM L-trp (+ L-trp) or not (- L-trp). (a) Fluorescent trace of switch activation. (b) PCR fluorescent traces of samples shown in a.

We proceeded to add EcB-SNAP to the test. We wanted to compare the activity with thrombin protease, without it as well as used GFP, L-trp and water as controls. Anxious that EcB-SNAP activity may be slower, we have chosen to pre-incubate it with L-trp synthesis reactants and thrombin protease for 90 minutes at 21 °C – the optimal temperature for the latter. It appears a range of EcB-SNAP has effect on the circuit start with greater protein concentration resulting in faster switching and then faster PCR (**Figure 94**). In comparison, no switch start nor PCR was observed for the negative controls (GFP, water). This suggests the switching and amplification are specific.

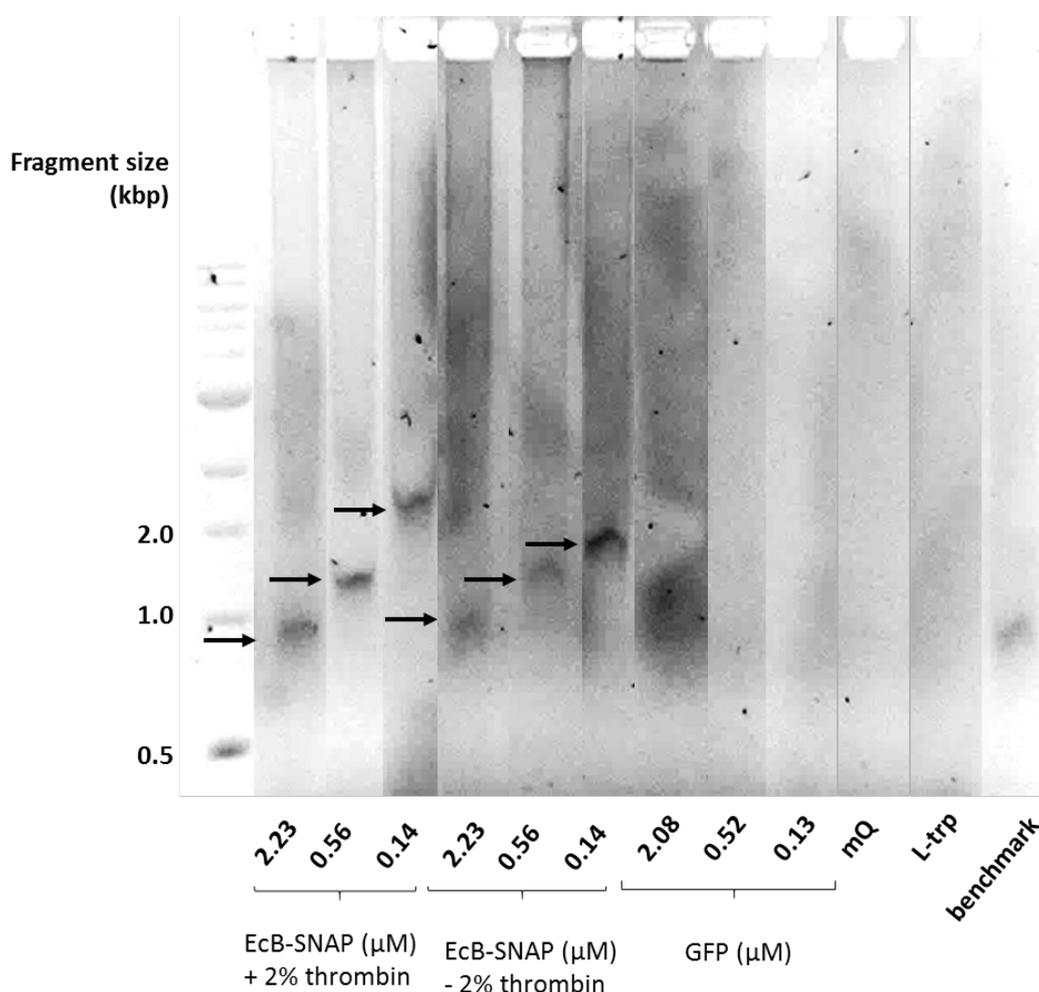
We also checked, whether EcB retains activity, when SNAP tag is attached (in which case use of thrombin protease would not be necessary, but only optional), and unlike experiences reported by Sieskind, this seemed to be the case. Both switching times and PCR curves appear comparable (**Figure 94**). Interestingly, unlike some other experimental runs, here resolution at the switch level (**Figure 94a**) appears finer. For example, a distinction can be drawn for 2.2 and 0.6  $\mu\text{M}$  of EcB-SNAP on the switch level, but not on the PCR, suggesting that primer concentration was probably saturating for both. Likewise, positive control (L-trp) can be seen starting the switch but the PCR curve is flat suggesting a quick transition from non-significant to saturating primer concentration. This could be addressed in the future with lower concentration of primer producing templates.



**Figure 94 Investigating activity of EcB-SNAP.** Range of EcB-SNAP with 2% thrombin (+2% thrombin) and without (-2% thrombin) as well as range of Gfp for control. “mQ” represents negative control with water, whilst “L-trp” – 1 mM of L-trp added for positive control. Protein samples were pre-incubated for 90 minutes at 21 °C with L-trp synthesis reactants and thrombin prior to addition of the molecular program and initiation of the experimental run. (a) Fluorescent trace of switch activation. (b) Real time PCR fluorescent traces of samples shown in a.

Generation of products was verified on the gel. The EcB band position was not very stable, but still confirmed the successful EcB-induced PCR. However, overall we concluded that we could move forward without use of the thrombin protease. Although it does not affect the circuit, it is not necessary either.

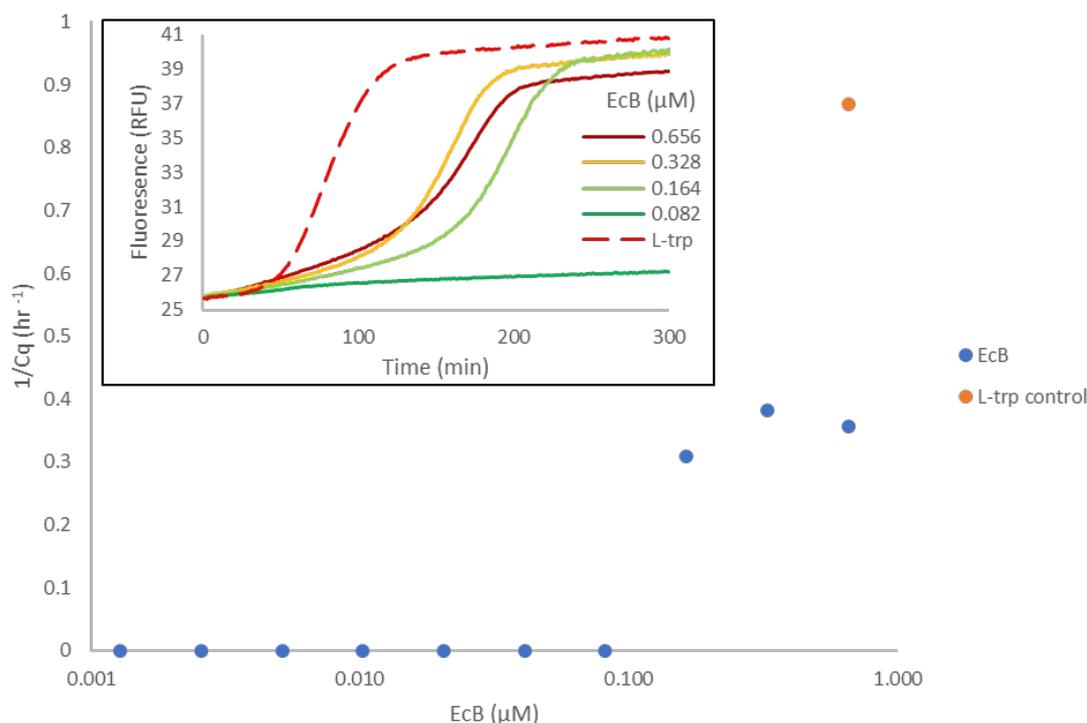
Nonetheless, it may also be useful to consider use of thrombin protease from the evolutionary context perspective. During the implementation of the full *in vitro* PEN-CSR unhindered enzyme activity is useful, but context and immediate environment are also important in defining selection conditions. It could be, that on a more subtle level, TrpB activity would somehow be impacted by presence of the SNAP tag and introduce an unknown bias. With such consideration, thrombin protease may still be useful in attempts to evolve the enzyme independently of the SNAP-tag context. Of course, this is limited by the fact, that the two are expressed in fusion.



**Figure 95 Amplification in response to EcB-SNAP.** Agarose gel electrophoresis of the products generated in **Figure 94**. Lanes were digitally rearranged for ease of interpretation. Black arrows highlight PCR products of interest. Range of EcB-SNAP with 2% thrombin (+2% thrombin) and without (-2% thrombin) as well as range of FFP for control. “mQ” represents negative control with water, whilst “L-trp” – 1 mM of L-trp added for positive control. Independently amplified fragment of the expected size is shown as “benchmark”.

Prior to moving to beads we wanted to insure that the concentration of EcB-SNAP they would carry would be sufficient. Our rationale was that we could generate THA beads with 10  $\mu\text{M}$  BG-maleimide. Making a rough assumption that the beads would arbitrarily represent 10 % of the sensing mixture we therefore needed to be able to detect 1  $\mu\text{M}$  and below. Hence, we tried a range of highly pure EcB-SNAP going down from 0.66  $\mu\text{M}$  in two fold dilutions (**Figure 96**). Interestingly, it seems our window of observation is quite narrow as the highest concentration slows down the switch, presumably due to oversaturation, whilst diluting it 8-fold sharply brings us below the limit of detection. Our attempts to make the switch less sharp by decreasing pskT or pT concentration (not shown) led to no avail as the stability of the system deteriorated sharply.

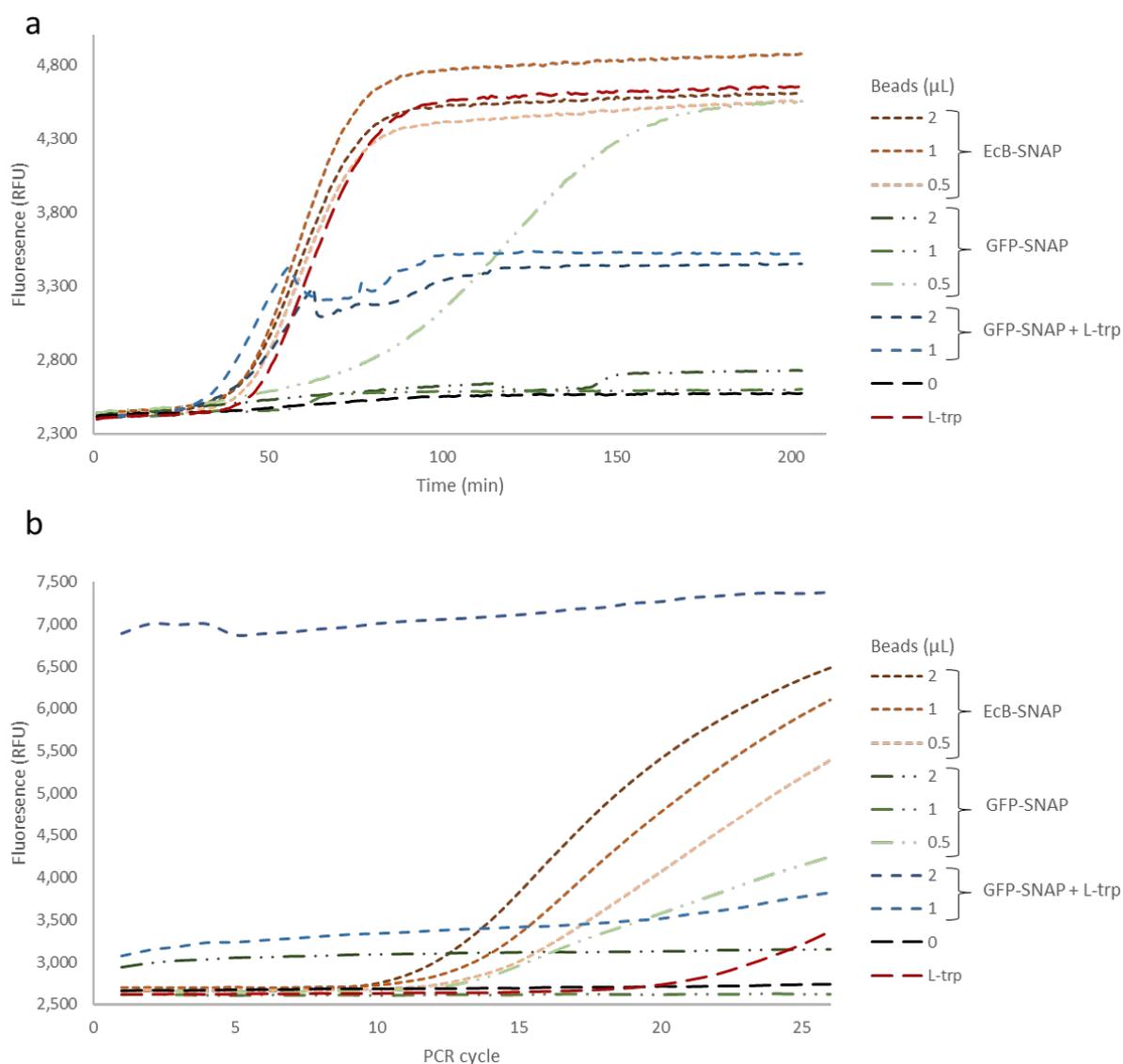
Nonetheless, the concentrations we detect are well below the 1  $\mu\text{M}$  limit we set, making the approach functional in droplets as is. If needed, we could also vary the temperature regime to first favour L-trp synthesis by keeping the temperature low (as we showed, PEN circuits can slow down significantly at low temperatures), followed by a normal incubation.



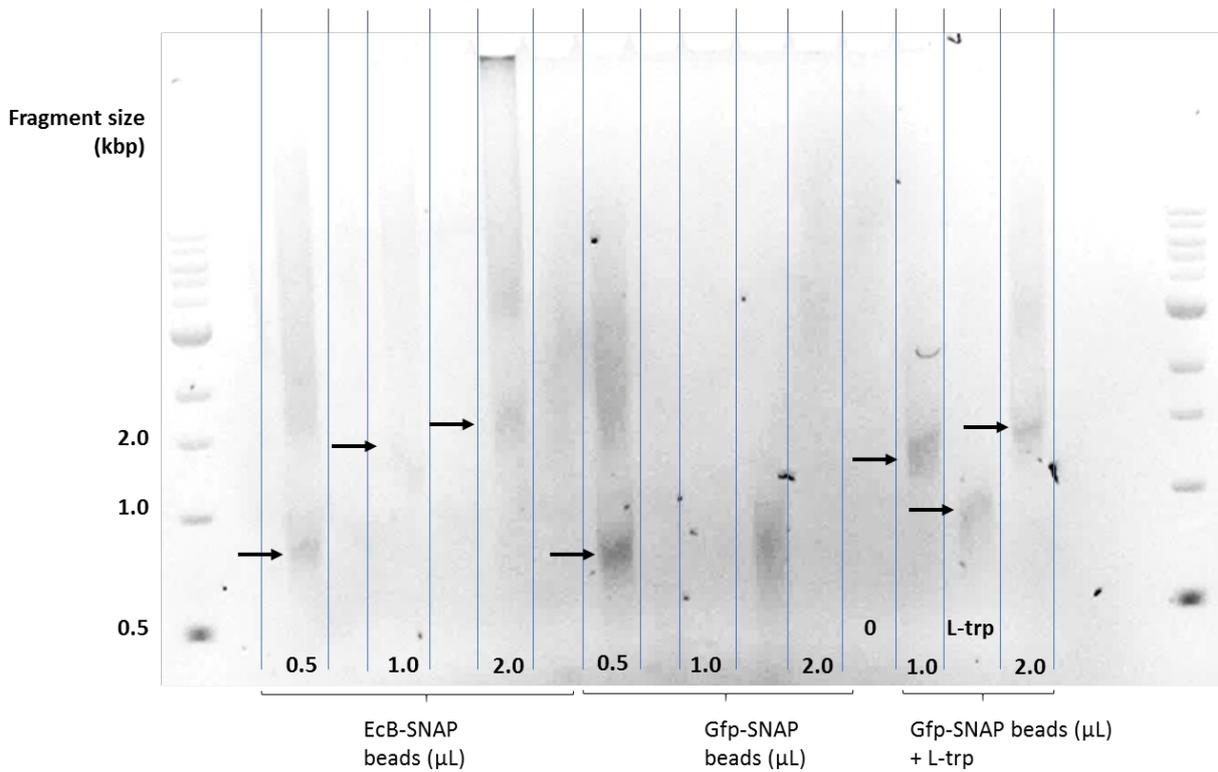
**Figure 96 Determining limit of detection for EcB-SNAP in IPA.** Range of EcB-SNAP concentrations (0.66-0.00  $\mu\text{M}$ ) against  $1/Cq$ . “L-trp control” series represents a positive control with 1 mM L-trp added. Inset: time-fluorescence traces for the control and three highest concentrations of EcB.

Thanks to careful instructions provided by Rémi Sieskind we produced THA beads and proceeded to immobilise the protein on them. To do so in a configuration close to the one that would happen in the final directed evolution protocol, we added beads produced to the IVTT mixture during EcB-SNAP expression in hope of translation and parallel immobilisation. Following a wash we added different quantities of beads covered with either EcB-SNAP or Gfp-SNAP (as a control). We assumed bead quantity would be a proxy for potential enzyme activity/concentration variations.

We observe that EcB-loaded beads successfully turn on the switch. So do the 1 mM L-trp control and GFP-bead-containing 1 mM L-trp samples (**Figure 97a**). However, GFP beads on their own do not, implicating that beads do not strongly affect our molecular program. 0.5  $\mu\text{L}$  of GFP beads appear to start with a large delay, and could be an outlier due to difficult pipetting of small volume of beads, hence inaccuracy in total volume. On the PCR-level (**Figure 97b**), the EcB concentrations are better resolved. Expectedly, the 0.5  $\mu\text{L}$  of GFP beads also amplify. We are not sure what occurred to the L-trp-containing GFP beads but we assume some mechanical motion in the tube could affect the readings in the machine that are unusual. Like with previous experiments, the verification on the gel suggests a specific PCR, although again, some bands are displaced (**Figure 98**).



**Figure 97 Investigating activity of EcB-SNAP on beads.** Range of EcB-SNAP or GFP-SNAP loaded beads quantities. “0” represents negative control with water, whilst “L-trp” – 1 mM of L-trp added for positive control. Total volume of the PCR reaction was 10  $\mu\text{L}$  (a) Fluorescent trace of switch activation. (b) PCR fluorescent traces of samples shown in a.



**Figure 98 Amplification in response to EcB-SNAP.** Agarose gel electrophoresis of the products generated in **Figure 97**. Lanes are highlighted with blue lines for clarity. Range of EcB-SNAP or Gfp-SNAP beads as well as range of Gfp-SNAP beads with 1 mM L-trp for control. "0" represents negative control with water, whilst "L-trp" – 1 mM of L-trp added for positive control.

## 2.4 Summary

Here we demonstrate that the sensor detecting L-trp, developed in Chapter 1 can be used for TrpB enzymatic activity selection. This activity of a metabolic enzyme can be detected *in situ* and translated into robust gene amplification. This holds true, even when the protein is immobilised on hydrogel beads as it retains its activity.

We first demonstrate that TrpB retains activity in the buffers, compatible with the molecular programs. We likewise demonstrate the setting when the molecular program is not disrupted by TrpB and its reactants. This concerns indole and TrpB itself that are disruptive to the molecular programs at high concentrations.

We proceeded with demonstration of isothermal primer amplification and PCR as a molecular program output to the TrpB input. In case of Pfb, this is more challenging as it binds DNA strongly, when heated. We explore the phenomenon to define the stoichiometric ratio. It appears achievable but potentially less convenient.

In the case of EcB, we demonstrate that a detection window is possible, where the timeframe of our sensor and enzymatic activity, driven by tolerable enzyme quantities overlaps. Likewise, we demonstrate that detectable concentrations of the enzyme are achievable when produced in IVTT and transferred, immobilised on beads. When in droplets, the product does not escape droplets, therefore avoiding crosstalk, whilst one of the reactants can be fed through the oil phase. Together, this means that our program can be used in emulsion within the *in vitro* CSR framework.

Improvements to the process are also possible. First, we demonstrate the improvement, decrease in the reporter template can have on the circuit. This is easy to implement without affecting our work as for observations in bulk, lower rT concentrations are sufficient. Primer production time, rate dynamic can also be optimised as we demonstrate here and in (3.3.7.1). Most of the time, primers were in excess for some of the samples. Avoiding this will make the response sharper and avoid non-specific PCR products.

Finally, improvements can be made to the gel analysis. It is possible that the product shift occurred due to some other assembly component remaining bound to the DNA. Purification, prior to running the gel would alleviate the issue. Likewise, nested PCR on diluted reaction mixture is possible to neatly amplify the gene of interest.

We believe we demonstrate here that our PEN DNA program can be coupled with CSR for directed evolution of TrpB. Furthermore, we hope it paves way to other small molecule catalysing proteins.

## Chapter 3 Exploring biosensor implementation.

### 3.1 Aims

In the chapters above, we demonstrate particular examples we took forward to showcase small-molecule detection and downstream processing. In this chapter we would like to share some of our insights about alternative sensor implementations we explored (**Figure 99**). We envisage they can be used to scale up future circuits.

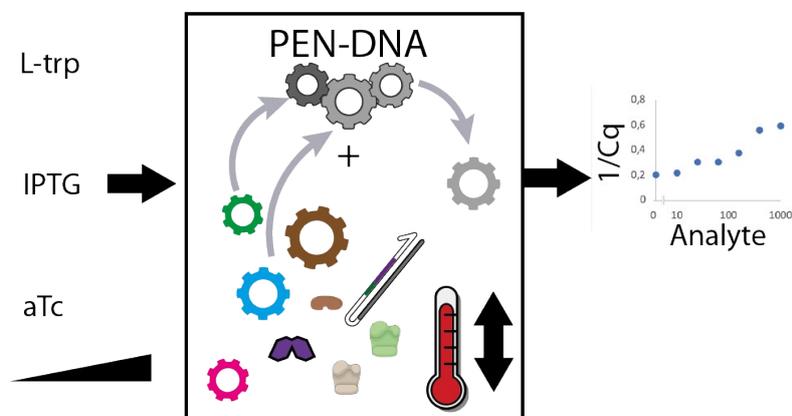
We will first talk about alternative architectures we have considered to overcome context dependence imposed by Nt.BstNBI. Nt.BstNBI, whilst instrumental for proof of principle work, at the same time made it evident, that “tailed” outputs can be problematic. The variation in sensing template efficiency dependent on the particular operator site sequence makes the design less predictable. At the same time, the tailed output is much less efficient: there is a thousand-fold difference in concentration of the sT and psTs used. This is especially relevant for the pskT, that has a weaker impact on the circuit.

Our approaches include a transcription-based system, one using an alternative nickase and finally chemical modifications altering the behaviour of the sensing modules. Combined, these provide a range of tools that will bring us closer to “plug and play” use of the transcription factors.

We will then follow by discussing the scalability of the circuit. Although scalability and multiplexing has already been demonstrated for PEN circuits, we wished to confirm that orthogonal switches can still be used in the same tube without cross talk, in the case of the 37°C PEN circuits.

We also explored potential distribution of the circuit into two orthogonal switches. This would allow to dedicate one to act as an interface and another as an output module with a communication device in-between. We also demonstrate that working at 37 °C opens the possibility to use additional (non-thermophilic) nickases, further expanding the possible sequence space for the switches.

We will also demonstrate linear output function and use of other transcription factors as well as other DNA polymerases. We will show that a circuit, similar to the circuit used in Chapter 2 can be run at a higher temperature of 45 °C. Finally, we will mention some smaller design aspects we found to have an impact on our circuits, that are useful to consider in future designs.

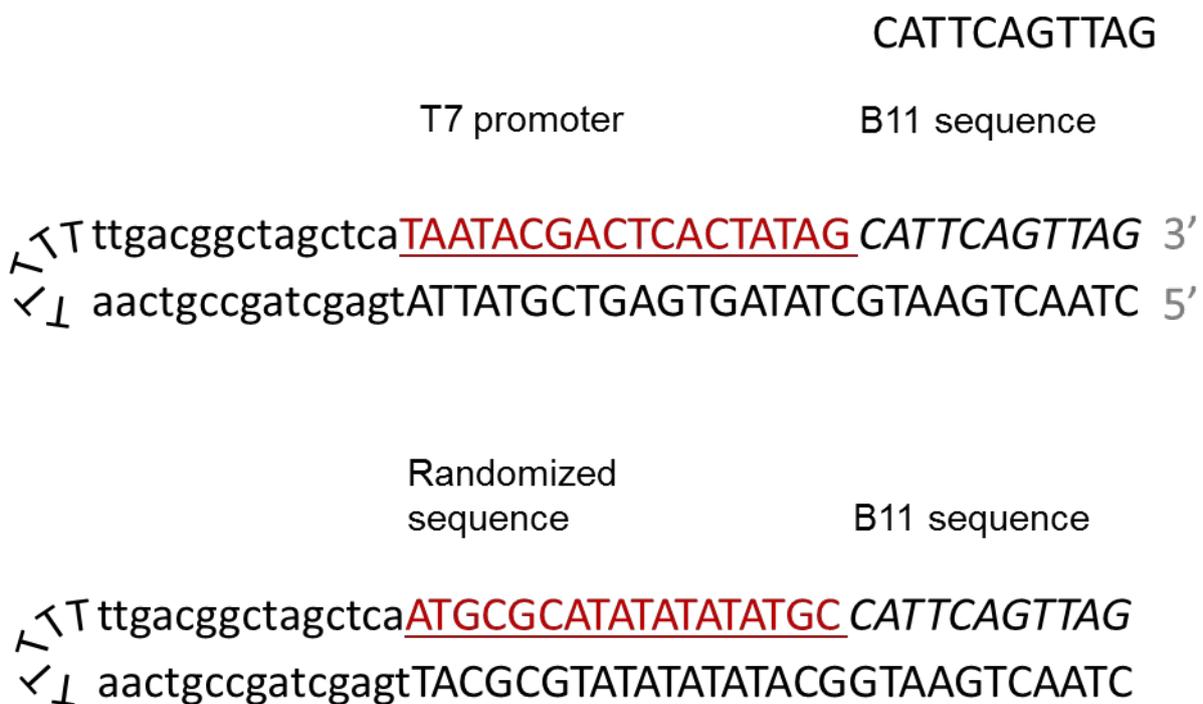


**Figure 99 Graphical abstract of chapter 3.** Various components are added or replaced in PEN-DNA, the temperature is varied. The reaction still responds to various analyte concentrations.

## 3.2 Alternative sensing architectures

### 3.2.1 T7 transcription system can start PEN circuits

Somewhat inspired by genelets circuits, described in (0.4), we thought we could use nature-like transcriptional control to drive the circuit. Instead of adapting the TF-promoter interaction to control DNA polymerase activity, the idea would be to include an RNA polymerase in the design, and then to use the produced transcripts as input to the PEN circuit. To this end we added T7 transcriptional components, namely T7 RNA polymerase and NTPs, to our miR-based circuit at 37°C and investigated, whether a T7 promoter, driving transcription of a trigger sequence could be used. In particular, we designed two psT-like hairpin structures, but replaced the nickase and the operator sites with a T7 promoter (**Figure 100**).



**Figure 100** Sequence-level representation of the T7 modules used as controls. Top – B11 trigger sequence is shown for reference. Middle – hairpin structure, containing T7 promoter (underlined), followed by the B11 sequence on the coding strand (*italic*). Bottom – hairpin structure, containing an arbitrary sequence (underlined), followed by the B11 sequence on the coding strand (*italic*).

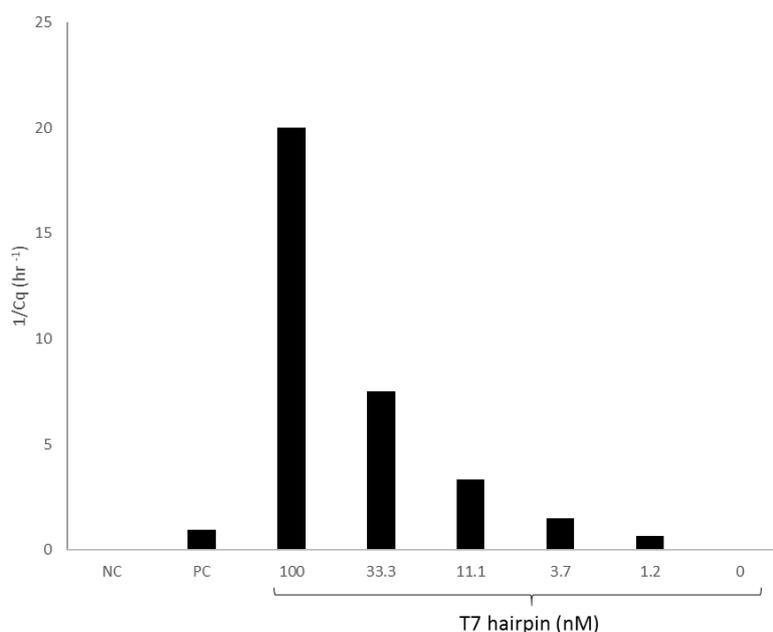
Here, we took our standard miR assembly (B11-based) as before and complemented it with 1% of NEB T7 RNA Polymerase ( $500 \text{ U ml}^{-1}$ ), 500 nM of ribonucleotides. We then added various concentrations of the T7 hairpin. As a control a tube contained 100nM of an hairpin of identical structure, except that the T7 promoter had been replaced by an arbitrary random sequence.

We could observe that increasing concentration of the hairpin induced faster start times, indicating that the T7 RNAP activity on its hairpin template was indeed detected by the switch. In the control, the randomised sequence did not trigger any start, confirming the T7-specific transcription mechanism (**Figure 101**).

We concluded that the production of small transcript by a hairpin containing a T7 promoter followed by a trigger sequence, in combination with T7 RNA polymerase, is indeed capable of starting the circuit in a way similar to pST.

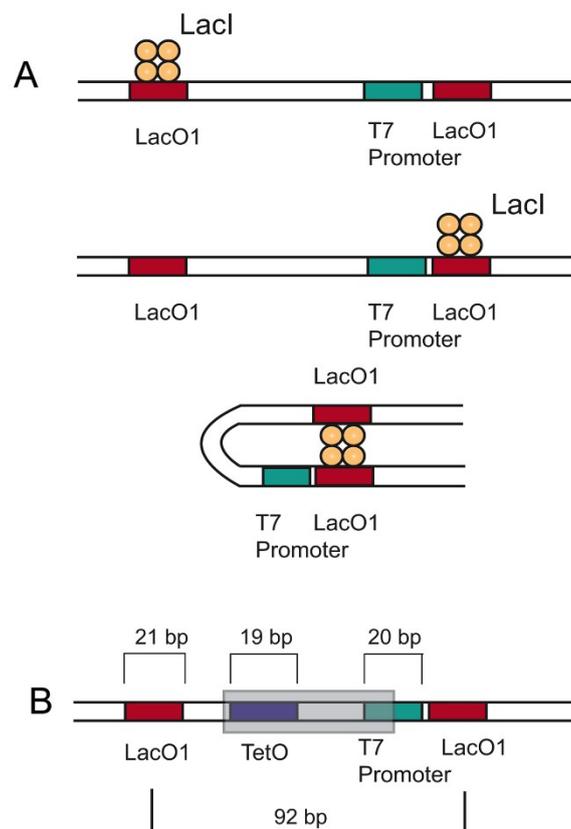
We observe that the concentration of the hairpin appears to correlate with faster start times, whilst the randomised sequence has no effect (**Figure 101**).

Therefore, we speculate, the hairpin containing a T7 promoter followed by a trigger sequence is indeed capable of starting the circuit akin to a pST.



**Figure 101 PEN circuit triggered by T7 RNA polymerase.** Range of concentrations of T7 promoter containing hairpins is shown (100-0 nM), along with negative control (NC), containing 100 nM of hairpin with a randomised sequence instead of the promoter and positive control (PC), containing 0.5 nM of sT.

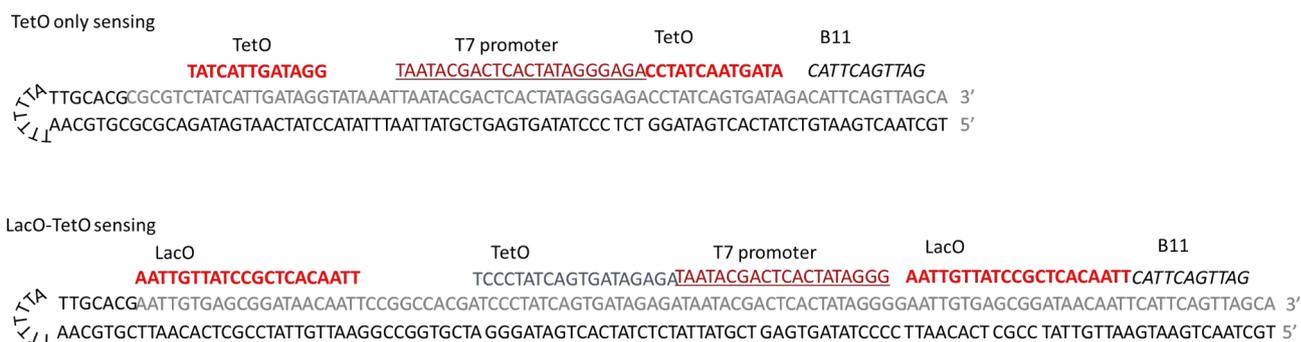
Encouraged by these preliminary results, we attempted to add the operator sequences, once again, replicating the psT approach at a transcriptional level. However, most existing transcription factor-based regulation processes are not directly competing with the T7 polymerase. We were inspired by recent work, where complex looping of DNA was used for *in vivo* and cell-free circuit control<sup>199</sup>. The authors propose the placement of two operator sites a pre-determined distance apart to allow TF-tetramer formation, as we mentioned in (0.2.6.2.2). This hides the T7 promoter, making it unavailable to T7 RNA polymerase. An additional operator site can be added in-between to provide multi-input control (**Figure 102**).



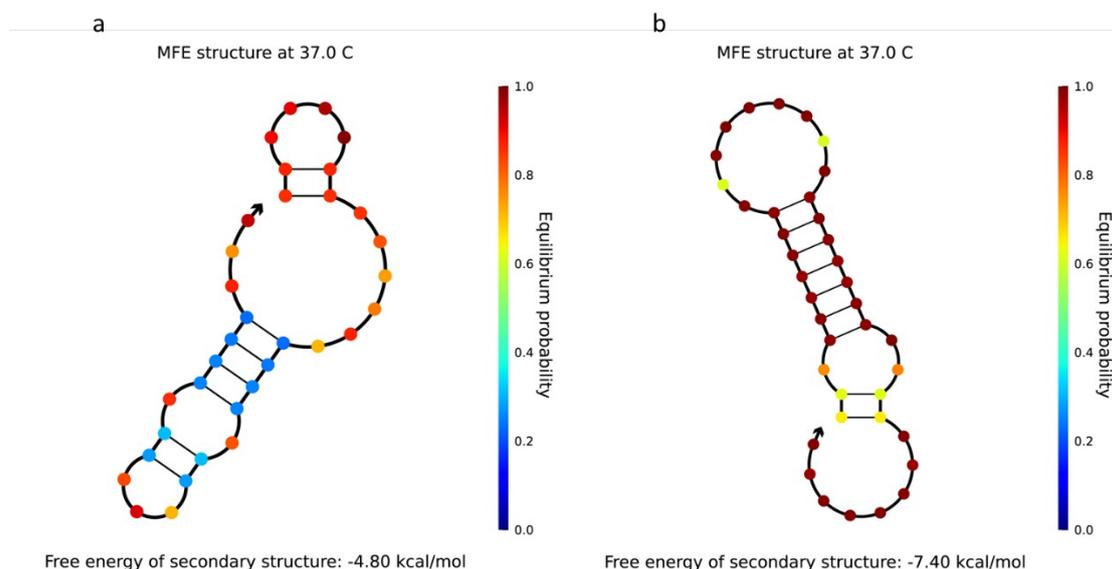
**Figure 102 Operator placement around T7 promoter.** As in<sup>199</sup>. **(A)** LacI tetramer is formed at either of the LacO1 sites flanking T7 promoter. A loop is formed once the second operator is bound. **(B)** As in **A** TetO site is added for additional input regulation.

We closely replicated these designs in our work (**Figure 103**). However, for some reason, even in the absence of transcription factors, these constructs were unsuccessful at starting the circuit (not shown) and therefore we decided not to proceed.

We speculate that formation of secondary structures within the transcript could be the source of difficulties (**Figure 104**). This is especially apparent in case of the longer variant. It could also be, that whilst RNA polymerase is able to transcribe short sequences, longer ones might not be produced due to premature RNA polymerase dissociation.



**Figure 103 T7-based sensor design.** Top: TetO-only containing sensor. Bottom: LacO-TetO based sensor. Non-chemically synthesised sequences are shown in grey and are expected to be completed by the DNA polymerase from the reaction assembly. Other sequences are aligned with the hairpins for ease of interpretation. The output is extended, compared to normal B11 to enhance complementarity to aT Here, a more complete version of the T7 promoter is used, copied from<sup>99</sup>. It includes some nucleotides that are transcribed.



**Figure 104 Simulation of secondary structures of transcripts from Figure 103.** Done with NUPACK<sup>200</sup>. (a) TetO only-sensing transcript. (b) TetO-LacI only-sensing transcript.

### 3.2.2 Nb.BsrDI for molecular program sensing

Our approach to modularity, outlined by us in (1.4.3), depends heavily on the nicking endonuclease used. The nickase defines the orientation of the recognition site relative to the operator and the output, therefore defining the actual output and the tail. We therefore wondered if we could find a better alternative to the Nt.BstNBI.

Of interest to us are nickases belonging to the N group such as the ones, typically used in the PEN toolbox. As might be inferred from the names, **Nb.BsmI** nicks the bottom strand of the site, whilst **Nt.BstNBI** – the top one. The natural nucleases of this group are typically related to heterodimeric Type II (T) endonucleases. These consist of a larger subunit with a recognition and a catalytic domains, and an additional, small subunit that only cleaves DNA in concert with the large one<sup>201</sup>. Inactivation of the small subunit, therefore, leads to nicking, instead of dsDNA cleavage<sup>202</sup>.

Alternative design strategies are possible for artificial nickases, such as decoupling of homodimer units<sup>203,204</sup> (for this group of endonucleases) and mutation of one of two cleavage sites in monomeric restriction enzymes<sup>205</sup>.

Our attention was captured by the fact, that in case of heterodimeric BsrDI and BstI enzymes, it was also possible to inactivate only the large subunit<sup>202</sup>. In context of our work, we found an old patent from 2002, speculating that Nt.BstNBI, functioning in the same way could also be reengineered to become a bottom-nicker<sup>206</sup> (**Figure 105**). This was tempting as it would allow to position the nickase recognition domain on the other side of the operator sequence, eliminating the long tail and making it context-independent.

N.BstNBI (B <sup>+</sup> C <sub>1</sub> <sup>+</sup> D <sub>1</sub> <sup>+</sup> ): natural monomer	GAGTCNNNN <sup>^</sup> N... CTCAGNNNNNN...
R.BstNBI (B <sup>+</sup> C <sub>1</sub> <sup>+</sup> D <sub>1</sub> <sup>+</sup> :C <sub>2</sub> <sup>+</sup> D <sub>2</sub> <sup>+</sup> ) hypothetical heterodimer	GAGTCNNNN <sup>^</sup> N... CTCAGNNNN <sup>^</sup> N...
N <sub>t</sub> .BstNBI (B <sup>+</sup> C <sub>1</sub> <sup>+</sup> D <sub>1</sub> <sup>+</sup> :C <sub>2</sub> <sup>-</sup> D <sub>2</sub> <sup>+</sup> ) top-strand nicker	GAGTCNNNN <sup>^</sup> N... CTCAGNNNNNN...
N <sub>b</sub> .BstNBI (B <sup>+</sup> C <sub>1</sub> <sup>-</sup> D <sub>1</sub> <sup>+</sup> :C <sub>2</sub> <sup>+</sup> D <sub>2</sub> <sup>+</sup> ) bottom-strand nicker	GAGTCNNNNNN... CTCAGNNNN <sup>^</sup> N...

