

UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE CIENCIAS
DEPARTAMENTO BIOLOGÍA
PROGRAMA DE MICROBIOLOGÍA



**Upgrading *Pseudomonas putida* as a Synthetic Biology chassis through
inter-operativity of genetic devices**

TESIS DOCTORAL

Memoria presentada para optar al grado de Doctor en Microbiología

Hüseyin TAŞ

Baccalaureus Scientiae in Biological Sciences and Bioengineering

Magister Scientiae in Computational Biology and Biophysics

DIRECTORES DE TESIS

Dr. Víctor de Lorenzo Prieto

Dr. Ángel Goñi Moreno

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC)

CENTRO NACIONAL DE BIOTECNOLOGÍA (CNB)

Madrid, 2020

To my mother and father without whose never failing support and sacrifices this Thesis would have never been possible.

Anneme ve babama, bu tez onların bitmek bilmeyen desteęi ve fedakarlıęı olmasa asla mümkün olmazdı.

ACKNOWLEDGEMENTS

I would like to mention all those whose limitless helps have been a support for me to complete this work that it is not only mine, but it belongs to those who were there for me when I needed them. A sentence would never be enough to show my appreciation, yet with the humble hope that it at least could be an indicator.

I must start with thanking both of my Thesis advisors:

Thank you Víctor for believing in me and for giving me the opportunity to build my scientific career at your laboratory. Your guidance on a path that has shaped and crafted my professional career was indispensable. This Thesis would not be as is without you, for that I cannot thank you enough.

Thank you Ángel. I must say that experiencing your energetic and scientific point of view of has added me irreplaceable value in my academic life. You showed me thinking out-of-the-box and turning the complex issues into basic building blocks. I cannot thank you enough.

I must continue with thanking to my family:

Without whose help and support this endeavor would be very difficult.

My dear sisters, Betül, Büşra and Merve, I thank you all for being on my side and listening to me all the time with long lasting patience and for continuing uninterruptedly the 'Kardeşler Reunion'. You have been there from the first moment to the end of it. Thank you all for being with me all the time.

I must thank to Laboratorio 30:

For making me feel like home. This has been an experience that would not be as is without the members of Lab30: I would like to mention some names yet this acknowledgement is not limited to these names, but belongs to a much bigger group of people. The ones who already left: Pablo Iván Nickel for being a friend and a colleague at the same time, Pavel Dvorák for sharing the aisle and many fun experiences together, Alberto Sánchez-Pascuala for keeping the constant support, Yamal Alramahi, Joseph Mougous, Jillian Silbert, Ilaria Benedetti, Danilo Rodrigo, Juhyun Kim and many more. And the ones that are still here: Esteban Martínez-Garcia for all the scientific discussions

and support, Tomás Aparicio for always encouraging me, Ángeles Hueso-Gil for all her support especially for the endless paperwork for PhD, Inés Merino for making the tricky bureaucracy easy pieces for us, Elena Algar, Elena Velázquez, Sofía, David, Luis and all previous and present members of Lab30.

I must thank to my university tutor:

Rafael Rivilla for guiding me on this Thesis.

I must thank to my collaborators and to those with whom I had a chance to exchange ideas:

Lewis Grozinger and Ruud Stoof from Newcastle University for fruitful collaborations & research together and accompanying me for the short but fun internship period I spent there.

Shuyi Zhang and Chris Voigt for welcoming me for a short stay to practice Cello system at Voigt Lab, MIT.

I must thank to my colleagues at European Synthetic Biology Society:

It was a privilege to be elected as the President of this Society and with the trust of all initiating an EU-based legal entity: 'European Synthetic Biology Society'. I thank all the steering committee members and the community of EUSynBioS.

I must thank to my friends:

Mustafa Anıl Koçak for being available for me, Murat İşbilen for not hesitating to get into big challenges together, Ahmet Hacıyakup for always supporting me, Ravish Patel for never forgetting me and always making me feel his support, Akos Nyerges for constructive science talks, Serkan Ağlaşan for his help and support, Gregor Steh for his never lasting friendship and Barış Yıldız for always listening to me.

I must thank to:

Everyone that are not mentioned here but always have helped me during this journey such as CNB administration and the employees, especially Human Resources department for taking care of the immense paperwork for me.

I must also thank to:

Madrid! It has been a great city, great place to share, live and learn. This period of my life has been invaluable to me, Madrid will always be in my heart.

Finally:

I would like to thank you Feyza, my beloved spouse. You took this challenge with me to the end of it and never left me alone. This journey turned into a much exciting experience with you being on the boat.

Funds and financial acknowledgements. This Thesis was funded by the projects of the European Union: ARISYS (ERC-2012-ADG-322797), EmPowerPutida (EU-H2020-BIOTEC-2014-2015-6335536) and MADONNA (H2020-FET-OPEN-RIA-2017-1-766975); and the project of Comunidad de Madrid - European Structural and Investment Funds - (FSE, FECER) InGEMICS-CM(S2017/BMD-3691); and the project of the Spanish Ministry of Science and Innovation: SETH (RTI2018-095584-B-C42 - AEI / 10.13039/501100011033) (MINECO/FEDER).

GENERAL INDEX

GENERAL INDEX.....	11
INDEX OF FIGURES.....	14
INDEX OF TABLES.....	16
ABBREVIATIONS.....	17
I. INTRODUCTION.....	21
II. OBJECTIVES.....	27
III. CHAPTER 1. Assets for gene circuit design in non-<i>E. coli</i> bacteria.....	31
1. Abstract.....	31
2. Results and Discussion.....	31
3. Conclusion.....	39
IV. CHAPTER 2. Portable functions.....	43
1. Abstract.....	43
2. Introduction.....	43
3. Results.....	44
3.1. Chassis-aware circuit design.....	44
3.2. Modifying thresholds.....	45
3.3. Gate compatibility.....	47
3.4. Portability of a NOR function.....	49
4. Discussion.....	50
5. Materials & Methods.....	52
5.1. Library of DNA gates.....	52
5.2. Experimental procedure.....	52
5.3. Flow Cytometry data calibration.....	52
5.4. Thresholding.....	52
5.5. Scoring system.....	53
V. CHAPTER 3. Contextual dependencies expand the re-usability of genetic inverters.....	57
1. Abstract.....	57
2. Introduction.....	58
3. Results.....	59
3.1. Generation of gate-context libraries.....	59
3.2. Effects of cross-context portability.....	61

3.3. Enhanced gate compatibility by fine-tuning contextual dependencies.....	63
3.4. Context-aware design rules for layered logic gates.....	65
4. Discussion.....	67
5. Materials and Methods.....	69
5.1. DNA and strain construction.....	69
5.2. Medium and experimental protocols.....	69
5.3. Flow cytometry analysis.....	70
5.4. Fluorescence data pre-filtered by cell size.....	70
5.5. Standard fluorescence measurements.....	70
5.6. Data fitting.....	71
5.7. Calculating compatibility.....	71
5.8. Computation of inverter chains.....	72
5.9. Similarity measure.....	72
5.10. Prediction.....	72
VI. CHAPTER 4. Tools for standardization of <i>Pseudomonas putida</i>	77
1. Abstract.....	77
2. Introduction.....	78
3. Results and Discussion.....	80
3.1. Reference promoter.....	80
3.2. RNA Polymerase immunocapture method.....	83
4. Conclusion.....	92
5. Materials and Methods.....	92
5.1. String-DB.....	92
5.2. Visual Molecular Dynamics analysis.....	93
5.3. Functional Analysis.....	93
5.4. Growth and Media.....	94
5.5. Formaldehyde Characterization.....	94
5.6. Lysis.....	95
5.7. Western Blotting.....	95
5.8. RNA Polymerase Pull-down.....	96
5.9. Sequencing.....	96
5.10. MALDI-TOF and LC-ESI-MS/MS.....	97
5.11. 1-3 Propanediol Characterization.....	97
5.12. Correlation and Outlier Analyses.....	98
5.13. Genetic Tagging in rpoC gene.....	98
VII. DISCUSSION.....	103
VIII. CONCLUSIONS.....	111

IX. REFERENCES.....	115
X. ANNEXES	125
I. Decomposing fluorescence and scattering.....	125
II. Compatibility tables for individual strains.....	126
III. Context Similarity	128
IV. Compatibility Scoring	132
V. Provided codes and data.....	134
VI. List of Supplementary Tables	135
1. Table for Function Portability	135
2. Table of primers	136
3. Table of Insulated gate response function parameters <i>P. putida</i> vs <i>E. coli</i>	138
4. Tables of constructs.....	140
4.1. List of gates from the Cello work.....	140
4.2. List of context dependent inverters.....	141
4.3. List of standardization and RNAP immuno-capture strains.....	145
VII. Summaries in other languages.....	146
1. Spanish Version.....	146
1.1. Presentación.....	146
1.2. Objetivos.....	148
1.3. Conclusiones.....	149
2. Turkish Version	150
2.1. Sunum	150
2.2. Hedefler.....	152
2.3. Sonuç.....	153