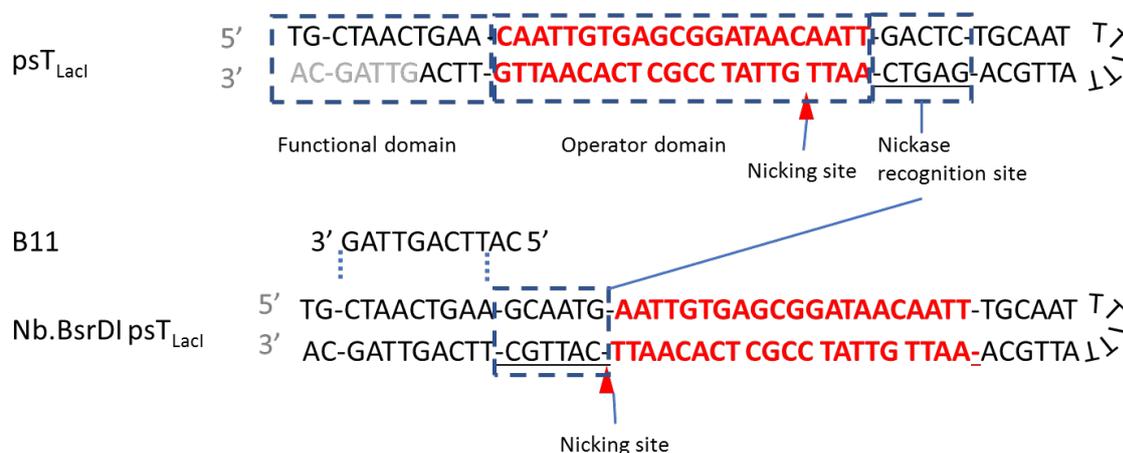
**Figure 105 Nt. BstNBI could be modified.** B – denotes a binding domain, C – a cleavage domain, D – dimerization domain. Active domains are labelled with a +, whilst inactive, with a -. Cleavage in DNA is denoted by a caret, whilst “N” represents any nucleotide.

To our surprise, we could not find a work that followed, attempting to generate Nb.BstNBI. However, a group in Russia seemed to have extensively studied BspD6I restriction enzyme. This protein is apparently an isoschizomer that is possibly related to the Nt.BstNBI. They have reported numerous characterisation experiments (for example,<sup>207,208</sup>), though the question of creating a bottom strand nicker remained. We reached one of the authors – Liudmila Abrosimova, who was kind enough to share her experience and direct us to little known work published in Russian<sup>209</sup>. In the paper they demonstrated that inactivation of the

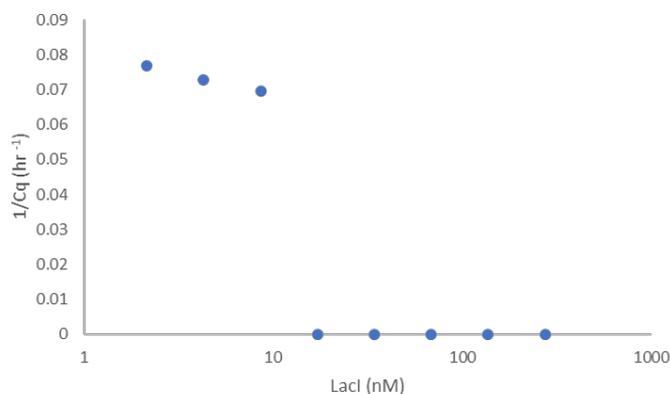
large subunit led to inactivation of the small one. We therefore decided to look for alternatives.

Our usual supplier (NEB) offered two bottom strand nickases, capable of nicking outside of the recognition domain: Nb.BtsI and Nb.BsrDI. Both require 6 bp recognition sequence, however, the latter was advertised to be functional at 65 °C, whilst the former at 37 °C. We chose to proceed with Nb.BsrDI as our insights could be generalised to other programs, running at higher temperatures.

We then designed the appropriate psT, basing the design on the original one but here, of course, the recognition site was located on the trigger side of the operator sequence. This results in a fixed-length short tail added to the trigger irrespective of the operator site or system used (**Figure 106**). A CG pair on the end the operator site (left hand side on the figure) was removed as it was originally incorporated into the design by accident although not essential.



**Figure 106 Nb.BsrDI-driven psT<sub>LacI</sub> in detail.** Here, original psT<sub>LacI</sub> is shown for comparison. Concerning Nb.BsrDI-psT<sub>LacI</sub>, going from right to left along a hairpin structure: polyT region and 6 bp, providing hairpin backbone; operator domain for TF binding (red); Nb.BsrDI recognition site (underlined), reaching into the operator site (red triangle); functional region consisting of the modified B11 signal sequence (shown standalone as well). The part of the sequence that was not chemically synthesised, but filled in by DNA polymerase upon reaction initiation, is shown in gray. Dashes are used for ease of interpretation.

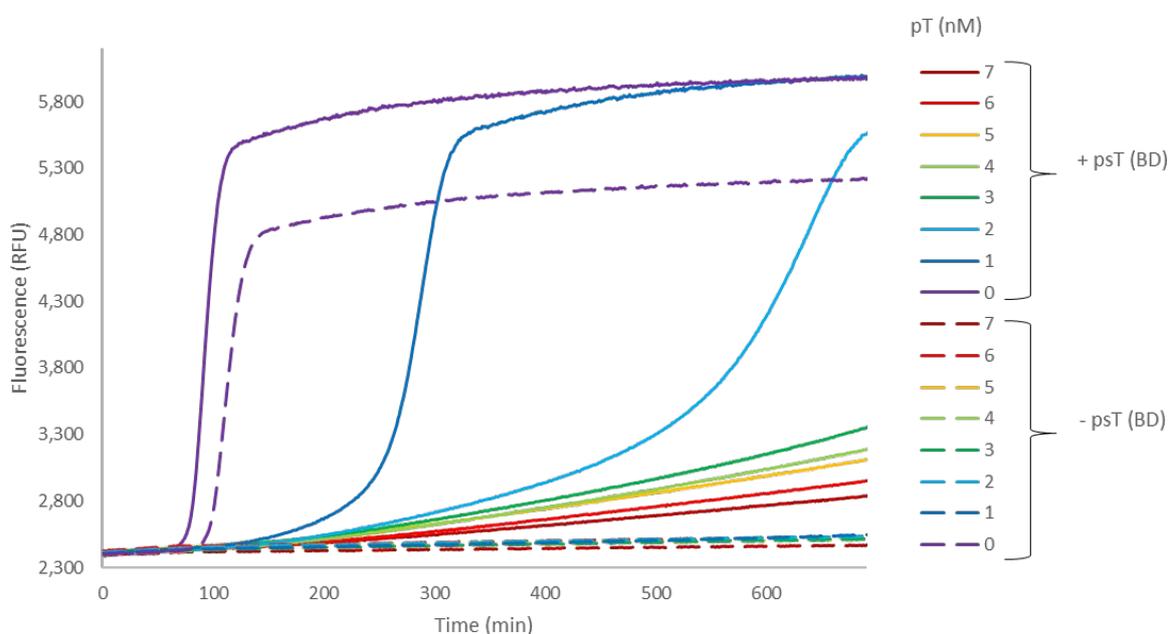


**Figure 107 Range of Lacl with a Nb.BsrDI-driven psT<sub>LacI</sub> switch.** Time-response to a range of Lacl dimer added to the switch with 5 nM of Nb.BsrDI-driven psT<sub>LacI</sub>.

We first verified that LacI binding still interfered with the nicking activity in the new orientation. We did that by assembling the a circuit with 5 nM of the Nb.BsrDI-driven psT<sub>LacI</sub>, 1 % (100 U ml<sup>-1</sup>, NEB) of the commercial Nb.BsrDI stock solution and a range of LacI concentrations.

Interference is observed and the concentration required is broadly similar to that required for our original psT. However, the start times in general are a lot slower. This could be due to slow activity of Nb.BsrDI at 37 °C, possibly related to non-specific binding to DNA. **(Figure 107)**

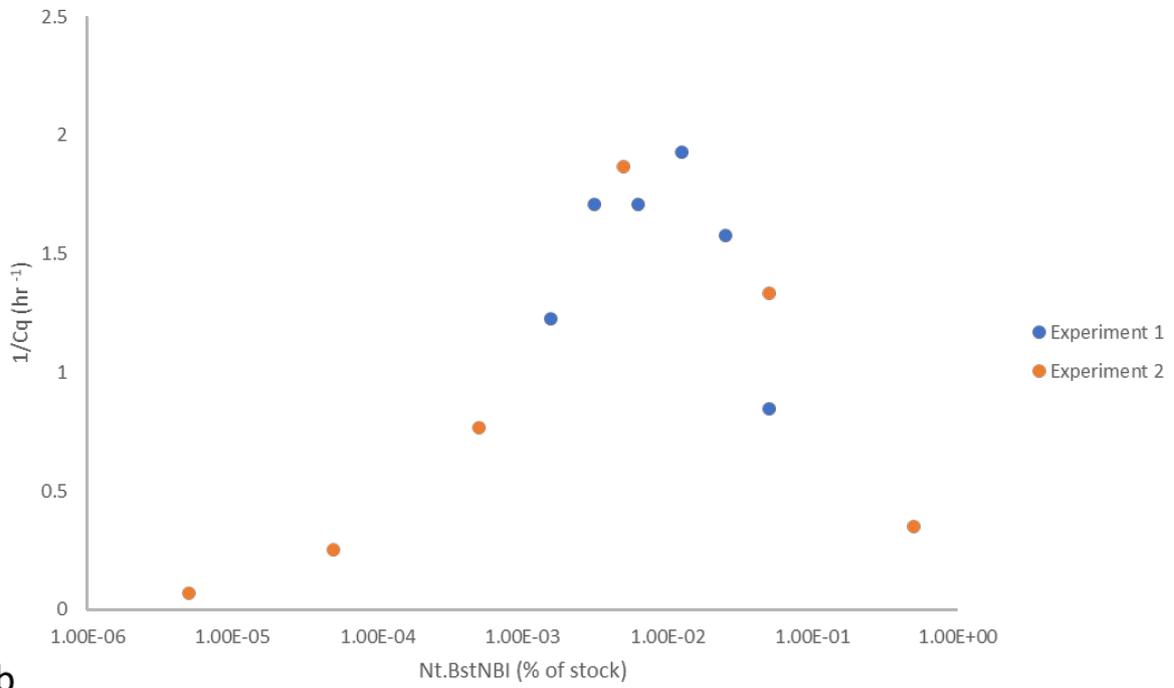
Speculating that Nb.BsrDI was slowing down the circuit, we tried to decrease the pT concentration, since this component can have a similar effect. Indeed, the system remains bistable at much lower concentrations of pT (1 nM vs 7 nM previously). This experiment confirmed that Nb.BsrDI indeed retards the entire circuit, decreasing the need for a sink **(Figure 108)**.



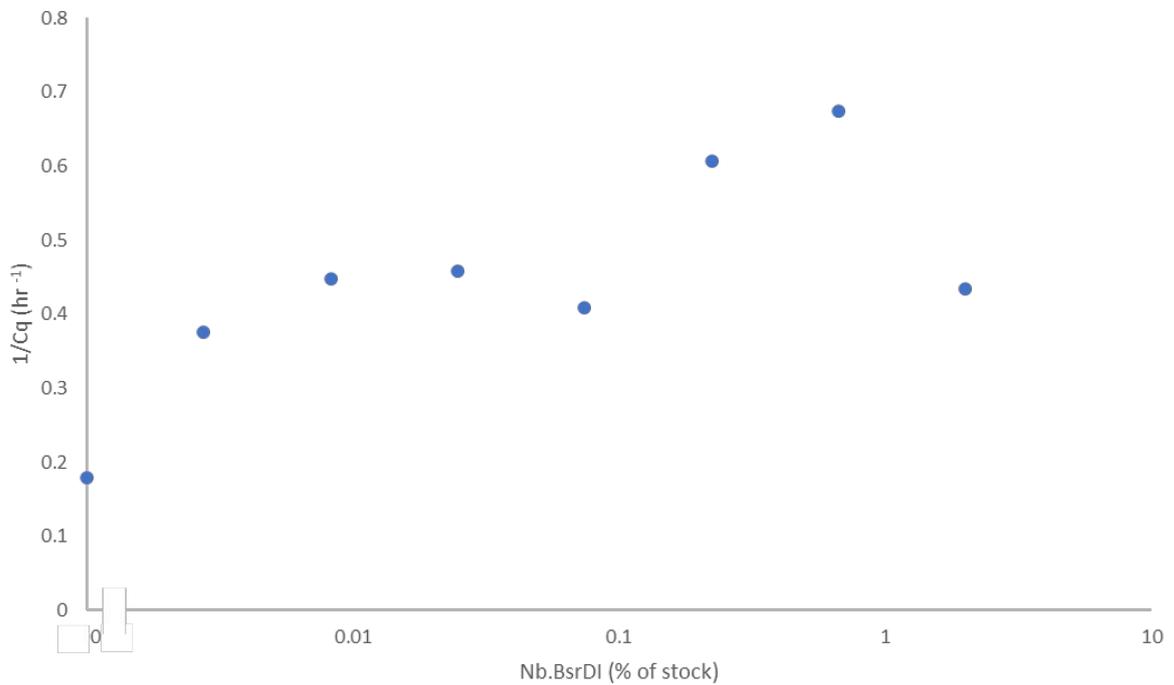
**Figure 108** Range of pT with a Nb.BsrDI-driven psTLacI switch. Cy5 fluorescent trace for pairs of samples either containing or not 5 nM of the psT, going from 7 to 0 nM of pT.

A more productive approach to avoid future instability problems was to optimise the concentration of Nb.BsrDI. We have observed before that the thermophilic Nb.BstNBI had an optimum concentration, below which and, more interestingly, above which, the circuit slowed down **(Figure 109a)**. The same seems true for Nb.BsrDI, although higher percentage of the commercial stock has to be used. Nonetheless, according to NEB, the stocks are equivalent in “units of activity” concentration **(Figure 109b)**. Here we have arbitrarily taken the 0.05% of the commercial stock concentration forward, as all samples apart from the negative control and 2% showed activity of comparable scale.

a

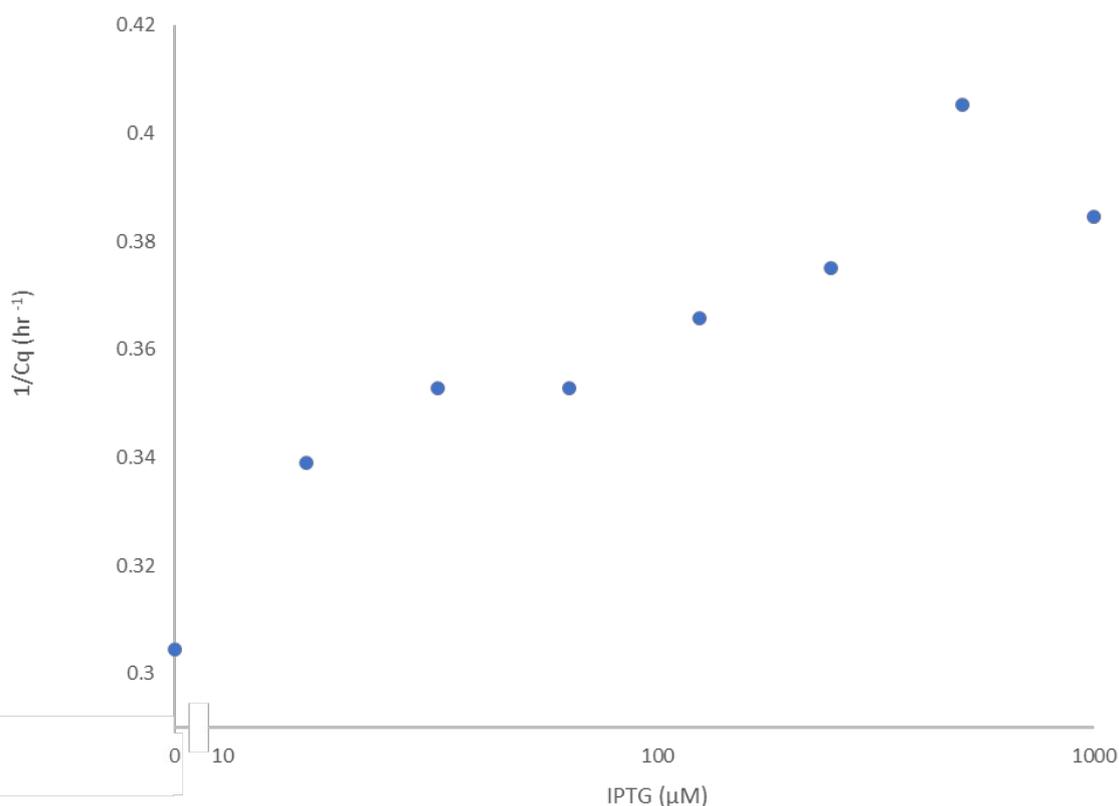


b



**Figure 109 Nickase concentration effect on the switch.** (a) Range of Nt.BstNBI concentrations expressed as percentages of the commercial stock (NEB). Two experiments were carried out in identical conditions for different concentration ranges. (b) Range of Nb.BsrDI concentrations expressed as percentages of the commercial stock (NEB).

Once the system was optimised, we tested if the sensitivity of the system to the binding transcription factor could again be used to sense its allosteric effector. The mix was therefore assembled with a constant concentration of Lacl (110 nM, as in prior experiments to maintain consistency) and we performed a range of IPTG. We were able to resolve IPTG concentrations, as in the case of Nt.bstNBI-based psT, demonstrating the portability of the approach (**Figure 110**).



**Figure 110 Nb.BsrDI-psT<sub>LacI</sub>-based IPTG detection.** Time response to a range of IPTG.

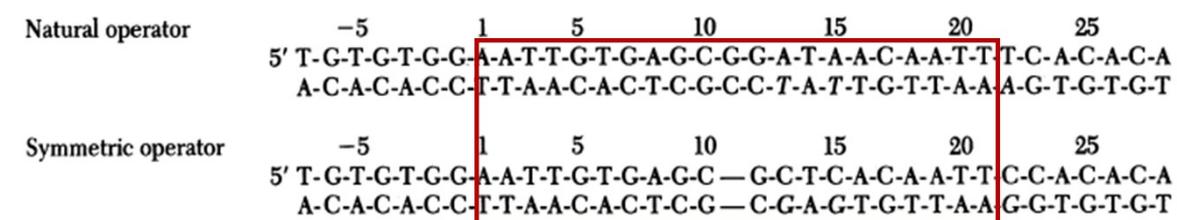
Although the Nb.BsrDI approach decreases the sequence constraints, and the requirements for long tails on the output oligonucleotides, actually it does not fully eliminate the context dependence. Indeed, it is known that the spacer bases between the recognition and the nicking sites in fact can have a strong impact on the nickase activity<sup>210</sup>. This is echoed by reports of Antoine Masurier and Guillaume Gines of our group, relating to Nt.BstNBI that are not yet published. However, once properly optimised we believe this design could be more promising than the one we have explored before.

### 3.2.3 Chemical optimisation of the pskT

An alternative to the architectural approaches described above, a potentially simpler way to address the slow production of the psT product, is chemical or sequence modification of the hairpin template to decrease the melting temperature of the output. Below we present several ideas (**Figure 112**).

First of all, we noticed that the poly-T tail, directly copied from the pT sequence is irrelevant in context of our pskTs. The operator sequence fragment on the tail of aB11 insures templated extension and irreversible binding of the trigger. We therefore removed these as well as a seemingly non-essential C-G base pair accidentally incorporated in the previous designs. This yielded “original S” design, S for “shortened”.

Then, we were tempted to increase the  $K_{of}$  of Lacl to the operator site by using an artificial sequence, reported to have a higher affinity for Lacl. In particular, it was noticed that Lacl operator site we are using is close to being symmetric. As Lacl binds as a dimer of two identical subunits it was hypothesized that one half of the operator could be better, than the other. Several variants were tried, and the inverted repeat of the left half of the operator indeed increased the tightness of binding by as much as 10-fold<sup>211</sup>. We have therefore replaced the natural Lacl operator sequence with the symmetrical one (original S (SymmO)).

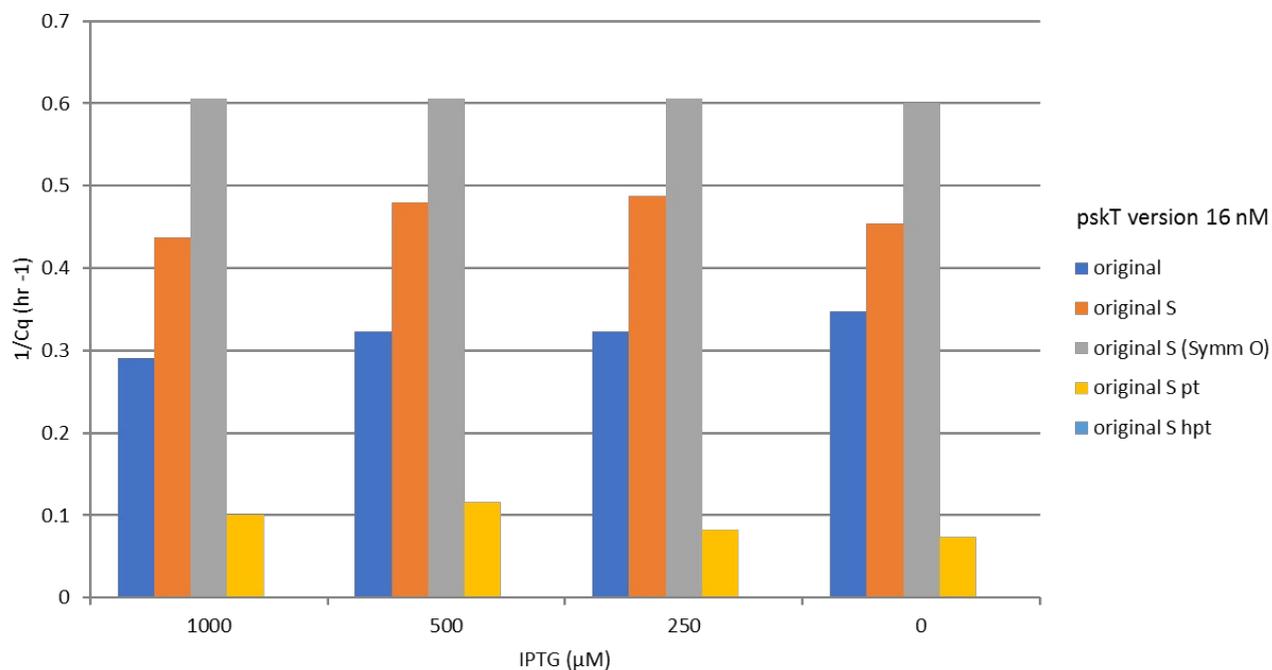


**Figure 111 Natural and symmetric Lacl operators.** Adapted from Sadler et al. Above: the natural operator sequence. Below: novel symmetric operator. Red box highlights the sections used in this work.

Furthermore, regarding chemical modifications it was reported that phosphorothioate modification of the nucleotides (PTO)<sup>212</sup>, readily available and used in this PEN circuit for 5' exonuclease protection, can decrease the melting temperature of the NA duplex. This was, for example, utilised in some other isothermal works<sup>213</sup>. We have therefore designed a sequence with such modifications on the template sequence of aB11 to facilitate dissociation (original S pt). We also explored a variant carrying some PTOs in the operator site, where we tried to avoid TF-interacting residues by visually inspecting the 3D structure of the bound TF available<sup>214</sup> (original S hpt).



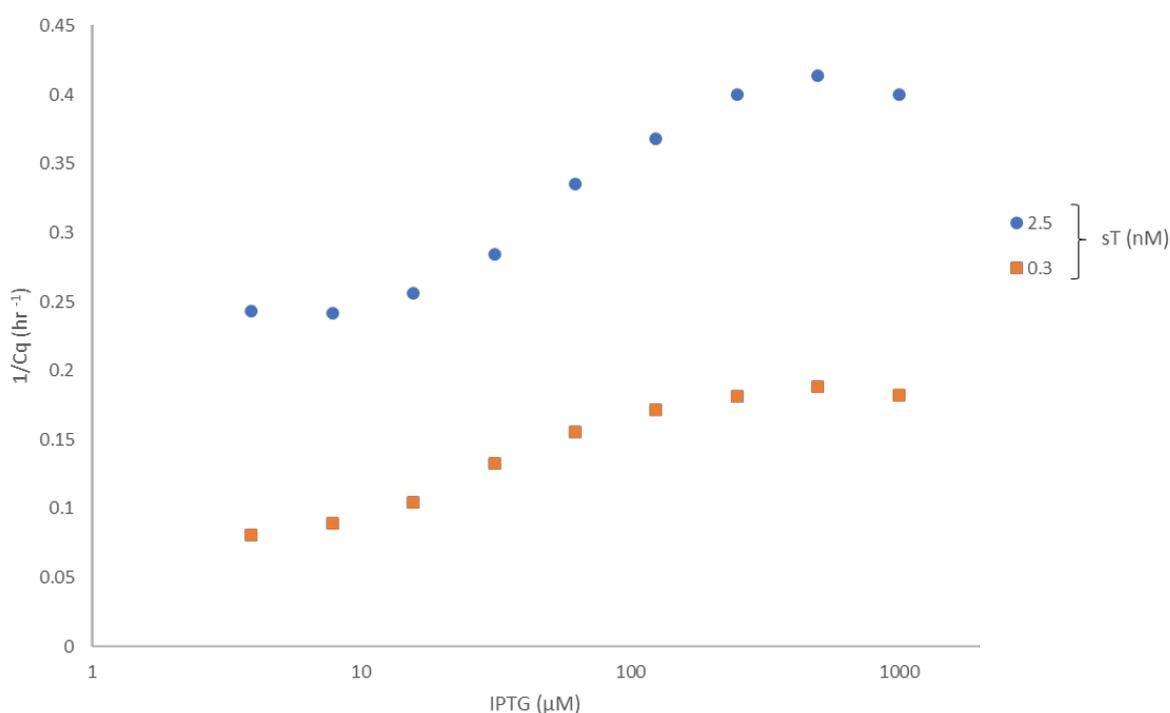
**Figure 112 pskT<sub>lacI</sub> variants tested in this work.** As before, underlined is the nickase recognition domain, with the nicking site shown by a red triangle. Operator site is in red, followed on the left by the aB11 sequence. Top to bottom: original pskT used in the rest of this work (original); pskT shortened by 3 bp between the operator and the trigger binding sequence (original S); as original S, bar with artificial symmetrical operator site (Symm O); as original S bar with phosphorothioate (PTO) modifications of the aB11 template strand, shown with asterisks (original S pt); as original S pt bar additional PTO modifications within the operator site (original S hpt). In original S hpt, slashes designate bases we designated as interacting with the transcription factor and therefore not modified them.



**Figure 113 Chemical variants of pskT<sub>lacI</sub> detecting IPTG.** IPTG concentrations of 1000 – 0 μM detected with various pskTs as shown in the legend.

To test these designs, we replaced the original pskT with all of the modified variants designed (at a relatively high concentration of 16 nM), otherwise leaving the reaction mixture as is and adding a range of IPTG. The modifications worked to various degrees of success. Interestingly, deletion of 3 nucleotides from the tail or change of the operator site appear to make the pskT less efficient at slowing down the circuit. However, in these conditions, the PTO variants seemed to be much stronger at slowing down or even completely killing the circuit (**Figure 113**). We therefore explored the strongest variant (“original S hpt”) further.

We proceeded to detect a range of IPTG with this variant (**Figure 114**). Compared with the results reported before, here 10 nM of pskT<sub>Lacl</sub> functioned with nM concentrations of sT (instead of 7 pM used in 1.3.2). In our opinion this demonstrates that chemical modifications also provide a way to optimize the design of the sensing module.



**Figure 114** IPTG detection by pskT<sub>Lacl</sub> “original S hpt”. The pskT acted against either 2.5 (circles) or 0.3 (squares) nM of sT.

### 3.3 Potential for circuit expansion

Previous scalability and predictability of PEN-circuits has been demonstrated before so we decided not to focus on these points. However, we wanted to verify that these properties are maintained at 37 °C. In this section, we will therefore report some additional components we have tested.

#### 3.3.1 B11b – an orthogonal switch

In parallel with the B11 trigger sequence-based switch that we have used throughout this text, an alternative – called B11b – was also tested (**Table 2**). This was originally done in

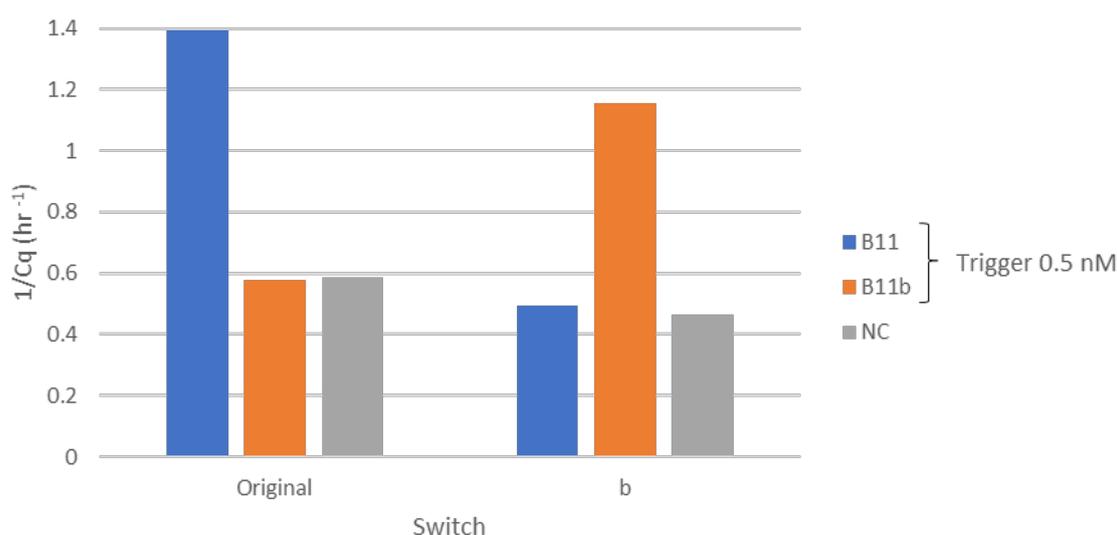
case the first switch would fail, or create some unexpected sequence constraints. However, as this did not happen, we used it in some explorative work.

B11	CATT <b>CAG</b> TTAG
B11b	CATT <b>CAAG</b> TTAG

**Table 2 Trigger sequences of B11 and B11b.** The sequence difference is highlighted.

The design space for generating trigger sequences is relatively constrained: small length (11 bp) is necessary for a reasonable melting temperature out of which 6 bp form the Nb.BsmI site.

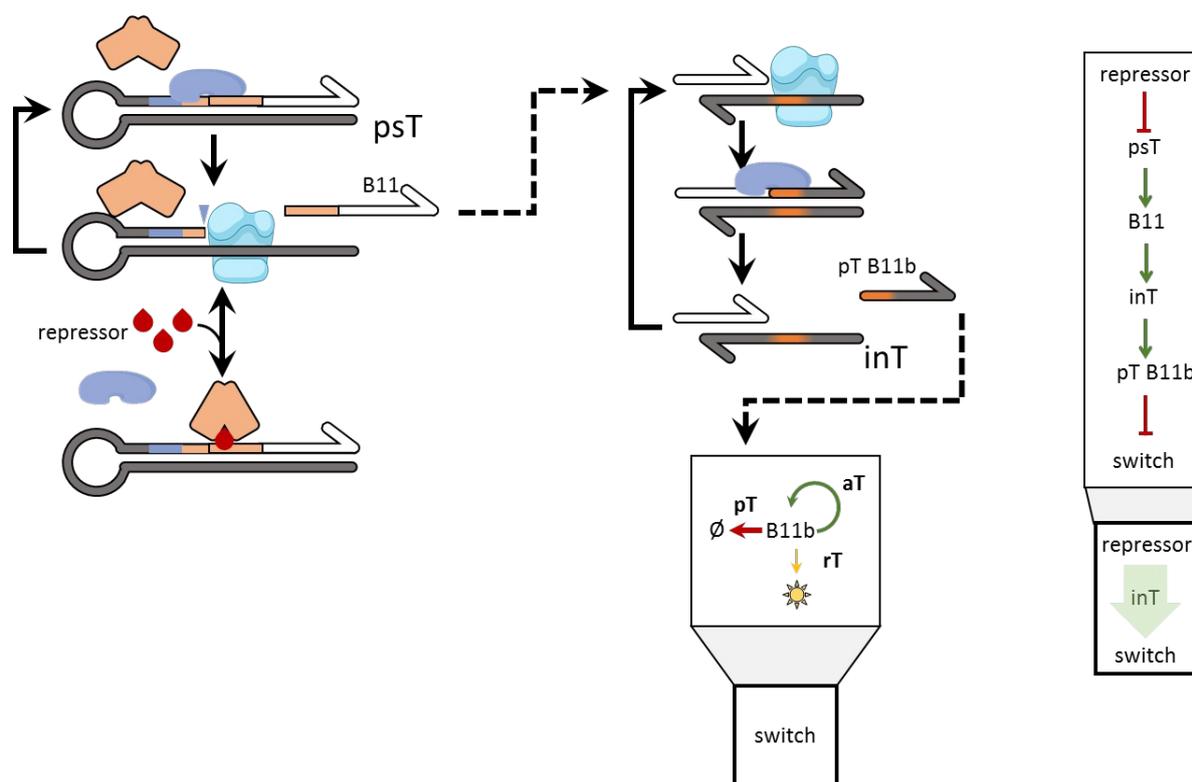
In the present case, B11b differs from B11 by 2 adjacent base pairs. We wondered, whether this would be sufficient to avoid cross talk. To this end we assembled both switches (see Materials for details) and added either the appropriate or the alternative trigger sequence at 0.5 nM final concentration. Satisfyingly, addition of the “wrong” trigger resulted in start times similar to those for adding water (**Figure 115**). This implies that the two switches are largely orthogonal at sequence level (they are not at machinery level, since they both use the same enzyme to function).



**Figure 115 B11 and B11b are orthogonal.** 1/Cq of B11 and B11b-based circuits similarly constructed. 0.5 nM of either B11, B11b trigger or water (NC) is added.

### 3.3.1.1 An alternative strategy for inverting the signal

In parallel to the creation and testing of pskTs, reported above, we considered an alternative approach to invert the signal. In this design, the small molecules would be sensed by a regular psT and the signal would then be inverted within the program by a separate template (**Figure 116**). Whilst involving additional parts, we thought this template, that we called the inverter template (inT), could allow for better modulation of the input sensing.

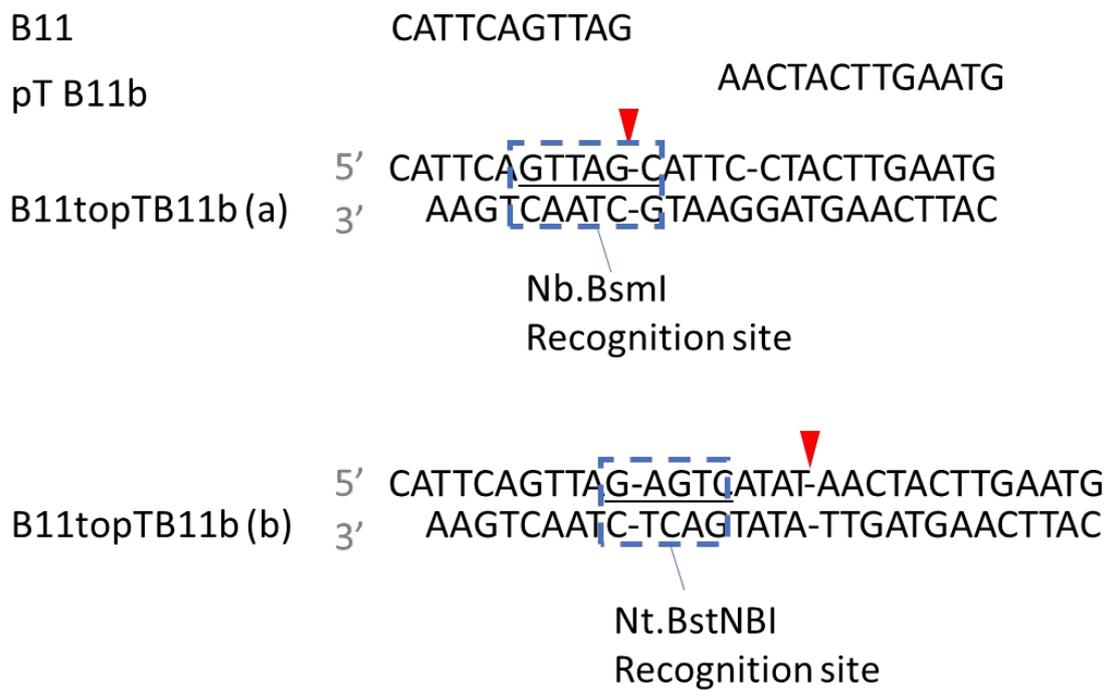


**Figure 116 Inverter template (inT) mode of action.** psT senses the absence of repressor, as in **Figure 47**. Trigger sequence (operator tail omitted for simplicity) binds the inverter template (inT). Cycles of extension and nicking lead to production of pT B11b that contributes to overall pT B11b concentration and inhibits the circuit. On the right, logic summary of the circuit.

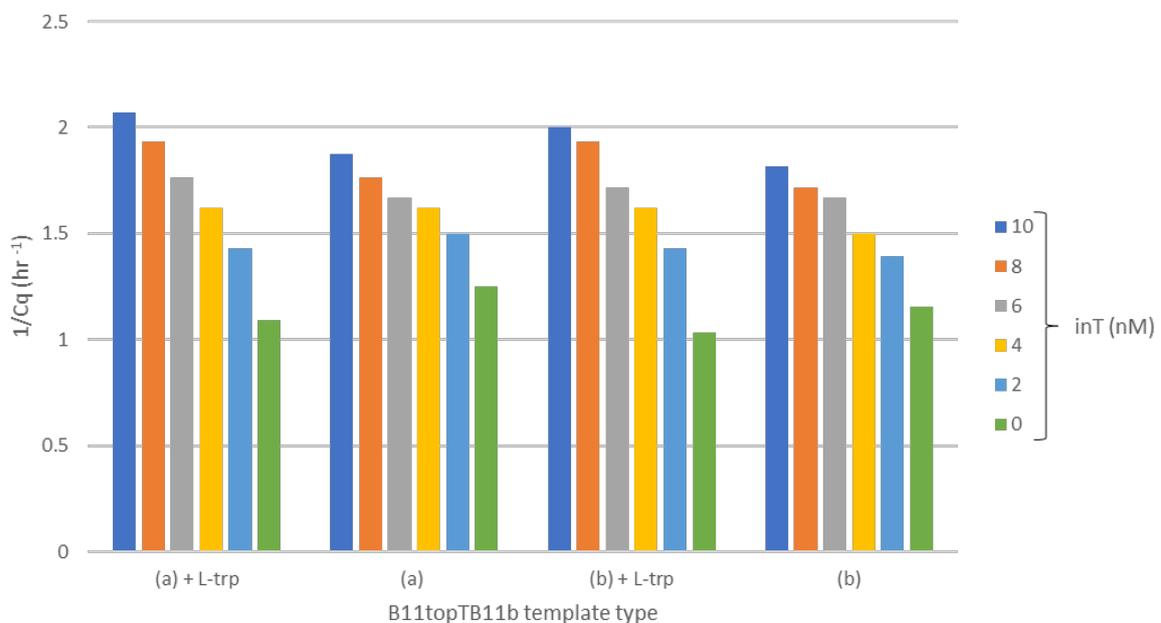
In particular, we first considered two designs based on either of the PEN-nickases driving the conversion (**Figure 117**). We denoted the templates “B11topTB11b” as they convert B11 signal to pT of B11b switch.

Variant “a” is Nb.BsmI driven: in this case the nicking site is mostly on the input signal potentially allowing to almost directly follow with a pT B11b sequence, although we chose to add a spacer. In this case, the B11 input could dynamically bind and unbind upon nicking cycles.

Alternatively, Nt.BstNBI used for variant “b” requires longer inverter template due to the need to incorporate the recognition site and a 4 bp spacer before the nicking site. This template is irreversible (extended inputs will not de-hybridize dynamically). This approach also uses orthogonal enzymes for signal conversion and driving on the downstream switch.



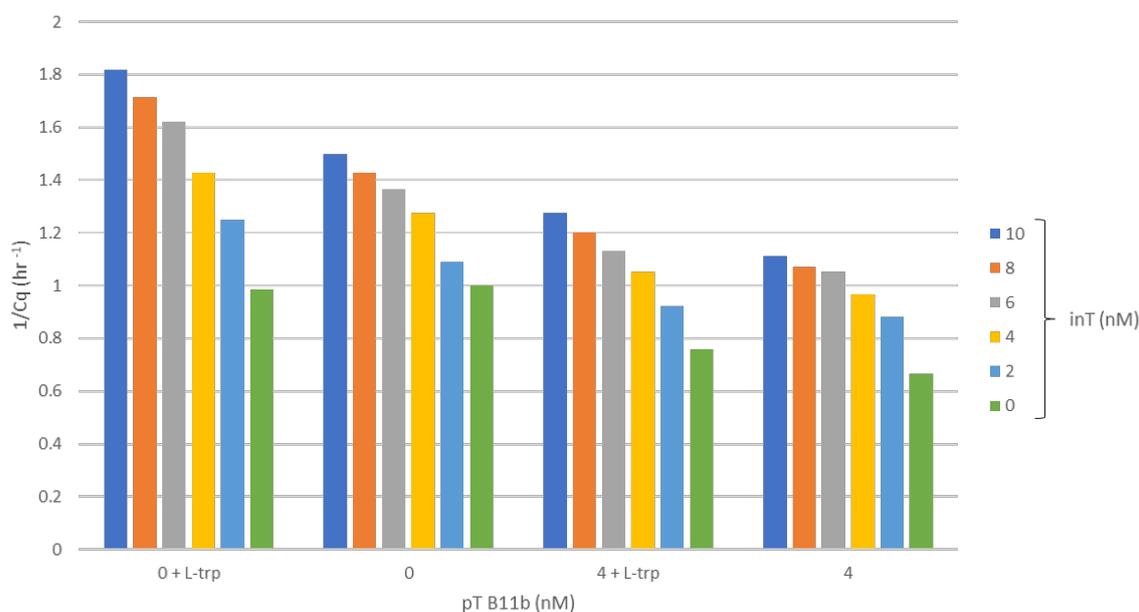
**Figure 117 Inverter template design variants.** Inverter templates (inTs) driven by Nb.BsmI (B11topTB11b (a)) and Nt.BstNBI (B11topTB11b (b)). B11 and pT for B11b-based circuit (pT B11b) are shown for reference. The (a) variant does not fully match pT due to an error in design. Nickase recognition domains are highlighted by dashed boxes, whilst the nicking sites are indicated by red triangles.



**Figure 118 Inverter template variant tests.** Ranges of concentrations of the inverter template (10-0 nM) are shown for B11topTB11b (a) and B11topTB11b (b), with (+ L-trp) and without 1 mM of L-trp added.

We tested both at a range of concentrations with 5 nM psT<sub>TrpR</sub> and 100 nM TrpR added but without pT B11b at this stage (therefore, just using aT B11b template) (**Figure 118**). This was done with and without 1 mM of L-trp to examine, whether these could be differentiated. We were perplexed to find that both templates increase the rate at which the circuit starts with increase in concentrations. Nonetheless, in both cases higher concentrations of inT allowed to distinguish negative and positive L-trp controls. The change appeared more noticeable for the “a” variant and we have therefore taken it forward.

We repeated the assay for “a” variant, this time comparing 0 and 4 nM of pT-B11b added, speculating that it could improve stability and resolution of the circuit (**Figure 119**). We replicated the increase in start time as a function of concentration of the inT, confirming the surprising prior result. We likewise still differentiated L-trp – containing tubes. Addition of pT-B11b expectedly slows down the B11b-based switch in all cases. However, the improvement in resolution of L-trp detection is non-evident. In particular ratio of negative to positive control start times, which we use as one of the scores to evaluate sensing circuits in the group changed for the worse (by -2.85 % average). When subtracting 1 and dividing such ratio by the positive control start time again (for normalisation and to favour faster-running sensors) the pT B11b also worsened the score. This is why we decided to focus on the pskT concept presented before. However, it is possible that better optimizing the inT approach could yield satisfactory results.



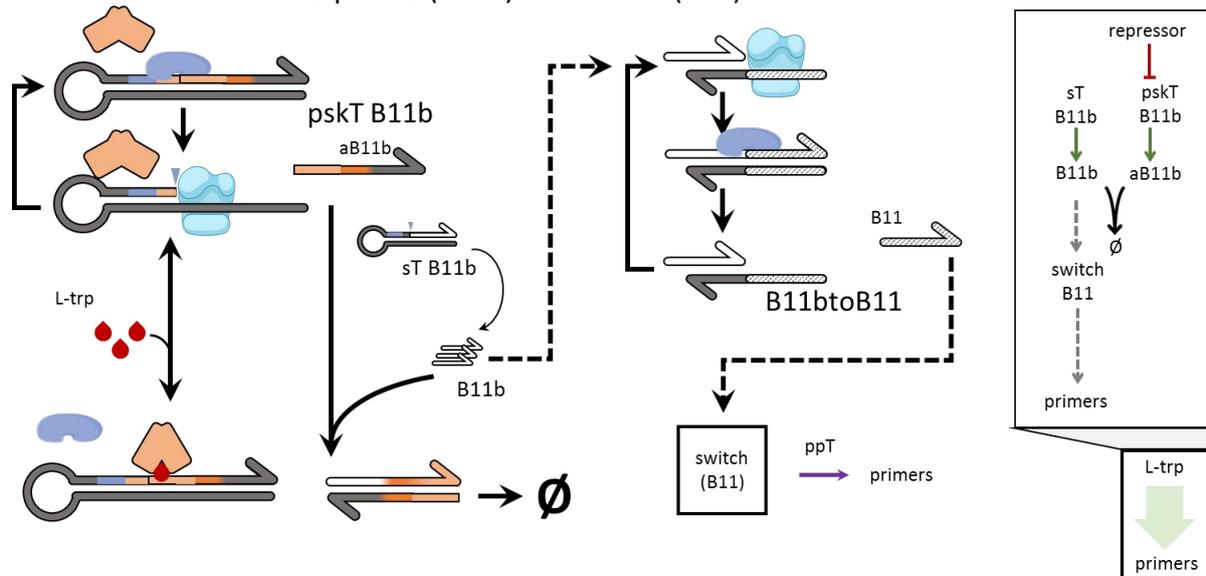
**Figure 119 B11topTB11b (a) test with pT.** Range of concentrations of the inT (10-0 nM) with either 0 or 4 nM pT B11b, with or without 1 mM L-trp.

### 3.3.1.2 Sensing separately from direct actuation

When using multiple components (rT, ppTs, aT, and pT) all binding to some extent to the same B11 sequence, the design optimisation space became increasingly difficult to navigate. A slight change in B11 dynamics, caused by sT/pskT alteration propagated to non-trivial changes affecting the other components all dependent on the B11. We therefore

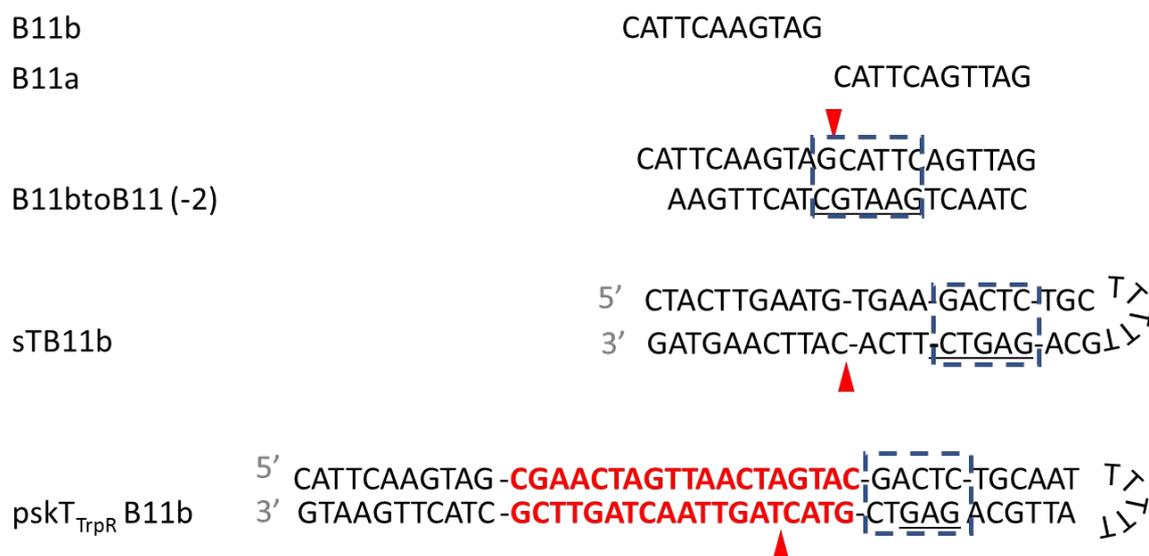
examined again the idea of using two switches when constructing the circuit, used in 2.3.3.1 for IPA control.

To implement this decoupling, we envisioned that a B11-based switch would control the downstream ensemble and would be fixed in terms of component concentration (**Figure 120**). However, its activation would depend on orthogonal B11b-based input, in turn dependent on L-trp activation. An intermediate converter is then necessary to transfer the information from one sequence (B11b) to the next (B11).



**Figure 120** B11b sensing circuit linked to B11 ppT. B11b-based  $pskT_{TrpR}$  competes with sT B11b as before (**Error! Reference source not found.c**). When L-trp is present,  $pskT_{TrpR}$  is suppressed and excess B11b binds B11btoB11. This produces B11 to trigger the switch that then produces primers as before.

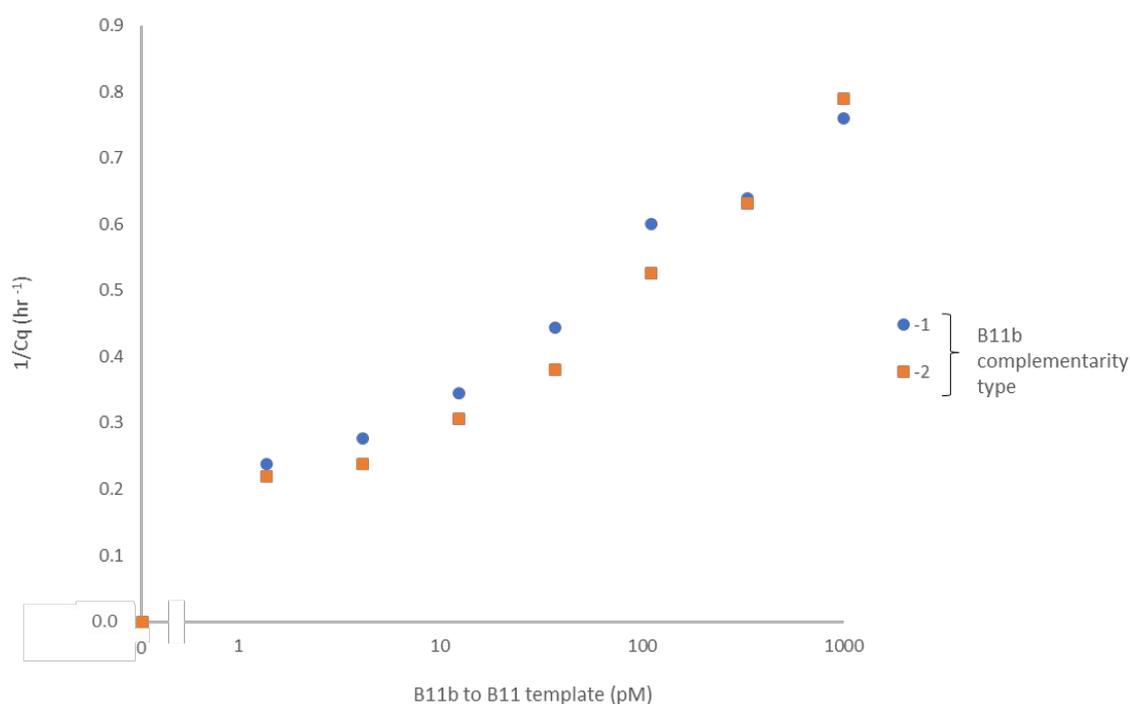
We designed the components missing to construct the circuit (**Figure 121**). Conveniently, only three new parts were required, two of which: the B11b-based sT and  $pskT_{TrpR}$ , only required small output modification. B11btoB11 was ordered in two variants: “-1” and “-2” (lacking 1 or 2 bp of complementarity accordingly) to disfavour B11b binding and direct the latter to aB11b instead.



**Figure 121 B11btoB11 design and accompanying templates.** B11btoB11 driven by Nb.BsmI, here with two nucleotides missing for full B11b complementarity (-2). sT B11b, generating B11b and pskT<sub>TrpR</sub> B11b are similar to those for B11 shown before. B11 and B11b are shown for reference. Nickase recognition domains are highlighted by dashed boxes, whilst the nicking sites are indicated by red triangles. TrpR binding site is shown in red.

We moved on to assemble the circuit in IPA buffer, as reported before. However, at the first stage of this process (adding the B11btoB11 template to the B11 switch) we noticed that even a concentration of about 1 pM of the template activated the B11 switch in absence of the B11b input (**Figure 122**). The starting rate appears to near saturation above 1 nM (less obvious in 1/Cq representation), whilst the decrease in either template concentration slows down the switching. “-2” variant was less leaky than “-1”, as we would expect in potential future implementations, but only slightly. It is not clear why the circuit exhibits such a high degree of leakiness.

This could be addressed by two means both unsuitable for us: either further addition of pT that would burden an already delicate B11 circuit or by decreasing the B11btoB11 template concentration. We were anxious of the latter option as concentrations below pM concentrations could lead to stochastic effects in emulsion. A rough estimation for a droplet of 5 pL (typical for a microfluidic device) containing 1 template molecule means around  $2 \times 10^8$  molecules in a litre, hence concentration of around  $3 \times 10^{-13}$  M or 0.3 pM. Of course, the real limit would need to be more carefully calculated and would be guided by Poissonian statistics. However, we decided to not take this design forward. Possibly, switches with greater orthogonality could be of use in the future.



**Figure 122 Range of B11btoB11 template concentrations in IPA buffer.** Results for “-1” and “-2” variants starting the circuit are shown with circles and squares accordingly.

### 3.3.2 Nb.BssSI can be used to drive a circuit

Nb.BsmI is typically used to drive PEN-toolbox switches for historical reasons. However, the expansion of nickase repertoire could allow for greater multiplexing within the same assembly. At the same time, when the 37 °C B11 and B11b-based switches were designed we wondered, whether Nb.BsmI, used at high concentrations in typical 50 °C PEN reactions would be sufficiently active at 37 °C. To this end, we considered Nb.BssSI – another commercially available nickase, as an alternative. Nb.BssSI is similar in properties to Nb.BsmI but the incubation temperature, recommended by the provider is 37 °C<sup>215</sup>.

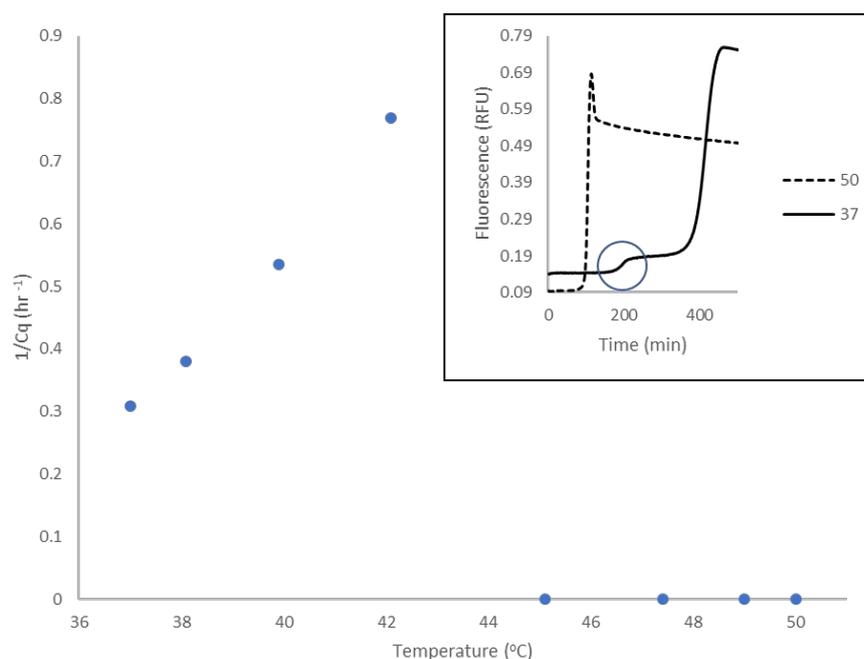
We applied identical design principles to those used in the other switches reported but replaced the 6 bp nicking site by that of Nb.BssSI (**Figure 123**). The trigger sequences were denoted as “Bss X” – an shorthand for the enzyme used, followed immediately by a variant number (X) as the amount of base pairs was not expected to change in other designs. Consequently, the aT and pT were denoted, for example, as “aT Bss 1”, in case of the aT to be used with Bss 1. As with the Nb.BsmI-driven switches, the sequences are constrained by the nicking site and the preference of lower melting temperature. This led to two very similar sequences.



**Figure 123 Nb.BssSI-driven switch components design.** Two sets of trigger (Bss), pseudo template (pT Bss) and autocatalytic template (aT Bss) driven by Nb.BssSI are shown. Dashed box represents the Nb.BssSI nickase recognition sequence, whilst red triangle marks the nicking site.

We tested the autocatalytic modules together with commercial Nb.BssSI to verify, whether amplification could be achieved at 37 °C or higher temperatures. To this end, we omitted the normal nickases in our otherwise usual miR mix. The only templates added was one of the aT Bss templates at 50 nM and Nb.BssSI at 2.5 % of stock or 500 U ml<sup>-1</sup> as reported by NEB (**Figure 124**). Whilst aT Bss 1 samples diverged without switching (not shown), aT Bss 2 demonstrated expected sigmoidal switching before diverging at temperatures up to 42.1°C.

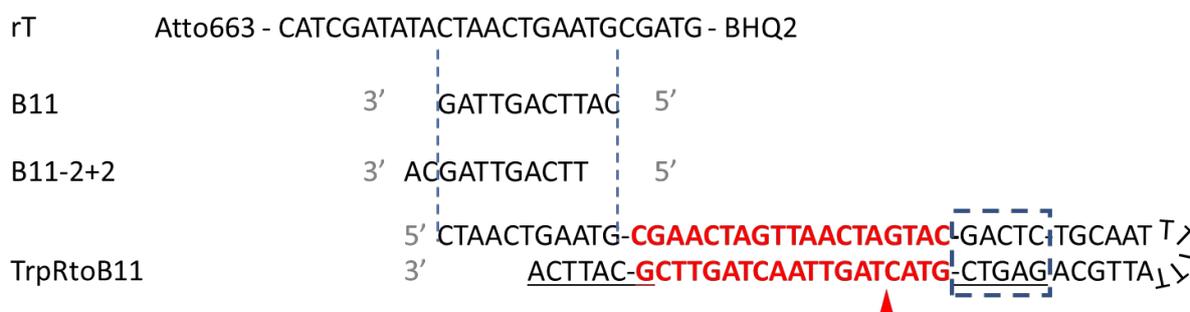
Although we did not proceed with addition of the pT to achieve bistability and further optimisation, we believe these results demonstrate a proof-of-principle. Later, Nb.BssSI-based circuits were reported by our colleagues<sup>216</sup>. An interesting property noted by Urtel *et al.* is that the nicking site is encoded only by 3 nucleotides (adenine missing in the nicked strand). This allowed them to remove dATP from an isothermal amplification circuit, compatible with PEN they built. Removal of dATP lead to suppression of uncontrollable divergence. In brief, the authors argue that untemplated amplification typically relies on extensive AT repeat sequences, formation of which is impossible in absence of dATP.



**Figure 124 aT Bss 2 can drive switching in miR circuits.** Main plot:  $1/Cq$  against temperature for 50 nM of aT Bss 2 with 500  $U\ ml^{-1}$  of Nb.BssSI.  $1/Cq$  was determined manually to avoid divergence amplification instead of the controlled one.  $1/Cq=0$  represents samples that diverged without observable switching. Inset: raw time-fluorescence curve of 37 (solid) and 50 (dashed) °C samples. Circle indicates the behaviour considered as switching.

### 3.3.3 Linear response to sensing

Although indirectly suggested by the work before, we wanted to verify that immediate linear output in response to the small molecule input was possible. Such system would involve a simple psT directly activating a fluorescent probe, for example. However, as our probe was designed to specifically match the full B11 trigger sequence and not the “-2+2” version (which we used because it has better complementarity for aT than for pT), we designed a separate psT “TrpRtoB11” (Figure 125). As the name would suggest, it responds to TrpR unbinding by producing full version of the B11. We assumed the mode of action would be similar for LacI and decided to not investigate it.

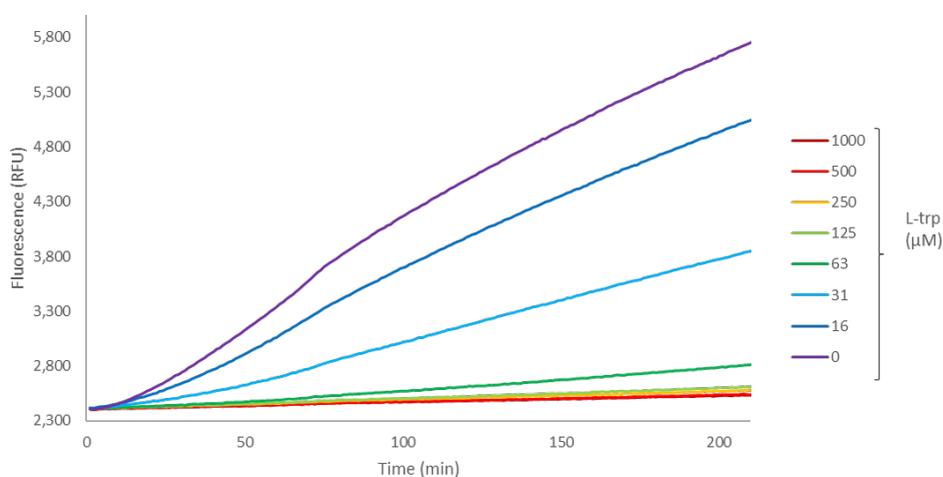


**Figure 125 Linear response template designs.** Concerning TrpRtoB11 as before in Figure 52 but the output sequence is identical to B11. rT carries Atto663 fluorophore modification on the 5' end and BHQ2 quencher on the 3'-end. On the diagram these are separated, as would happen upon trigger sequence binding and extension. Dashes are used for ease of interpretation.

When 500 nM of rT is combined with 5 nM of TrpRtoB11 in a miR assembly lacking other templates, linear increase in fluorescence of various gradients is observed in response to varying L-trp concentrations as we would expect (**Figure 126**). This explicitly confirms, that we can complement switch-like response of normal PEN-circuits with linear elements, if needed.

This can be of relevance in case of small molecule detection unlike in previous uses of PEN. First, let us consider previous PEN-circuits with nucleic acid inputs, for example, micro RNA. The detecting template is designed to have around 20 bp complementarity ensuring strong binding of the miRNA to the template (i.e., low  $K_d$  in sub nanomolar region). Once bound, linear output is established (in proportion to the quantity of miRNA bound) and once a threshold is passed, the signal is amplified to become detectable. Here, exponential amplification is essential to increase the initial signal, possibly consisting of pM concentrations of linear output, to a fluorescently detectable output.

In our case, the initial linear output is governed by the transcription factor behaviour, determined by the binding curve to the ligand, and the respective  $K_d$ . The response can be affected by the ratio of the transcription factor to the sensing template to some extent as before. Also, sufficient concentrations of the sensing and reporter templates are necessary. However, as the system is intrinsically limited by the  $K_d$  of the transcription factor, exponential amplification is not imperative in this case. Hence, alternative selection functions are possible in contrast to the switch-based, that could often lead to a more “yes/no” kind of selection.



**Figure 126 Linear response to L-trp sensing.** Time-fluorescence trace in Cy5-channel as response of TrpRtoB11-rT system to a range of L-trp (1000-0  $\mu$ M).

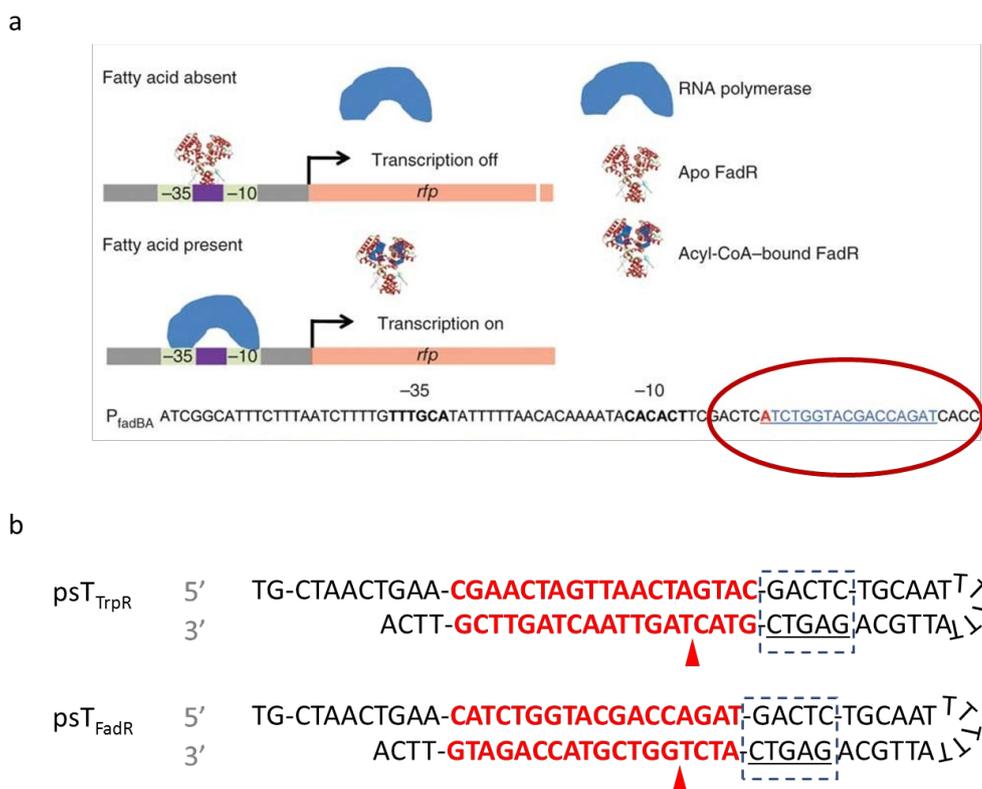
### 3.3.4 Using other transcription factors

We primarily focused on the use of TrpR and LacI as a proof of principle, however, we consider our approach generalisable to other transcription factors. TrpR was chosen from a table of examples of TFs used in biotechnological applications *in vivo* reported by Mahr and Frunzke<sup>217</sup>. Admittedly, not all of the entries were suitable for our *in vitro* circuits: for example, an AraC-based sensor, reported by Chou and Keasling<sup>218</sup>, like the native promoter relies on multiple spaced out operator sites for DNA looping. We considered it impractical to synthesise a construct of hundreds or base pairs. Some other sensors detected metabolites, impractical

to handle outside of the host due to stability concerns. However, we did explore FadR and TetR TFs to various degrees of success.

### 3.3.4.1 *FadR*

FadR sensor was used *in vivo* by Zhang *et al.* to detect acyl-CoA as a proxy for fatty acid synthesis<sup>219</sup>. With a single binding site,  $K_d$  of 0.2 nM<sup>220</sup> and inducer response similar to that of LacI it seemed a perfect candidate. In fact, the authors illustrated the principle of action similarly to us (**Figure 127a**). We have therefore used the sequence reported in our design, otherwise reusing the psT architecture (**Figure 127b**).

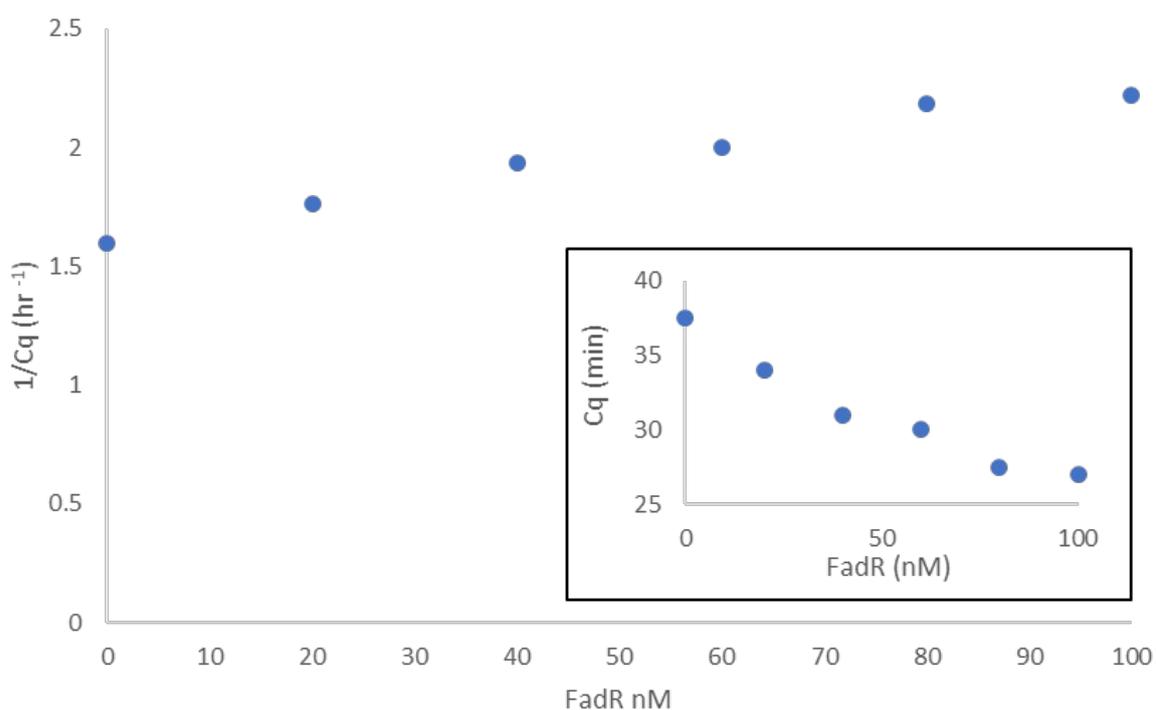


**Figure 127 Design of psT<sub>FadR</sub>.** (a) Illustration of the mode of action of FadR transcription factor along with a native FadR-regulated promoter sequence –  $P_{fadBA}$ . Circled and underlined is the operator sequence. Adapted from <sup>219</sup>. (b) psT architectures with TrpR operator (as before, shown for reference) and FadR. The operator sequences are in red, Nt.BstNBI recognition sequence is dashed boxes, whilst nicking sites are indicated by red triangles.

We took our miR-based switch assembly as before and added  $\text{psT}_{\text{FadR}}$  at a concentration of 5 nM as well as a range of FadR repressor, sourced from *E. coli* (see Methods for details) expecting to detect interference. Adding up to 100 nM of FadR dimer appears to marginally increase the rate of switching (**Figure 128**). This trend is against the expectation. However, when considering the decrease in non-inverted Cq it is only on a scale of minutes.

In this particular experiment, FadR protein stock was just concentrated enough to prepare the 100 nM reaction, whilst lower concentration reactions were prepared by diluting the stock in various amounts of 1x miR buffer. As the protein stock only contained Tris-HCl and NaCl (see 4.1.1), compared to miR buffer that contained other PEN-toolbox oriented components, for example dNTPs, this trend could be considered an artefact caused by this inequivalence in the TF carrier buffer.

We are unsure as to why the transcription factor did not interfere as expected. We did not investigate this further but we speculate, that the particular orientation of the operator site and therefore the transcription factor on the hairpin can impact its capacity of shielding the site from the nickase. *In vivo* some TFs have an opposing effect depending on the operator site position.

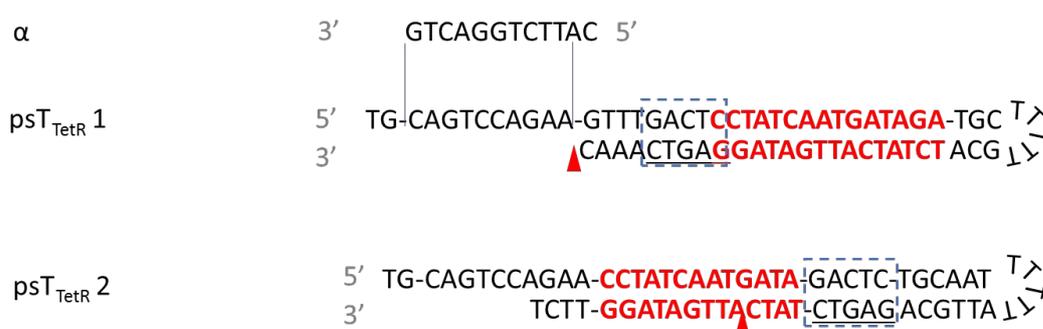


**Figure 128 FadR effect on  $\text{psTFadR}$  driven sensor.** An average for two replicates by different experimenters is shown. Main plot:  $1/Cq$  of a range of FadR concentrations added. Inset: same data but Cq values shown instead.

### 3.3.4.2 *TetR*

We used the tetracycline repressor as our original proof of principle model. Like LacI and FadR, it binds the operator in absence of the inducer molecule – tetracycline. This system is commonly used for artificial *in vivo* systems control. As we wanted to verify the feasibility of the transcription factor-based approach before we designed switches to work at 37 °C, we re-used the switch<sub>50</sub> we have spoken about in **Figure 40**. As it showed some capacity to discriminate sT-induced start from self-start down to 42 °C, we have taken this temperature forward as the lowest permissible. Therefore, in this subsection, unless otherwise indicated, the experiments were carried out at 42 °C, in a modified enzyme/template assembly as found in (4.2.1 same as in **Figure 40**). The difference which should note between the miR assembly used throughout the rest of the text is the use of Vent (exo-) polymerase instead of the *Bst* DNA Polymerase, Large Fragment. We will discuss this later in this section as well as in the following one.

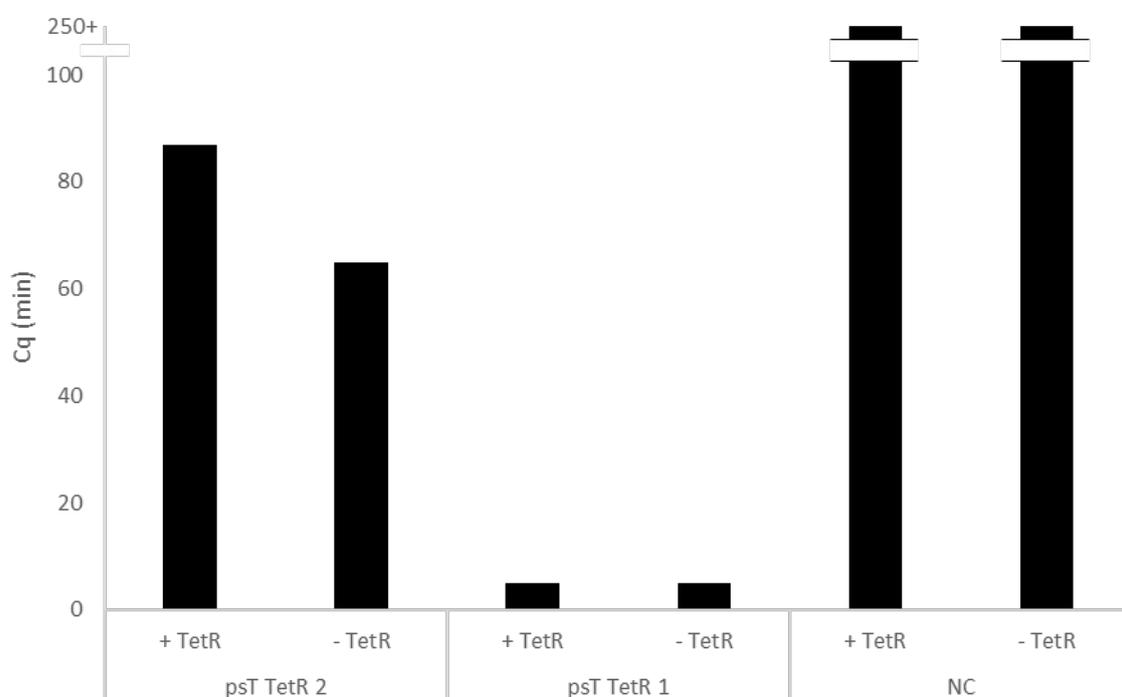
When designing the first psT, we wondered, what orientation the operator site should have relative to the nicking site. If the operator site orientation is fixed, two options are possible (**Figure 129**). Placement of the nicking site closer to the output sequence allows for “tailless” output, given appropriate spacer length. However, there is little overlap between the operator and nickase-related sites: in case of TetR, there happened to be a possibility of one nucleotide overlap but this might not be the case for other operator sequences. It was not clear whether proximity-rather than a real overlap would be sufficient in the design. We called this design psT<sub>TetR</sub> 1. An alternative is placing the nicking site further away from the output, thus cleaving in the operator site and creating an output with a tail. Such tail could hamper dissociation of the output but this arrangement creates a clear overlap between the operator site and the nicking site. We called this variant psT<sub>TetR</sub> 2.



**Figure 129 Design of psT<sub>TetR</sub>.** α sequence is shown for reference along with two psT<sub>TetR</sub> designs differing in nicking site position. TetR operator site is shown in red, Nt.BstNBI recognition site is highlighted by a dashed box.

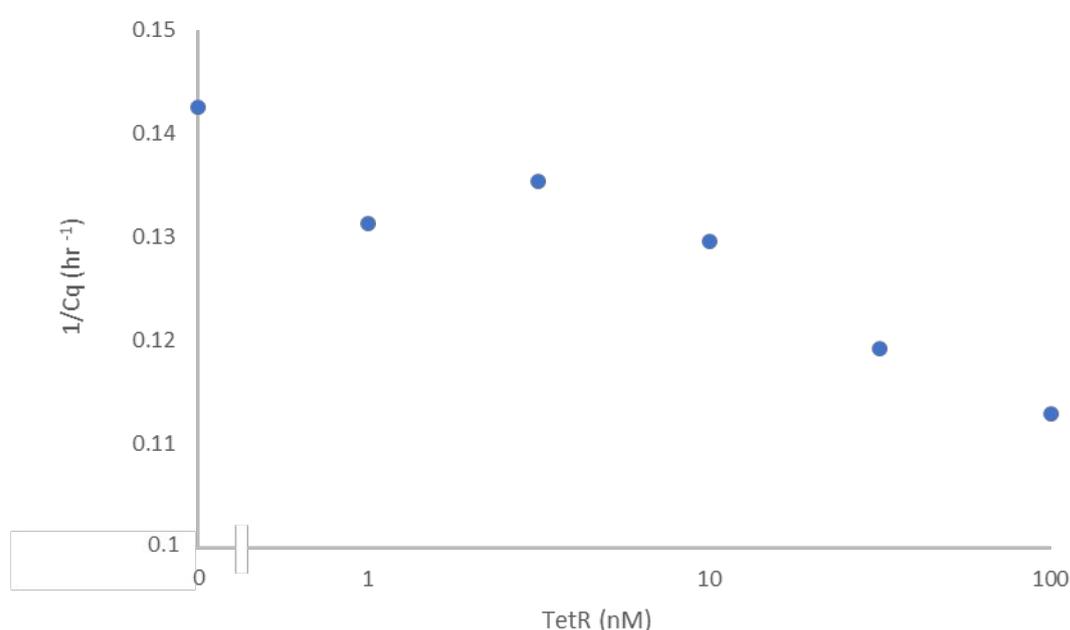
We checked both designs by adding either template at a concentration of 0.5 nM along with negative controls containing water. To check for interference, we added 100 nM of TetR dimer to half of the samples in expectation it will slow down the start of the switch in comparison to the samples containing water instead (**Figure 130**).