INDEX OF FIGURES

Figure 1: Schematic representation of BHR genetically-encoded inverters.....	33
Figure 2: Reference plasmids.	35
Figure 3: Experimental protocol.....	36
Figure 4: Behavior of genetically-encoded inverters in <i>P. putida</i> vs. <i>E. coli</i>	38
Figure 5: Parts and parameters	45
Figure 6: Threshold multiplier change and corresponding compatible gate pairs for <i>E. coli</i> and <i>P. putida</i>	46
Figure 7: Asymmetric thresholds for maximum compatible pairs	47
Figure 8: Pair compatibility and Scoring comparison for 20 gates.	48
Figure 9: <i>In silico</i> to <i>in vivo</i> NOR gate application in <i>P. putida</i>	49
Figure 10: Generating a library of gate-context devices.	60
Figure 11: Non-linear effects in the cross-context portability of inverters.	62
Figure 12: Comparing inverter compatibility and similarity across contexts.	65
Figure 13: Calculation of maximum circuit depth as a result of layering inverters. ...	66
Figure 14: Promoter characterization under 96 well plate growth condition.....	81
Figure 15: Promoter characterization under flask growth.	82
Figure 16: Promoters comparative strengths under 3 growth conditions.....	82
Figure 17: Synthetic promoters sequences.	83
Figure 18: Genetic tagging verifications of <i>rpoC</i>	84
Figure 19: Representation of an RNA Polymerase core enzyme.	85
Figure 20: Colony morphology.	85
Figure 21: Experimental procedure.	86
Figure 22: Pull-down of RpoC interacting proteins under various environmental conditions.....	87
Figure 23: Fold changes for heat shock and solvent stress conditions.....	89
Figure 24: Quantitative comparison of Exponential state, Heat Shock and Solvent Stress conditions.	91
Figure 25: Purification factor.	93
Figure 26: Formaldehyde characterization in <i>Pseudomonas putida</i>	95
Figure 27: Sanger sequences of <i>in vivo</i> tags.....	96
Figure 28: 1,3-Propanediol characterization in <i>Pseudomonas putida</i>	97

Figure 29: Genetic Tagging in <i>rpoC</i> Gene.	99
Figure S 1: Example of filtering of flow cytometry data.	125
Figure S 2: Compatibility tables for the library in different hosts.....	127
Figure S 3: Similarity scores heatmap for 20 gates in 7 contexts.....	131
Figure S 4: Compatibility score heatmaps for gates in the 7 contexts.	133

INDEX OF TABLES

Table 1: Collection of SEVA constructed NOT gates.....	34
Table S 1: Calculation of Sensor OFF/ON activities.	135
Table S 2: Characterization and measurement.	135
Table S 3: Complete set of primers used.....	136
Table S 4: Insulated gate response function parameters <i>E. coli</i> vs <i>P. putida</i>	139
Table S 5: 12 main gates and 8 variants that were obtained from the Cello work. ...	140
Table S 6: List of all libraries for context dependent inverters.	141
Table S 7: Standardization and RNAP immuno-capture Strains.....	145


ABBREVIATIONS

1-3-PD	1-3 Propanediol
3MBz	3-Methylbenzoate
A	Absorbance
ACN	Acetonitrile
Ap	Ampicillin
ATP	Adenosine triphosphate
BHR	Broad Host Range
bp	Base pair
CAD	Computer-Aided Design
CHCA	α -Cyano-4-hydroxycinnamic acid
FA	Formaldehyde
FC	Fold Change
IPTG	Isopropyl β -d-1-thiogalactopyranoside
Km	Kanamycin
PoPS	Polymerase per second
PSMs	Peptide-to-spectrum matches
RBS	Ribosomal Binding Site
RNAP	RNA Polymerase
SD	Shine-Dalgarno Sequence
SEVA	Standard European Vector Architecture
TFA	Trifluoroacetic acid
WB	Western Blot

I. INTRODUCTION

I. INTRODUCTION

Current synthetic biology is known for the efforts in designing regulatory networks and implementation of genetic circuits. The typical methodology to this end is assigning biological counterparts of input and output states via an analogy made from computational sciences i.e. computer science and electronics engineering. Inputs are bio- or physico- chemical entities (e.g. temperature, radiation, a pollutant or any type of inducers), operational machineries as being the cellular effectors (e.g. proteins, designed RNA or DNA sequences, cellular metabolites) that contain an intelligent design behind with a layer of computation and outputs as being the action/product of interest (e.g. induction of targeted taxis, release of molecules, induction of gene expression, programmed cell death, relocation of DNA information etc.) [1, 2]. Memory storage, complex finite machines, biosensors for environmental bioremediation and health applications, or intracellular control of metabolites [3-5] are some of the Boolean logic based operations in genetic circuit design. The urge of more complex circuit designs [6-9] resulted in the interest of their automation and design which is achieved with a recent milestone e.g. CelloCAD study [4] – a predictable and automated biosensor design tool with applications of complex calculation layers composed of inverter systems (NOT logic gates) of repressible promoters. Cello design is initially done for a strain of *Escherichia coli* and the success rate is depicted as: 92% correctly predicted states and over 70% correctly predicted biosensor functions, which is pioneering within the contemporary levels. Still, this design tool is limited within the organism it is designed for [4, 10]. Libraries used for characterization of host are subject to optimization for each new host – that is, they are not suitable for 'as is' use. The inherent approach of host-specific genetic part optimization has limited this design tool and several others, and resulted them in to be abstained from widespread use. Within this view, small alterations even selecting a different strain within the same species, may fluctuate the design efficiency and limit the tool for single organism. The bottleneck here could be solved with the presence of large enough libraries (to be characterized without requiring additional optimization steps) and ready to implement for various hosts.



Digital side of electronics engineering with implementations of basic properties of circuits as ON / OFF states had an impact on design ability of biology via genetic expression systems – that is to say, genetic Boolean circuits [4] that have the information processing component as logic operation gates. Constructing these gates in a row make up for the computational layer [4, 11]. The information processing can be converted in such a way that desired functions [12] such as diagnostic implementations [13], metabolic production [14] or bioremediation [15, 16] to memory storage [17] can be retrieved as outputs. Yet, as initial proof of concepts are successfully achieved, still compared to their electronics equivalents the biological systems are more prone to alterations over time due to several parameters e.g. intrinsic noise, evolution or orthogonality [18]; thus, *in vivo* ambiance challenges the idea of isolated *in silico* Boolean logic operators. Subsequently, biocircuit design implementations share a similar endeavor that is intracellular background.


The challenge synthetic biology has faced resulted from analogy of electronics is now requiring an out of the box approach to overcome with – that is, oversimplification of sole DNA parts being sufficient for accounting the performance of biocircuits. This angle is one of the upgrades highlighted in this Thesis. The host genetic texture has been recently suggested as playing a central role on robustness of cellular circuits [19] – a perspective until now stayed partially revealed. Further attempts have been suggested to overcome the burdens resulting from DNA focused viewpoint – host-awareness (a term recently coined) explains the mutual effect between context and genetic circuits [20-22] and interoperability of synthetic biology devices [23].

Although many known biocircuit CADs are constructed on the idea of host-optimized parts [4, 10], in order to make use of the parts developed to this point, reutilization of genetic parts engineered for one host towards other hosts (e.g. device portability) may be a competent approach. Following contemporary attempts reported on interactions between host and genetic devices, it may be safe to think that a way towards device portability may be by considering contextual effect. Consequently, contextual effect could be formalized to tackle the notion of re-configurability.

The urging interest for spreading synthetic biology applications upon new frontiers requires better fitting parts to chassis. A chassis of interest may be defined by the task

to be implemented. Yet, most synthetic biology applications prioritize use of one chassis. Whereas, some applications may be enabled by benefiting variation of host cells [24] i.e. using *Pseudomonas putida* for environmental applications by borrowing the biocircuits designed in *Escherichia coli*. However, it is probable that designs optimized for one host may perform functions differently in another host – emphasizing the genetic context as a key concept. An intrinsic value of biology, context-variability, is argued here to be used as a potential gain for genetic circuit design instead of regarded it as *impedimentum*. Consequently, function performed by a genetic device would be considered as a combination of its DNA sequence and the context introduced in. Accordingly, consideration of context could be used as an approach to optimize genetic devices, which would yield engineering of reconfigurable circuit elements that are containing same DNA information yet performing distinctive functions [25].

Reproducible communication of research outputs is one of the central questions in synthetic biology i.e. building standards for sharing results and standardized tools plus chassis for specific purposes. Gene expression is one topic where standardization efforts are visible [26, 27], yet would benefit some refinements. A standard chassis with a reference promoter having a reproducible constitutive expression in altered environments, together with known numbers of actual RNA Polymerase at a given time (PoPS) could facilitate a solid standard strain. *Pseudomonas putida*, the soil bacterium, is a known chassis famed with its large-scale applications in environment and industry due to its metabolic pathways enduring solvent stresses and robust biochemical reactions [15, 28, 29]. Thus, it converts into a wholesome candidate to engineer as a synthetic biology chassis for standardization. Tools developed for one purpose could have multi-uses, as one for pull-down of RNAP to facilitate mapping its interactome could also be used for PoPS calculations. Extensive research has been done for utilizing *P. putida* in biotechnological applications [28-35], and its transcriptional regulation is intriguing curiosity and strengthening its position as a microbial chassis. RNAP is at the center of transcriptional regulation and its interactions could be the starting point for unveiling novel expression systems – besides already known tight regulation systems of *P. putida* e.g. *XylS/Pm*, *AlkS/PalkB* etc. [36, 37]. RNAP in the course of action interacts with DNA, RNA and proteins. It transcribes RNA as a result of its interaction with DNA that is directed by



regulatory proteins. During the course of interaction, RNAP interacting proteins such as transcriptional regulators, sigma and anti-sigma factors play key importance. These interactions could be arranged to rewire genetic networks [38].

Sigma factors could be of potential use for engineering cellular machineries. Yet, identifying them in *P. putida* requires novel tools (as had been the case for other chassis [33, 39]) that may help to understand RNAP interactome. *P. putida* has guaranteed its central role in synthetic biology [28, 40], however the endeavor of upgrading *P. putida* may be excelled by eliminating the unknowns and understanding its whole especially where limited information is a challenge e.g. transcriptional factors [41]. In this study efficient tools were developed and tested under biologically relevant stress and growth conditions towards *à la longue* achieving standardization of both gene expression units and RNAP quantification that is also extensively exploited for a downstream application e.g. RNAP interactome mapping.