Negative controls lacking a psT did not start. psT<sub>TetR</sub> 1 samples start the circuit instantly, whilst as expected, psT<sub>TetR</sub> 2 is slower to act. For psT<sub>TetR</sub> 2 we observe a difference in start times with the transcription factor-containing sample being slower. We believe this is because of TetR interfering with the nickase. Therefore we have chosen design 2 to be taken forward (simply psT<sub>TetR</sub> from now on) and based other psT and pskT on it. We would also note that the slowing down of the start for variant 2 is probably due to the trigger sequence struggling to dissociate as it is burdened by the operator fragment tail, increasing the overall melting temperature.



**Figure 130 Evaluation of psT<sub>TetR</sub> designs.** psT<sub>TetR</sub> 2, psT<sub>TetR</sub> 1 (0.5 nM) or mQ (NC) are run either with (+ TetR) or without (- TetR) 100 nM of TetR dimer. NC samples did not start for over 250 minutes observed.

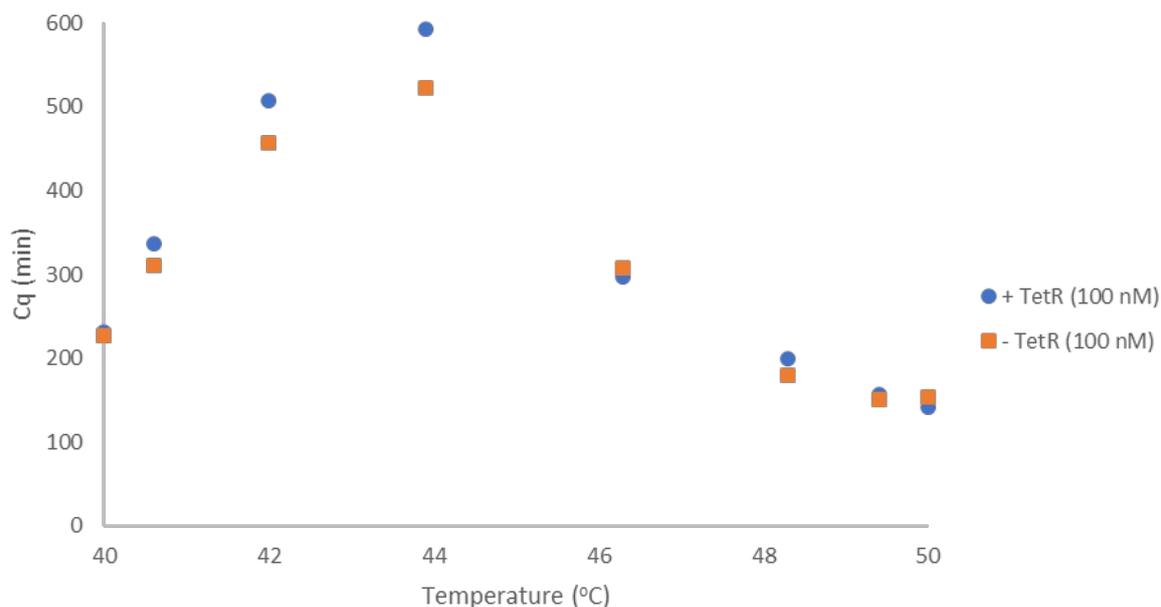
The next step was to determine whether the circuit is responsive to varying concentrations of TetR repressor. We took  $psT_{TetR} 2$  at 0.5 nM, as before but this time added TetR at varying concentrations (**Figure 131**). We observe a log-linear trend in start times. This is because concentration of TetR affects its ability to interfere with the circuit. We would expect that the addition of various quantities of tetracycline to a fixed concentration of TetR would replicate this result. However, we noticed that in this experiment the fastest sample (0 nM TetR) took 421 minutes to start whilst the slowest (100 nM TetR) – 531 minutes. In waiting of the 37 °C switches, we therefore asked ourselves, whether the running temperature could be optimised. We hoped that since TetR appeared to withstand 42°C in our reactions, it could be functional at even greater temperatures.



**Figure 131** Range of TetR concentrations with  $psT_{TetR}$ -driven circuit. TetR concentrations are plotted against 1/Cq values.

As mentioned above, we wondered if we could go to higher temperatures. We bore in mind that at lower temperatures the self-start of the system became too rapid for the system to function (as shown in **Figure 40**) but it was not clear, at what temperature TetR would cease to bind. Therefore, we carried out an experiment where we tested only the presence/absence of TetR dimer (0 or 100 nM) and varied the temperature instead.

We observe, as before, that increase in temperature from 40 to 44 °C increases the gap between negative and positive control samples. However, this time the gap is not due to the absence or presence of the  $psT$  but due to obstruction by TetR. Beyond that point there is a sharp drop in both resolution of positive/negative control but also in delay to start the switch. We believe at higher temperatures TetR is either unable to bind due to higher temperature or due to the process of denaturation, leading to the loss of function. The increase in switching rate is expected as the circuit was originally optimised for 50°C. We have therefore noted that temperatures above 44 °C should be avoided.



**Figure 132 TetR regulation of a circuit at a range of temperatures.** Identical samples containing 100 nM TetR dimer (+ TetR) or not were incubated the temperature indicated.

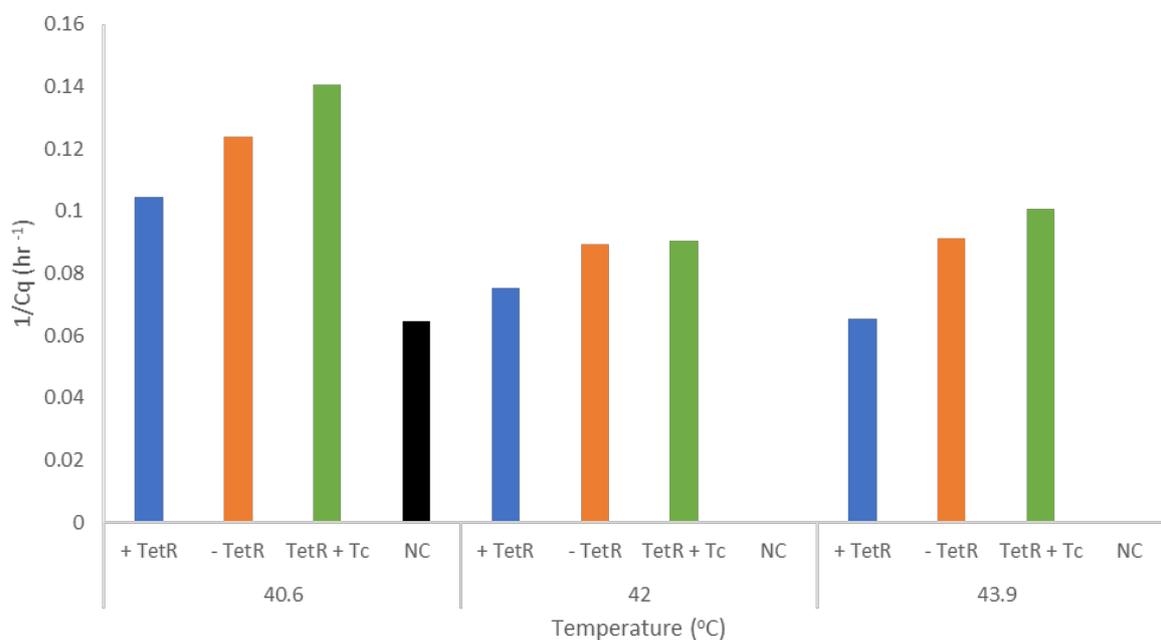
Knowing that:

1. TetR can interfere with the nickase-driven starting of the switch;
2. Decrease in binding-capable TetR concentration increases the starting rate;
3. Temperatures above 44 °C should be avoided,

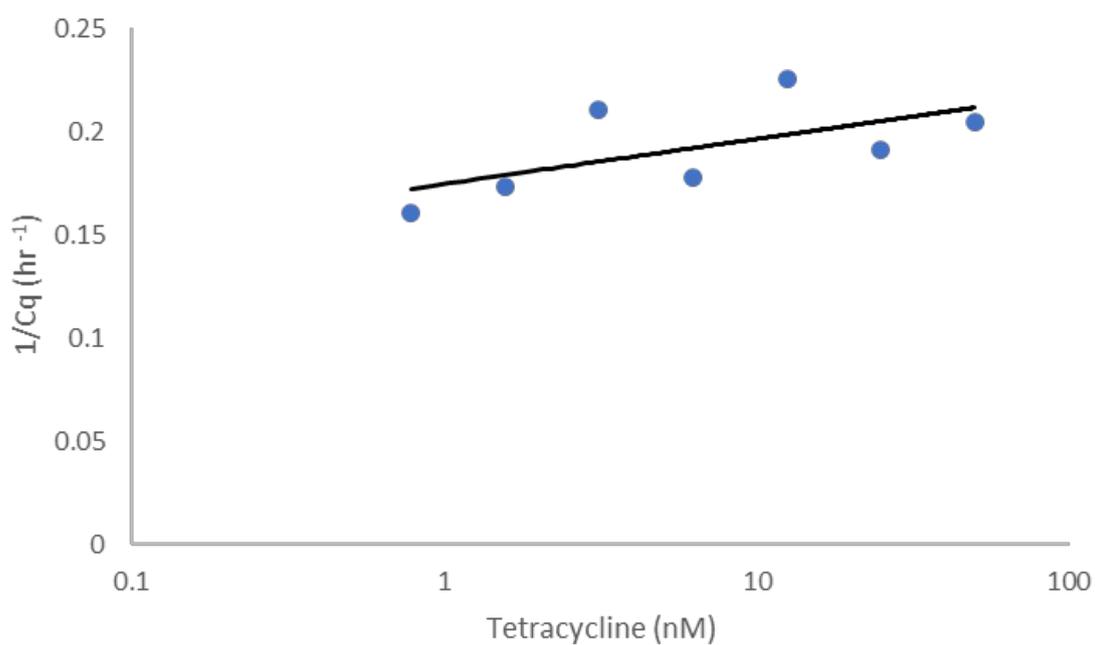
We decided to proceed with tetracycline detection. In particular, we expected excess of tetracycline added to TetR-containing sample to replicate the behaviour of the experiment not containing TetR at all. This is because, as we mentioned, addition of tetracycline strongly decreases the affinity of TetR to the operator.

Therefore, we ran yet another variation of the experiments above, this time making samples: containing TetR (100 nM) and the psT (0.5 nM); same as previous bar with tetracycline added (100 nM); one containing psT-only; and one lacking all of the components above (**Figure 133**).

We observe in all cases that tetracycline-containing samples are the fastest to start, usually similar to those, lacking TetR. As before, addition of TetR in absence of tetracycline slows down the switching. In turn, absence of psT keeps the switch off with the exception of 40.6 °C, where self-start occurs late in the run, in line with our previous experience. Overall, 43.9 °C provided the largest distinction between TetR samples containing or lacking tetracycline, so we have taken this temperature forward.



**Figure 133 Detection of tetracycline at a range of temperatures.** For each temperature, samples containing psT and TetR (+ TetR), containing psT only (- TetR), containing psT, TetR and tetracycline and lacking all three (NC) are shown.



**Figure 134 Range of tetracycline detection.** Concentrations of tetracycline (100-0 nM) were added to otherwise equivalent samples at 43.8 °C with 0.5 nM psT<sub>TetR</sub> 2 and 100 nM TetR dimer.

We moved on to explore a range of tetracycline concentrations at 43.8 °C with conditions as before, to demonstrate the first proof of principle of small molecule sensing with the PEN toolbox. In particular, we went down in concentration from 100 nM arbitrarily chosen before.

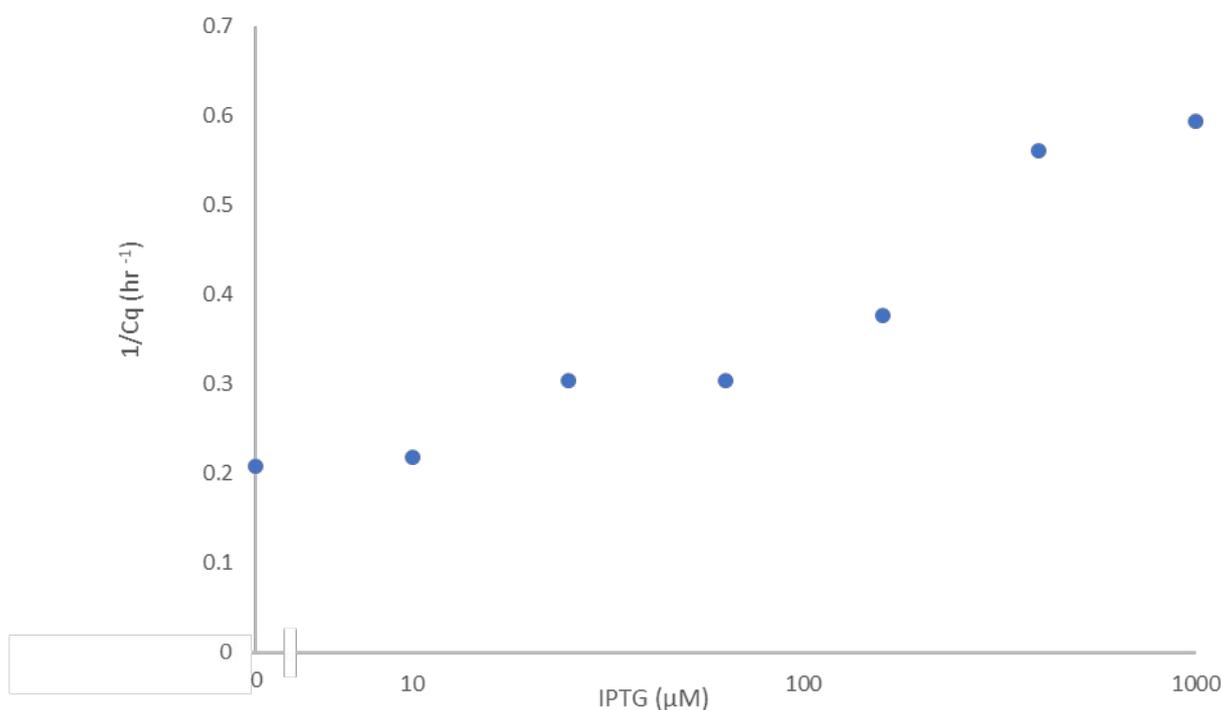
We observe that decrease in concentration slows down the switch (**Figure 134**). We attribute the two above-trendline points to experimental execution discrepancy. Overall, the set of experiments related to TetR demonstrates in principle that a transcription factor can be used with PEN-toolbox to detect a range of enzyme concentrations. However, the performance of this particular switch was suboptimal at temperatures suitable for proper TF function. This is why, satisfied with this proof of principle we decided to commit ourselves to the 37 °C systems, being confident that TF-based psT module will function.

### 3.3.5 Using different polymerases to drive the circuit at 37 °C

Our initial attempts to run the new B11 and B11b circuits at 32-42°C (where we simply used 50 nM of aT, similar to how it is described in 3.3.2), used an enzyme assembly driven by Vent (exo-) polymerase at 4%. However, in absence of pT, the assemblies failed to start in the first 800 minutes. Moreover, some of the samples did not start instantly, even when 5 mM of trigger sequence was added (not shown, see discussion related to **Figure 143** for another example). We speculate this could be due to lower temperature regime at which thermophilic enzymes might lose activity. We therefore replaced Vent (exo-) polymerase with Bst Polymerase Large Fragment. This proved to be much more successful at lower concentrations of the enzyme. Apart from different temperature adaptations – Vent is extremely stable, whilst BstLF is recommended for use at 65°C, we believe BstLF strand displacing activity plays a role. Its ability to displace a nicked strand, even when dissociation due to temperature is unlikely makes the system more robust. This is especially true when considering the use of long-tailed signals produced by psTs. An additional advantage we observed, when we moved onto isothermal primer amplification experiments followed by PCR, was that 1% Vent (exo-) needed for the PCR did not affect the behaviour of the assembly, previously optimised with BstLF only (not shown). However, the usage of BstLF came with greater instances of non-specific amplification or sporadic starts, compared to Vent (exo-).

Taking advantage of the targeted 37 °C temperature regime, we also considered Klenow fragment – the *E. coli* well known analogue of Bst Large Fragment. Apart from being mesophilic, the properties are broadly identical: it possesses a strand-displacing activity and lacks a 5' → 3' exonuclease activity. In context of the PEN-toolbox, it was examined by Gabriel Ragala – an intern I had the pleasure of working with. The process of optimisation was broadly similar to the one done for other PEN-toolbox circuits as previously reported with Klenow exhibiting behaviour, similar to Bst. Furthermore, Gabriel did not have time during his internship to perfect the circuit. Nonetheless, he demonstrated detection of IPTG with ps<sub>T<sub>lacI</sub></sub>-based circuit similar to the one presented in 1.3.2 (**Figure 135**).

Taken together with other results above, we consider this a reasonable demonstration that PEN-toolbox is not only modular in the sense of DNA fragments, but also, in terms of substituting the enzymes driving the circuit.



**Figure 135** Detection of IPTG with a Klenow-driven  $psT_{lacI}$  circuit. Note the discontinuous horizontal axis. Here standard miR assembly was used with BstLF substituted by Klenow Fragment (3'→5' exo-) 0.1% of stock (1 U/ml<sup>l</sup>, NEB), driven by 5 nM of  $psT_{lacI}$ .

### 3.3.6 TrpR-sensing is possible at higher temperatures

When we initiated our work on TrpB, the thermophilic PFB variant appeared to be a more promising candidate, as previously reported by Arnold and colleagues. We were therefore concerned if it would exhibit sufficient activity at 37 °C, as previous assays were done at 70°C. More importantly, increasing the working temperature to just 45°C would have opened us access to the use of numerous parts already existing in our group. Therefore, we assayed, whether TrpR is stable at higher temperatures, our circuits can be run at higher temperatures, and finally, whether our approach can be ported into an existing switch, designed to run at higher temperatures. We have so far mostly spared the reader of the details of how we arrived at the specific compositions of our miR and IPA assemblies. However, we will use some examples in this subsection to discuss several aspects of the said optimisation.

#### 3.3.6.1 Running the existing IPA circuit at higher temperatures

We first wanted to verify, whether TrpR retains functionality after being heated to 45 °C on its own.

We prepared a simple assay: the IPA mix was prepared as before with either 1 mM L-*trp* or water for positive and negative controls but Bst polymerase driving the circuit was excluded. The samples were incubated at a range of temperatures between 37 and 45 °C for 30 minutes, after which the polymerase was added and the temperature adjusted to 37°C for all samples. Our assumption was that TrpR would be subjected to the temperature treatment

in the same buffer environment, whilst the switch would be run at the temperature previously optimised for. Therefore, any change in performance would be attributed specifically to effects on TrpR. The remainder of the enzymes have been used at higher temperatures without decrease in efficiency, whilst we considered effect of nickase action in absence of the polymerase negligible. However, none of the samples start. We speculated that this was due to ttRecJ 5' exonuclease being rather efficient at breaking down the source templates not carrying PTO modification protecting other templates. Usually, the hairpin templates are protected due to polymerases completing the partially double stranded structures. Once double-stranded, ttRecJ does no longer exhibit the 5' exonuclease activity towards them as it a single stranded DNA exonuclease.

We thus quickly adjusted the approach by excluding both BstLF polymerase and ttRecJ during the pre-incubation stage. We also removed Vent polymerase as we considered the fact that it is widely used to drive 45 °C circuits and therefore could impact the performance at different temperatures.

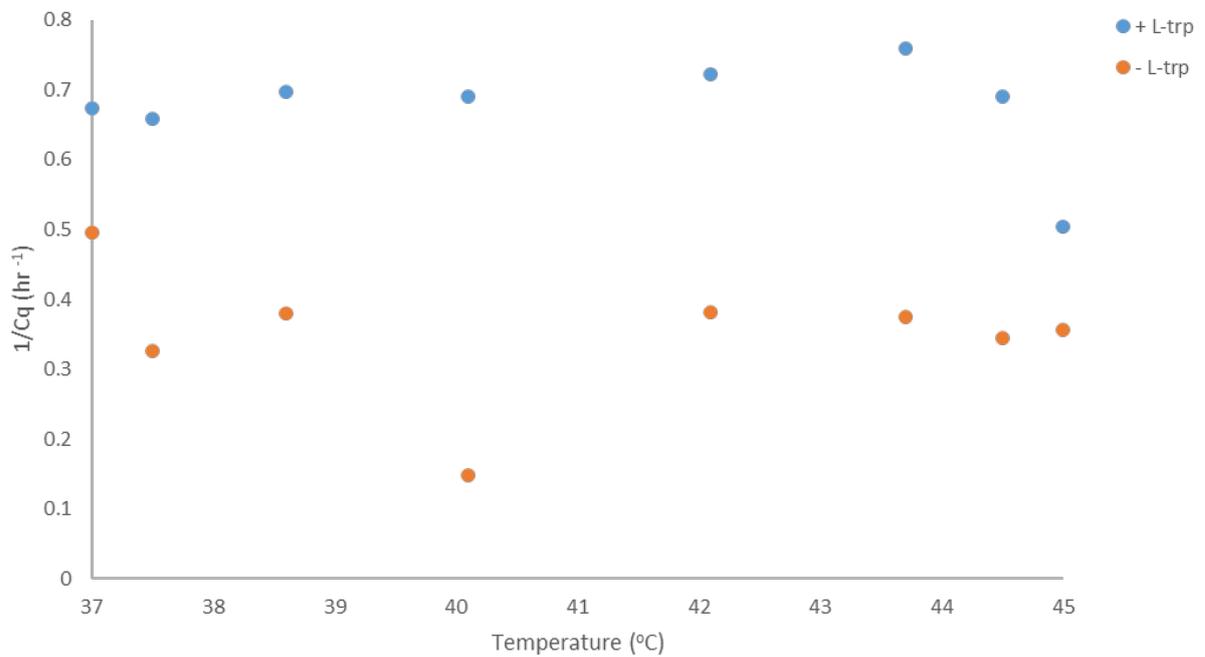
It appears that all samples behave similarly across the range of preincubation temperatures (**Figure 136**). Although the results were relatively noisy (for example the 40.1 °C sample is uncharacteristically slow which we would discard as an experimenter's error) we overall considered that TrpR remains intact during the heat shock and therefore decided to proceed to testing the behaviour of the entire circuit.

We did this by simply raising the temperature at which we run the main reaction (**Figure 137**). The system slows down first, increasing the separation of the positive and the negative controls. However, starting with the 40.1 °C point, the relative order of positive and negative control start times is swapped. This result has been replicated (not shown) and the sets of controls are loaded into the machine in a single strip, so this is unlikely to be an error. We speculate, that at higher temperatures the stability of the switch is lost, leading to non-specific starts of the negative control samples. We are not sure what causes the change of behaviour of the positive controls at higher temperatures .

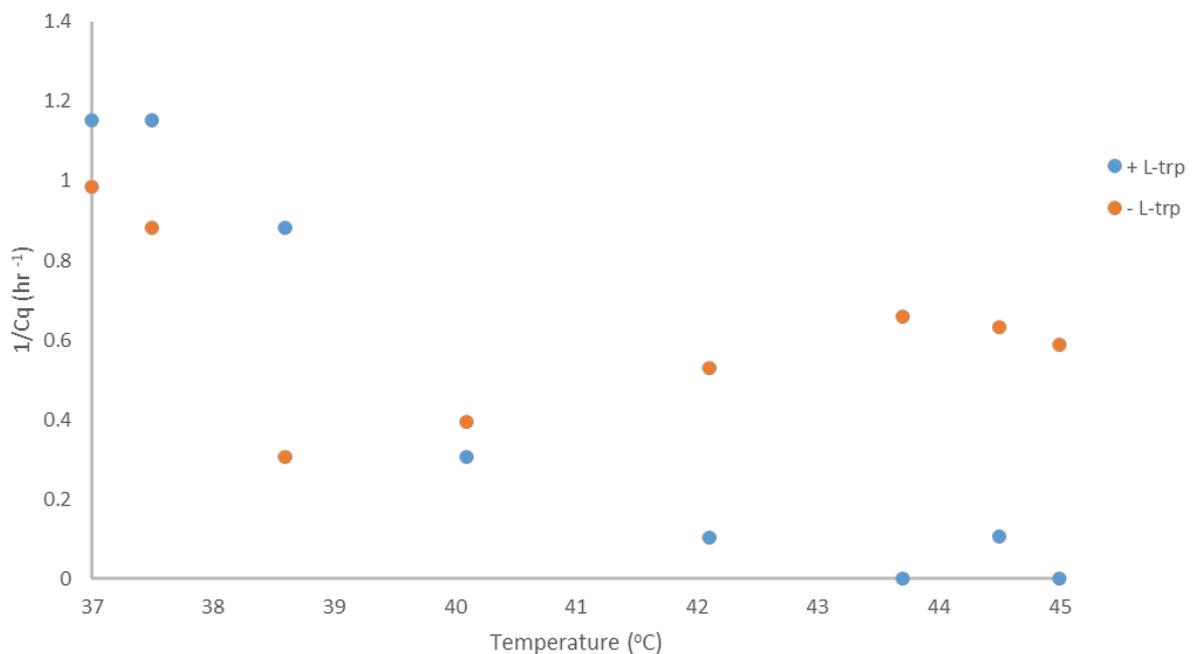
As we hypothesised that the circuit was self-starting, we wondered if additional pT could help, raising the threshold. We tried adding a linear range. As the mix already contained 10 nM of pT, we have therefore gone from 10 to 17 nM in strip of 8 tubes. As most other times, we used 1 mM L-trp as a positive control and 0 mM as a negative. We chose to examine two temperatures: 40 and 42 °C as at these temperatures the system was just starting to fail and we therefore hope it could be rescued.

At 40 degrees all negative controls start at the same time, whilst for positive controls the pT slows down the start and slightly changes the gradient of amplification (**Figure 138**). We believe here, even for lowest pT concentrations the system is bistable, and the apparent start corresponds to divergence (confirmed in the EvaGreen channel). Divergence and production of non-specific DNA happens irrespective of the circuit components. We noticed that for different circuits divergence (mostly driven by polymerase type and concentration) tends to happen at a relatively constant time point, but that this process is sensitive to the buffer used with the IPA buffer being more sensitive than the miR one. Here 150 minutes therefore sets a limit to the possible time of observation, after which the sensing assembly "expires".

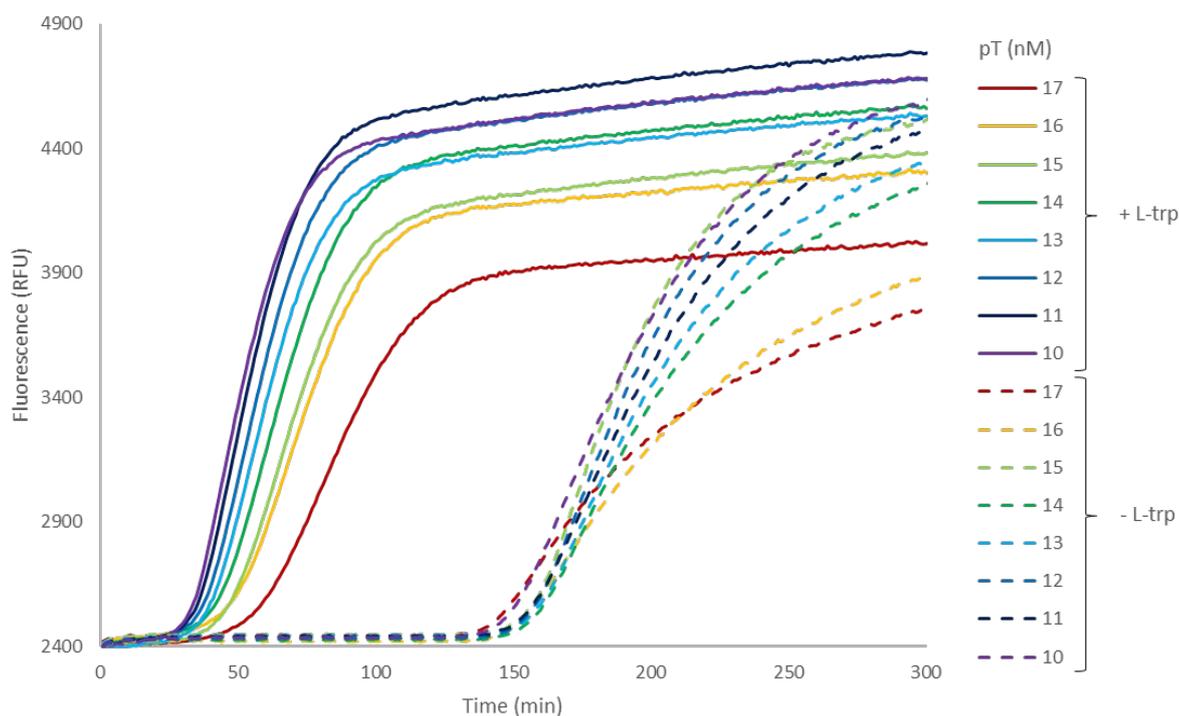
For the "ON" samples, the impact is limited, with only the highest pT concentrations slowing them down.



**Figure 136** TrpR incubated at a range of temperatures senses L-trp in IPA. Samples with 1 mM L-trp (+ L-trp) or without (- L-trp) incubated for 30 minutes at different temperatures (37-45 °C). prior to running IPA mixture completed with Bst, Vent polymerases and ttRecJ at 37°C.



**Figure 137** IPA reaction running over a range of temperatures. Samples with 1 mM L-trp (+ L-trp) or without (- L-trp) were incubated at different temperatures (37-45°C).

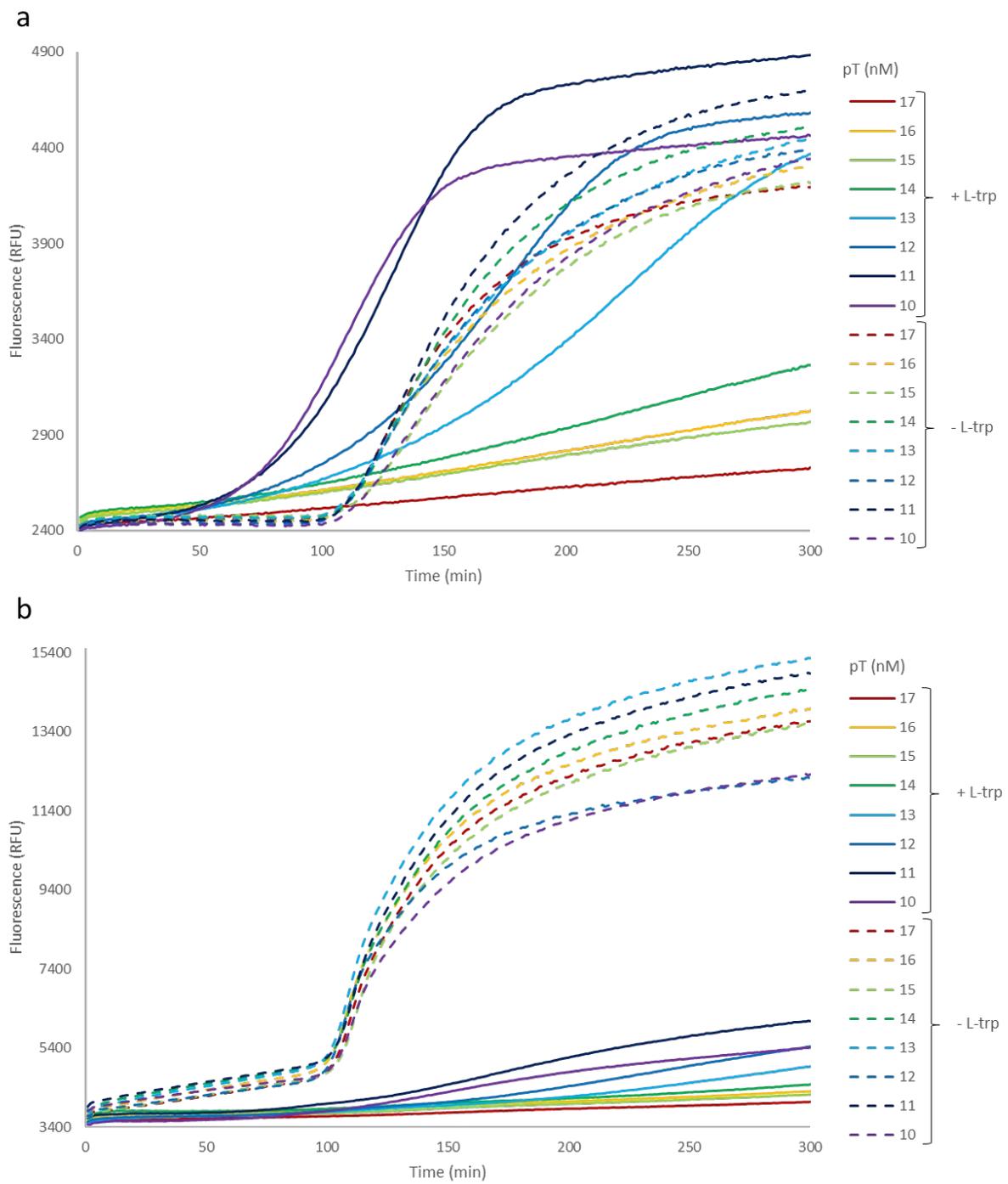


**Figure 138** Cy5 fluorescent traces of IPA reaction at 40 °C. Range of pT concentrations (10-17 nM) was added to samples with 1 mM L-trp (+ L-trp) or without it (- L-trp).

At 42 °C, the behaviour is roughly similar for the negative control (**Figure 139**). They diverge simultaneously, even faster than at the lower temperature. The positive controls become more sensitive to the increase in pT, which is an expected phenomenon. Interestingly the cognate start and pT load, even quite weak, is enough to inhibit the divergence mechanism.

This confirms that higher temperatures do not provide an appropriate window of observation for the switch, because it induces premature non-specific divergence. Overall, we determined in this section that TrpR retains its capacity to impact the circuit after being heated to 45 °C. Despite this, our B11-based IPA system was unstable above 40 °C. Therefore we decided to try another set of templates.

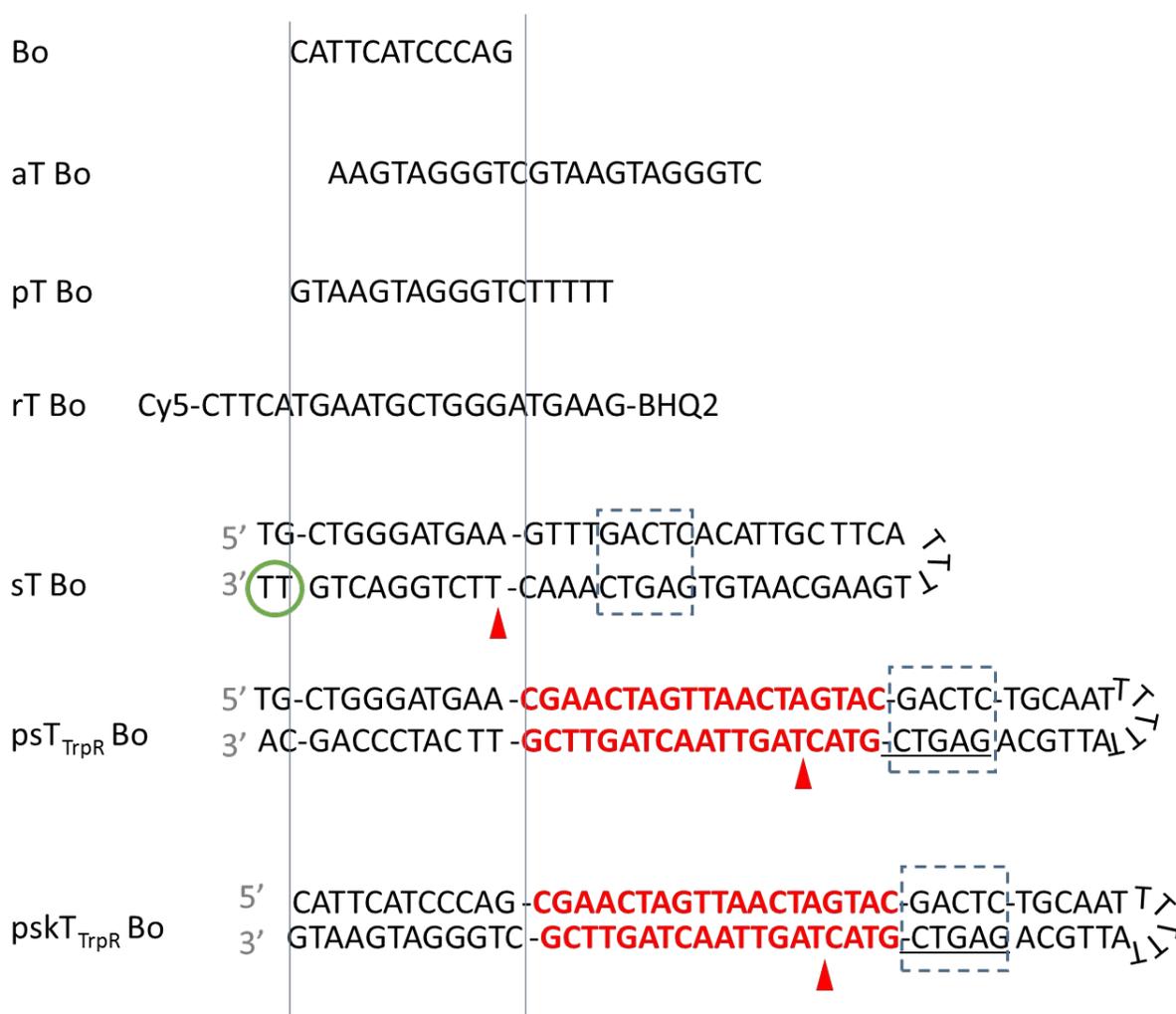
We would note that we managed to replicate our detection and amplification of Pfu, similar to **Figure 73** bar with 40 °C incubation. We are not showing the results as they are broadly similar.



**Figure 139** Fluorescent traces of IPA reaction at 42°C. Range of pT concentrations (10-17 nM) was added to samples with 1 mM L-trp (+ L-trp) or without it (- L-trp). (a) Cy5 fluorescent trace. (b) Non-specific FAM fluorescent trace.

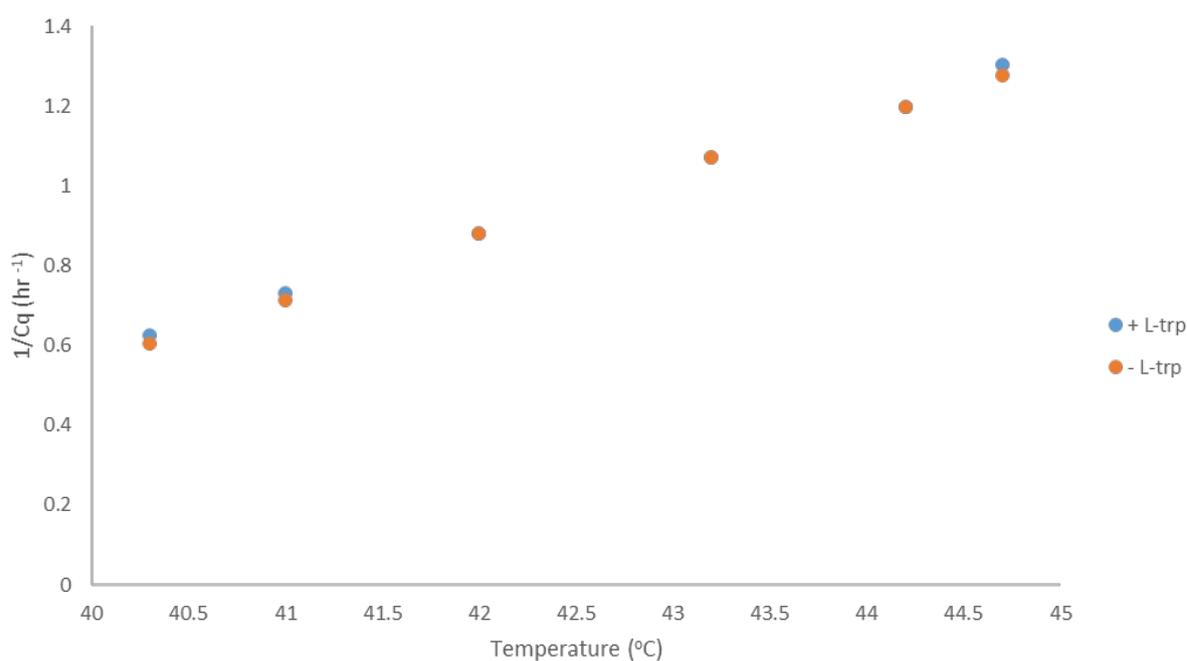
### 3.3.6.2 Transferring the architecture into a different switch.

We have taken the Bo set of oligonucleotides, kindly provided by Guillaume Gines, as they were known to function at 45°C, and designed complementary sensing templates (**Figure 140**). The basic switch components are similar to the other switches we have worked with. Although, there are some small differences. Here, the pT carries a longer polyT tail (5 bp). The rT is “-2” like the aT, to decrease rT competition during switch activation. The sT carries a small defect – the last two nucleotides are not identical to that of a -2+2 trigger; however, this is alleviated during the first round of nicking and subsequent extension in accordance with the correct templating sequence. Our new psT<sub>TrpR</sub> Bo is a simple chimera of psT<sub>TrpR</sub> with the output as Bo sequence, whilst for pskT<sub>TrpR</sub> Bo the output is, of course replaced by the antitrigger sequence. Below, we provide some details on the process of getting this new design for sensing L-trp.



**Figure 140 Bo switch components used in this work.** Dashed boxes show Nt.BstNBI recognition sites, whilst red triangles show the restriction sites. In red is the TrpR operator site. Vertical blue lines align the Bo sequence with other components. Green circle highlights a defect in the design which is not a typing error.

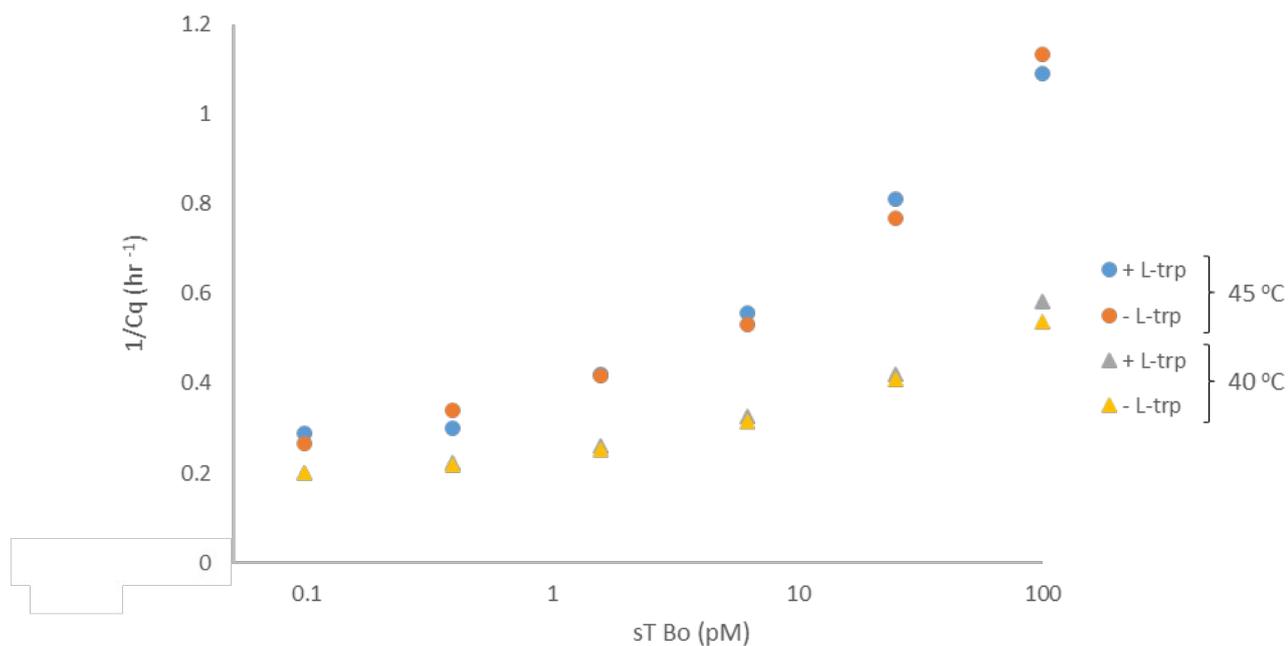
We first attempted to take the enzyme/template concentrations suggested by Gines as a baseline (see 4.2.2), which is adapted to Vent (exo-) polymerase only. We replaced miR buffer with IPA, added 100 pM sT Bo and 10 nM psk<sub>TrpR</sub> Bo along with the usual concentration of TrpR and examined whether our usual positive and negative control (1 and 0 mM L-trp) could resolved at any temperature between 40 and 45 °C (**Figure 141**). All fluorescent traces appeared regular (not shown) and there is a clear correlation between the temperature and the starting time. However, the controls could not be discriminated. We concluded that optimisation was solicited.



**Figure 141 Bo-based switch detecting L-trp at different temperatures in IPA buffer.** 100 pM sT Bo ,10 nM psk<sub>TrpR</sub> Bo along with 100nM TrpR dimer, see 4.2.2 for full assembly composition. Samples containing 1 mM L-trp (blue) or 0 mM L-trp (orange) are shown.

We first hypothesised that the start clearly occurs, however, the input of  $\text{pskT}_{\text{TrpR}} \text{Bo}$  is dwarfed by the  $\text{sT Bo}$ . We therefore made a range of  $\text{sT Bo}$  concentrations, whilst keeping the rest as before, including the positive and negative controls. This time we only considered 45 and 40 °C to preserve reagents (**Figure 142**).

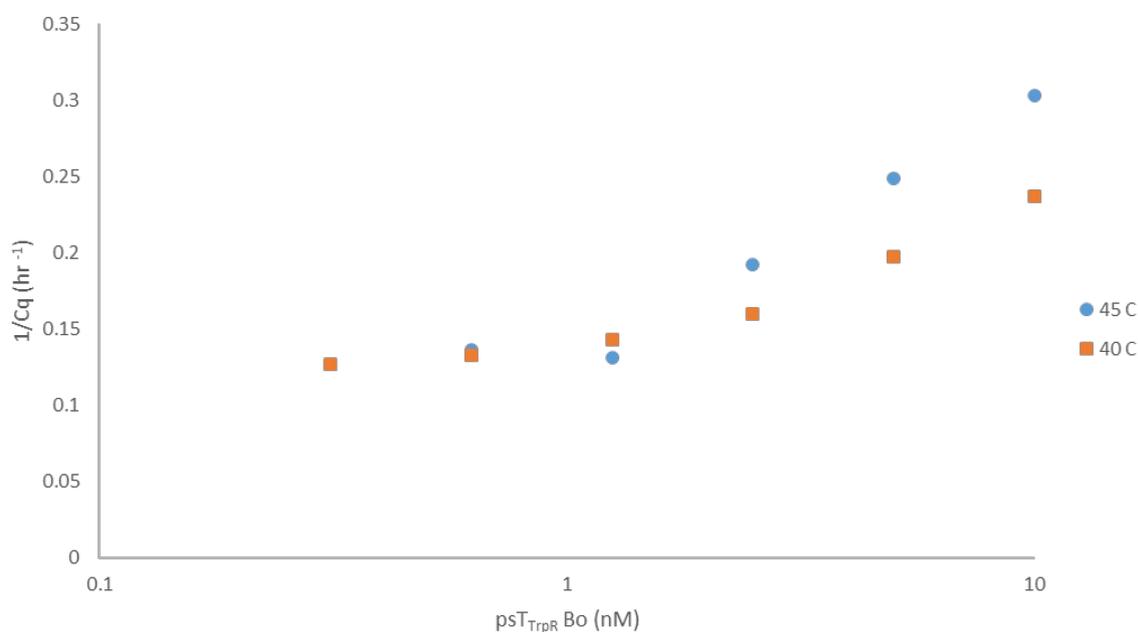
We observed that the circuit starts slower at lower temperature as before. It is similarly slowed down by decrease in  $\text{sT Bo}$  concentration. However, the positive and the negative controls are still indistinguishable at both temperatures.



**Figure 142 Range of  $\text{sT Bo}$  at different temperatures in IPA buffer.** Conditions similar to **Figure 141** but  $\text{sT Bo}$  concentration is varied. Reactions were run at either 45 (circles) or 40 (triangles) degrees Celcius, containing either 1 or 0 mM L-trp.

Our next hypothesis was that in the current setting, the  $pskT_{\text{TrpR}}$  was inefficient in comparison to all other components of the circuit. We therefore switched to use  $psT_{\text{TrpR}} \text{Bo}$ , that would act upstream of the aT Bo, and hence, the L-trp input would be amplified. We replaced 100 pM sT Bo and the  $pskT$  with up to 100 pM  $psT_{\text{TrpR}} \text{Bo}$ , compared to previous runs. This led to very slow start times (not shown) even when we tested higher concentrations in a subsequent experiment (**Figure 143**).

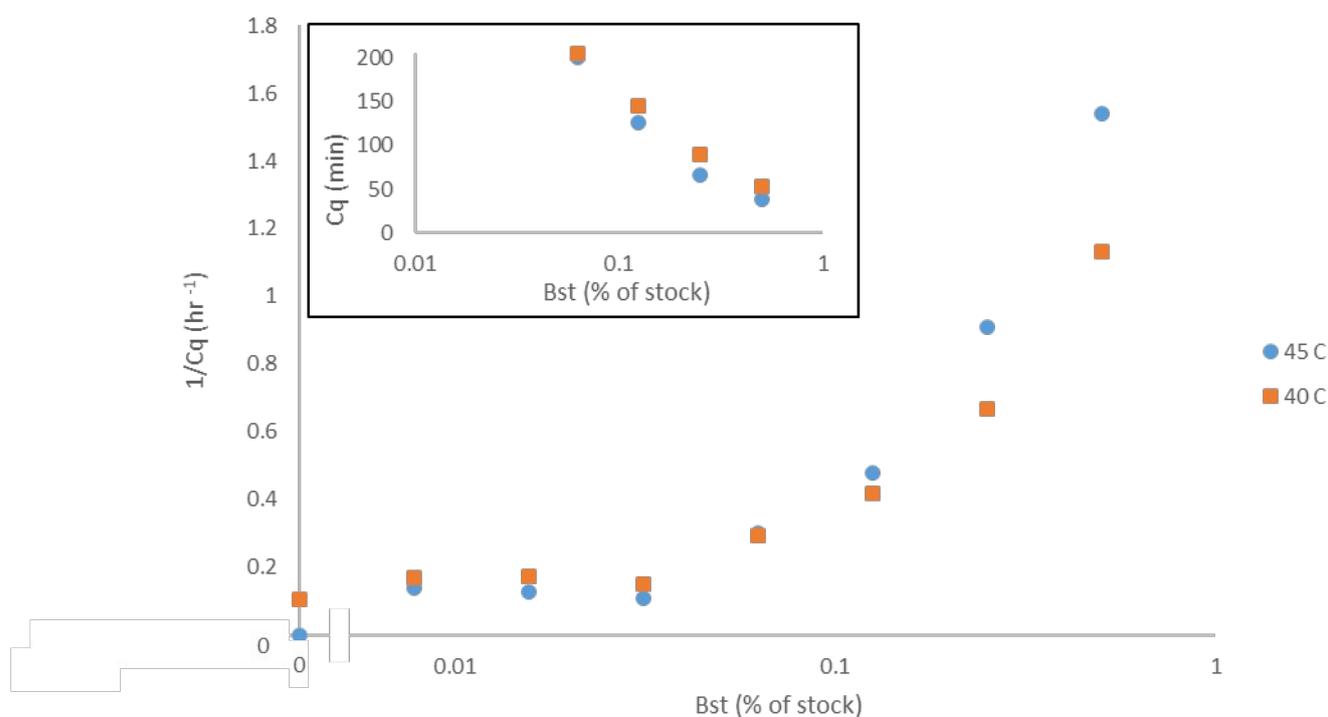
We concluded that indeed both sensing templates are inefficient in this setting. We supposed that, despite the increase in temperature, the long operator tail prevents the output of the sensing template from dissociating. Vent(exo-) polymerase, as we mentioned, has limited strand displacement activity. Hence, only components with favourable binding parameters (the temperature-specific Bo switch components) actively participate in the reaction.



**Figure 143** Range of  $psT_{\text{TrpR}} \text{Bo}$  at different temperatures in IPA buffer. Conditions similar to **Figure 142** bar<sub>TrpR</sub>TBo concentration is varied in absence of sT/ $pskT$ . Reactions were run at either 45 (circles) or 40 (squares) degrees Celsius.

We switched to Bst polymerase (LF) due to its strand displacement activity. We halved the concentration of Vent, to decrease the total amount of polymerase in the mixture and added a range of Bst DNA polymerase Large Fragment along with 2.5 nM  $\text{psT}_{\text{TRP}} \text{Bo}$  – a concentration that seemed convenient to work with in the future (**Figure 144**).

At the lowest concentrations of BstLF we observe a plateau of start times, suggesting that Bst (LF) has little impact in that setting and we are still observing self-start. However, concentrations above 0.06% of commercial stock to total volume quickly speeds up the switching. We hypothesise that, once at sufficient concentration, Bst efficiently drives production of the trigger sequence from  $\text{psT}$  by strand displacement. The same strong strand displacement activity could explain why this circuit is not sensitive to temperature variation. Oligonucleotides are forced off the templates irrespective of their likelihood of spontaneous dissociation.



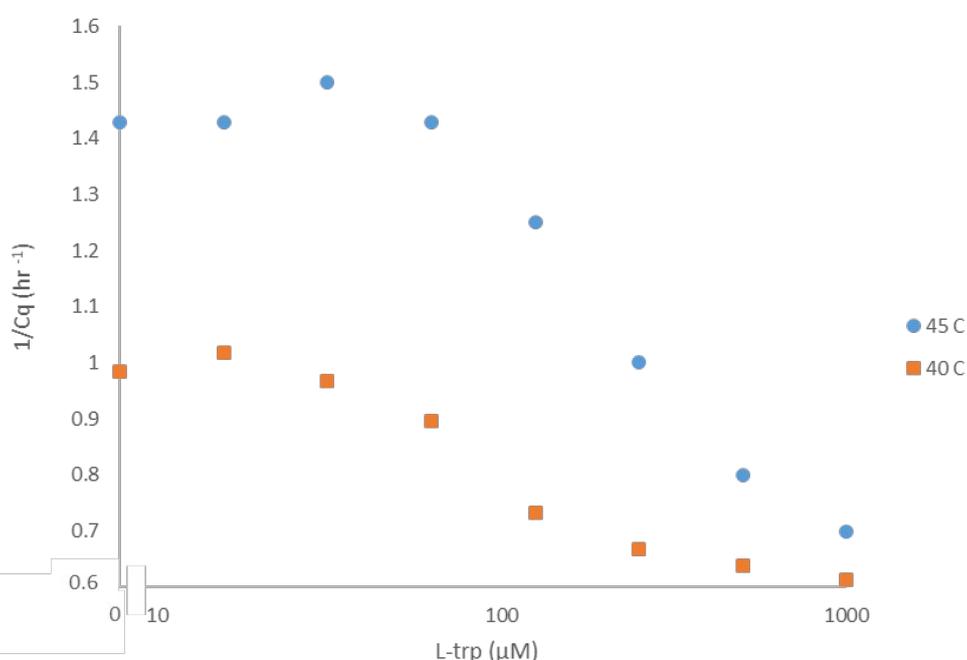
**Figure 144** Range of Bst polymerase added to Bo-based switch. Here concentration of Vent is halved, compared to **Figure 143** and the switch is driven by 2.5 nM  $\text{psT}_{\text{TRP}} \text{Bo}$ . The samples were run at either 45 (circles) or 40 (squares)°C. Inset: Cq for 4 highest concentrations of Bst is shown.

We then proceeded to use this psT-driven circuit to verify if TrpR can still sense L-trp and serve effectively as an interface. We fixed the concentration of Bst LF at 0.45% of the commercial stock and kept the remaining as before (**Figure 145**). We observed a sigmoidal response of the start time to L-trp, similar to what we have demonstrated before. This highlights that the sensor works at 45 °C. Whilst the slowest samples take the same amount of time to start, the lowest concentrations of L-trp are quicker at higher temperature.

This confirms that our sensor can be used at 45 °C and be ported into yet another circuit. Furthermore, in this configuration, 45 °C is optimal, compared to 40 °C.

Encouraged by the fact that TrpR retains activity within a circuit running at 45 °C we went back to focus on our original objective of composing an inverted, pskT-based sensor. Thinking addition of Bst resolved prior constraints we tried replacing the psT<sub>TrpR</sub> in the assembly just before with 10 nM of pskT<sub>TrpR</sub> and various concentrations of sT, going down from 1 nM to 10 pM. Only the negative control with sT excluded showed a small delay, though L-trp containing and lacking samples were not in expected order of starting (not shown).

We followed with ranges of pskT<sub>TrpR</sub> up to a 1000 nM against 10 pM (first experiment) or 1 pM (subsequent experiments) of sT. This time only at 40 °C, expecting that the pskT will have more time to produce sufficient amounts of antitrigger at this temperature to impact the circuit. This time some samples did not start (**Figure 146**). We observe two clusters of start times, around 50 and 150 minutes with sometimes an “in-between curve”, starting Cy5 signal increase at 50 minutes but only plateauing with the 150 minute samples (**Figure 146** inset). Whilst decrease in sT slows down the fastest starters it has little effect on the slow samples.

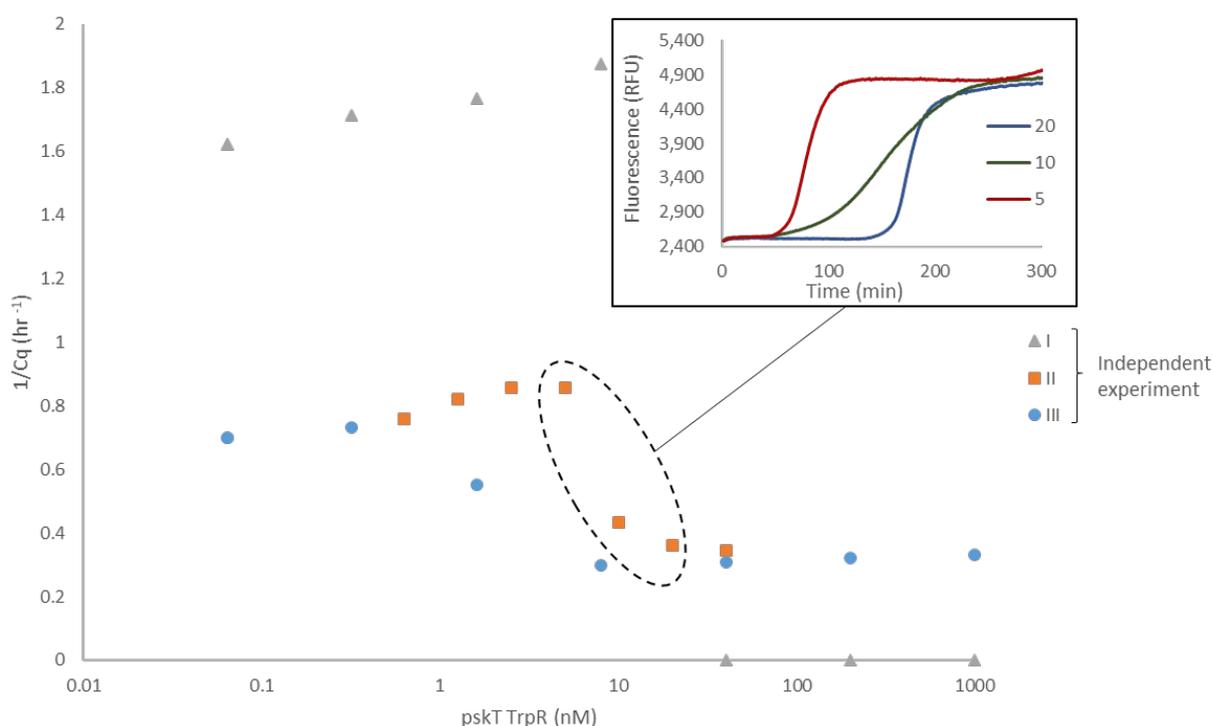


**Figure 145** Range of L-trp detected by psT<sub>TrpR</sub> Bo-driven switch. Conditions used as in **Figure 144** with Bst fixed at 0.45 % of the commercial stock. The samples were run at either 45 (circles) or 40 (squares) °C.

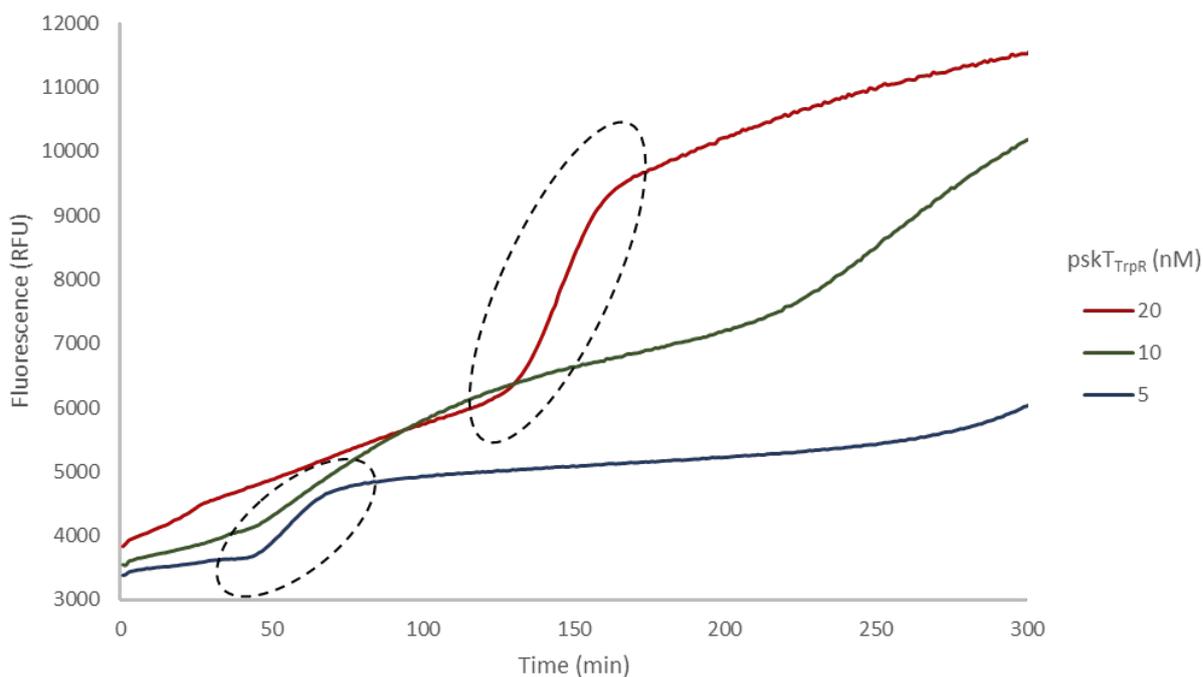
We would like to use these data to demonstrate another aspect of discovering the window of observation for a circuit. Once the circuit is optimised for bistability, it tends to either the “on” or the “off” state. This presents a certain challenge for sensing a range of analyte: for greatest resolution, the pskT concentration should be such, that the circuit is in minimally “off” state by default. With addition of pskT repressor, in case of TrpR, the activity of pskT decreases from maximal, thus tending to the “on” state. Different concentrations of the repressor would only be resolved, if the intermediate states can be attained. As they do so by tilting the curve of the early stages of the exponential amplification (where pskT burdens aT, by acting as a sink for newly produced triggers), the response is non-linear. Hence, the log-linear plots we presented throughout. In addition, the response of the transcription factors itself tends to only be linear over a small range, outside of which it is switch-like. This imposes the limit as to which concentrations can be reasonably resolved.

Here, we determined that concentrations above about 10 nM of the pskT should be sufficient and have decided to use this in our attempt to detect L-trp with it.

On a side note, in the FAM channel for the non-starting samples we observe atypical linear growth in fluorescent signal, prior to the non-specific start “bump” (**Figure 147**). We think this is due to antitrigger-trigger complex formation giving rise to EvaGreen fluorescence due to dsDNA intercalation.



**Figure 146 Ranges of pskT<sub>TrpR</sub> Bo.** Conditions as in **Figure 145** bar psT is replaced by a range of pskT. Experiments were run at 40 °C with no L-trp added and either 10 pM (I) or 1 pM (II-III) of sT. Inset: Cy5 time-fluorescence traces for the points indicated by the dashed circle.

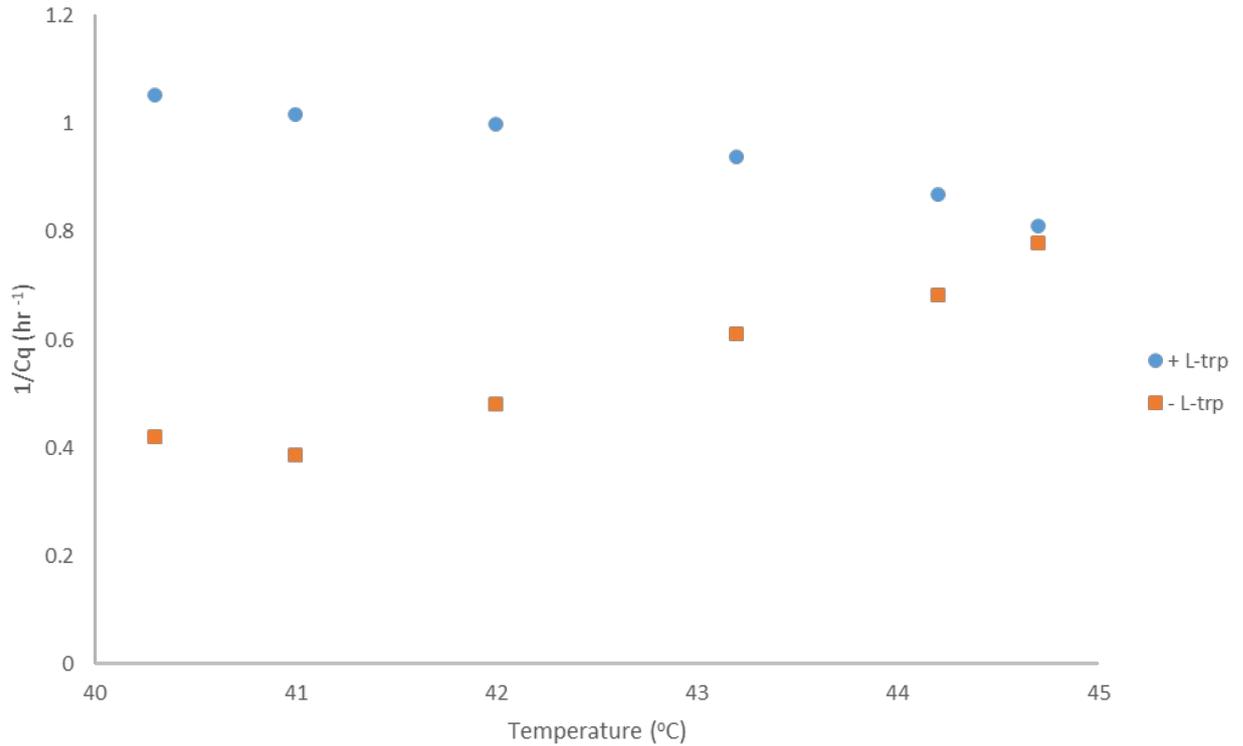


**Figure 147** FAM fluorescent traces for some samples in **Figure 146**. Samples shown are indicated on **Figure 146** with a dotted oval. Here dotted ovals highlight steep regions of the curves, corresponding to Cy5 probe opening.

Having discovered the conditions at which  $\text{pskT}_{\text{TrpR}}$  can have an impact on the circuit, we have taken the concentration of 15 nM forward and went back to our investigation of temperature effect on L-trp detection. We have made an assay over the range of temperatures to assess this as before, using 1 mM L-trp positive and 0 mM negative controls (**Figure 148**).

We observe the L-trp samples to be well-resolved at 40°C. However, the delay of negative control start notably diminishes with increasing temperatures, whilst the positive controls slow down slightly. At 45 °C, they are almost simultaneous.

This is contrary to the trend in **Figure 145**. There samples cluster in the “off” state, unless turned on before self-start. Here, all positive controls start more or less simultaneously (the difference in start times plotted arise mostly due to difference in the steepness of the fluorescent curves due to temperature). However, negative control starts are delayed by the  $\text{pskT}$ , therefore creating the gap in start times. We are unsure why  $\text{pskT}$  efficiency decreases with increase in temperature. This could be due to the fact that the 2 bp difference in complementarity of the trigger towards the aT and the antitrigger becomes less significant. It could also be, that aT performs more efficiently at higher temperatures. Either way, we concluded that although TrpR itself can be used as an interface for PEN-circuits at higher temperatures, this particular configuration may require further optimisation.



**Figure 148**  $pskT_{TrpR}$  Bo detecting L-trp over a range of temperatures. Samples containing 1 mM L-trp (circles) and 0 mM are shown.

### 3.3.6.2.1 pT effects on the window of observation.

We would like to complement our discussion on creating the window of observation with  $pskT$  by also discussing further the effect of pT. We consider a toy model, where the behaviour of the system can be described with the following:

$$\left\{ \begin{array}{l} \alpha'(t) = \frac{aT(\alpha)}{1 + a\alpha(t)} - \frac{pT(\alpha)}{1 + b\alpha(t)} - kt\beta(t)\alpha(t) + psT - RecJ\alpha(t), \\ \beta'(t) = -kt\beta(t)\alpha(t) + pskT - RecJ\beta(t). \end{array} \right.$$

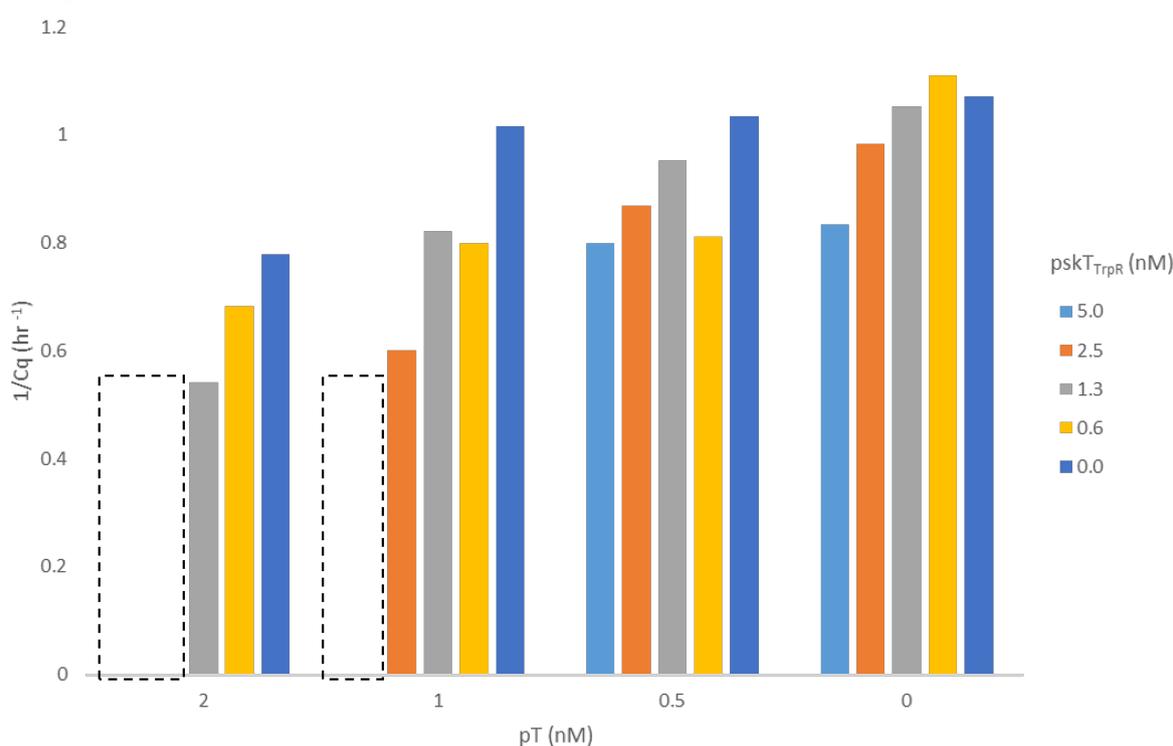
Here,  $\alpha$  represents trigger, whilst  $\beta$  – antitrigger. Going from right to left for both equations they are both: broken down by  $ttRecJ$ ; linearly produced by the corresponding sensing templates (carrying the TF-response function or being simply sT); and annihilate each other through binding each other with an affinity constant  $k$ . Furthermore, the trigger sequence is subject to amplification by aT and deactivating by pT, where the equivalent terms in the equation represent concentration-affinity product, whilst constants  $a$  and  $b$  – saturability.

From this, we can infer, that although in some instances, as seen in **Figure 146**,  $pskT$  can have an effect on the circuit on the edge of bistability, pT is posed to have greater impact.

Increase in the concentration of pskT results in the increase in linear production of the antitrigger, however, this is mitigated by the ttRecJ. pT is not affected by such constraint: when not saturated in early stages of amplification increase in pT proportionally increases the sink capacity. Furthermore, as we have seen, the simple inactivating architecture of pT is more efficient than the pskT, which have to produce the long-tailed antitrigger. Therefore, pT capacity to act as sink per nM of oligo is greater. This can be useful, when considering the overall quantity of the DNA participating in the molecular program.

We briefly demonstrate the concept by showing an optimisation step of the pskT<sub>TrpR</sub>-based circuit used in 1.3.2 (**Figure 149**). Here, we varied the concentration of pskT and the pT simultaneously.

Here, in absence of pT, pskT<sub>TrpR</sub> has little noticeable effect. However, increasing pT concentration by as little as 0.5 nM or 1 nM, brings the highest pskT<sub>TrpR</sub> concentrations from the “on” state to the “off” state sharply. The converse argument is also true: increase in pskT at well-picked pT concentration can also result in such transition (as we have shown above). However, we note that there is a difference in concentrations concerned, making pT a more favourable candidate for approaching bistability, which is then to be exploited by pskT for sensing.



**Figure 149** Optimisation step of the pskT<sub>TrpR</sub>-based circuit used in 1.3.2. Concentrations of the pT are varied (2-0 nM) together with those of pskT<sub>TrpR</sub> (5-0 nM) as shown in the legend. Dashed boxes highlight samples that remained “off” and therefore are plotted as 1/Cq = 0.

### 3.3.7 Other aspects to consider

#### 3.3.7.1 ppT design: trigger affinity impacts switch behaviour.

During our work we considered several variants of the primer-producing templates (ppTs). We were inspired by the approach of Sieskind here as well and therefore took his

design as a starting point (**Figure 150**). Sieskind determined that short primers Fwd51 and Rev51 are sufficient for context-independent amplification of the gene of interest from the SNAP-tag containing plasmid. At the same time, the length allowed them to dissociate relatively easily, which was important in case of his circuit relying on Vent(exo-) polymerase.

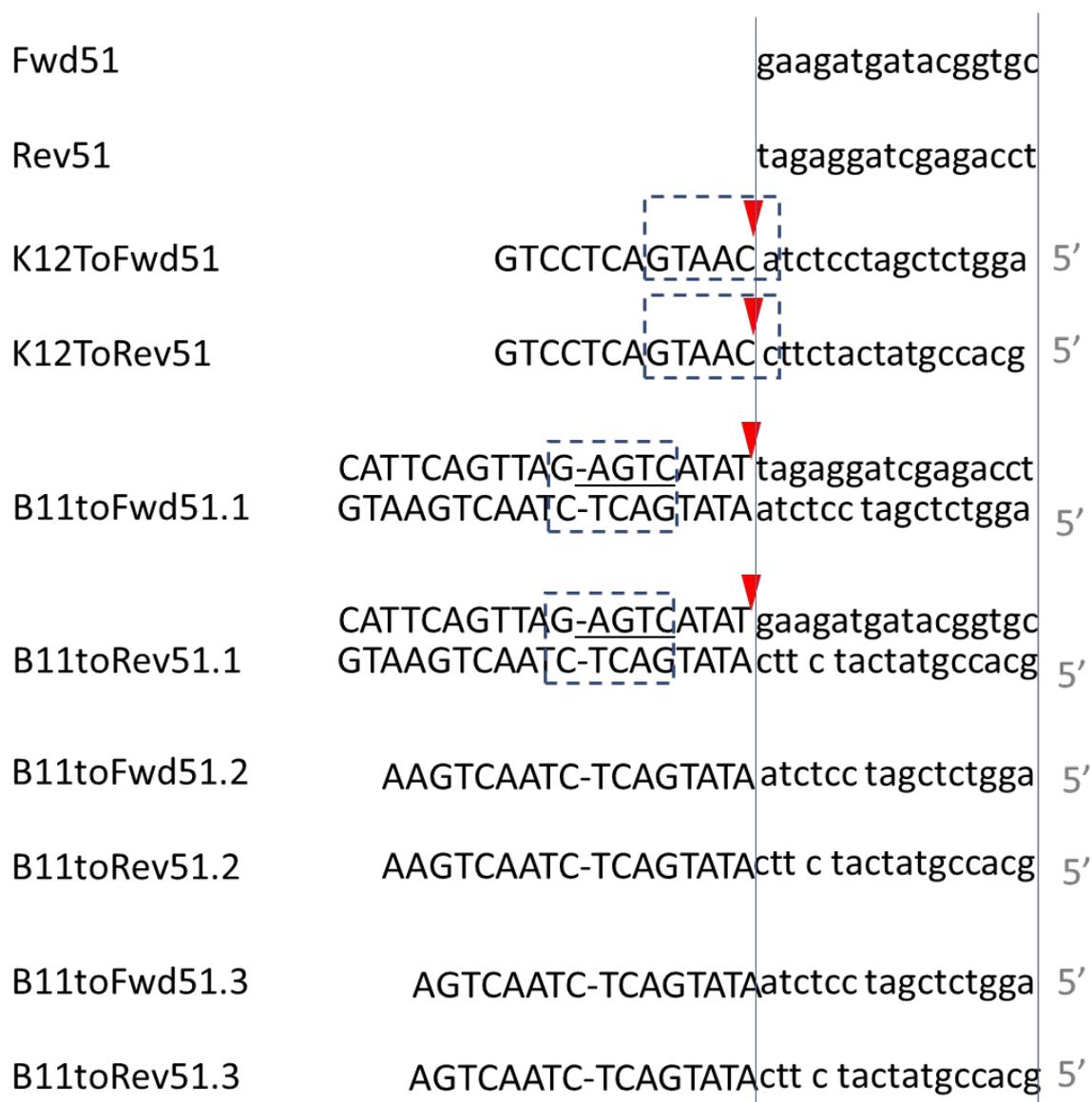
His circuit was designed to detect Nb.BsmI activity. Therefore the roles of the two nickases were inverted, compared to our circuit: the aT module was driven by Nt.BstNBI, whilst Nb.BsmI was driving the sT and primer generation. This allowed to design primer producing templates driven by Nb.BsmI nicking site that was conveniently completed by the first nucleotide of the primer.

In our case, we kept the primer sequences (along with the vector) but chose to drive the primer generation with Nt.BstNBI. This allowed us to mirror the use of the nickases and Nt.BstNBI appeared more efficient. Most importantly, however, this allowed for more modular design: independent of the input and output sequences full Nt.BstNBI recognition site can be added (though in our case there was a 1 bp overlap), followed by a spacer. The addition of the 8 base pairs also leads to irreversible binding of the trigger sequence. Once the primer production is “ON” it remains so. Coupled with the switching behaviour upstream this should have a sharper response in primer production initiation.