II. OBJECTIVES

II. OBJECTIVES

The **general objective** of this Thesis is:

Upgrading *Pseudomonas putida* as a synthetic biology chassis for genetic circuit designs, investigating interoperability concept, developing tools for standardization and quantitative assets for engineering of transcription regulatory mechanisms.

This general objective includes the following **specific partial objectives**:

1. To design an *ad hoc* broad-host-range biocircuitry tool set for implementation of design automation in non-*Escherichia coli* Gram-negative bacteria.
2. To investigate automated circuit design in *Pseudomonas putida*.
3. To establish contextual variability as an influence on portability of genetic devices and using it for improving inter-operativity concept.
4. To engineer a chassis fortified with incorporated standardization tools within the genome.
5. To develop a pull-down method for revealing transcription regulation machinery and for quantitative interactome mapping of *Pseudomonas putida*.

III. CHAPTER 1

ASSETS FOR GENE CIRCUIT DESIGN IN NON-*ESCHERICHIA COLI* BACTERIA

III. CHAPTER 1. Assets for gene circuit design in non-*E. coli* bacteria

*The content of this chapter has been published as:

Huseyin Tas, Ángel Goñi Moreno, Víctor de Lorenzo (2020) A standardized broad host range inverter package for genetic circuitry design in Gram-negative, at BioRxiv.org with the following link:


<https://doi.org/10.1101/2020.07.14.202754>

1. Abstract

Genetically encoded logic gates, especially inverters—NOT gates—are the building blocks for designing circuits, engineering biosensors or decision-making devices in synthetic biology. However, the repertoire of inverters readily available for different species is rather limited. In this work, a large whole of NOT gates that was shown to function previously in a specific strain of *Escherichia coli*, was recreated as broad host range (BHR) collection of constructs assembled in low, medium and high copy number plasmid backbones of the SEVA (standard European Vector Architecture) collection. The input/output function of each of the gates was characterized and parameterized in the environmental bacterium and metabolic engineering chassis *Pseudomonas putida*. Comparisons of the resulting fluorescence cytometry data with those published for the same gates in *Escherichia coli* provided useful hints on the portability of the corresponding gates. The hereby described BHR inverter package (20 different versions of 12 distinct gates) thus become a toolbox of choice for designing genetic circuitries in a variety of Gram-negative species other than *E. coli*.

2. Results and Discussion

Design and implementation of genetic circuits is one of trademarks of contemporary synthetic biology [42]. The archetypal approach involves abstracting biological cues (e.g. effectors, metabolites, proteins) and physicochemical signals (e.g. inducers, temperature) as inputs. These are processed through a more or less complex computation layer (most often assembled with regulatory parts: transcriptional



factors, riboregulators etc) which then releases another biological signal as the output. This in turn can be used as the input for another node of the circuit and so on [1, 2]. Further abstraction of biological circuits as wholes of Boolean logic gates enables a superior level of complexity, as shown by a suite of examples involving rewiring of stress responses, detection of environmental contaminants, implementation of cellular calculators and others [3-5]. Yet, as the demand for increasingly complex circuits grows [4, 6, 8], so does the interest in automation of their design and execution. One major landmark in this direction was the publication in 2016 of the CelloCAD platform [4], a complete operative system for virtual assembly and eventual implementation of logic circuits in *E. coli* through combination of series of NOR gates—based themselves on a large collection of well characterized promoter/repressor pairs. This tool affords automated design and simulation in seconds of complex genetic networks with successful prediction of > 70% of all states shown by the actual DNA constructs once synthesized and knocked in *E. coli* [4, 43]. Yet, due to the material nature of the building blocks, CAD tools such as Cello are inherently restrictive to single strains or species. While these systems may work well in a given organism, the parameters and general behavior can change very significantly when passed to another hosts. This is a considerable issue when circuits are desired to compute signals under environmental and industrial conditions for which *E. coli* is not an optimal chassis.

On this basis we set out to engineer a robust and easy-to-use package of NOT gates that could be used for circuit design in Gram-negative bacteria other than *E. coli* and that—once characterized in the host of interest—could benefit from the CelloCAD software [4] for automatic assembly of the cognate DNA.

To this end, we recreated the DNA sequence encoding each of the inverters available in the Cello platform and pass them to standardized vectors of the broad host range SEVA collection with low, medium and high copy numbers. The general organization of the constructs is sketched in **Figure 1**. Note that the business part of each plasmid is flanked by upstream and downstream terminators added by the SEVA structure to mitigate potential read-through from vector promoters. The plasmids backbones were retrieved from SEVA database and repository [44].

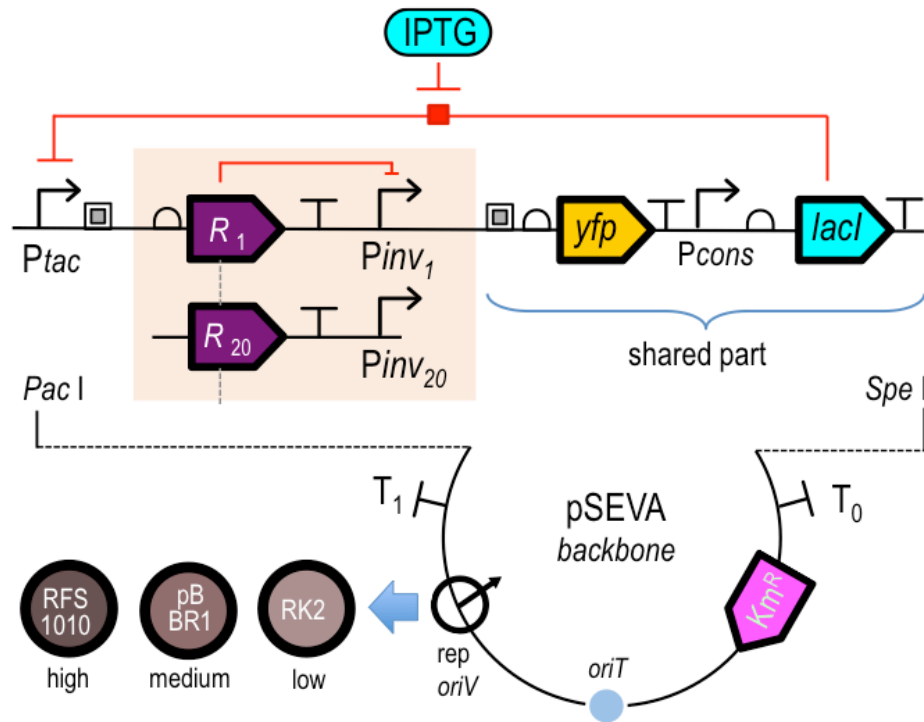


Figure 1: Schematic representation of BHR genetically-encoded inverters.

The organization of 20 variants of functional parts assembled following the SEVA standard is sketched (components not to scale). Similarly to the earlier collection for *E. coli* [4] there are segments shared by all construct i.e. the $P_{Tac}/lacI/IPTG$ -dependent expression system for each of the repressor/SD combinations (R_1 to R_{20}) and the *yfp* gene used for fluorescent readout of inverter performance. Segments coming from SEVA backbone (T_1 and T_0 terminators, origin of conjugal transfer *oriT*) and the *KmR* resistance gene are common to all constructs as well. The variable parts include [i] the DNA encoding the SD, the repressor gene (*R*) and the cognate repressible promoter (*P_{inv}*) upstream of the reporter *yfp* and [ii] the BHR origin of replication: RK2, pBBR1 or RFS1010, each of them supporting a different copy numbers as indicated.

In practical terms, each of the segments encoding the gates was amplified from the collection of *E. coli* NEB10 β strains bearing p15A/*Km^R* vectors (~ 15 copies) inserted with the cognate DNA [4]. Primers (Merck Sigma Aldrich, Inc) were designed for adding SEVA-compatible *PacI* and *SpeI* restriction sites to the extremes of the amplicons generated with the proofreading DNA polymerase of Q5 High-Fidelity (New England BioLabs, Inc.). Primers used for amplification of the DNA of the NOT gates and verification of the constructs are listed in Annex **Table S3**. Following PCR, the resulting fragments were separately cloned in pSEVA221 (RFS1010*oriV*, ≥ 50 copies) and first captured, verified and re-sequenced in the host strain *E. coli* CC118 λ pir. The complete catalogue of constructs is shown in **Table 1**. Note that some gates (those designated BM3R1-B1, PhIF-P3, QacR-Q2, SrpR-S2 and SrpR-S3) could not be cloned in the higher copy number vectors, presumably due to the toxicity

effects of the encoded repressor proteins. The collection (**Table 1**) includes a total of 12 inverters, some of them bearing two or more variants of the Shine-Dalgarno sequence of the repressor gene that changes its expression levels. For example, AmeR-F1 (AmeR is the gate and F1 is the SD) has only 1 version of the ribosomal