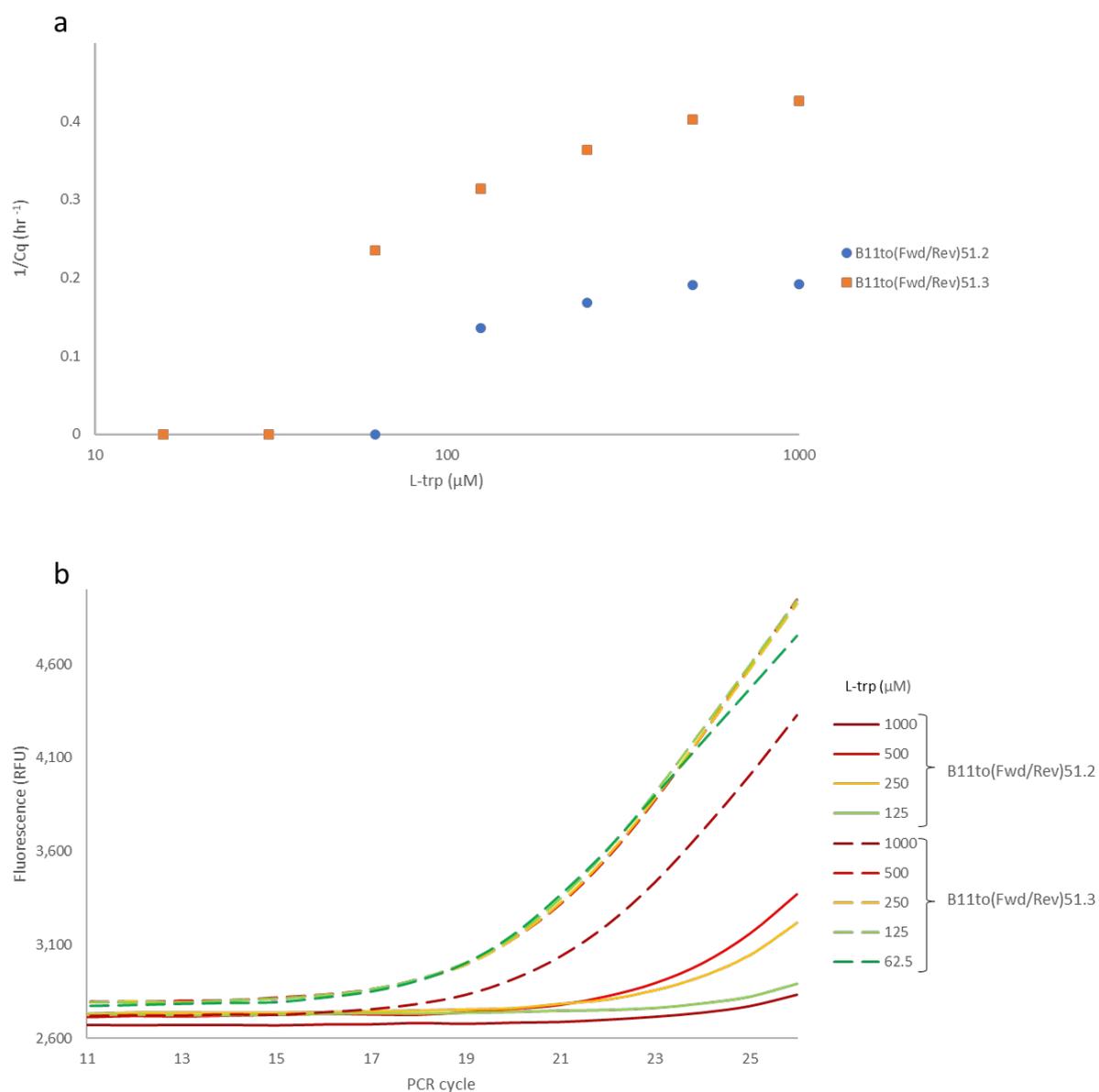
As ppT bind trigger sequences and are added at 10 nM concentrations (arbitrarily determined), we were concerned whether it could act as an additional sink for the trigger sequence in early stages of amplification. This would linearize early stages of the response and saturate the ppTs gradually, rather than in a switch like fashion. We have therefore gradually decreased the affinity of ppTs to our B11 trigger by removing 2 or 3 complementary bases at the 3'-end. Similar to other modules, we called ppTs to describe the input (B11) and the output (Fwd51), followed by a version number (1-3). One may notice that our designation for Fwd and Rev are contrary to that of Sieskind but we believe this not to affect the behaviour of a given pair.

We used version 2 of our ppTs for all of the experiments reported (also previously mentioned, as ppT<sub>Fwd</sub>/ ppT<sub>Rec</sub>). Version 1 was abandoned during early optimisation stages. However, we did compare version 2 with version 3 (**Figure 151**). This was done by adding the same range of L-trp to both and observing switching and amplification. 1 bp difference appears to facilitate earlier exponential amplification in line with our hypothesis. This, in turn leads to earlier PCR amplification.

Our speculation is that sequential consumption of trigger, first by the detection circuit and the by ppT (only once exponential amplification had occurred), allows the former to start earlier, in turn saturating the ppTs earlier as well. We believe this to be a useful insight for future designs that would be both fitter for quick switching and for easy designation of the moment all ppTs were saturated. However, this experiment was done towards the end of the project and no other experiments were carried out to exploit this insight.



**Figure 150** Examples of different ppT designs. Top to bottom: Fwd/Rev51 primer pair, K12ToFwd/Rev51 ppT pair, B11toFwd/Rev51.X ppT pair versions. Blue vertical lines are shown for ease of alignment. Lowercase letters represent primer output sequence. Dashed boxes show nickase recognition domains, whilst red triangle – the nicking site.



**Figure 151 L-trp detection and amplification response with different ppTs in IPA.** (a) IPA reactions containing various concentrations of L-trp (1-0 mM) were ran, containing 10 nM of a pair of B11to(Fwd/Rev)51.2 (circles) or B11to(Fwd/Rev)51.3 (squares). (b) PCR cycling fluorescent trace of samples shown in a. Only curves that did not stay flat are plotted. The highest concentrations of L-trp are correctly plotted as not being the fastest, we are unsure why.

### 3.3.7.2 *Transcription factor formulation strongly affects the behaviour of the sensor*

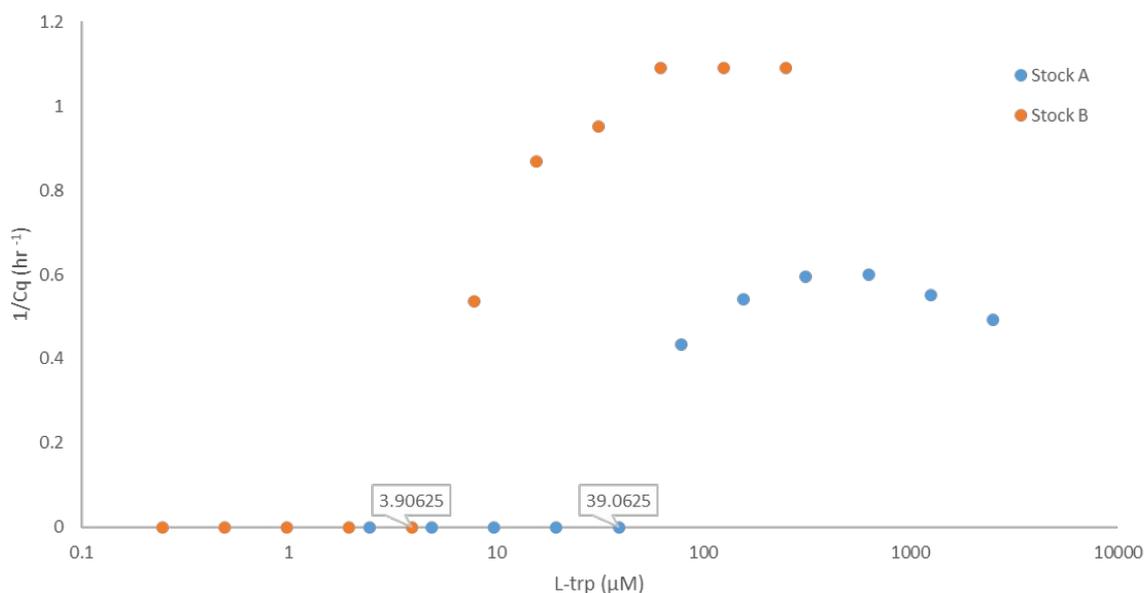
Transcription factors are at the heart of the sensing circuits that we have built. We noticed that their concentration and purity can strongly affect the circuit. Below we will describe some observations we have made relating to both.

#### 3.3.7.2.1 Concentration

We have previously mentioned that TrpR stock dilution just by two-fold can have an impact on the detection limit of the circuit (**Figure 50**). This investigation was prompted by discrepancies during our attempts to reproduce results of a detection of a fixed range of analyte, for example L-trp, with a different transcription factor stock.

We have noticed the results can vary in how many points are observed before escaping the detection limit for seemingly equivalent concentrations of the transcription factor in otherwise equivalent experiments (**Figure 152**). Since protein stocks were the only variation between the experiments, we first quickly excluded the potential exonuclease contamination (discussed in 0 below): the background rise in Cy5 fluorescence was rather similar. This meant the difference came from a discrepancy in the actual TF concentration, despite apparent equivalence.

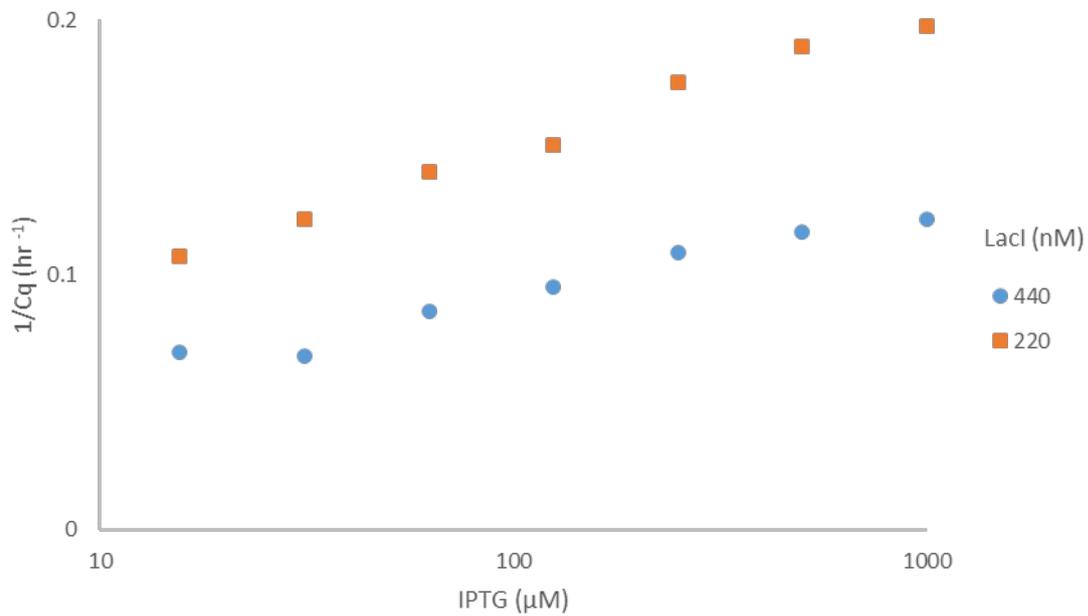
Coupled with prior observation that large TF excess can strongly boost sensitivity in such systems<sup>167</sup> we concluded that we were observing the same effect. The scale of the effect was surprising: it happened to be exactly 10-fold change, especially given that apparent lower protein concentration gave better sensitivity. This highlights the inaccuracy of our concentration approximation using  $A_{280}$  measurements. We tended to evaluate and normalise new batches of the transcription factors through tests within our circuits. Approaches beyond SDS-PAGE and  $A_{280}$  could make this step redundant. Alternatively, large batches could be produced to avoid frequent recalibration.



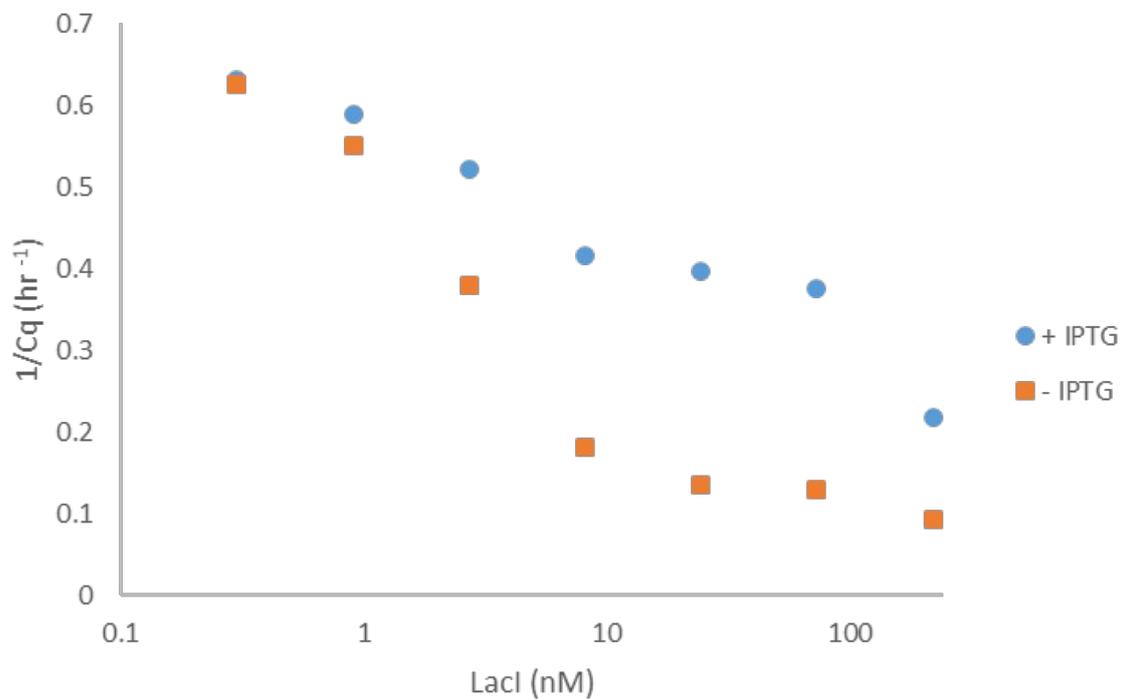
**Figure 152 Effect of change of TrpR stock on detection capacity of pskT-based switch.** Stock A was thought to be used at 630 nM, whilst Stock B at 500 nM measured through  $A_{280}$ . Identical molecular programs contained 10nM pskT and 7pM sT.

Similar to TrpR, we explored the effect of Lacl concentration on the sensing assembly in hope that increase in concentration could improve the detection limit (**Figure 153**). It appears, increasing concentrations of Lacl shift the entire response-curve, without clearly improving the resolution. Further increase in start delay could start pushing the points out of the window of observation within which the switch is stable. Therefore, we decided to keep Lacl concentration at minimally sufficient 110 nM. This effect could be driven by Lacl non-specific DNA binding disrupting the circuit or weak transient binding to the psT in absence of IPTG.

At concentrations significantly lower, the Lacl response is also lost, as there is insufficient amount of Lacl for interference (**Figure 153**). This is in line with our expectations from **Figure 44**.



**Figure 153 Effect of Lacl concentration on IPTG resolution.** Experiments were run as in Figure 46 bar with different concentrations of Lacl: 440 nM of dimer (circles), 220 nM of dimer (squares).



**Figure 154 Loss of IPTG detection of low Lacl concentrations.** Samples containing 1 mM IPTG (circles) or not (squares) are shown. Conditions as in Figure 153, other than the concentration of Lacl is varied instead.

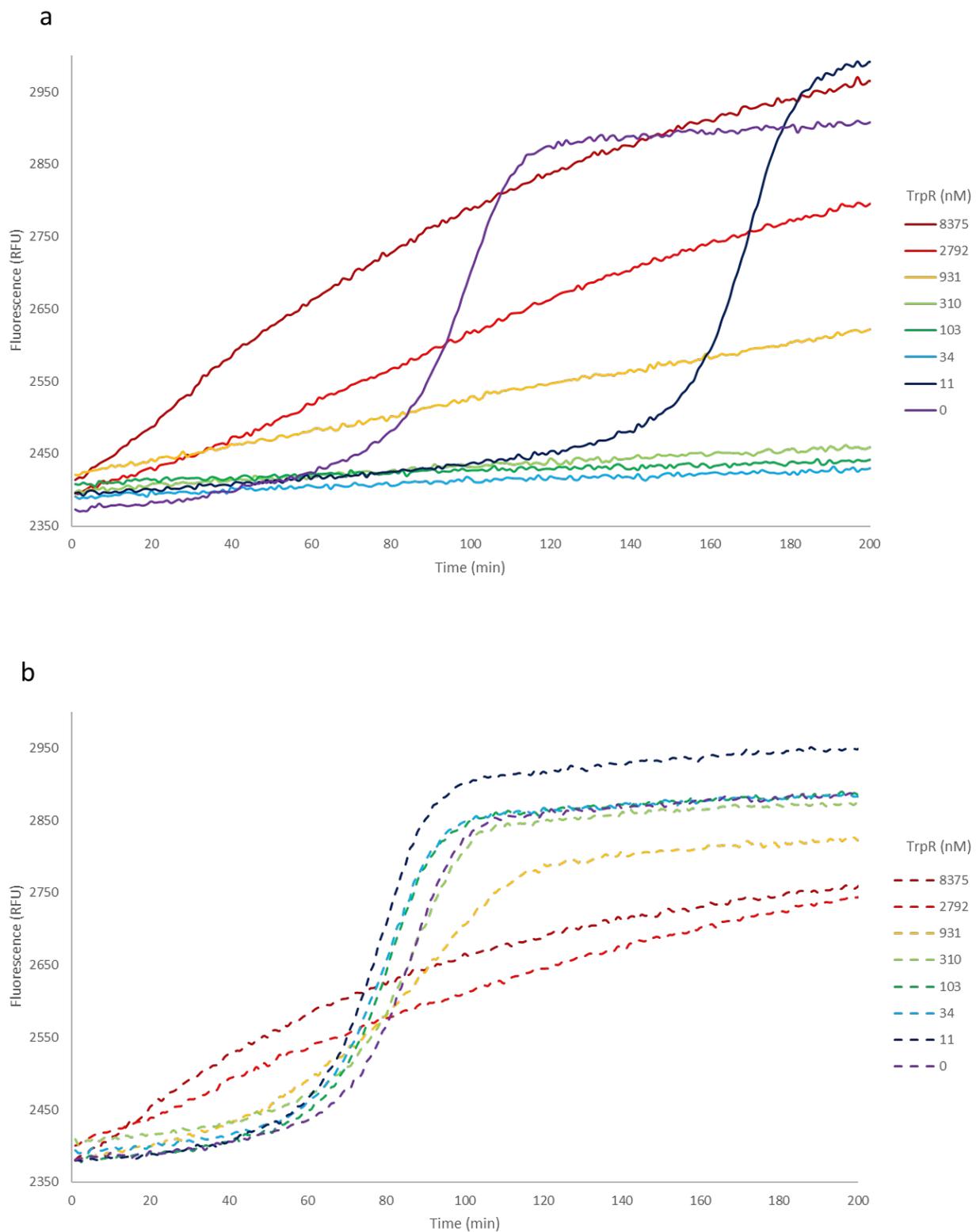
### 3.3.7.2.2 Purity

Early on, we observed that highly concentrated protein preparation would give rise to disruption of the switch, as in **Figure 155**, for example. Flat line or usual switching would be observed in the FAM channel (not shown). However, this is combined with non-sigmoidal rise in Cy5 fluorescence for high protein concentration samples.

We believe this is due to contamination with non-specific DNAses breaking down program components.

Like one would expect, without TrpR the switch loses sensing capacity. With increase in concentration of TrpR within the permitted region the window between positive and negative control increases, until bistability is achieved. For practical purposes, we would therefore empirically pick a concentration of the repressor slightly above the minimal required for bistability and take it forward, governed by frugality and concern for potential nuclease effects. Indeed, in the region between 310 nM and 34 nM the sensor would appear to behave in a rather similar manner.

As discussed in the methods, we purified our transcription factors using standard protocols for His-tag purification. Optimisation of all of the steps of such protocols could improve consistency of the batches and decrease exonuclease contamination. However, simple recommendations, suggested by the column manufacturer, such as increase of NaCl concentration in purification buffer to 1 M, as well as increase in imidazole concentrations and additional washes on the column did not suffice. We suspect an alternative expression strain and protein expression induction regime could be tried. The most robust solution, however, would be the use of other chromatographic methods, such as size exclusion chromatography, for purification.



**Figure 155** Cy5 fluorescence for a range of TrpR protein affecting psT-based switch. TrpR concentration approximated via spectrophotometry is given in nM in the legend. Solid lines correspond to samples with 1000 mM L-trp added (**a**), whilst dashed lines represent the same sample bar without L-trp (**b**).

### 3.4 Summary

In this chapter we demonstrate that our transcription factor-based approach is not limited in use to the particular systems we have reported in chapters 1 and 2.

The “genelet” approach demonstrated that the DNA switches can be connected to a transcriptional activity, but this concept was not successful for small molecule sensing. We also showed that pST/pskT architectures can be further modularised by the use of distal bottom-strand nicking nucleases. The operator “tail” constraint can be likewise alleviated through simple chemical modifications of the template.

Echoing the works at higher temperatures<sup>152</sup> we demonstrate that switches, driven by different nickases or simply orthogonal through minor sequence alterations can be designed. However, distributed circuit inversion relying on such switches may solicit more careful designs to insure complete orthogonality.

The sensing interface itself is not constrained to LacI and TrpR transcription factors, although the plug and play approach sometimes fails unexpectedly. Therefore, caution has to be exercised. On the other hand, a transcription factor shown to be functional once (TrpR, for example), can be ported into other switches and temperatures regimes. On the other side, the switches need to be carefully designed for a particular temperature.

Finally, we have shared some observations useful for future expansion of the work, concerning the optimisation of the circuit in IPA reactions and ensuring proper preparation of the transcription factors in order to integrate optimally in the design.

## Discussion and conclusion

This work demonstrates that we can engineer modular converters to interface DNA-based molecular programs with small molecule inputs. This is significant, because once the molecular signal is forwarded to the DNA layer, a wide range of on-the-shelf molecular programming techniques and modules can be used for signal processing. Here, we used signal amplification, thresholding, shaping, inversion and integration. Both in Chapters 1 and 3 we demonstrate that this approach is modular and tunable, as DNA reactions can be designed with limited cross-over, between themselves or with other chemistries. It is thus possible to add a branch – or replace one – with the operator sequence and the TF of interest, without the need to redesign the other components. This, we have no doubt, in turn paves way to scalability of future circuits in line with those, already reported for nucleic acids only inputs.

A limit of this design, however, may lie in its dependence on existing allosteric DNA-binding proteins. Fortunately, large libraries of natural and artificial TFs are readily available, due to their common use in a variety of applications<sup>166,217,221</sup>. Faulon *et al.* for example demonstrated the modular use of various transcription factors for similar signal integration in bacterial context<sup>222,223</sup>, and compiled a catalogue containing over 4600 entries of characterized TFs<sup>224</sup>. This is complemented with a CAD tool suggesting enzymatic pathways for small molecule conversion to those detectable by transcription factors<sup>129</sup>. As we demonstrate in Chapter 2, this is also applicable to our approach: non-detectable analytes, such as indole can be converted into detectable ones, such as L-tryptophan.

As we have discovered, additional small molecules and proteins can interfere with the circuit. However, in case of small molecules as we show for indole, for this to occur the concentration should be considerable, compared to the  $K_d$  of most transcription factors. Similarly, we think most metabolic enzymes have evolved to be compatible with the intracellular environment of the host and other reactions within. Therefore, interference with simpler PEN context is unlikely. We consider the case of *Pyrococcus furiosus* tryptophan synthase subunit B exceptional: extreme heating is required to bind DNA. We would also note that, the aforementioned CAD tool takes advantage of enzymes, functional in mesophilic settings.

Concerning non-biologically relevant molecule sensing, a number of studies have reported a surprising mutational plasticity of natural allosteric transcription factors<sup>225</sup>. TrpR for example, can be repurposed from L-trp to specific bromotryptophan sensing with only a few mutations<sup>226</sup>. Likewise, the logic of the response of the transcription factors has been reported to be inverted in some cases. Recent work has demonstrated that LacI can even display band-stop filter behavior with only two mutations<sup>227</sup>. The retargeting of transcription factors to sensing non-natural molecules then in turn expands our scope for directed evolution goals. A transcription factor and the cognate enzyme (often adjacent in bacterial genomes) can be evolved in pair to work with a novel molecule, given potential for promiscuity exists. First, the transcription factor is modified through a binding assay-based directed evolution to specifically sense the modified, possibly non-natural, product of interest. In parallel, another variant could be evolved to invert the response to the original metabolite, although inversion can also be done on the DNA layer, as we have shown. Second, the new transcription factor(s)

can be used to evolve the cognate enzyme to catalyze previously secondary reaction, whilst penalizing for the wild-type activity.

We have mostly spoken about reactions, where a certain product is sought. At the same time, some settings may focus on situations where not generation but rather a breakdown of a substrate is required. In our group, Margarida Gomes currently works on enzymatic breakdown of plastics. These are solid, so the exciting approaches explored by her for directed evolution can exploit compartmentalisation of materials needed for replication, released subject to sufficient activity. This is not an option for soluble substrates. Although not explored in detail, industrial waste products come to mind as well as endocrine disrupting substances released from households into the sewage.

Something we did not work towards but should be achievable is the evolution of enzymes for biomedical applications. The simplest case is enzyme replacement therapy, where the patients are subjected to biweekly infusions of the enzymes, they are lacking due to genetic abnormalities. These could be made more stable or efficient, decreasing the periodicity of infusions and the cost accordingly. The latter is due to decrease in the amount of enzyme needed to produce. A more advanced approach explored by Martin Fussennger's group is the use of what they call prosthetic gene networks<sup>228</sup>. A semi-permeable compartment with transgenic cells is implanted into the host with the cells being able to compensate for the function the host is lacking. As the cells are usually from the host, they can persist for long periods of time whilst carrying out their function. Once again, they can be made more efficient, making the prosthetic compartment smaller. Minimal adaptation of the *in vitro* PEN-CSR will be required and will concern replacement of the cell free protein expression system with that, suitable for mammalian genes. Fortunately, such systems are available<sup>229–231</sup>.

When used in-emulsion, molecular programs offer massive assay parallelization at little extra cost, provided all operations are performed autonomously in the closed microcompartment<sup>232</sup>. We therefore envisage the use of our sensors in applications involving high-throughput screening, given that the desired product does not travel between droplets. Here, in Chapter 2 we demonstrate an application related to enzymatic library variant screening through coupling the sensing to the relevant downstream actuation. However, our colleagues exploring biomedical applications: Guillaume Gines<sup>155</sup> and Antoine Masurier use *in emulsio* molecular programs for biomedical screening. For example, they evaluate single cell microRNA or surface marker profiles and use molecular programs to specifically obtain information about the cells of interest. This approach could be complemented by high throughput metabolome analysis. Transcription factors are usually used by the organisms to control their metabolic flux *in vivo*. This therefore guarantees a library of transcription factors that can be used to create highly multi-dimensional metabolome profiling networks. Some transcription factors sense or regulate multiple pathways or work in complexes; however, this can be alleviated thanks to their evolvability we have mentioned before.

Antoine Masurier's work is also of note, as in another project he uses antibodies for specific detection of proteins with the signal being converted onto the molecular program level, providing access to the proteome exploration.

Although we have focused here on small organic molecules, a number of transcription factors are sensitive to metal ions<sup>233</sup>, pH<sup>234</sup> or physical signal<sup>235</sup>, widening the range of potential inputs. Finally, the protein mediated strategy explored here is not limited to

transcription factors. Split DNA- or RNA-processing enzymes, for example RNA polymerase<sup>236</sup> or CRISPR-Cas9<sup>237</sup>, can be made sensitive to a variety of ligands, and mediate their presence to a DNA/RNA circuit. Similarly, optogenetics explores strategies to modulate genetic processes using light<sup>238</sup>.

Although this work mainly addresses the chemical interface upstream of the DNA circuit, the downstream connection is equally important for “real-world” applications<sup>239</sup>. For example, the actuation of enzymes with DNA has been demonstrated<sup>240</sup> and implemented in particular as an output of PEN-DNA modules<sup>241</sup>. This can be used to create non-nucleic acid outputs, for example, metabolites, contributing to the exploration of artificial cell systems.

The *in vitro* PEN-CSR and molecular programming approaches have their limitations. Despite apparent simplicity of the individual elements, combining them into a functional pipeline can be challenging. Sieskind and Dramé-Maigné dedicated years to extensive optimization of their respective processes to arrive to mock library selections with repeated cycles posing an additional challenge. Processes inside living organisms are highly dependent on their context and had millions of years to co-evolve, whilst we only hope to make an educated guess to replicate dynamics of individual components of such systems in isolation. As we have illustrated in Chapter 3, even PEN-DNA systems of two or three oligonucleotides may require several experiments to optimize. This was possible thanks to prior personal expertise on the matter as well as extensive experience of the inventors of the approach. In this sense, rationally optimizing a 33-component mixture based on two fluorescent outputs to achieve behaviors reported in Chapter 2 can be considered a remarkable feat.

It is evident that as complexity of the pipeline increases, so does the training required to master all of the concepts. Likewise, the overlap between acceptable sections of the parameter spaces constricts, making it increasingly hard to add new components. This is also impacted by the fact that only a limited amount of readouts is possible for monitoring, usually just the intended output. To this extent, the vast non-functional regions of the design space are frequently assigned by us as “toxic”, with only an educated guess regarding the cause.

In light of this, Thibault Di Meo and Taro Furubayashi of our group explored an IVTT-based directed evolution approach aiming to use autonomous replication systems. Here, the IVTT provides a pre-optimized environment for enzyme expression and activity. Provided the artificial autonomous replication system is coupled to the desired activity (through a host of existing tools, portable from *in vivo* approaches), it offers a one-pot directed evolution system. In comparison *in vitro* PEN-CSR requires DNA amplification, protein expression, PEN-detection and PCR, involving the production of emulsion three times. One-pot reaction also means the system can be fed new empty compartments, to be populated during transient emulsion disruption, making the directed evolution process continuous. Therefore, this approach appears promising. Interestingly, transcription factor interference is one of the strategies they consider, for replication control.

Overall, two and a half years of experimental work only permit to explore so much. We believe great expansion awaits the PEN toolbox in years to come, possibly with the some of the approaches we demonstrate in Chapter 3. We believe this is realistic as the concepts we had time to take forward proved functional.

## Materials and Methods

### 4.1 Molecular biology

#### 4.1.1 Transcription factor production

Transcription factors were expressed in *E. coli* T7 express strain from NEB as per manufacturer's instructions in a volume of 1L of LB from pR063 plasmid kindly provided by Rémi Sieskind (see 4.3.3). pR063 is a pIVEX-based plasmid containing a T7 promoter, RBS as well as a thrombin protease site followed by 6xHis-tag and KlenTaq gene in the middle that was replaced with a TF of choice via Gibson assembly (NEBuilder HiFi DNA Assembly Cloning Kit standard protocol). Similar 2-part Gibson assembly was used throughout. TFs were sourced from NEB 5 alpha genome (GenBank: CP017100.1) by PCR apart from TetR, kindly provided by Jure Tica on the HP\_p15A-K07dg1-10-dM13 plasmid. Primers used are listed in **Table 3**.

A pellet was then collected and purification carried out using His GraviTrap TALON per manufacturer's instructions. Lysis was carried out by freeze-thawing once and sonication with Branson Sonifier 250 for 10 minutes alternating 30 second on/off cycles at 40% duty cycle and Output Control 4 using the larger probe provided. Binding, wash and elution buffers were modified to contain 500 mM NaCl (all) and 20, 30, 500 mM imidazole. After elution fractions were run on NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 15-well in MOPS buffer and stained with SimplyBlue™ SafeStain as per "Microwave procedure" to evaluate product purity, as in **Figure 156**. CLARIOstar LVis Plate was used to approximate protein concentration by using 260nm absorbance readings. For long term storage 20mM Tris-HCl buffer at pH 8 with 500mM NaCl and 50% glycerol was used.

Primer name	Sequence
GlacI FWD	CTTTAAGAAGGAGATATACCATGGTGAATGTGAAACCAGTAAC
GlacI REV	AAGCTGCCACGCGGCACCAGCTGCCCGCTTCCAGTC
GtetR FWD	CTTTAAGAAGGAGATATACCATGAGCCGTCTGGACAAGAG
GtetR REV	AAGCTGCCACGCGGCACCAGGCTACCGCTCTCGCATTTCA
GTrpR FWD	CTTTAAGAAGGAGATATACCATGGCCCAACAATCACCCCTA
GTrpR REV	AAGCTGCCACGCGGCACCAGATCGCTTTTCAGCAACACCT
GFadR FWD	CTTTAAGAAGGAGATATACCATGGTCATTAAGGCGCAAAG
GFadR REV	AAGCTGCCACGCGGCACCAGTCGCCCTGAATGGCTAAAT
iR280	CTGGTGCCGCGTGGCAGCTtacATCCCGGGGGGGTTCT
REV_020	GGTATATCTCCTTCTTAAAGTTAAAC

**Table 3 Primers used for transcription factor production.** All primer names include the name of the corresponding transcription factor. The remaining primers were used for pIVEX backbone amplification.

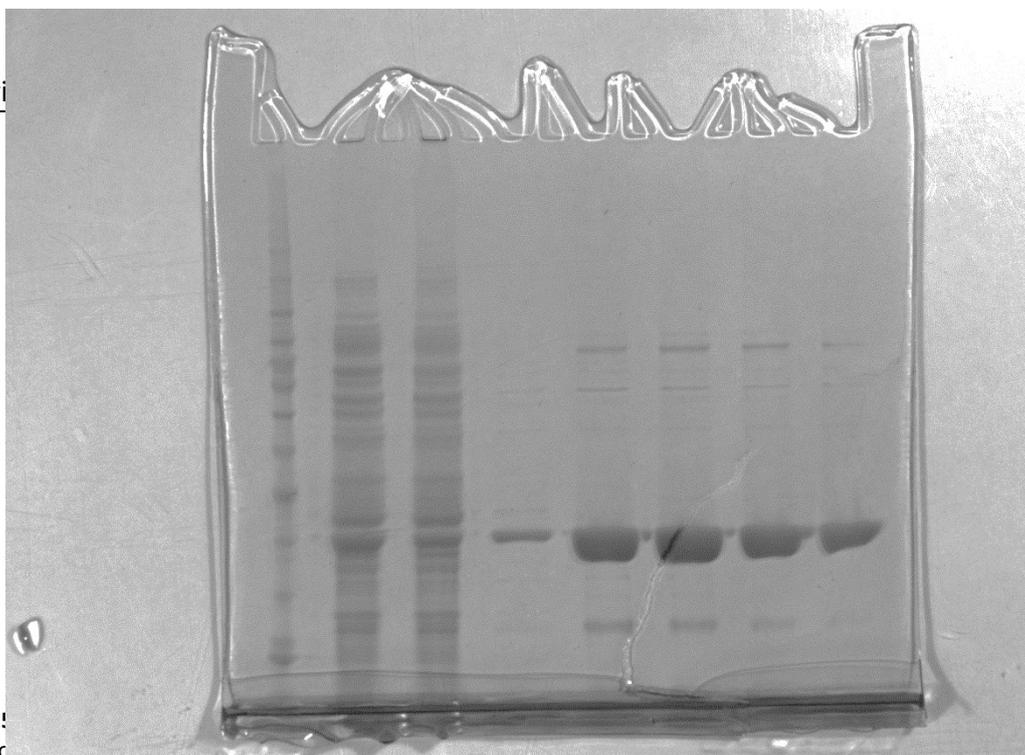


Figure 1:  
flow thro

ill lysate,

#### 4.1.2 Enzyme purification

*E. coli* based TrpB CDS was extracted as TF before whilst *P. furius* variant was ordered from IDT using the UNIPROT ID Q8U093 entry with adjustments suggested by IDT for ease of synthesis and expression (see 4.3.3). The genes were cloned as before ( specific primers in **Table 4**) and expressed in Promega KRX cells as per manufacturer's instructions.

Name	Sequence
PfB2 REV	AAGCTGCCACGCGGCACCAGCACATTACCGCTGACTTTCA
PfB2 FWD	CTTTAAGAAGGAGATATACCATGTGGTTCGGTGAGTTTGG
EcB REV2	AAGCTGCCACGCGGCACCAGGATTTCCCCTCGTGCTT
EcB FWD	CTTTAAGAAGGAGATATACCATGACAACACTTAACCCCT

**Table 4** Primers used for enzyme cloning. The name of the primer indicates correspondence to the enzyme.

#### 4.1.3 Site-directed mutagenesis

Point mutations were incorporated using Q5 Site-Directed Mutagenesis Kit as per manufacturer's instructions using primers shown in **Table 5**.

Name	Sequence
T292S_R	TTGCCCCCTCCTCGTCCTG
T292S_F	ATCAAGCCAAGCCACTCGATTGCGCCGG
M149T_F	GTTTTTCGTACCCGCTTAATGGGTGCGGAAG
M149T_R	GTTAGGCGACTGGCGTTC
N171D_F	AGATGCCTGTGATGAGGCGCTGC
N171D_R	TTCAGCGTCGCGGAACCG

**Table 5** Primers used for site directed mutagenesis. The name of the primer represents the mutation it incorporates.

#### 4.1.4 SNAP-tagging TrpB

EcB and Pfb variants were cloned into SNAP-containing p070 vector to replace sfGFP, using the same forward primers, as before. The reverse primers for the inserts and backbone are shown below (Table 6).

Name	Sequence
PfB2 SNAP2 REV	GTCGCCATCACCGGTCTGTACCACATTACCGCTGACTTTCA
EcB Snap2 REV	GTCGCCATCACCGGTCTGTACGATTTCCCCTCGTGCTT
iR398	GTACAGACCGGTGATGGCGAC

**Table 6 Primers, specifically used for SNAP-tag addition.** The names of the primers indicate the corresponding enzymes, whilst iR 398 is the reverse primer for the backbone, to be used with REV\_020.

## 4.2 Molecular programming

Similar sets enzymes were used with two buffer recipes: the miR buffer optimised for detection only and IPA mod 2 (mentioned as “IPA buffer” everywhere in the text), that also permits PCR.

### 4.2.1 The miR buffer

Reactions were performed in a buffer containing 20 mM Tris-HCl pH8.9, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 40 mM KCl, 10 mM  $\text{MgSO}_4$ , 10 mM NaCl, 2  $\mu\text{M}$  Netropsin (Sigma-Aldrich), 200  $\mu\text{g ml}^{-1}$  BSA (New England Biolabs, NEB), 0.1% Synperonic F108 (Sigma-Aldrich), 0.4X EvaGreen dye (Biotium) and dNTPs (25  $\mu\text{M}$  each). All experiments also contained Bst DNA polymerase, large fragment, the nicking endonucleases Nb.BsmI and Nt.BstNBI (all from NEB) respectively used at 36, 300 and 0.5 U  $\text{ml}^{-1}$  (respectively 0.45, 3 and 0.005% final dilutions of the commercial stock solutions). The thermophilic 5' -> 3' exonuclease ttRecJ was purified in the laboratory, stored in Diluent A + 0.1% Triton X-100 at a concentration of 1.53  $\mu\text{M}$  and used at 22.95nM. This is based on a recipe previously developed by Gines, Rondelez and colleagues. 50 nM aT, 50 nM rT and 7 nM pT was used, unless otherwise specified.

Experiment in **Figure 40** and those performed in 3.3.4.2 used the reaction buffer that was identical to the one reported in the above. However, the enzyme composition was different. Vent (exo-), the nicking endonucleases Nb.BsmI and Nt.BstNBI (all from NEB) were respectively used at 80, 300 and 100 U  $\text{ml}^{-1}$  (respectively 4, 3 and 1% final dilutions of the commercial stock solutions). The thermophilic 5' -> 3' exonuclease ttRecJ was purified in the laboratory, stored in Diluent A (NEB)+ 0.1% Triton X-100 at a concentration of 1.53  $\mu\text{M}$  and used at 22.95 nM as before. Unless otherwise indicated, aT $\alpha$  was present at 50 nM, pT $\alpha$  at 20 nM, rT $\alpha$  at 50 nM.

### 4.2.2 The IPA buffer

Based on a recipe kindly developed by Rémi Sieskind and colleagues, reactions were performed in a buffer containing 20 mM Tris-HCl pH8.9, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20 mM KCl, 7 mM MgSO<sub>4</sub>, 5 mM NaCl, 2 μM Netropsin (Sigma-Aldrich), 1% Triton X-100, 200 μg ml<sup>-1</sup> BSA (New England Biolabs, NEB), 0.4X EvaGreen dye (Biotium) and dNTPs (225 μM each). All experiments also contained Bst DNA polymerase, large fragment, Vent (exo-) DNA polymerase, the nicking endonucleases Nb.BsmI and Nt.BstNBI (all from NEB) respectively used at 28, 20, 200 and 50 U ml<sup>-1</sup> (respectively 0.35, 1, 2 and 0.5% final dilutions of the commercial stock solutions). The thermophilic 5' → 3' exonuclease ttRecJ was purified in the laboratory, stored in Diluent A + 0.1% Triton X-100 at a concentration of 1.53 μM and used at 22.95nM. Unless otherwise indicated, where appropriate, L-tryptophan synthesis components were also added in form of 5mM L-serine, 80 nM pyridoxal 5'-phosphate (PLP), 2.5 mM indole as well as TrpR to a concentration of 100nM, where detection capacity was tested. Similarly, 90 nM aT, 50 nM rT, 15 nM pT, 8 pM sT and 10 nM ppT (each) were used, unless otherwise specified.

### 4.2.3 Oligos used

Oligos were typically ordered from Eurofins unless dye modifications were required in which case Biomers was the preferred source. The list for B11 and B11b switches is presented in **Table 7**. The remainder is reported directly throughout the text. B11b switch was used with 50 nM aT and 8 nM pT, where appropriate.

Part name	Sequence
B11	CATTCAGTTAG
aT	C*T*A*ACTGAATG-CTAACTGAA
pT	A*A*-C*TAAGTGAATG
rT	Atto663-C*A*T*CGATATACTAACTGAATGCGATG-BHQ2
B11b	CATTCAGTTAG
aT B11b	C*T*A*CTTGAATG-CTACTTGAA
pT B11b	AA-CTACTTGAATG

**Table 7 Oligos used in molecular programs.** Full oligo names followed by the sequence. \* denotes PTO modification. Dyes are shown where appropriate.

### 4.2.4 Running the reactions

Reactions were assembled in a total volume of 10 μl and run at 37 °C in a CFX96 real-time PCR detection system (Biorad). As reported for polymerase/nickase amplification schemes<sup>154</sup>, we sometimes observed the emergence of template-independent amplification products after at longer incubation times. These parasitic reactions—that may occur from the high or low state of the switches—are easily distinguished from their fluorescent signature and were not considered for the analysis.

Reactions were monitored non-specifically from the fluorescence signal produced by a double-stranded DNA intercalating dye (EvaGreen). A template conjugated to Atto663 fluorophore on the 5'-end and BHQ2 quencher on 3'-end (rT) was used to specifically monitor switch activation. The signal of the fluorophore is quenched when the template is in a hairpin conformation altered upon trigger binding. Fluorescence signals were acquired every minute in the channels corresponding to the two dyes present in solution. "Start time" was derived from Cy5 time traces through a workflow implemented in R. In brief, each trace is normalised so that fluorescence values lie between 1 and 0; differences between 2 points 10 minutes apart are then calculated (last 10 minutes are excluded) and the lower time-point corresponding to maximum difference value is extracted. The data is manually curated to identify flat traces with no distinctive switch-driven fluorescence increase<sup>174</sup> and are shown as "No start".

### 4.3 Microfluidics and directed evolution workflow

The process largely follows that proposed by Rémi Sieskind and colleagues and therefore we provide an extract from his thesis<sup>182</sup>.

#### 4.3.1 Hydrogel bead generation

##### 4.3.1.1 Hyaluronic acid thiolation

First, hyaluronic acid thiolation was carried out. For this, we prepared 500mL of MES buffer (0.1M, pH 4.75) from MES hydrate (ref.M2933-25G, Sigma-Aldrich) and hydrochloric acid (ref. 258148-500ML, Sigma-Aldrich). We then dissolved 250mg sodium hyaluronate (ref. HA40K-1, Lifecore biomedical) in 25mL of MES buffer. After total dissolution, we added first 50mg of PDPH (ref. 22301, Thermofisher) and again, after total dissolution, 300mg of EDC-HCl (ref. 8510070025, Merck Millipore). The mixture was then constantly stirred and incubated at room temperature (around 25°C) for 6 hours. We then dialysed the mixture against ultra-pure water (milliQ) overnight at 4°C using a dialysis cassette MWCO 3.5 kDa (ref. 66130 Thermofisher). The day after, we retrieved the solution and added 100mg of Tris(2-carboxyethyl)-phosphine (TCEP) (ref. C4706-2G, Sigma-Aldrich). The solution was then incubated and stirred for 3 hours at room temperature. We prepared 3L of a solution of NaCl (0.1M, pH 3.5) from NaCl powder (ref. S9888-1KG, Sigma-Aldrich) and hydrochloric acid and dialysed the mixture against this saline solution three times three hours at room temperature using a new identical dialysis cassette MWCO 3.5 kDa. We then dialysed three more times (3 hours) in milliQ at room temperature before retrieving the mixture and freeze-drying it overnight.

##### 4.3.1.2 Microfluidic generation

The two masks used for microfabrication were designed using the student version of AutoCAD and then ordered from Selba S.A. (on-film laser photoplotting at 50800dpi, negative polarity, mirror orientation).

The 2-layer mould of the Mixing-Millipede was obtained using classical microfabrication techniques based on SU-8 2000 permanent epoxy negative photoresist spread on silicon wafers. The resins and protocols were chosen with the objective of making

a first 5  $\mu\text{m}$ -high motif and a second aligned 50  $\mu\text{m}$ -high one. To do so we chose the SU-8 2005, spun at 3000 RPM and the SU-8 2050, spun at 3250 RPM. After characterization using a mechanical profilometer, we found a height of 4.5  $\mu\text{m}$  for the nozzles (first lower motif) and a height of 45  $\mu\text{m}$  for the inner (aqueous phase) and outer (oil phase) channels (second motif).

To prepare microfluidic devices, we poured a mix of 40g of liquid PDMS elastomer with 4g of the curing agent (Silicone Sylgard 184, Dow Corning, ref. 1317318) on the mould edged with an aluminium foil. We then degassed the mixture in a vacuum chamber (until all bubbles disappeared) and incubated it at 70°C for 2 hours. The formed PDMS slab was then delicately detached from the mould. The inlets and outlet were punched using a 1.5mm biopsy punch (Integra Miltex, ref. 33-31A-P/25), and the mixing chamber was punched using a 4mm biopsy punch (IntegraMiltex, ref. 33-34-P/25). The core sample of this last punching was kept for further use. The PDMS slab and a microscopy 1mm-thick glass slide were washed using isopropanol and ultra-pure water and taped on all faces to remove dusts. In a white room, we functionalized the PDMS slab and the glass slide in a plasma cleaner and bonded them covalently. The result was then incubated at 200°C for 5 hours. A 2mm-large magnetic stirring bar (Bel-Art Spinbar, ref. F37119-0002) was inserted in the 4mm-wide hole. The core sample extracted earlier was cut transversally in two halves. One half was then used to fill the hole above the stirring bar without hindering its movement. We mixed then 1g of the PDMS elastomer with 0.1g of the curing agent, degassed it and then recovered the filled hole to guarantee tightness even under pressure during experiments. The device was then cooked for 2 hours at 70°C. Finally, a hydrophobic electronic grade coating (3M Novec 1720) was injected by the outlet and then removed before a last bake at 90°C for 1 hour. This last step was re-done each time we wanted to use a new device to ensure perfect hydrophobicity.

Prior to any microfluidic manipulation, the two THA and PEGDVS solutions (see Table M.3) were filtered using syringe filters presenting 0.2  $\mu\text{m}$  pore size (Corning, ref. 431212).

We filled the two identical 250  $\mu\text{L}$  syringes (Hamilton ref. 81120) with 75  $\mu\text{L}$  of pure Novec 7500 fluorinated oil (to compensate the dead volume in the needle and tubing) and with 150  $\mu\text{L}$  of either the THA or the PEGDVS solution. We then centrifuged by hand the syringe to force the oil to be at the bottom (near the piston), the aqueous phase above and finally the air (near the outlet). We then adapted 20-30cm of PTFE tubing (1.6mm outer diameter, 0.5mm inner diameter, BOLA, ref. S1810-09) on two 25Gx1" needles (Terumo, ref. AN\*2425R1). We adapted the needles on the syringes and filled the tubing with the solution. To increase the amount of solution in the syringe, it was also possible to suck more solution up from the end of the PTFE tubing, making sure that no air bubble enter the system. We then placed the syringes on the syringe pump and inserted the tubing into the PDMS punched inlets.

We performed the beads generation using a mixed setup containing a syringe pump, set at 3  $\mu\text{L}/\text{min}$ , to impose the flowrate and the 1:1 composition for the two aqueous inlets (THA and PEGDVS solutions), a pressure pump, much more reactive, set at around 200mbar, for the oil phase, and a magnetic stirrer, set at maximum speed to make the micro-barrel mix efficiently the two aqueous solutions. We prepared the oil phase by dissolving 2%  $w/w$  of the Fluosurf surfactant in the Novec 7500 fluorinated oil (3M). The droplets were gathered in a 1mL pipet tip at the outlet of the millipede device. After transfer in a 2mL centrifuge tube, the emulsion was incubated at 30°C for the time allowing to get the proper amount of amplified DNA. We then broke the emulsion and recovered the newly formed beads.

### 4.3.1.3 Emulsion break and washing

To break emulsions, we first removed the oil/surfactant mixture under the emulsion and then add a large amount of fresh oil without surfactant. After flipping the tube a few times, we left it on a holder so that the droplets could decant. Once the oil was fully clarified, we removed the oil again and add a few hundred microliters of perfluoro-octanol (PFO) (Sigma-Aldrich, ref. 370533-25G). The addition of a large volume of an aqueous buffer (PBS-Tween20<sub>0.1%</sub> most of the time for us) prior to PFO addition is necessary when breaking emulsions containing beads and not for classical emulsions. We vortexed extensively afterwards and then centrifuged with a bench-top microcentrifuge for a few minutes. The aqueous phase (less dense than the fluorinated organic phase, and thus above) is then recovered using a pipette.

To wash the beads, we centrifuged first the tube to pellet the beads at its bottom. We then removed carefully the supernatant using a pipette, avoiding to disrupt the pellet. Finally, we added a large volume of the washing buffer and resuspend the pellet by vortexing extensively. This procedure is considered as one wash. The beads were always washed at least three or four times, so that the preceding buffer and eventual contaminants got sufficiently diluted.

### 4.3.2 IVTT protein expression and immobilisation

*In vitro* transcription translation of proteins was done with 1nM of template plasmid for 2 hours in the IVTT mixture (Table 8) with THA beads added, where necessary. A positive control with p070 sfGFP plasmid was placed into a CFX machine to ensure the mix was functional. S17 Cell extract, energy mix and the amino acid mix were kindly provided by Rémi Sieskind.

S17 Cell extract	30% v/v
Potassium Glutamate	80 mM
Amino acid mix	3 mM
Energy mix	1x
Magnesium Glutamate	4 mM
PEG 8000	2% w/w
Maltodextrin	1.26% v/v
DNA template	1 nM

Table 8 Cell extract-based *in vitro* transcription/translation (IVTT) mixture used.

Our beads lacked RCA components and therefore the recipe was simpler. We dissolved THA and PEGDVS crosslinker separately to 36 mg/ml and 20 mg/ml accordingly in 1x PBS. The THA also contained 10 nM DyLight405-maleimide (for easier tracking of the beads), 20  $\mu$ M of benzylguanine-maleimide and 50  $\mu$ M of TCEP. They were then combined in 1 to 1 ratio as described above.

### 4.3.3 Vectors and genes

We present here the sequences, not published and specific to this work.

**Table 9 Vectors and genes used in this work unavailable un public databases.**

Name	Sequence
PfB ordered as gBlock	<p>ATGTGGTTTCGGTGAGTTTGGTGGGCAGTATGTCCCGGAAACGCTTATTGAACCGCTGAA  GGAGCTGGAAAAGGCATATAAGCGGTTCAAAGATGATGAAGAGTTCAACCGCCAACCTCA  ACTATTACTTAAAAACATGGGCTGGTCGCCAACTCCGCTGTATTATGCTAAGCGTCTTAC  GGAAAAGATTGGGGGTGCAAAAATCTACCTGAAGCGTGAAGACTTGGTGCACGGGGGC  GCTCATAAAACAAATAACGCGATTGGCCAGGCTTTGCTCGCAAATTTATGGGTAAAGACC  CGCTTAATTGCTGAGACAGGCGCAGGTCAACATGGCGTAGCCACAGCGATGGCTGGCGC  GTTATTAGGCATGAAGGTAGATATCTATATGGGTGCTGAAGACGTAGAGCGCCAGAAGA  TGAACGTTTTTCGCATGAAGTTGTTGGGGGCGAATGTAATTCCAGTCAACAGTGGCTCAC  GCACTTTGAAAGATGCCATTAATGAAGCATTACGGGACTGGGTAGCCACCTTCGAGTAC  ACGCACTACCTGATTGGCTCTGTCGTCGGCCCTCATCCGTATCCTACTATTGTGCGCGACT  TCCAGAGCGTCATTGGCCGTGAAGCCAAAGCACAAATCCTGGAGGCAGAGGGCCAGTT  GCCTGACGTCATTGTCGCCTGTGTTGGGGGCGGGTCAAACGCTATGGGGATCTTTTACCC  GTTTGTAAATGATAAAAAAGTCAAACCTCGTTGGCGTGGAAGCTGGCGGCAAGGGGCTTG  AGAGTGGCAAACATTCCGCAAGCTTAAATGCTGGGCAGGTTGGGGTATTTTCATGGCATG  TTATCTTACTTCTTACAGGACGAGGAGGGGCAAATCAAGCCAACACACTCGATTGCGCCG  GGGCTGGATTACCCTGGTGTGCGGGCCGGAACACGCTTACCTGAAAAAATCCAACGTGC  GGAGTACGTGACTGTCACAGATGAGGAAGCGCTCAAAGCCTTCCACGAATTATCACGTA  CCGAAGGGATCATCCCGGCTCTGGAGTCAGCACATGCCGTGCGGTACGCTATGAAGTTG  GCGAAAGAGATGTCACGTGATGAAATCATTATTGTAACTTGAGTGGTTCGGGGCGATAA  AGACCTCGACATCGTCCTGAAAGTCAGCGGTAATGTG</p>
p063 His-pIVEX vector used in this work	<p>CTGGTGCCGCGTGGCAGCTTACATCCCGGGGGGGTCTCATCATCATCATCATTAA  TAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGGCTGCTAACAAA  GCCCCAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT  TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGTGAAAGGAGGAACTATATCCGGAT  ATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGA  AGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTGCGACAGTGTCCGAGAACGGGTGC  GCATAGAAATTGCATCAACGCATATAGCGTAGCAGCACGCCATAGTACTGGCGATGC  TGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATG  CCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTTCAT  ACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATC  GATGATAAGCTGTCAAACATGAGAATTCGTAATCATGTCATAGCTGTTTCTGTGTGAAA  TTGTTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTG  GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA  GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGC  GGTTTGCGTATTGGGCGCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGTTT  GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA  GGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGT  AAAAAGGCCGCTTGTGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAA  AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT  TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTG  TCCGCCTTTCTCCCTCGGGAAGCGTGGCGTTTTCTCATAGCTCACGCTGTAGGTATCTCA  GTTCCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCCC</p>

ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT  
CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC  
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTAT  
CTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA  
ACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAA  
AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACG  
AAAACCTCACGTTAAGGGATTTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC  
TTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA  
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCC  
ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC  
CCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA  
AACCAGCCAGCCGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCCTCCAT  
CCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCG  
CAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTC  
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAA  
AGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTATC  
ACTCATGGTTATGGCAGCACTGCATAATTCTTACTGTCATGCCATCCGTAAGATGCTTT  
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG  
TTGCTCTTGCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT  
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCCTGTTGAG  
ATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTTACTTTACCA  
GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC  
GACACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTGAAGCATTATCAG  
GGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAATAAGGG  
GTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATG  
ACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCCTTCGGTGAT  
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGC  
GGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGCG  
GGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTAAGTGTGAGAGTGCACCATATATGCG  
GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCAT  
TCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAG  
CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCC  
AGTCACGACGTTGTAACACGACGGCCAGTGCCAAGCTTGCATGCAAGGAGATGGCGCCC  
AACAGTCCCCCGCCACGGGGCCTGCCACCATAACCCACGCCGAAACAAGCGCTCATGAG  
CCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGGCGATATAGGCGCCAGCAA  
CCGACCTGTGGCGCCGGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGAT  
CTCGATCCCGCGAAATTAACGACTCACTATAGGGAGACCACAACGGTTTTCCCTCTAGA  
AATAATTTGTTTAACTTTAAGAAGGAGATATACC

p070 plasmid provided by Rémi Sieskind, containing sfGFP	CTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGGTTTTTGTGAAAGGAGG AACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTG GCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGCCAAAGCGGTCCGACAGTGCTC CGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGTAGCAGCACGCCATAG TGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCAT AACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCAT TGTTAGATTTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGC ATTAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTCGTAATCATGTCATAGCTGT TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAA AGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC TGCCCCGTTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACCG CGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGC GCTCGGTGTTTCGGTGCAGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTA TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAG CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACG AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG ATACCAGGCGTTTCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTT ACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCT GTAGGTATCTCAGTTCGGTGTAGGTGTTGCTCCAAGCTGGGCTGTGTGCACGAACCC CCGTTCCAGCCCGACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAA GACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGAC AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC TTGATCCGGCAAAACAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGAT TACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC TCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTT CACCTAGATCCTTTTAAATTAATAAAGTAAATCAATCTAAAGTATATATGAGTAA ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT TTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTA TCCGCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCCGCAAGTT AATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTT GGTATGGCTTCATTCAGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATG TTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAAGTTGGCC GCAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC GGCGACCGAGTTGCTCTTGGCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGA ACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTA CCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTCAGCATCTT TTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAAAG GGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTGA AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAAT AAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGACGTCTAAGAAAC CATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGG CGTTTTCGGTGTGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTACAGC TTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGGTCAGCGGGTGT GGCGGGTGTGCGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGACTGAGAGTGC ACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGC
--	--

GCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG  
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC  
CAGGGTTTTCCAGTCACGACGTTGTA AACGACGGCCAGTGCCAAGCTTGCATGCAAG  
GAGATGGCGCCCAACAGTCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACA  
AGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATAT  
AGGCGCCAGCAACCGCACCTGTGGCGCCGGTGTGCGGCCACGATGCGTCCGGCGTA  
GAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACG  
GTTTCCCTCTAGAAATAATTTTGT TTA ACTTTAAGAAGGAGATATAACCATGGAGCTTTTCA  
CTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT CAGC  
GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTG  
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCTCGTGACCACCTGACCTACGGCGT  
GCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT  
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA  
CCC GCGCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG  
CATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA ACTACAACA  
GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG  
ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC  
CCCCATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGC  
CCTGAGCAAAGACCCCAACGAGAAGCGCGATCATGGTCTGCTGGAGTTCTGTGACCG  
CCGCCGGGATCGTACAGACCGGTGAATTCACCATGGACAAAGACTGCGAAATGAAGCGC  
ACCACCCTGGATAGCCCTCTGGGCAAGCTGGAAGTGTCTGGGTGCGAACAGGGCCTGCA  
CCGTATCATCTTCTGGGCAAAGGAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCC  
AGCCGCCGTGCTGGGCGGACCAGAGCCACTGATGCAGGCCACCGCTGGCTCAACGCCT  
ACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCTGTGCCAGCCCTGCACCACCCAGTGT  
TCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAACTGCTGAAAGTGGTGAAGTTC  
GGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACCGC  
CGCCGTGAAAACCGCCCTGAGCGGAAATCCCCTGCCATTCTGATCCCCTGCCACCGGGT  
GGTGCAGGGCGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTG  
GCTGCTGGCCACGAGGGCCACAGACTGGGCAAGCCTGGGCTGGGTCTGCAGGCGGA  
TCCGCGTTTTAACTCGAGGTTAATTAATCCCGGGGGGGTTCTCATCATCATCATCAT  
TAATAAAAGGGCGAATTCAGCACACTGGCGGCCGTTACTAGTGATCCGGCTGCTAAC  
AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA

## Appendix A Structure inference-based methods for rational design of proteins.

Here we present an overview of the structure inference-based methods for rational design of proteins mentioned in 0.3.1.2. All of the approaches are frequently reviewed so we describe each of them only in brief<sup>242</sup>.

*Ab initio* methods could be considered to be the most versatile as these can be deployed based on sequence only. They rely on thermodynamics that stipulate that the native structure of a protein is the result of arriving at a global free energy minimum as proposed by Anfinsen<sup>243,244</sup>. The global free energy is calculated based on the amino acid sequence, giving rise to properties that allow to evaluate the fold through potential energy functions. The folding problem is thus a global optimisation problem where the minimum energy of the protein conformation is sought through identifying appropriate values of such variables as torsion angles and atom positions. Although existing structures are not directly used, they are employed for empirical determination of parameters to be used in potential energy functions. Despite the benefit of independence from PDB, it is apparent as outlined by Levingthal that large protein structures lead to an explosion of degrees of freedom, thus making exploration difficult<sup>245,246</sup>. An appropriate geometric representation thus has to be taken to avoid all-atom simulation yet preserve essential characteristics. The potential function also has to be chosen. It can be either derived from PDB-based training set, capturing all observed behaviours, even those not fully understood. Alternatively, it can be derived from physically-based parameters determined from small molecules or quantum mechanics calculations. The latter has the advantage of predicting folds that were not observed yet. Finally, various surface search techniques are available to choose from. We would remark that *Ab initio* simulations also come in hand in molecular docking problems: where structures of two molecules are known, most likely binding orientation can be predicted<sup>247,248</sup>.

Fragment-based methods rely on the observation that even when a new fold is discovered, it usually consists of motifs – short amino acid fragments – already known. Thus, if short amino acids sub sequences of the novel protein align with already known ones, structural fragments can be hypothesised and combined to predict the overall structure. As before, the overall structure is then assembled based on scoring functions and optimisation algorithms, for example, via simulated annealing<sup>249,250</sup>. We would note the key assumption here is that a local sequence gives rise to the local structure. Even so, in nature, similar sequences frequently give rise to different structures due to complex environment. Scoring functions are frequently used to determine fragments that are more likely to be found together, usually based on known proteins<sup>251</sup>. Overall, the process somewhat resembles the *Ab initio* methods in being able to predict new folds, yet it reduces the conformational search space as the fragments already achieved their minimum potential energy. Ironically, though, the combinatorial nature of the method gives rise to another dimension of the conformation search space. Evaluating regions with several possible combinations of fragments is also non-trivial.

Fold recognition and threading methods rely on the fact that the structure is usually more conserved than the sequence. Hence, a novel sequence can be projected onto known folds in hope of fitting. This is especially attractive, given that the overall number of folds is thought to be limited<sup>252–254</sup>. The process involves selecting appropriate templates and then

projecting the target sequence onto them. In threading methods, pioneered by Jones et al.<sup>255</sup>, three-dimensional templates of similar fold are identified. These can have evolutionary relationship (homologues) or not (analogues)<sup>256</sup>. Therefore, compared to *Ab initio* methods, quality of fit dependent on pair potentials is optimised instead of global minimal energy of the fold. In fold recognition methods, profile-based approach is usually used. Information about the templates and the target is represented in a linear form, capturing amino acid properties. These can be as simple, as hydrophobicity or as detailed as evolutionary conservation of given residues. For both fold recognition and threading follows the process of constructing a template library based on PDB. Frequently, simplified structures are used. A scoring function is then defined to evaluate fits, followed by actual search within the possible fit space.

Meta servers also exist to pool together predictions from different methods. These allow to benefit from strengths of several models, especially, when coupled with model-evaluation programs. We would name 3D-Jury<sup>257</sup> as one of the most popular meta-servers with reviews of others being available<sup>258</sup>. Similarly, an array of model-evaluation programs exists<sup>259</sup>, with QMEAN being of the latest ones<sup>260</sup>.

Finally, homology-based modelling or comparative modelling uses close alignment between the target sequence and that of a template protein with a known structure. This is usually possible due to evolutionary relationships between the proteins with both sequences and structures being conserved. Frequently, even if an amino acid is replaced, the residue still possesses similar properties and does not strongly alter the fold. Atomic-resolution accurate models are therefore possible: an existing closely matching structure is identified and only select divergent residues then have to be modelled. Either pairwise or multiple sequence alignment are possible to identify the template with FASTA and BLAST being well-known examples of the former. Multiple sequence alignment methods were mentioned above. Where possible, comparative modelling is used due to its accuracy and possibility to easily evaluate reliability of the structure. The obvious limitation, however, is that prediction of new folds is impossible and reasonable homology is necessary for the method to work.

Apart from the TrpB example, real-life applications of homology-based modelling include work Choi et al. and Qi et al.<sup>261,262</sup>. In the former case, mutagenesis of epoxide hydrolase (EH) from a fish *Mugil cephalus* was guided based on the structure templated on EH of *Aspergillus niger*, a fungus, sometimes known as the “black mold”. The initial reaction rate was boosted 35-fold. Qi and colleagues worked on glycerol dehydratase (GDH) from *Klebsiella pneumoniae* and used modelling to predict the possible effect of mutations, using PopMuSiC package. They achieved around 2 times increase in pH stability as well as a boost in activity.

The variety of approaches for structure prediction can be overwhelming. Critical Assessment of protein Structure Prediction (CASP) has thus been proposed by John Moult to evaluate the prediction methods in a variety of disciplines and according to a number of criteria once every two years. Latest results can therefore be useful to select a method of choice<sup>263</sup>. For example, in CASP13 (2018), most recent one at the time of writing AlphaFold system competed remarkable well in the ‘template free modelling’ (FM) category. That is, it predicted well previously unseen folds. Interestingly, the training of the neural networks of this system relied heavily on examining coevolutionary patterns.

## Appendix B L-trp synthesis detection through absorbance readings.

As briefly mentioned in 2.3.2, we struggled to obtain an absorbance reading that would convince us of L-trp synthesis.

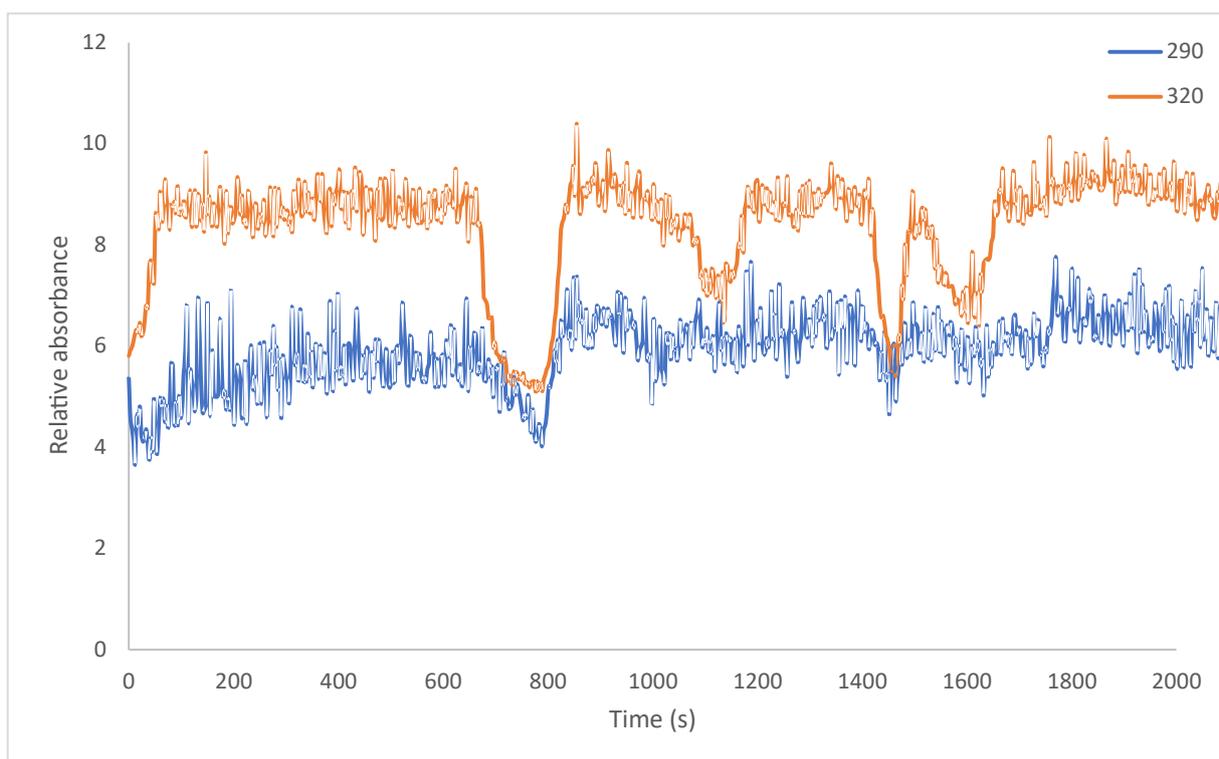
We took absorbance spectra of first rough test runs (not shown). Although taken at fixed timepoints between which sample evaporation was possible, it appeared absorbance beyond 310 nm was evolving.

We therefore decided to trace absorbance at 290 nm and 320 nm over time. For example, as in **Figure 157**. 18  $\mu\text{M}$  of TrpB was mixed with 3 mM indole, 3 mM L-serine, 40  $\mu\text{M}$  PLP in 50 mM phosphate buffer pH 8 and incubated at 37  $^{\circ}\text{C}$  within the absorbance plate reader. The assays were always performed in uncovered Greiner 96 well plate, black, half area, standard clear bottom in 30  $\mu\text{L}$  of volume.

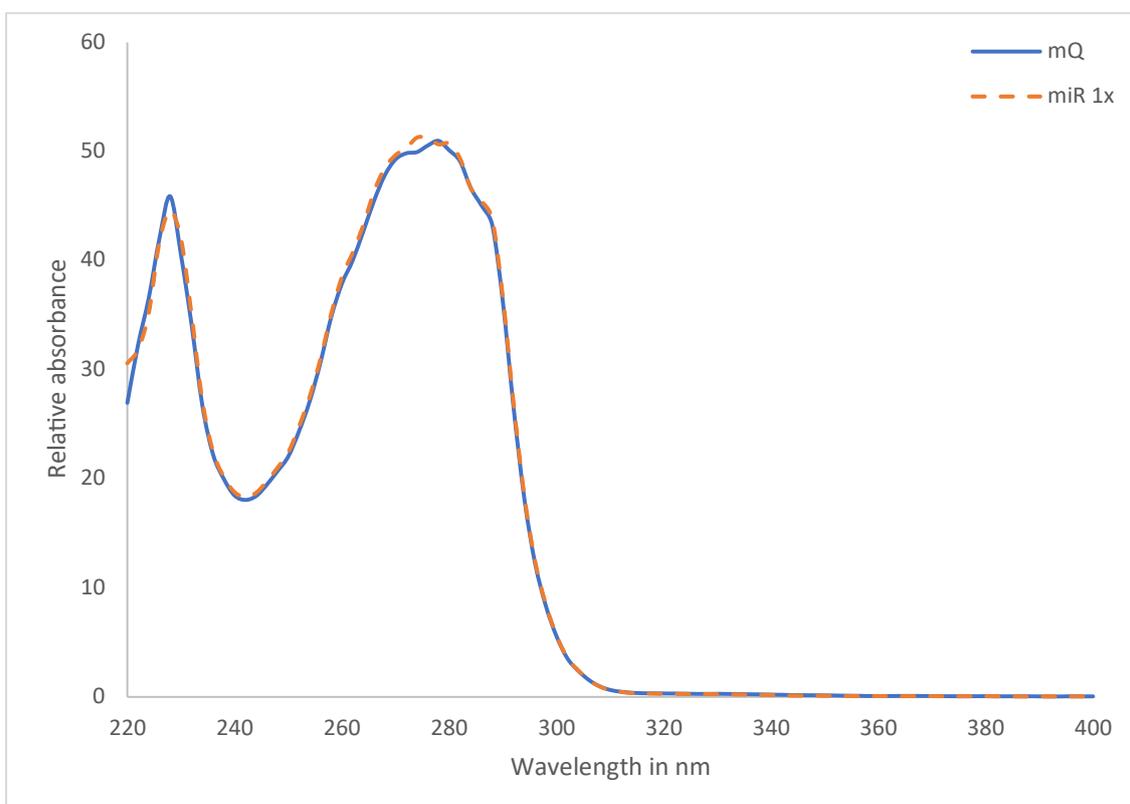
We see relatively flat traces in both channels that sometimes show correlated fluctuations. We tried addressing the problem in several ways. We varied buffers (miR, Tris, phosphate as reported by Arnold) in concern they could be unexpectedly affecting the enzyme or provide strong background absorbance. However, at the end we quickly checked our stock of L-trp for abnormal spectra or spectral shift in miR buffer. We observed no difference between the spectrum in pure water and 1x miR buffer (**Figure 158**), in turn both matching the one reported by the manufacturer of the machine used (**Figure 67**).

We also varied enzyme concentrations (from tens of  $\mu\text{M}$  to pM) and time windows of observations up to multiple hours (not shown) out of concern that we could be missing the reaction was happening too fast or too slow to observe.

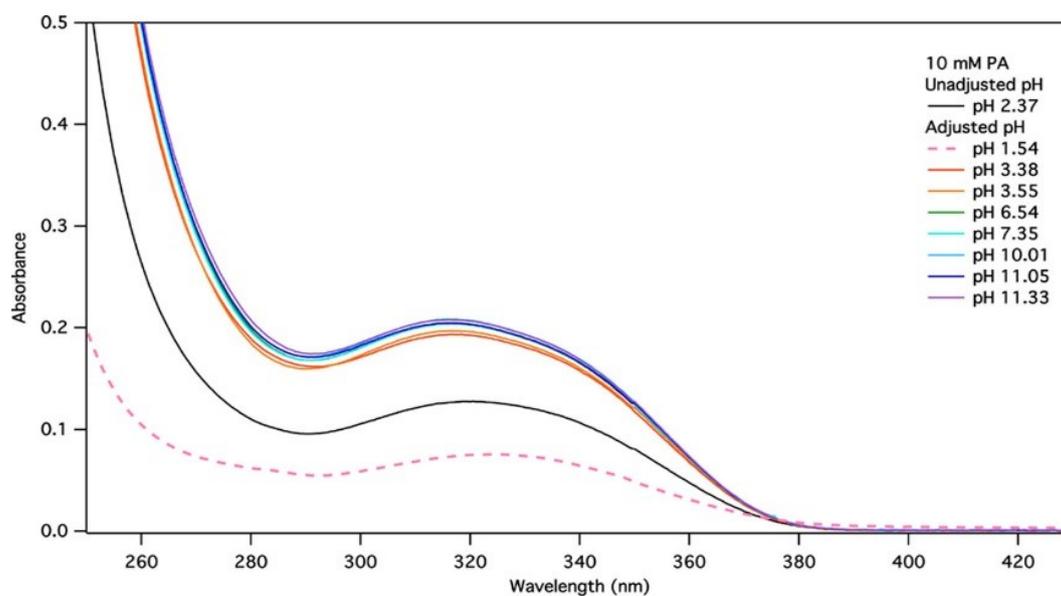
It is uncertain what the cause could be of 290 and 320 nm absorbance correlation. We speculate, that in part it could be due to some mechanical motion of the liquid or the machine as perturbations can be observed at seemingly plateau sections of the curves, which are similar to both traces. It is possible for this sensitivity to arise due to minimal reaction volume to cover the bottom of the plate well we were using – 30  $\mu\text{L}$  was barely enough. This was driven in part by frugality regarding the enzyme use. However, preliminary tests (not shown) also showed that at mM concentrations and longer path lengths indole and L-tryptophan saturated the absorbance readings of the machine. Thus, some increase in absorbance could be hidden within the data. However, we are more inclined to believe that at least in part L-serine was consumed to produce pyruvate, which has been reported as a side product<sup>264,265</sup> of TrpB and fittingly does seem to absorb at 320 nm strongly<sup>266</sup> (**Figure 159**).



**Figure 157** Characteristic absorbance readings at 290 nm (blue) and 320 nm (orange) during L-trp synthesis reactions. Blank correction applied using an identical sample lacking TrpB. Reactions ran at 37 degrees in a total volume of 30  $\mu$ L, here in phosphate buffer with 0.1% NEB stock of BSA as well as 3mM of L-serine, Indole, 40  $\mu$ M of PLP and 18  $\mu$ M of EcB.



**Figure 158** Absorbance spectra of L-tryptophan used in this work. In water (blue) and in 1x miR buffer (orange dashed).



**Figure 159 UV-vis absorption spectra of pyruvic acid over a range of pH.** 10mM pyruvic acid solutions in pH 1.54 through 11.33. As reported by Rebecca Rapf and colleagues<sup>66</sup>.