Table 1: Collection of SEVA constructed NOT gates

Inverters	Low Copy pSEVA221	Medium Copy pSEVA231	High Copy pSEVA251
AmeR-F1	✓	✓	✓
AmtR-A1	✓	✓	✓
BetI-E1	✓	✓	✓
BM3R1-B1	✓	✓	-
BM3R1-B2	✓	✓	✓
BM3R1-B3	✓	✓	✓
HlyIIR-H1	✓	✓	✓
lcaRA-I1	✓	✓	✓
LitR-L1	✓	✓	✓
LmrA-N1	✓	✓	✓
PhIF-P1	✓	✓	✓
PhIF-P2	✓	✓	✓
PhIF-P3	✓	✓	-
PsrA-R1	✓	✓	✓
QacR-Q1	✓	✓	✓
QacR-Q2	✓	✓	-
SrpR-S1	✓	✓	✓
SrpR-S2	✓	✓	-
SrpR-S3	✓	✓	-
SrpR-S4	✓	✓	✓
Autofluorescence	✓	✓	✓
Standardization	✓	✓	✓
Promoter Activity	✓	✓	✓

binding site of the repressor gene; however, SrpR has 4 versions (SrpR-S1, SrpR-S2, SrpR-S3, SrpR-S4 (SrpR is the gate and S1, S2, S3 and S4 the 4 different SDs). Note that the arrangement of functional parts of the gates in the SEVA vectors is identical to the one originally adopted in the p15A/KmR vectors of the Cello platform (**Figure 1**). In addition to the plasmids bearing 20 inverters, we built a set of reference constructs allowing relative promoter units (RPU) to be converted and compared between different conditions. These references (**Figure 2**) include [i] autofluorescence plasmids (recreating the business parts of pAN1201 [4]) and consisting of each of the backbone plasmids but without any insert e.g. missing the repressor/target promoter

segments highlighted in **Figure 2 [ii]** RPU standard plasmids derived from pAN1717 [4] in which *yfp* is expressed through the constitutive promoter J23101 and which are used in combination with the autofluorescence plasmid for converting YFP readouts of each inverter into RPU (see Equation 1 below) and **[iii]** promoter activity plasmid

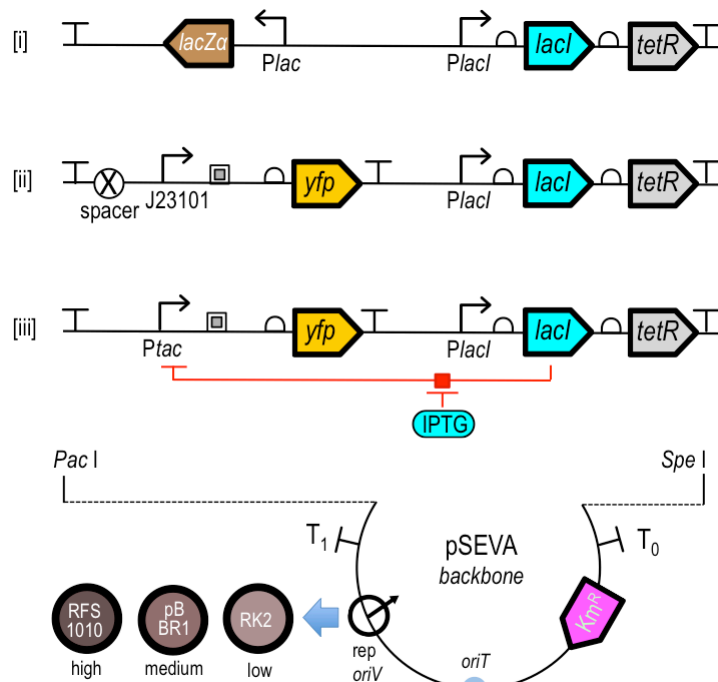


Figure 2: Reference plasmids.

[i] Autofluorescence plasmid: Schematic representation of BHR autofluorescence plasmid. Autofluorescence plasmid is used as the circuit backbone as published previously [4]. Sensors LacI and TetR are encoded in the same operon at the downstream of *PlacI* constitutive promoter. *lacZα* gene is expressed by constitutive *Plac* promoter. The terminator at the downstream of *Plac* is L3S2P21 terminator and *tetR* gene is insulated with native AraC terminator. Both terminator are used for

insulating the transcriptional read-throughs. The SEVA backbone terminators (T_0 and T_1) are kept to insulate the circuit encoded in-between. Km resistance gene is used as a part of SEVA backbone. The BHR origin of replication is prepared in 3 versions as seen in the figure, RK2, pBBR1, RFS1010. [ii] RPU standard plasmid: Schematic representation of BHR RPU standard plasmid. Sensors LacI and TetR are encoded in the same operon at the downstream of *PlacI* constitutive promoter. *yfp* gene is expressed by standard constitutive J23101 promoter. There is a 15 bp spacer at the upstream of J23101 promoter, and the terminator upstream of it together with the terminator at the downstream of *yfp* is L3S2P21 terminator that is used for insulating the transcriptional read-throughs. *tetR* gene is insulated with native AraC terminator. The SEVA backbone terminators (T_0 and T_1) are kept to insulate the circuit encoded in-between. Km resistance is used as a part of SEVA backbone. The BHR origin of replication is prepared in 3 versions as seen in the figure, RK2, pBBR1, RFS1010. [iii] Promoter activity plasmid: Schematic representation of BHR promoter activity plasmid. Sensors LacI and TetR are encoded in the same operon at the downstream of *P_{lacI}* constitutive promoter. The sensors are to be used with corresponding promoter in this case *P_{Tac}* is used. *P_{Tac}* is regulated by LacI expressed in the operon and by the inducer IPTG. RiboJ ribozyme is used as a promoter context insulator between *P_{Tac}* and *yfp* gene. By regulating IPTG levels the activity of the *P_{Tac}* promoter is measured by YFP levels. The BHR origin of replication is offered in 3 versions as seen in the figure, RK2, pBBR1, RFS1010.

recreating pAN1818 [4], which is used for measuring the promoter activity (P_{Tac} - YFP) based on inducer concentrations (**Figure 2**). The library of constructs listed in **Table 1** is expected to ease utilization of CelloCAD as a genetic programming tool in

a suite of Gram-negative bacteria. Note that users have a choice to pick the same gate borne by plasmids with different copy numbers, what may be critical to avoid potential toxicity. Adoption of the standard SEVA format gives two additional advantages. While the construct library was built on vectors with a Km resistance gene, the modularity of the SEVA format makes its replacement by an alternative antibiotic marker for selection [44, 45] easy. Also, the BHR nature of the standardized construct affords their implantation in diverse Gram-negative hosts, thereby giving a

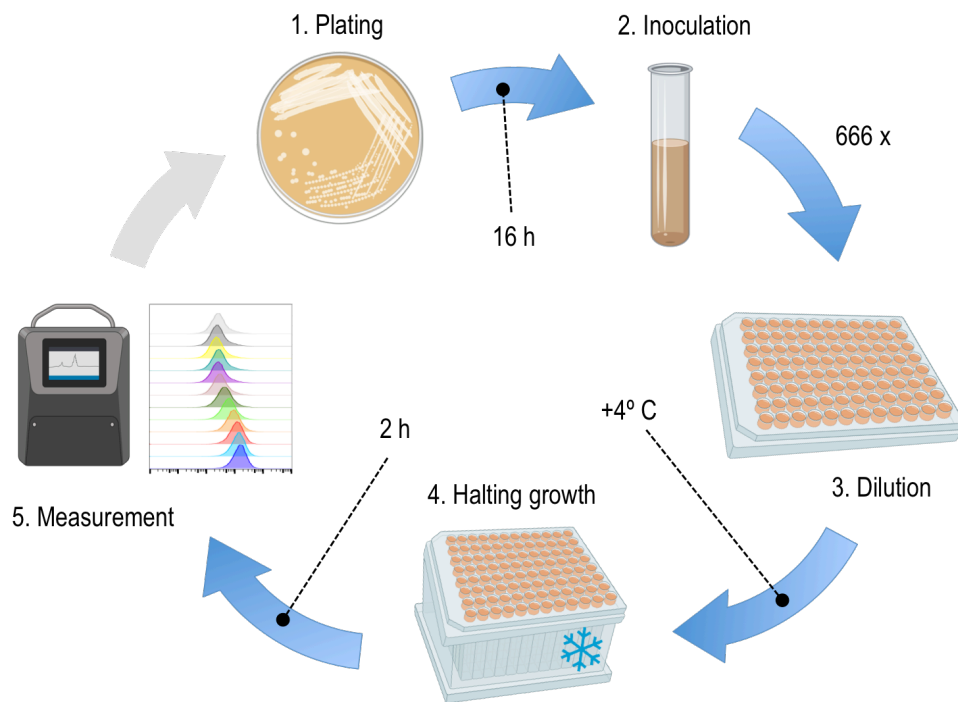


Figure 3: Experimental protocol.

Experimental protocol used for performance measuring of the genetic inverters listed in Table 1 in *P. putida* KT2440. In all the experiments bacteria were grown in M9 medium adapted for growth of *P. putida* (250 ml of liquid culture contained 25 ml x10 M9 salts, 500 μ l of 1M MgSO₄, 2.2 ml of 20% citrate and milliQ-H₂O to volume). 50 μ g ml⁻¹ kanamycin were added to secure plasmid retention (all of them Km^R.) For the experiments, saturated overnight cultures were diluted ~ 600-fold in a microtiter plate with 200 μ l per well, added with IPTG concentrations ranging 0 to 1000 μ M and incubated at 30 °C with shaking for 24 h. Cultures (typically reaching OD₆₀₀ ~ 0.2-0.3) were then kept in the cold for the rest of the procedure. YFP fluorescence distribution of each sample was measured with a Miltenyi Biotec MACS flow cytometer at channel B1 with an excitation of 488 nm and emission of 525/50 nm. For each sample 30 thousand events were collected with singlet gating. Calibration was done by using MACSQuant Calibration Beads (see text for explanation).

chance to compare their performance in different species and thus learn about interoperability and context dependencies in circuit designs.