## Bibliography

1. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond)*. **27**, 113–159 (1928).
2. AVERY, O., MACLEOD, C. & McCARTY, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES: INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J. Exp. Med.* **79**, 137–158 (1944).
3. Hershey, A. D. & Chase, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.* **36**, 39–56 (1952).
4. Thaplyal, P. & Bevilacqua, P. C. Experimental Approaches for Measuring pKa's in RNA and DNA. in *Physiology & behavior* **176**, 189–219 (2014).
5. Jones, S., Shanahan, H. P., Berman, H. M. & Thornton, J. M. Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res.* **31**, 7189–7198 (2003).
6. Duguid, J. G. & Bloomfield, V. A. Electrostatic effects on the stability of condensed DNA in the presence of divalent cations. *Biophys. J.* **70**, 2838–2846 (1996).
7. Cuculis, L. *et al.* Divalent cations promote TALE DNA-binding specificity. *Nucleic Acids Res.* **48**, 1406–1422 (2020).
8. Levene, P. The structure of yeast nucleic acid. *J. Biol. Chem.* **40**, 415–424 (1917).
9. Johnson, J. L. 1 Determination of DNA Base Composition. *Methods Microbiol.* **18**, 1–31 (1985).
10. Elson, D. & Chargaff, E. On the desoxyribonucleic acid content of sea urchin gametes. *Experientia* **8**, 143–145 (1952).
11. CHARGAFF, E., LIPSHITZ, R. & GREEN, C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* **195**, 155–60 (1952).
12. Rudner, R., Karkas, J. D. & Chargaff, E. Separation of *B. subtilis* DNA into complementary strands. 3. Direct analysis. *Proc. Natl. Acad. Sci.* **60**, 921–922 (1968).
13. Meselson, M. & Stahl, F. W. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **44**, 671–682 (1958).
14. Pray, L. A. Discovery of DNA Structure and Function: Watson and Crick. *Nature Education* **1**, 100 (2008).
15. Mikkola, S. *et al.* The mechanism of the metal ion promoted cleavage of RNA phosphodiester bonds involves a general acid catalysis by the metal aquo ion on the departure of the leaving group. *J. Chem. Soc. Perkin Trans. 2* 1619–1625 (1999). doi:10.1039/a903691a
16. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7287–7292 (2004).
17. Maderspacher, F. Theodor Boveri and the natural experiment. *Curr. Biol.* **18**, R279–R286 (2008).
18. Bordenstein, S. R. Genomic and cellular complexity from symbiotic simplicity. *Cell* **158**, 1236–1237 (2014).
19. Han, K. *et al.* Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.* **3**, 3–9 (2013).
20. Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**,

- 365–368 (2004).
21. Birky, C. W. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. *Annu. Rev. Genet.* **35**, 125–148 (2001).
  22. Lederberg, J. Cell Genetics and Hereditary Symbiosis. *Physiol. Rev.* **32**, 403–430 (1952).
  23. Plasmid. Available at: [https://commons.wikimedia.org/wiki/File:Plasmid\\_\(numbers\).svg](https://commons.wikimedia.org/wiki/File:Plasmid_(numbers).svg). (Accessed: 10th November 2020)
  24. LaMorte, W. W. Chromosomes Contain Our Genetic Code. Available at: <https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH/DNA-Genetics/DNA-Genetics2.html>. (Accessed: 10th November 2020)
  25. Gahl, W. Mitochondrial DNA. *National Human Genome Research Institute* Available at: Talking Glossary of Genetic Terms. (Accessed: 10th November 2020)
  26. Fragkos, M., Ganier, O., Coulombe, P. & Méchali, M. DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015).
  27. Del Solar, G. & Espinosa, M. Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* **37**, 492–500 (2002).
  28. Riera, A. *et al.* From structure to mechanism— understanding initiation of DNA replication. *Genes Dev.* **31**, 1073–1088 (2017).
  29. Wu, W. J., Yang, W. & Tsai, M. D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat. Rev. Chem.* **1**, 1–16 (2017).
  30. Zenkin, N., Naryshkina, T., Kuznedelov, K. & Severinov, K. The mechanism of DNA replication primer synthesis by RNA polymerase. *Nature* **439**, 617–620 (2006).
  31. Brody, L. C. *No Title*.
  32. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci.* **59**, 598–605 (1968).
  33. Ogawa, T. & Okazaki, T. Discontinuous DNA replication. *Annu. Rev. Biochem.* **49**, 421–457 (1980).
  34. Masur. Depiction of DNA replication with replication fork, strands and okazaki-fragments. Available at: [https://en.wikipedia.org/wiki/File:Replication\\_fork.svg](https://en.wikipedia.org/wiki/File:Replication_fork.svg). (Accessed: 10th November 2020)
  35. Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. & Khorana, H. G. Studies on polynucleotides. XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* **56**, 341–361 (1971).
  36. Mullis, K. B. The unusual origin of the polymerase chain reaction. *Sci. Am.* **262**, 56–65 (1990).
  37. Enzoklop. No Title. Available at: [https://commons.wikimedia.org/wiki/File:Polymerase\\_chain\\_reaction.svg](https://commons.wikimedia.org/wiki/File:Polymerase_chain_reaction.svg).
  38. Bernstein, H. & Bernstein, C. Circular and branched circular concatenates as possible intermediates in bacteriophage T4 DNA replication. *J. Mol. Biol.* **77**, 355–361 (1973).
  39. Ali, M. M. *et al.* Rolling circle amplification: A versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.* **43**, 3324–3341 (2014).
  40. Gane, A. The future of DNA amplification. (2020). Available at: <https://www.cytivalifesciences.com/en/us/news-center/the-future-of-dna-amplification-10001>. (Accessed: 23rd November 2020)

41. Isothermal Amplification, Application overview. *NEB website* Available at: <https://international.neb.com/applications/dna-amplification-pcr-and-qpcr/isothermal-amplification>. (Accessed: 22nd November 2020)
42. Van Ness, J., Van Ness, L. K. & Galas, D. J. Isothermal reactions for the amplification of oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4504–4509 (2003).
43. Walker, G. T., Little, M. C., Nadeau, J. G. & Shank, D. D. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 392–396 (1992).
44. Qian, C., Wang, R., Wu, H., Ji, F. & Wu, J. Nicking enzyme-assisted amplification (NEAA) technology and its applications: A review. *Anal. Chim. Acta* **1050**, 1–15 (2019).
45. Zhou, Y., Zhang, X. & Ebright, R. H. Random mutagenesis of gene-sized DNA molecules by use of pcr with taq DNA polymerase. *Nucleic Acids Res.* **19**, 6052 (1991).
46. PALERO, F. & CRANDALL, K. Phylogenetic Inference Using Molecular Data. 67–88 (2009). doi:10.1201/9781420092592-c5
47. Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P. & Tyson, G. W. Shining a Light on Dark Sequencing: Characterising Errors in Ion Torrent PGM Data. *PLoS Comput. Biol.* **9**, (2013).
48. CRICK, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–63 (1958).
49. Nabel, C. S., Manning, S. A. & Kohli, R. M. The Curious Chemical Biology of Cytosine: Deamination, Methylation, and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol.* **7**, 20–30 (2012).
50. Pribnow, D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 784–788 (1975).
51. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202 (2015).
52. BRENNER, S., JACOB, F. & MESELSON, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **190**, 576–581 (1961).
53. Das, J. K., Majumder, A., Choudhury, P. P. & Mukhopadhyay, B. Understanding of Genetic Code Degeneracy and New Way of Classifying of Protein Family: A Mathematical Approach. *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016* 262–267 (2016). doi:10.1109/IACC.2016.57
54. CRICK, F. H. C., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. General Nature of the Genetic Code for Proteins. *Nature* **192**, 1227–1232 (1961).
55. BioNinja. (2020). Available at: <http://www.vce.bioninja.com.au/aos-3-heredity/molecular-genetics/transcription.html>. (Accessed: 9th November 2020)
56. Wong, E. V. *Cells - Molecules and Mechanisms*. (Axolotl Academica Publishing, 2020).
57. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34–38 (1975).
58. HOAGLAND, M. B., STEPHENSON, M. L., SCOTT, J. F., HECHT, L. I. & ZAMECNIK, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231**, 241–57 (1958).
59. Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* **19**, 526–541 (2018).
60. Lobanov, A. V., Turanov, A. A., Hatfield, D. L. & Gladyshev, V. N. Dual functions of

- codons in the genetic code. *Crit. Rev. Biochem. Mol. Biol.* **45**, 257–265 (2010).
61. Szavits-Nossan, J., Ciandrini, L. & Romano, M. C. Deciphering mRNA Sequence Determinants of Protein Production Rate. *Phys. Rev. Lett.* **120**, (2018).
  62. Vickers, T. No Title. Available at: [https://en.wikipedia.org/wiki/Amino\\_acid#/media/File:Amino\\_acid\\_zwitterions.svg](https://en.wikipedia.org/wiki/Amino_acid#/media/File:Amino_acid_zwitterions.svg). (Accessed: 9th November 2020)
  63. A Brief Guide to the Twenty Common Amino Acids. (2014). Available at: <https://www.compoundchem.com/2014/09/16/aminoacids/>. (Accessed: 9th November 2020)
  64. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
  65. Reusch, Wi. Peptides & Proteins. *Online* (2013). Available at: <https://www2.chemistry.msu.edu/faculty/reusch/virtTxtJml/protein2.htm>. (Accessed: 8th December 2020)
  66. Voet, Donald; Voet, J. G. *Biochemistry*. (John Wiley & Sons, Inc., 2010).
  67. Cooper, J. The Ramachandran Plot. (1995). Available at: [http://www.cryst.bbk.ac.uk/PPS95/course/3\\_geometry/rama.html](http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html). (Accessed: 10th November 2020)
  68. Alberts, B; Johnson, A; Lewis, J. et al. *Molecular biology of the cell*. (Garland Science, 2002., 2002).
  69. OpenStax. Microbiology. *LibreTexts* (2020). Available at: [https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A\\_Microbiology\\_\(OpenStax\)/07%3A\\_Microbial\\_Biochemistry/7.04%3A\\_Proteins](https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(OpenStax)/07%3A_Microbial_Biochemistry/7.04%3A_Proteins). (Accessed: 11th November 2020)
  70. Wheeler, R. 1GZX Haemoglobin. (2007). Available at: [https://commons.wikimedia.org/wiki/File:1GZX\\_Haemoglobin.png](https://commons.wikimedia.org/wiki/File:1GZX_Haemoglobin.png). (Accessed: 11th November 2020)
  71. Zephyris. Carboxypeptidase catalysis. (2006). Available at: [https://en.wikipedia.org/wiki/File:Carboxypeptidase\\_catalysis.png](https://en.wikipedia.org/wiki/File:Carboxypeptidase_catalysis.png). (Accessed: 13th November 2020)
  72. Zephyris. Serine protease catalysis. Available at: [https://en.wikipedia.org/wiki/File:Serine\\_protease\\_catalysis.png](https://en.wikipedia.org/wiki/File:Serine_protease_catalysis.png). (Accessed: 13th November 2020)
  73. Watkins-Dulaney, E., Straathof, S. & Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* (2020). doi:10.1002/cbic.202000379
  74. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**, 1–5 (2011).
  75. FISCHER, E. H. & KREBS, E. G. Conversion of phosphorylase b to phosphorylase a in muscle extracts. *J. Biol. Chem.* **216**, 121–32 (1955).
  76. Dalbey, R. E. & Kuhn, A. Protein Traffic in Gram-negative bacteria - how exported and secreted proteins find their way. *FEMS Microbiol. Rev.* **36**, 1023–1045 (2012).
  77. Mougous, J. D., Gifford, C. A., Ramsdell, T. L. & Mekalanos, J. J. Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nat. Cell Biol.* **9**, 797–803 (2007).

78. Silverman, J. M. *et al.* Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Mol. Cell* **51**, 584–593 (2013).
79. Russell, A. B. *et al.* Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343–349 (2011).
80. Schrader, E. K., Harstad, K. G. & Matouschek, A. Targeting proteins for degradation. *Nat. Chem. Biol.* **5**, 815–822 (2009).
81. Luo, S., McNeill, M., Myers, T. G., Hohman, R. J. & Levine, R. L. Lon protease promotes survival of Escherichia coli during anaerobic glucose starvation. *Arch. Microbiol.* **189**, 181–185 (2008).
82. Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42 (2011).
83. Mitarai, N. & Pedersen, S. Control of ribosome traffic by position-dependent choice of synonymous codons. *Phys. Biol.* **10**, (2013).
84. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).
85. Geyer, C. R., Colman-Lerner, A. & Brent, R. ‘Mutagenesis’ by peptide aptamers identifies genetic network. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8567 (1999).
86. Ruscito, A. & DeRosa, M. C. Small-molecule binding aptamers: Selection strategies, characterization, and applications. *Front. Chem.* **4**, 1–14 (2016).
87. Yang, S. *et al.* Oligonucleotide Aptamer-Mediated Precision Therapy of Hematological Malignancies. *Mol. Ther. - Nucleic Acids* **13**, 164–175 (2018).
88. Histidine. Trp operon attenuation. (2011). Available at: [https://commons.wikimedia.org/wiki/File:Trp\\_operon\\_attenuation.svg](https://commons.wikimedia.org/wiki/File:Trp_operon_attenuation.svg). (Accessed: 12th November 2020)
89. Narlikar, G. J., Fan, H. & Kingston, R. E. Cooperation between Complexes that Regulate Chromatin Structure and Transcription Chromatin structure creates barriers for each step. *Cell* **108**, 475–487 (2002).
90. Lan, X., Glass, C. K. & Rosenfeld, M. G. Coactivator and corepressor complexes in nuclear receptor function. *Curr. Opin. Genet. Dev.* **9**, 140–147 (1999).
91. Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.* **100**, 79–90 (2016).
92. Joachimiak, A., Kelley, R. L., Gunsalus, R. P., Yanofsky, C. & Sigler, P. B. Purification and characterization of trp aporepressor. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 668–672 (1983).
93. MONOD, J. & JACOB, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).
94. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
95. Jacob, F. & Monod, J. On the Regulation of Gene Activity. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 193–211 (2012).
96. Gilbert, W. & Muller-Hill, B. Isolation of the Lac Repressor. *Proc. Natl. Acad. Sci.* **56**, 1891–1898 (1966).
97. Lewis, M. A tale of two repressors- a historical perspective. *J. Mol. Biol.* **409**, 0–1 (2011).
98. Juers, D. H., Hakda, S., Matthews, B. W. & Huber, R. E. Structural Basis for the Altered Activity of Gly794 Variants of Escherichia coli  $\beta$ -Galactosidase. *Biochemistry* **42**,

- 13505–13511 (2003).
99. Quinn, J. Y. *et al.* SBOL Visual: A Graphical Language for Genetic Designs. *PLoS Biol.* **13**, 1–9 (2015).
  100. Barkley, M. D., Jobe, A., Bourgeois, S. & Riggs, A. D. Interaction of Effecting Ligands with Lac Repressor and Repressor-Operator Complex. *Biochemistry* **14**, 1700–1712 (1975).
  101. Swigon, D., Coleman, B. D. & Olson, W. K. Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9879–9884 (2006).
  102. Jeeves, M., Evans, P. D., Parslow, R. A., Jaseja, M. & Hyde, E. I. Studies of the Escherichia coli Trp repressor binding to its five operators and to variant operator sequences. *Eur. J. Biochem.* **265**, 919–928 (1999).
  103. Xie, Y. & Reeve, J. N. Regulation of tryptophan operon expression in the archaeon Methanothermobacter thermoautotrophicus. *J. Bacteriol.* **187**, 6419–6429 (2005).
  104. Karr, E. A., Sandman, K., Lurz, R. & Reeve, J. N. TrpY regulation of trpB2 transcription in Methanothermobacter thermoautotrophicus. *J. Bacteriol.* **190**, 2637–2641 (2008).
  105. Cafasso, J. *et al.* Preliminary crystallography confirms that the archaeal DNA-binding and tryptophan-sensing regulator TrpY is a dimer. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1493–1495 (2010).
  106. Čuboňová, L., Sandman, K., Karr, E. A., Cochran, A. J. & Reeve, J. N. Spontaneous trpY mutants and mutational analysis of the TrpY archaeal transcription regulator. *J. Bacteriol.* **189**, 4338–4342 (2007).
  107. Poluri, K. M. & Gulati, K. *Protein engineering techniques: Gateways to synthetic protein universe. SpringerBriefs in Applied Sciences and Technology* (2017). doi:10.1007/978-981-10-2732-1
  108. Boville, C. E. *et al.* Engineered Biosynthesis of  $\beta$ -Alkyl Tryptophan Analogues. *Angew. Chemie - Int. Ed.* **57**, 14764–14768 (2018).
  109. Steipe, B., Schiller, B., Pluäckthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *Journal of Molecular Biology* **240**, 188–192 (1994).
  110. Lehmann, M. *et al.* The consensus concept for thermostability engineering of proteins: Further proof of concept. *Protein Eng.* **15**, 403–411 (2002).
  111. Watanabe, K., Ohkuri, T., Yokobori, S. I. & Yamagishi, A. Designing thermostable proteins: Ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J. Mol. Biol.* **355**, 664–674 (2006).
  112. Sullivan, B. J., Durani, V. & Magliery, T. J. Triosephosphate isomerase by consensus design: Dramatic differences in physical properties and activity of related variants. *J. Mol. Biol.* **413**, 195–208 (2011).
  113. Ehren, J., Govindarajan, S., Morón, B., Minshull, J. & Khosla, C. Protein engineering of improved prolyl endopeptidases for celiac sprue therapy. *Protein Eng. Des. Sel.* **21**, 699–707 (2008).
  114. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **6**, 807–813 (2010).
  115. Thompson, J. N. *The coevolutionary process.* (University of Chicago Press, 1994).

116. Horner, D. S., Pirovano, W. & Pesole, G. Correlated substitution analysis and the prediction of amino acid structural contacts. *Brief. Bioinform.* **9**, 46–56 (2008).
117. Fodor, A. A. & Aldrich, R. W. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins Struct. Funct. Genet.* **56**, 211–221 (2004).
118. Yip, K. Y. *et al.* An integrated system for studying residue coevolution in proteins. *Bioinformatics* **24**, 290–292 (2008).
119. Socolich, M. *et al.* Evolutionary information for specifying a protein fold. *Nature* **437**, 512–518 (2005).
120. Chen, Z., Meyer, W., Rappert, S., Sun, J. & Zeng, A. P. Coevolutionary analysis enabled rational deregulation of allosteric enzyme inhibition in *Corynebacterium glutamicum* for lysine production. *Appl. Environ. Microbiol.* **77**, 4352–4360 (2011).
121. Tsai, C. J., Del Sol, A. & Nussinov, R. Protein allostery, signal transmission and dynamics: A classification scheme of allosteric mechanisms. *Mol. Biosyst.* **5**, 207–216 (2009).
122. Labrou, N. E. Random Mutagenesis Methods for In Vitro Directed Enzyme Evolution. *Curr. Protein Pept. Sci.* **999**, 1–12 (2009).
123. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **175**, 1946–1957.e13 (2018).
124. Q5® Site-Directed Mutagenesis Kit. *NEB website* Available at: [https://international.neb.com/products/e0554-q5-site-directed-mutagenesis-kit#Product Information](https://international.neb.com/products/e0554-q5-site-directed-mutagenesis-kit#Product%20Information). (Accessed: 22nd November 2020)
125. Ruff, A. J., Dennig, A. & Schwaneberg, U. To get what we aim for - Progress in diversity generation methods. *FEBS J.* **280**, 2961–2978 (2013).
126. Q5® Site-Directed Mutagenesis Kit. *NEB website*
127. George P., S. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-. )*. **228**, 1315–1317 (1985).
128. Daugherty, P. S., Iverson, B. L. & Georgiou, G. Flow cytometric screening of cell-based libraries. *J. Immunol. Methods* **243**, 211–227 (2000).
129. Delépine, B., Libis, V., Carbonell, P. & Faulon, J. L. SensiPath: computer-aided design of sensing-enabling metabolic pathways. *Nucleic Acids Res.* **44**, W226–W231 (2016).
130. Buller, A. R. *et al.* Directed evolution of the tryptophan synthase  $\beta$ -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
131. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R. & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
132. Herger, M. *et al.* Synthesis of  $\beta$ -Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **138**, 8388–8391 (2016).
133. Buller, A. R., Van Roye, P., Murciano-Calles, J. & Arnold, F. H. Tryptophan Synthase Uses an Atypical Mechanism To Achieve Substrate Specificity. *Biochemistry* **55**, 7043–7046 (2016).
134. Rix, G. *et al.* Scalable continuous evolution for the generation of diverse enzyme

- variants encompassing promiscuous activities. *Nat. Commun.* **11**, 5644 (2020).
135. Hagiya, M. From molecular computing to molecular programming. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **1716**, 89–102 (2001).
  136. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science (80-. )*. **314**, 1585–1588 (2006).
  137. Zhang, D. Y., Turberfield, A. J., Yurke, B. & Winfree, E. Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science (80-. )*. **318**, 1121–1125 (2007).
  138. Genot, A. J., Fujii, T. & Rondelez, Y. In vitro regulatory models for systems biology. *Biotechnol. Adv.* **31**, 789–796 (2013).
  139. Tyson, J. J. & Novák, B. Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61**, 219–240 (2010).
  140. Rondelez, Y. Competition for catalytic resources alters biological network dynamics. *Phys. Rev. Lett.* **108**, 1–5 (2012).
  141. Genot, A. J., Fujii, T. & Rondelez, Y. Computing with competition in biochemical networks. *Phys. Rev. Lett.* **109**, 1–5 (2012).
  142. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011).
  143. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).
  144. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
  145. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
  146. Kim, J., White, K. S. & Winfree, E. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**, (2006).
  147. Kim, J. & Winfree, E. Synthetic in vitro transcriptional oscillators. *Mol. Syst. Biol.* **7**, 1–15 (2011).
  148. Weitz, M. & Simmel, F. C. Synthetic in vitro transcription circuits. *Transcription* **3**, 87–91 (2012).
  149. Baccouche, A., Montagne, K., Padirac, A., Fujii, T. & Rondelez, Y. Dynamic DNA-toolbox reaction circuits: A walkthrough. *Methods* **67**, 234–249 (2014).
  150. Fujii, T. & Rondelez, Y. Predator - Prey molecular ecosystems. *ACS Nano* **7**, 27–34 (2013).
  151. Rondelez, Y. *et al.* Synthesis and materialization of a reaction–diffusion French flag pattern. *Nat. Chem.* **9**, 990–996 (2017).
  152. Rondelez, Y. & Gines, G. Multiplex Digital MicroRNA Detection Using Cross-Inhibitory DNA Circuits. *ACS sensors* **5**, 2430–2437 (2020).
  153. Ogata, N. & Miura, T. Creation of genetic information by DNA polymerase of the thermophilic bacterium *Thermus thermophilus*. *Nucleic Acids Res.* **26**, 4657–4661 (1998).
  154. Irvine, B. *et al.* Specific versus Nonspecific Isothermal DNA Amplification through Thermophilic Polymerase and Nicking Enzyme Activities † . *Biochemistry* **47**, 9987–9999 (2008).
  155. Gines, G. *et al.* Isothermal digital detection of microRNAs using background-free molecular circuit. *Sci. Adv.* **6**, 1–9 (2020).

156. Westbrook, A. M. & Lucks, J. B. Achieving large dynamic range control of gene expression with a compact RNA transcription-translation regulator. *bioRxiv* 055327 (2016). doi:10.1101/055327
157. Chen, Y., Kim, J. K., Hirning, A. J., Josić, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Science (80-. )*. **349**, 986–989 (2015).
158. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-. )*. **249**, 505–510 (1990).
159. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
160. Fabry-Wood, A. *et al.* Microcompartments for Protection and Isolation of Nanoscale DNA Computing Elements. *ACS Appl. Mater. Interfaces* **11**, 11262–11269 (2019).
161. Yang, K. *et al.* High-Affinity Nucleic-Acid-Based Receptors for Steroids. *ACS Chem. Biol.* **12**, 3103–3112 (2017).
162. Porchetta, A., Vallée-Bélisle, A., Plaxco, K. W. & Ricci, F. Using distal-site mutations and allosteric inhibition to tune, extend, and narrow the useful dynamic range of aptamer-based sensors. *J. Am. Chem. Soc.* **134**, 20601–20604 (2012).
163. Yang, X. *et al.* Characterization and application of a DNA aptamer binding to l-tryptophan. *Analyst* **136**, 577–585 (2011).
164. Song, S., Wang, L., Li, J., Fan, C. & Zhao, J. Aptamer-based biosensors. *TrAC - Trends Anal. Chem.* **27**, 108–117 (2008).
165. McConnell, E. M., Nguyen, J. & Li, Y. Aptamer-Based Biosensors for Environmental Monitoring. *Front. Chem.* **8**, 1–24 (2020).
166. Li, J. W., Zhang, X. Y., Wu, H. & Bai, Y. P. Transcription Factor Engineering for High-Throughput Strain Evolution and Organic Acid Bioproduction: A Review. *Front. Biotechnol.* **8**, 1–10 (2020).
167. Zhao, G. *et al.* A novel strategy to analyze l-tryptophan through allosteric Trp repressor based on rolling circle amplification. *Biosens. Bioelectron.* **71**, 103–107 (2015).
168. Yao, Y. *et al.* Development of small molecule biosensors by coupling the recognition of the bacterial allosteric transcription factor with isothermal strand displacement amplification. *Chem. Commun.* **54**, 4774–4777 (2018).
169. Klig, L. S., Crawford, I. P. & Yanofsky, C. *Nucleic Acids Research.* **15**, 5339–5351 (1987).
170. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3581–3584 (1973).
171. Platt, T., Files, J. G. & Weber, K. Lac repressor. *J. Biol. Chem.* **248**, 110–121 (1973).
172. Romanuka, J. *et al.* Specificity and Affinity of Lac Repressor for the Auxiliary Operators O2 and O3 Are Explained by the Structures of Their Protein – DNA Complexes. *J. Mol. Biol.* **390**, 478–489 (2009).
173. Falcon, C. M. & Matthews, K. S. Glycine insertion in the hinge region of lactose repressor protein alters DNA binding. *J. Biol. Chem.* **274**, 30849–30857 (1999).
174. Montagne, K., Gines, G., Fujii, T. & Rondelez, Y. Boosting functionality of synthetic DNA circuits with tailored deactivation. *Nat. Commun.* **7**, 1–12 (2016).
175. Zhang, R. -g. *et al.* The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature* **327**, 591–597 (1987).
176. Arvidson, D. N., Bruce, C. & Gunsalus, R. P. Interaction of the Escherichia coli trp

- aporepressor with its ligand, L-tryptophan. *J. Biol. Chem.* **261**, 238–243 (1986).
177. Hurlburt, B. K. & Yanofsky, C. trp Repressor/trp operator interaction. Equilibrium and kinetic analysis of complex formation and stability. *J. Biol. Chem.* **267**, 16783–16789 (1992).
  178. Montagne, K., Plasson, R., Sakai, Y., Fujii, T. & Rondelez, Y. Programming an in vitro DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 1–7 (2011).
  179. Baccouche, A., Montagne, K., Padirac, A., Fujii, T. & Rondelez, Y. Dynamic DNA-toolbox reaction circuits: A walkthrough. *Methods* **67**, 234–249 (2014).
  180. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5**, 1–7 (2009).
  181. Lu, M. S., Mauser, J. F. & Prehoda, K. E. Ultrasensitive synthetic protein regulatory networks using mixed decoys. *ACS Synth. Biol.* **1**, 65–72 (2012).
  182. Sieskind, R. In vitro enzyme directed evolution using DNA artificial networks. (Université de Paris, 2019).
  183. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci U S A* **98**, 4552–4557 (2001).
  184. Dramé-Maigné, A. Compartmentalized directed evolution of enzymes using molecular programs as self-selection function. (Sorbonne Paris Cité, 2018).
  185. LEDERBERG, J. Cell genetics and hereditary symbiosis. *Physiol. Rev.* **32**, 403–430 (1952).
  186. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, 666–675 (2013).
  187. Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & De Lorenzo, V. SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43**, D1183–D1189 (2015).
  188. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34–38 (1975).
  189. Ito, K., Uno, M. & Nakamura, Y. A tripeptide ‘anticodon’ deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
  190. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202 (2015).
  191. Casamassimi, A. & Ciccodicola, A. Transcriptional regulation: Molecules, involved mechanisms, and misregulation. *Int. J. Mol. Sci.* **20**, (2019).
  192. Atkinson, T. J. & Halfon, M. S. Regulation of gene expression in the genomic context. *Comput. Struct. Biotechnol. J.* **9**, e201401001 (2014).
  193. Kim, J., Darlington, A., Salvador, M., Utrilla, J. & Jiménez, J. I. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations. *Curr. Opin. Biotechnol.* **62**, 29–37 (2020).
  194. Hage, D. S. *et al.* Chapter 1 - Affinity Chromatography. in *Liquid Chromatography* (eds. Fanali, S., Haddad, P. R., Poole, C. F., Schoenmakers, P. & Lloyd, D.) 1–23 (Elsevier, 2013). doi:<https://doi.org/10.1016/B978-0-12-415806-1.00001-2>
  195. Juillerat, A. *et al.* Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chem. Biol.* **10**, 313–317 (2003).

196. Johnsson, K. Neb expressions. *NEB expressions* **3**, 2 (2008).
197. Gröne, Mark; Maurer, F. *Tryptophan quantification using UV fluorescence measurements OH.* (2014).
198. Chen, H., Xu, Z., Wang, C., Liao, Y. & Cen, P. Preliminary study on preparation of E. coli cell-free system for protein expression. *Front. Chem. Eng. China* **2**, 224–229 (2008).
199. Iyer, S., Karig, D. K., Norred, S. E., Simpson, M. L. & Doktycz, M. J. Multi-Input Regulation and Logic with T7 Promoters in Cells and Cell-Free Systems. *PLoS One* **8**, 1–12 (2013).
200. Zadeh, J. N., Wolfe, B. R. & Pierce, N. A. Nucleic acid sequence design via efficient ensemble defect optimization. *J. Comput. Chem.* **32**, 439–452 (2011).
201. Zheleznaya, L. A. *et al.* Nicking endonucleases. *Biochem.* **74**, 1457–1466 (2009).
202. Xu, S. Y. *et al.* Discovery of natural nicking endonucleases Nb.BsrDI and Nb.BtsI and engineering of top-strand nicking variants from BsrDI and BtsI. *Nucleic Acids Res.* **35**, 4608–4618 (2007).
203. Besnier, C. E. & Kong, H. Converting MlyI endonuclease into a nicking enzyme by changing its oligomerization state. *EMBO Rep.* **2**, 782–786 (2001).
204. Xu, Y., Lunnen, K. D. & Kong, H. Engineering a nicking endonuclease N.AlwI by domain swapping. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12990–12995 (2001).
205. Zhu, Z., Samuelson, J. C., Zhou, J., Dore, A. & Xu, S. Y. Engineering Strand-specific DNA Nicking Enzymes from the Type IIS Restriction Endonucleases BsaI, BsmBI, and BsmAI. *J. Mol. Biol.* **337**, 573–583 (2004).
206. Heiter, Daniel; Lunnen, Keith; Wilson, G. G. A method for engineering strand-specific, sequence-specific, DNA-nicking enzymes. (2003).
207. Abrosimova, L. A. *et al.* Peculiarities of the interaction of the restriction endonuclease BspD6I with DNA containing its recognition site. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**, 1072–1082 (2016).
208. Abrosimova, L. A. *et al.* A study on endonuclease BspD6I and its stimulus-responsive switching by modified oligonucleotides. *PLoS One* **13**, 1–19 (2018).
209. Yunusova, A. K. . A. R. I. . P. T. A. . A. L. A. . & L.A1, K. G. S. . Z. Small subunit of restriction endonuclease BspD6I is inactive in the presence of catalytic-deficient large subunit. *Mol. Biol. Genet.* **18**, 200–208 (2017).
210. Rosa, J., Fernandez-Gonzalez, E., Ducani, C. & Högberg, B. BtsCI and BseGI display sequence preference in the nucleotides flanking the recognition sequence. *PLoS One* **13**, 1–15 (2019).
211. Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 6785–9 (1983).
212. Boczkowska, M., Guga, P. & Stec, W. J. Stereodefined phosphorothioate analogues of DNA: Relative thermodynamic stability of the model PS-DNA/DNA and PS-DNA/RNA complexes. *Biochemistry* **41**, 12483–12487 (2002).
213. Jung, C. & Ellington, A. D. A primerless molecular diagnostic: phosphorothioated-terminal hairpin formation and self-priming extension (PS-THSP). *Anal. Bioanal. Chem.* **408**, 8583–8591 (2016).
214. Bell, C. E. & Lewis, M. Crystallographic analysis of Lac repressor bound to natural operator O1. *J. Mol. Biol.* **312**, 921–926 (2001).
215. Nb.BssSI Product information. Available at:

- [https://international.neb.com/products/r0681-nbbsssi#Product Information](https://international.neb.com/products/r0681-nbbsssi#Product%20Information). (Accessed: 9th December 2020)
216. Urtel, G., Van Der Hofstadt, M., Galas, J. C. & Estevez-Torres, A. REXPAR: An Isothermal Amplification Scheme That Is Robust to Autocatalytic Parasites. *Biochemistry* **58**, 2675–2681 (2019).
  217. Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.* **100**, 79–90 (2016).
  218. Chou, H. H. & Keasling, J. D. Programming adaptive control to evolve increased metabolite production. *Nat. Commun.* **4**, 1–8 (2013).
  219. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30**, 354–359 (2012).
  220. van Aalten, D. M. F. The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.* **20**, 2041–2050 (2001).
  221. Engstrom, M. D. & Pfleger, B. F. Transcription control engineering and applications in synthetic biology. *Synth. Syst. Biotechnol.* **2**, 176–191 (2017).
  222. Pandi, A. *et al.* Metabolic perceptrons for neural computing in biological systems. *Nat. Commun.* **10**, 1–13 (2019).
  223. Voyvodic, P. L. *et al.* Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *Nat. Commun.* **10**, 1–8 (2019).
  224. Koch, M., Pandi, A., Delépine, B. & Faulon, J. L. A dataset of small molecules triggering transcriptional and translational cellular responses. *Data Br.* **17**, 1374–1378 (2018).
  225. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids Res.* **48**, 1–14 (2020).
  226. Ellefson, J. W., Ledbetter, M. P. & Ellington, A. D. Directed evolution of a synthetic phylogeny of programmable Trp repressors article. *Nat. Chem. Biol.* **14**, 361–367 (2018).
  227. Tack, D. *et al.* The genotype-phenotype landscape of an allosteric protein. 1–28 (2020). doi:10.1101/2020.07.10.197574
  228. Chassin, H. *et al.* Sensing and responding to allergic response cytokines through a genetically encoded circuit. *Nat. Commun.* **8**, 1–13 (2017).
  229. Kopniczky, M. B. *et al.* Cell-Free Protein Synthesis as a Prototyping Platform for Mammalian Synthetic Biology. *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.9b00437
  230. Brödel, A. K., Sonnabend, A. & Kubick, S. Cell-free protein expression based on extracts from CHO cells. *Biotechnol. Bioeng.* **111**, 25–36 (2014).
  231. Klymenko, O. V., Shah, N., Kontoravdi, C., Royle, K. E. & Polizzi, K. M. Designing an Artificial Golgi reactor to achieve targeted glycosylation of monoclonal antibodies. *AIChE J.* **62**, 2959–2973 (2016).
  232. Genot, A. J. *et al.* High-resolution mapping of bifurcations in nonlinear biochemical circuits. *Nat. Chem.* **8**, 760–767 (2016).
  233. Rutherford, J. C. & Bird, A. J. Metal-Responsive Transcription Factors That Regulate Iron, Zinc, and Copper Homeostasis in Eukaryotic Cells. *Eukaryot. Cell* **3**, 1–13 (2004).
  234. Rollins, J. A. & Dickman, M. B. pH signaling in *Sclerotinia sclerotiorum*: Identification of a pacC/RIM1 homolog. *Appl. Environ. Microbiol.* **67**, 75–81 (2001).

235. Janmey, P. A., Wells, R. G., Assoian, R. K. & McCulloch, C. A. From tissue mechanics to transcription factors. *Differentiation* **86**, 112–120 (2013).
236. Pu, J., Zinkus-Boltz, J. & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol.* **13**, 432–438 (2017).
237. Schmelas, C. & Grimm, D. Split Cas9, Not Hairs – Advancing the Therapeutic Index of CRISPR Technology. *Biotechnol. J.* **13**, 1–12 (2018).
238. Repina, N. A., Rosenbloom, A., Mukherjee, A., Schaffer, D. V. & Kane, R. S. At light speed: Advances in optogenetic systems for regulating cell signaling and behavior. *Annu. Rev. Chem. Biomol. Eng.* **8**, 13–39 (2017).
239. Green, L. N. *et al.* Autonomous dynamic control of DNA nanostructure self-assembly. *Nat. Chem.* **11**, 510–520 (2019).
240. Janssen, B. M. G., Engelen, W. & Merkx, M. DNA-directed control of enzyme-inhibitor complex formation: A modular approach to reversibly switch enzyme activity. *ACS Synth. Biol.* **4**, 547–553 (2015).
241. Meijer, L. H. H. *et al.* Hierarchical control of enzymatic actuators using DNA-based switchable memories. *Nat. Commun.* **8**, (2017).
242. Dorn, M., E Silva, M. B., Buriol, L. S. & Lamb, L. C. Three-dimensional protein structure prediction: Methods and computational strategies. *Comput. Biol. Chem.* **53**, 251–276 (2014).
243. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN. *Proc. Natl. Acad. Sci.* **47**, 1309–1314 (1961).
244. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (80- )*. **181**, 223–230 (1973).
245. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
246. Karplus, M. The Levinthal paradox: Yesterday and today. *Fold. Des.* **2**, 69–75 (1997).
247. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
248. Lengauer, T. & Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, 402–406 (1996).
249. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
250. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein Structure Prediction Using Rosetta. in *Bioinformatics: Tools and Applications* **383**, 66–93 (2004).
251. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Genet.* **57**, 702–710 (2004).
252. Richardson, J. S. The Anatomy and Taxonomy of Protein Structure. in *Structure* 167–339 (1981). doi:10.1016/S0065-3233(08)60520-3
253. Li, H., Helling, R., Tang, C. & Wingreen, N. Emergence of preferred structures in a simple model of protein folding. *Science (80- )*. **273**, 666–669 (1996).
254. Wang, Z. X. A re-estimation for the total numbers of protein folds and superfamilies. *Protein Eng.* **11**, 621–626 (1998).
255. Jones, D. T., Taylor, W. R. & Thornton, J. M. A new approach to protein fold

- recognition. *Nature* **358**, 86–89 (1992).
256. Russell, R. B., Saqi, M. A. S., Bates, P. A., Sayle, R. A. & Sternberg, M. J. E. Recognition of analogous and homologous protein folds - Assessment of prediction success and associated alignment accuracy using empirical substitution matrices. *Protein Eng.* **11**, 1–9 (1998).
257. Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics* **19**, 1015–1018 (2003).
258. Fischer, D. Servers for protein structure prediction. *Curr. Opin. Struct. Biol.* **16**, 178–182 (2006).
259. Kihara, D., Chen, H. & Yang, Y. Quality Assessment of Protein Structure Models. *Curr. Protein Pept. Sci.* **10**, 216–228 (2009).
260. Benkert, P., Tosatto, S. C. E. & Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins Struct. Funct. Genet.* **71**, 261–277 (2008).
261. Choi, S. H., Kim, H. S. & Lee, E. Y. Comparative homology modeling-inspired protein engineering for improvement of catalytic activity of Mugil cephalus epoxide hydrolase. *Biotechnol. Lett.* **31**, 1617–1624 (2009).
262. Qi, X., Guo, Q., Wei, Y., Xu, H. & Huang, R. Enhancement of pH stability and activity of glycerol dehydratase from *Klebsiella pneumoniae* by rational design. *Biotechnol. Lett.* **34**, 339–346 (2012).
263. Kryshchak, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins Struct. Funct. Bioinforma.* **87**, 1011–1020 (2019).
264. Crawford, I. P. & Ito, J. SERINE DEAMINATION BY THE B PROTEIN OF ESCHERICHIA COLI TRYPTOPHAN SYNTHETASE. *Proc. Natl. Acad. Sci.* **51**, 390–397 (1964).
265. LANE, A. N. & KIRSCHNER, K. The Catalytic Mechanism of Tryptophan Synthase from *Escherichia coli* Kinetics of the Reaction of Indole with the Enzyme—l-Serine Complexes. *Eur. J. Biochem.* **129**, 571–582 (1983).
266. Rapf, R. J., Dooley, M. R., Kappes, K., Perkins, R. J. & Vaida, V. PH Dependence of the Aqueous Photochemistry of  $\alpha$ -Keto Acids. *J. Phys. Chem. A* **121**, 8368–8379 (2017).

## RÉSUMÉ

---

Les circuits moléculaires *in vitro*, basés sur des chimies programmables par ADN, peuvent remplir une gamme croissante de fonctions de haut niveau. Cependant, la plupart des démonstrations ne peuvent accepter que des acides nucléiques comme signaux d'entrée. Dans cette thèse, nous introduisons une stratégie générale pour interfacer les circuits basés sur l'ADN avec des signaux non ADN, stratégie basée sur des modules de traduction d'entrée. Nous concevons ces modules pour qu'ils soient réglables et modulaires. Ils peuvent être utilisés pour transmettre ou inverser la réponse associée à la présence d'une petite molécule donnée. Ceci est démontré en les combinant pour construire des circuits de détection qui fournissent une réponse temporelle quantitative fluorescente à la concentration d'entrée, avec une bonne spécificité et sensibilité.

La programmabilité de la couche d'ADN est exploitée pour effectuer une inversion logique, une modulation de signal et une tâche de classification sur deux entrées. Nous démontrons également la compatibilité avec les conditions biochimiques standard en effectuant une détection intégrée « one-pot » d'une enzyme grâce à son activité métabolique native. Ceci est approfondi en ajoutant une sortie PCR au circuit.

Nous prévoyons que cette stratégie sensible de conversion de petites molécules en ADN jouera un rôle essentiel dans les futures applications des circuits au niveau moléculaire, y compris les méthodes de crible fonctionnel et d'évolution dirigée *in vitro*.

## MOTS CLÉS

---

Programmation moléculaire, facteur de transcription allostérique, détection de petites molécules, évolution dirigée.

## ABSTRACT

---

*In vitro* molecular circuits, based on DNA-programmable chemistries, can perform an increasing range of high-level functions. However, most reported demonstrations can only accept nucleic acid as input signals. In this thesis, we introduce a general strategy to interface DNA-based circuits with non-DNA signals, based on input-translating modules. We design these modules to be tunable and modular. They can be repurposed to either transmit or invert the response associated with the presence of a given input. This is demonstrated by combining them to build sensing circuits that provide a fluorescent quantitative time-response to the concentration of their small-molecule input, with good specificity and sensitivity.

The programmability of the DNA layer is leveraged to perform logical inversion, signal modulation and a classification task on two inputs. We also demonstrate compatibility with standard biochemical conditions by carrying out one-pot detection of an enzyme through its native metabolic activity. This is taken further by adding PCR output to the circuit.

We anticipate that this sensitive small-molecule-to-DNA conversion strategy will play a critical role in the future applications of molecular-level circuitry, including functional screens and *in vitro* directed evolution methods.

## KEYWORDS

---

Molecular programming, allosteric transcription factor, small molecule sensing, directed evolution.