In order to validate these features we tested and parameterized the whole low-copy number versions of the inverter library of **Table 1** in the environmental bacterium and Synthetic Biology chassis *Pseudomonas putida* KT2440 [29, 46] (Annex, **Table S6** - Tas193 - Tas212). The experimental workflow to this end (**Figure 3**) was designed considering the specific needs of *P. putida* for growth as detailed in the legend of the figure. Once strains bearing each of the constructs were generated, transformants were grown and IPTG concentrations of (5-1000 μ M) added to activate each of the devices, for a total 24 h period. YFP fluorescence emission detected with a flow cytometer was recorded after 24 of IPTG addition. Data were then analyzed with FlowJo software (<https://www.flowjo.com/>). An important detail was that the auto-gating option of the software was set considering at least 50% of the events covered while Forward and Side scatters were plotted. The same gating conditions were applied to all specimens in the same group and repeated for all the samples.

On the basis of the thereby produced data we quantified the output of the individual devices at each condition as standard RPU (relative promoter units). This was done by characterizing the fluorescence values emitted by the bacterial population of the cultures exposed to IPTG levels covering an induction ranges from none to saturation (in our case 12 points of growing effector concentrations). On this basis, the RPU value at each point can be calculated with the equation below:

Equation 1

$$RPU = \frac{\langle YFP \rangle - \langle YFP \rangle_{\text{autofluorescence}}}{\langle YFP \rangle_{\text{standardization}} - \langle YFP \rangle_{\text{autofluorescence}}}$$

Where $\langle YPF \rangle$ is the median fluorescence value from the gate of interest that is to be converted into RPU (from either the inverter-bearing plasmids or the control promoter activity plasmid), $\langle YPF \rangle_{\text{autofluorescence}}$ is the median fluorescence value of auto-fluorescence plasmid, $\langle YPF \rangle_{\text{standardization}}$ is the median fluorescence value from the standardization plasmid. Next, the output vs input values of each inverter were represented in a response function plot generated with Hill fits and utilizing the RPU figures as the input to the calculations. Specifically, Hill equation parameters were estimated by plotting each inducer levels with their corresponding standardized median fluorescence values (Supplementary Data Raw) by means of the equation below:

Equation 2

$$y = y_{min} + \frac{(y_{max} - y_{min}) k^n}{k^n + x^n}$$

where y is the output promoter activity, y_{min} is the minimum observed value for promoter activity, y_{max} is the maximum observed value for promoter activity, k is the input value for which half maximum value for output promoter is reached, n is the Hill coefficient. Experimental response data were thus entered in the above Hill equation.

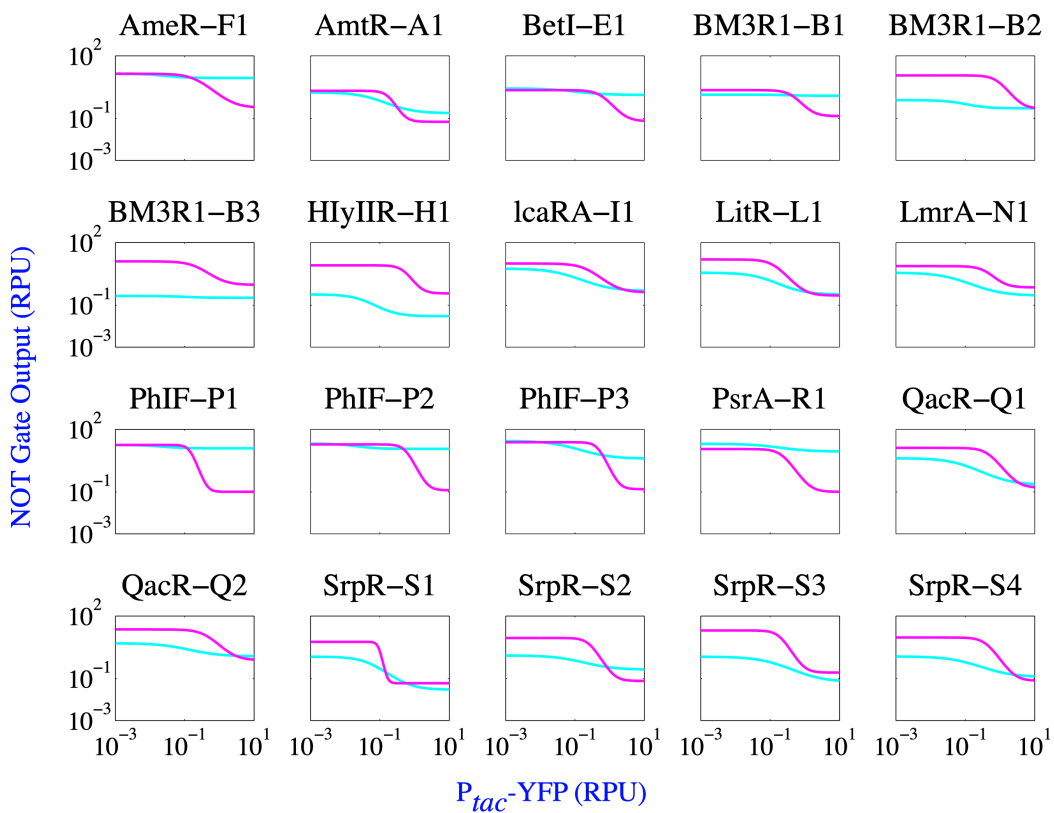


Figure 4: Behavior of genetically-encoded inverters in *P. putida* vs. *E. coli*.

The panel shows the comparisons of the Hill fits of the same inverters in *E. coli* NEB10 β (magenta; data retrieved from [4]) and their behavior in *P. putida* KT2440 (cyan; experimental from this study). X-axes correspond to the activity of the IPTG-inducible pTac promoter and Y-axes indicate the activity of corresponding inverters. Both axes indicate YFP expression in RPU.

The corresponding Hill parameters are provided in Supplementary **Table S4** for both *E. coli* (retrieved from www.cellocad.org) and *P. putida* values (this study). Fitting was performed with MATLAB using the scripts provided in Supplementary Data Scripts. The resulting characterization of the 20 inverters in *P. putida* based on the protocols and calculations above is shown in **Figure 4**. The data showed a range of

divergences in the behavior of the inverters in either host. In some cases, the patterns were comparable both in terms of the dynamic ranges of the input/output, the contour of the response curves and the specific values of promoter strengths. In other cases, the shape of the curve was kept but the boundaries changed very significantly. And finally, in yet another series of devices there was little if any similarity in their input/output transfer functions between the two types of bacteria. While these data expose the limitations of just exporting genetic devices from one species to the other, the large repertoire of gates also enable users to pick the ones whose parameters are compatible with the CelloCAD tool for automation of circuit designs[4]. Furthermore, the dataset associated to **Figure 4** encrypts valuable, quantitative information on the interoperability of parts and circuits between different biological recipients of the same constructs, an issue hardly tackled thus far in the Synthetic Biology literature [47-49].

In sum, we believe that the hereby described collection of inverters available in a BHR format will help to expand the possibilities of genetic programming towards bacteria other than *E. coli* but still interesting from a SynBio-inspired biotechnological perspective. Furthermore, their testing and parameterization in various hosts may deliver general portability rules that thus far rely on a mere trial-and-error exercise. The whole collection of constructs is available through the SEVA database and vector repository at <http://seva.cnb.csic.es>.

3. Conclusion

In conclusion, the limitation in the accessibility of available logic gates for high throughput applications (e.g. BHR) has driven the desire for engineering a easy to use library for automation of circuit designs. Our library is suitable with CelloCAD tool. It is based on SEVA structure which gives a chance to automated circuitry designs not only in *E. coli*, but in any given Gram-negative bacteria. By introducing this fast circuitry automation methodology, a limiting factor on one of the powerful design tools is partially alleviated. Now it is accessible broad host range, whose application is shown in *P. putida*. Our system may be a handy approach to respond to important bottleneck and a need in implementations of circuitry designs.

IV. CHAPTER 2

PORTABLE FUNCTIONS


IV. CHAPTER 2. Portable functions

1. Abstract

Boolean logic operations are commonly preferred in biocircuit design applications. Both intracellular and external factors impact upon *in vivo* implementation of genetic circuits. Host genetic material is a key point on how circuits perform. According to the host used a circuit may turn into inoperable or confront efficacy problems. That is, a biocircuit with the very same DNA sequence may alter its performance according to the chassis - portability phenomenon. This underpins the attempts at illumination of interoperability concept. In this study, a NOR gate circuit was designed and implemented in *Pseudomonas putida* by using a known automated biosensor design tool CelloCAD. Additionally, notional improvements were tested in thresholding and scoring guidelines of the Cello system together with constructs introduced e.g. *P. putida* originated expression systems. A new scoring system allowed *in silico* comparison of how (in)compatible two given gates are - giving opportunity for assigning a degree of functionality to circuit connections. These scores are built upon the concept of thresholding, which is of central importance for obtaining well-defined Boolean circuit outputs. Thresholding parameters affected the total number of compatible gate pairs *in silico*. The experimental data was converted into Relative Promoter Unit (RPU), as a standard expression unit. Cello promoters (P_{Tac} , P_{Tet} , P_{BAD}) and additional 2 *P. putida* promoters (P_m and P_{alkB}) were all characterized as RPU in *P. putida*. Our results indicated a circuit with same genetic material functions in *P. putida* does not necessarily execute the same way in *E. coli* - underlining the lack of DNA material interoperability, but function portability.

2. Introduction

Circuit design applications in synthetic biology borrow numerous approaches from Boolean logic synthesis. Intracellular noise, evolution and host genetic material remain challenging points which frustrate stable circuit performances [22]. Yet, host specific optimization of parts and devices is a common practice [4] to achieve similar device behaviors in different chassis. Portability as an unsolved issue causes limitations in the spread of biocircuits, as remaking new circuits to obtain the same function in new organisms is time-consuming. When a circuit is moved into new



hosts, its performance may encounter phenotypic alterations that result in powerful circuits to become impractical ones [47]. The need for formalizing this issue gives the motivation to explore conceptual novelties in portability.

The idea of applying engineering principles to biology, that is, merging basics of electronic circuits with fine-tuning ON and OFF states of expression systems in biology, has resulted in DNA based biocircuitry of complex computational layers [4]. Linking promoters and their regulators in a way that accounts for a logical operation makes a biological Boolean circuit [11]. These connections are used for intelligent design of solutions which synthetic biology applies to e.g. DNA memory [17], bioremediation [15, 16], biocomputing [4, 18] or diagnostics [13]. Yet, due to intrinsic characteristics of cellular structures genetic circuits are more prone to alterations in performance and susceptible to outside effects in comparison to their electronic counterparts that offer standardized and idealized digital computing. This underlines the central role of host structure while contemplating circuit performance.

While applying engineering principles, synthetic biology may confront interoperativity as a key issue. Several studies elaborate on its contribution to the implementation of portable backbones and libraries [50], investigate its parameters [48] and reformulate the issue [51]. Though interoperability is discussed in many aspects, it is mostly focused on transfer of circuits and of performances among hosts. Consequently, interoperability would not only consist of circuit transfer but also the idea of moving its function. In the light of this approach, without requiring fine-tunings portable functions could be achievable and could allow new phenotypes in different chassis.

3. Results

3.1. Chassis-aware circuit design

Moving a genetic construct from one context to others has an impact on device behavior. Context could be defined with parameters such as host genetic pool, circuit-carrying backbone. In order to highlight the problems with disregarding contextual dependencies [51] the term host-awareness [48, 52] was coined - that is, to emphasize the interaction between the biocircuit and the host, and to emphasize the failure when

deciding phenotypic features of a genetic circuit solely based on its DNA sequence. A Boolean NOT logic device (**Figure 5A**) was used to question interoperability limits of the very basic building block of logic based genetic operations. A BHR library made for multi-host circuit design purposes was used to test circuit phenotypes for both *E. coli* and *P. putida* chassis [50]. The impact of backbones (**Figure 5B**) was insignificant, and the effect of intra-species context switch (**Figure 5C**) was limited, when compared to inter-species influence (**Figure 5D**) on dynamic behavior of the inverter performance among two chassis - emphasizing the importance of taking into account the chassis aware circuit designs.

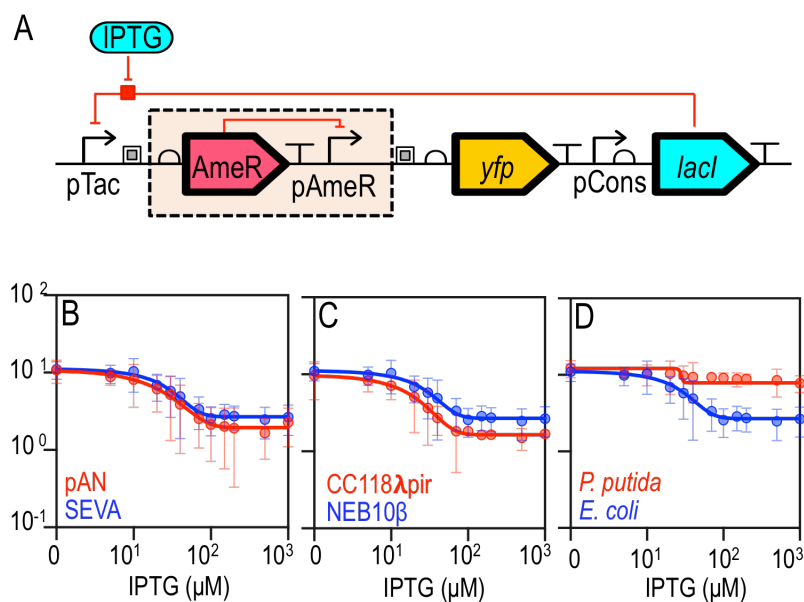


Figure 5: Parts and parameters

Gate schematic and comparison of parameter impacts. The strain used is *E. coli* NEB10 β and the inverter is AmeR-F1 NOT gate, unless otherwise is mentioned. Results are from at least 3 experimental repeats with 2 technical replicates. y -axis is normalized fluorescence value of YFP and x -axis is the molar concentration (μM) of the inducer IPTG. **A.** The schematics of gate characterization device. The gates used in this study are NOT gates based on promoter-repressor based expression. X represents any of the 20 gates. **B.** Response curve showing how plasmid backbone pAN (**red**) vs pSEVA (**blue**) affects the response given to the inducer. **C.** Response curve showing the Genotype effect between CC118 λ pir (**red**) vs NEB10 β (**blue**). **D.** Response curves showing the interspecies effect between *P. putida* KT2440 (**red**) vs *E. coli* NEB10 β (**blue**).

3.2. Modifying thresholds

Thresholding of the NOT gate response curves is required for the interpretation their input and output expression levels as Boolean signals. Thresholds discretize the continuous-valued input and output ranges into regions. Two of these regions

represent logical signals – expression levels that lie within these regions are interpreted as either digital ON or OFF. These regions are separated by a third, within which inputs and outputs are regarded as ambiguous, and which represents circuit failure.

The purpose of the ambiguous region is to adequately separate the ON and OFF signals, such that intrinsic or extrinsic noise cannot (to some confidence level) lead to misinterpretation of a ON or OFF signal, which is another failure mode of the circuit. The optimum size of the ambiguous region is therefore related to the signal-to-noise ratio at both ON and OFF expression levels. If signal-to-noise ratio differs between contexts, it may then be prudent to adapt the size of the ambiguous region accordingly. In our calculation of thresholds levels, which is based on the definition given in [4], we introduce a parameter, the ‘threshold multiplier’ (described below in Methods), which determines the size of the ambiguous region for a given NOT gate. Our *in silico* analysis of the effect of this parameter, shown in **Figure 6**, reveals that the number of compatible gate pairs can depend significantly on its value.

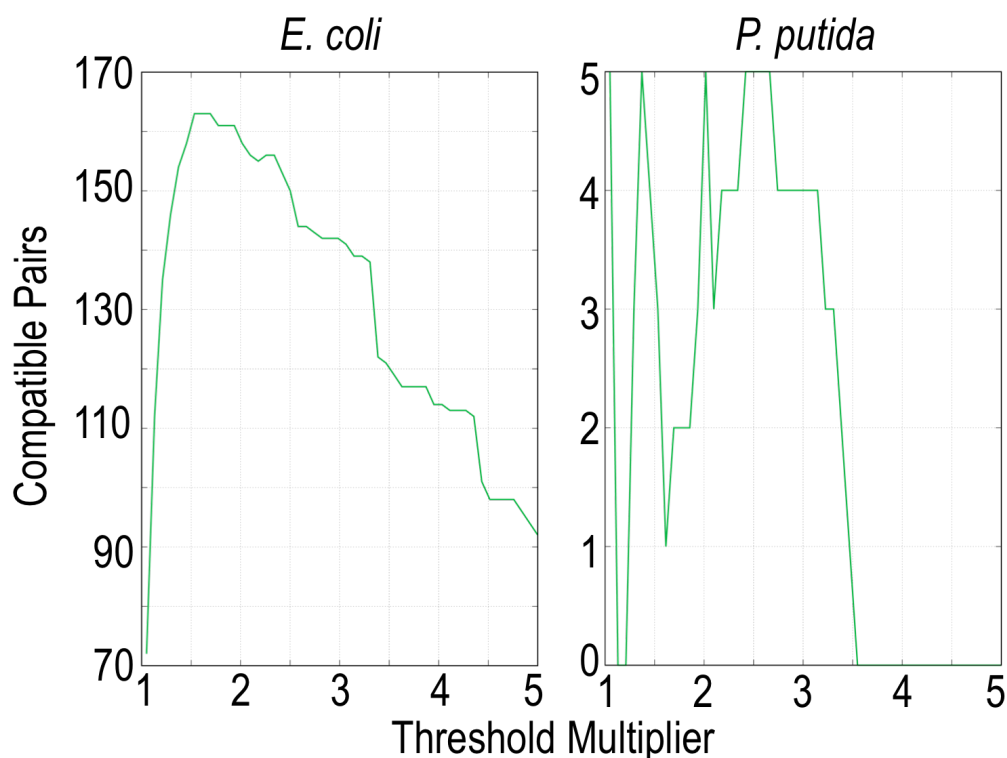


Figure 6: Threshold multiplier change and corresponding compatible gate pairs for *E. coli* and *P. putida*.

Threshold multiplier is changed gradually and compatible gate numbers are reflected. *E. coli* (left) results suggest many options for candidate circuit designs, whereas *P. putida* (right) outcomes are bound with less compatibility.

If signal-to-noise ratio differs between expression levels, it may also be appropriate to adjust the skewness of the ambiguous region. For example, if low expression levels (those corresponding to an OFF state) exhibit higher noise than high expression levels, the OFF region may be expanded to account for greater fluctuations in expression level in this regime, while the ON region may be contracted, since large fluctuations are not expected in ON expression levels. This situation is typical for the stochastic genetic processes employed in synthetic biological logic gates. For this case, our calculation of threshold levels introduces two parameters, an upper and lower ‘threshold multiplier’, which can be used to perform asymmetric thresholding. In silico, the maximum compatible gate pairs could potentially be increased from 5 to 10 using these asymmetric thresholds (**Figure 7**).

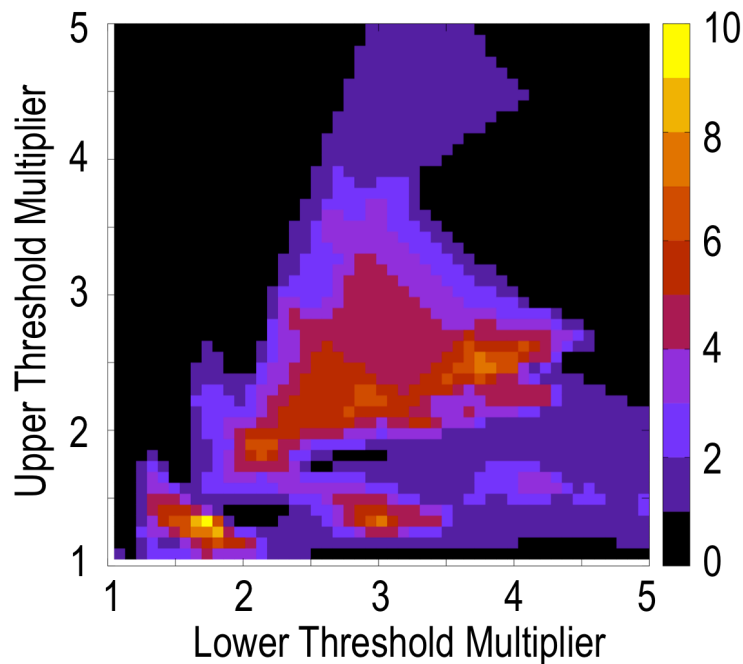


Figure 7: Asymmetric thresholds for maximum compatible pairs

An asymmetric threshold approach is applied to reach maximum compatible pairs. Heat-map shows the gradient in number of compatible pairs among the BHR library of 20 NOT gates without sacrificing inverter functions. Theoretically, maximum number that could be achieved is seen as 10.

3.3. Gate compatibility

Evaluation of gate responses in *P. putida* based on written rule set of Cello in *E. coli* [4] resulted in seeing 5 compatible gates (**Figure 8**), out of 362 combinations. This number is 162 in *E. coli* of Cello. Originally, the DNA circuitry parts used were optimized for *E. coli*. With the help of a BHR library, automated biosensor design

was tested in *P. putida* too. As a conclusion, a greatly limited set of pairs was obtained from which computational layers could be designed.

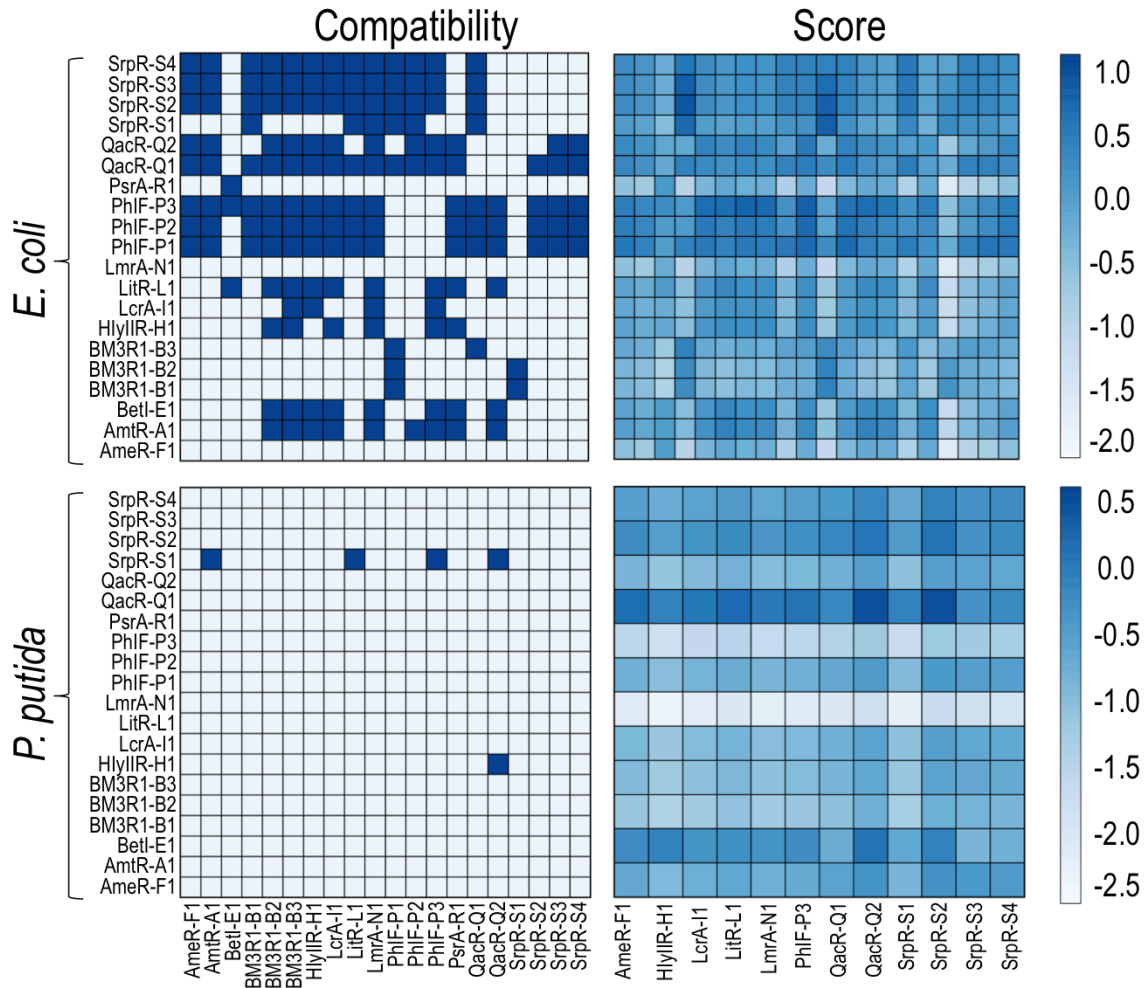


Figure 8: Pair compatibility and Scoring comparison for 20 gates.

Compatibility and scoring matrices given for *E. coli* and *P. putida*. Binary color code in compatibility matrices shows compatibility of pairs with each other (dark blue compatible, white incompatible). Number of compatible gates in *E. coli* is much greater than *P. putida* equivalent. Scoring shows how far gates are from being (in)compatible pairs to each other. A heatmap score >0 indicates compatible and <0 indicates not compatible. Magnitude of color code indicates how (in)compatible gates are.

Our scoring system attempts to identify candidate pairs that are likely to work when connected *in vivo*. We intended to use the circuit-scoring system of Cello study to score gate performances in *P. putida* (Figure 8, bottom right) in a comparison to *E. coli* (top right) counterpart. We noted that this system does not necessarily indicate compatibility (two incompatible gates may still obtain a higher score than two compatible ones), but rather, when two pairs of gates are both compatible, the score indicates which pair is 'more compatible' in sense. In a sample set that is composed

solely of compatible gates, this makes sense e.g. in Cello, where the scoring is applied to circuits that are by definition, constructed from compatible gates. However, this does not account for incompatible gates, which resulted in considering an approach that reflects how far away from being (in)compatible a pair of gates are. With a simple but novel approach (see Materials and Methods) we defined a new scoring metric holding compatibility information *per se*. **Figure 8** shows a comparison of compatibility and this new scoring metric applied to the library of gates in *E. coli* and *P. putida*. Compatibility scores so defined may assist to focus efforts of strain specific optimizations.

3.4. Portability of a NOR function

Characterization of 20 NOT gates to automate the circuit design process via CelloCAD was done in *P. putida*. In addition to 3 Cello promoters, host specific

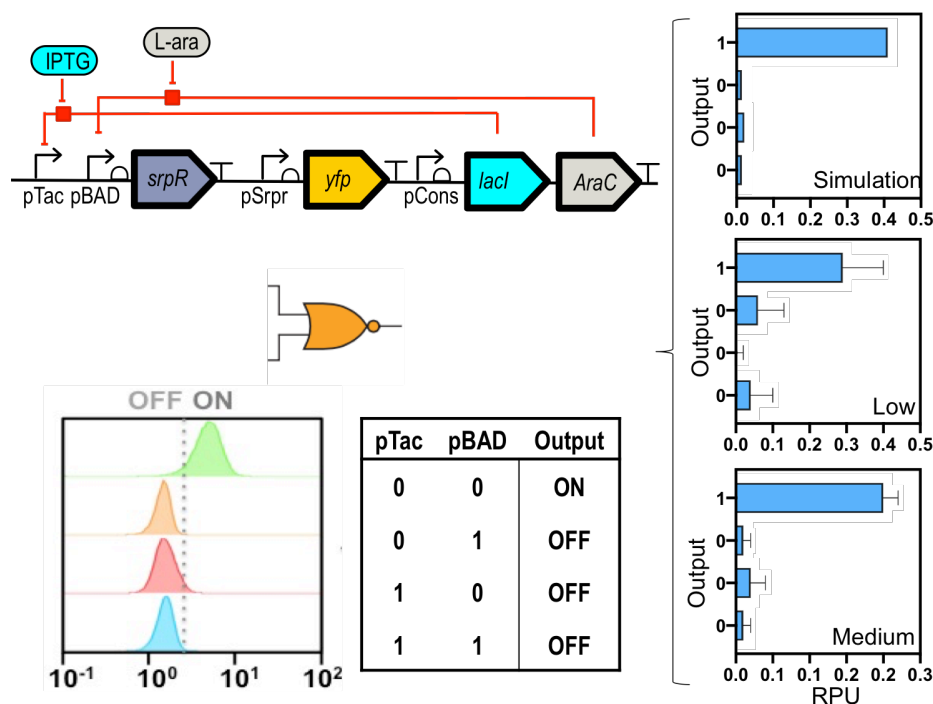



Figure 9: *In silico* to *in vivo* NOR gate application in *P. putida*

NOR gate application. Schematic representation shows IPTG and L-arabinose inducible circuit. Simulation indicates *in silico* values received from CelloCAD software and low copy and medium copy values show experimental results obtained from flow cytometer in *P. putida*.

promoters P_{alkB} and P_m were also characterized (see **Table S1**). Eventually, CelloCAD was used with experimental data derived from the new chassis, *P. putida*. A NOR gate design was retrieved from the program with a simulation of possible outcomes



(**Figure 9**). After synthesis of the NOR gate sequence suggested by the program the performance of low copy and medium copy versions were tested. Despite working with a limited design space in *P. putida*, the design of a NOR gate was achieved. P_{BAD} promoter that is not allowed being at the second position (downstream of another promoter) in Cello design due to roadblocking phenomenon [4] was used in this design as a succeeding promoter. That is, a circuit that would putatively not work in *E. coli* has shown good performance in another chassis.

4. Discussion

The motivation of this study is to prioritize understanding of portability. Here a circuit design automation library of NOT gates, originally optimized for an *E. coli* strain, was used for another chassis, *P. putida*. This comparison showed that the performance of genetic logic gates can depend heavily on the chassis organism. Although the DNA sequences of the genetic parts are the same, the observed results vary. That is, a design that would work in one chassis would not work in another one, and vice versa. It is also true that a circuit that may not perform its function in one chassis could execute it in another one. To take advantage of this, we calculated the impacts and designed a device for *P. putida*. This is a NOR gate device, and is predicted to not comply with roadblocking rules of *E. coli*, but shown to work in *P. putida*. That is, the function was portable, but not the device. This approach highlights that a change in our perspective towards portability of functions would benefit automation endeavors for general design purposes.

All data received from *P. putida* experiments were converted into RPU accordingly. Using standardized units NOT gate response functions were overlaid based on hill equations [4, 51]. This pool of data was used not only for compatibility analysis, but also for investigating how thresholding might affect the size of the design space. Consequently, relaxed and asymmetric thresholding showed that the number of compatible gates could potentially be increased. Data was encoded in a machine-readable JSON constraint file (UCF) made for *P. putida* and used for the cellocad.org design process.

The scoring system developed here assesses how compatible two gates are. This enabled a comparison among pairs of gates and assignment of a degree of

functionality, in addition to having a view of categorized library of parts for functional studies. In RPU units, a comparative promoter activity and dynamic range study was done. Regular Cello promoters were supplemented with two new promoters, P_m and P_{alkB} , originating from pseudomonad studies. The increased number of promoters could benefit automated designs by expanding the design space. Bistable gate function is central for operational NOT gates, which require a clear distinction between expression levels for each state. A steep transition between ON and OFF states ensures diminishing leakage upon receiving the signal. Thresholding therefore is another key to well-defined circuit outputs. With calculated adjustments, classical thresholding approach was challenged here and as a result over 200% increase in compatible gate pairs was achieved.

Comparison of circuit simulation of NOR gate received from CelloCAD and experimental results indicated that there is a close proximity. This supports the idea that genetic background of this chassis is less likely to cause unpredicted impacts over circuitry implementations. This eases applying results from *in silico* calculations into the chassis and increases the chances to overlay design outputs with real life applications. Using an automated sensor design tool for implementation of a NOR gate design in *P. putida* showed that as a chassis *P. putida* is ready to be counted in for complex cellular computation designs.

A library of BHR inverters, primed to implement quick automation studies for biosensor design, was used in gram-negative *P. putida*. A limited number of gate pairs were identified as compatible. This was enough to build a NOR gate - a fundamental building block of complex computational layers. CelloCAD was used for the design purposes of the NOR gate. Even though the DNA sequences of parts were originally designed for *E. coli*, the Cello-given design was not suitable to perform in *E. coli* due to the roadblocking phenomenon. The very same sequence performed a NOR gate function in *P. putida*. That is, a circuit may not work in one chassis, yet may execute in another chassis, concluding that the portability of functions.



5. Materials & Methods

5.1. Library of DNA gates.

Library of DNA gates was borrowed from the study at Tas et al., 2020 [50]. The gate collection used in this study contains the pSEVA221-based NOT logic gate collection from the mentioned study composed of 20 NOT gates together with 3 plasmids to calculate autofluorescence, standardization and promoter activity. In order to characterize more promoters in *P. putida* new calibration plasmids were prepared and a complete list of DNA constructs used in this study is shared at supplementary material **Table S6**. All the constructs are reserved at SEVA bank at CNB-CSIC, Madrid, Spain and ready for distribution for research purposes.

5.2. Experimental procedure.

Experimental procedure was adapted based on the procedure Tas et al., 2020 [51]. O/N liquid cultures with corresponding M9 minimal medium were inoculated from freshly prepared LB-agar plates. Following this, cultures were used for inoculation of 96 well plates with corresponding inducer concentrations and to the final of 666 times dilution. As the cultures reached to a late exponential stage they were removed from the air shaker for flow cytometer measurements. Throughout the measurements plates were kept on cold platform to halt growth.

5.3. Flow Cytometry data calibration.

Fluorescence and scattering values are correlated to filter possible variability based on Stoof *at al* procedure [53]. Fluorescence is considered in two sub-part: scattering influenced and experimentally defined. Context average is used for evaluating scattering parts in order to succeed a comparativeness among experiments (see Annex).

5.4. Thresholding.

The four thresholds, OL, OH, IL and IH were obtained according to the following formulae:

$$OL = y_{min} \times t_{lower}$$

$$OH = \frac{y_{max}}{t_{upper}}$$

$$IL = \left(\frac{k^n y_{max} (t_{lower} - 1)}{y_{max} - t_{lower} y_{min}} \right)^{1/n}$$

$$IH = \left(\frac{k^n (y_{max} - t_{lower} y_{min})}{t_{lower} y_{min} - y_{min}} \right)^{1/n}$$

Where y_{max} , y_{min} , n and k are parameters obtained by fitting the hill equation to the experimental data for each gate, and where t_{lower} and t_{upper} are the ‘threshold multipliers’. The interpretation of these multipliers is straightforward for the output thresholds OL and OH. For input thresholds IL and IH, these multipliers set the lower (upper) thresholds as the input values that produce outputs that are multiples (fractions) of $y_{min}(y_{max})$. E.g. if $t_{lower}(t_{upper})$ is 2, then IL (IH) is the input value that would produce an output that is double (half) the value of $y_{min}(y_{max})$.

5.5. Scoring system.

Two NOT gates A and B are considered compatible if $OL_A < IL_B$ and $OH_A > IH_B$, meaning that the definition of A’s output states agrees with the definitions of B’s input states. The compatibility score s , used in this study is based upon the margin of error in these definitions, and is defined as:

$$s = \min\left(\ln\left(\frac{IL_B}{OL_A}\right), \ln\left(\frac{OH_A}{IH_B}\right)\right)$$

This metric is positive for compatible pairs and negative otherwise. Compatibility is interpreted here as the ability to connect the output of A to the input of B, with the result that if the input to A is ON (OFF), then the output of B is also ON (OFF).

V. CHAPTER 3

CONTEXTUAL DEPENDENCIES EXPAND THE RE-USABILITY OF GENETIC INVERTERS

V. CHAPTER 3. Contextual dependencies expand the re-usability of genetic inverters

*The content of this chapter has been published as:

Huseyin Tas, Lewis Grozinger, Ruud Stoof, Víctor de Lorenzo and Ángel Goñi-Moreno (2020) Contextual dependencies expand the re-usability of genetic inverters, at BioRxiv.org with the following link:

<https://doi.org/10.1101/2020.07.15.204651>

1. Abstract

The design and implementation of Boolean logic functions in living cells has become a very active field within synthetic biology. By controlling networks of regulatory proteins, novel genetic circuits are engineered to generate predefined output responses. Although many current implementations focus solely on the genetic components of the circuit, the host context in which the circuit performs is crucial for its outcome. Here, we characterize 20 genetic NOT logic gates (inverters) in up to 7 bacterial-based contexts each, to finally generate 135 different functions. The contexts we focus on are particular combinations of four plasmid backbones and three hosts, two *Escherichia coli* and one *Pseudomonas putida* strains. Each NOT logic gate shows seven different logic behaviors, depending on the context. That is, gates can be reconfigured to fit response requirements by changing only contextual parameters. Computational analysis shows that this range of behaviors improves the compatibility between gates, because there are considerably more possibilities for combination than when considering a unique function per genetic construct. Finally, we address the issue of interoperability and portability by measuring, scoring, and comparing gate performance across contexts. Rather than being a limitation, we argue that the effect of the genetic background on synthetic constructs expand the scope of the functions that can be engineered in complex cellular environments, and advocate for considering context as a fundamental design parameter for synthetic biology.



2. Introduction

The abstraction of gene regulatory signals into on (high) and off (low) values allows for the design and implementation of genetic Boolean circuits [4] inspired by digital electronics. Such devices result from assembling two or more genetic logic gates [4, 11]—the basic unit for processing information in genetic circuits based on Boolean logic. A core objective of Synthetic Biology is the building of new regulatory circuits to compute inputs into outputs according to predefined logical functions [12], which are then used in a number of applications, ranging from bioproduction [14] to medical diagnosis [13]. Although this approach has been relatively successful, genetic logic gates are way more fragile and less reliable than their electronic counterparts as their signals are rarely constant and often fluctuate over time. Consequently, the large-scale control of gene regulation based on Boolean logic alone is challenging. The central underlying issue is that a number of features intrinsic to biological systems, such as gene expression noise, analogue signaling and evolutionary dynamics, make the intracellular environment an unsuitable domain for engineering idealized Boolean logic [18].

A fundamental challenge for the design of robust synthetic circuits, which underpins this work, is the oversimplified model that DNA elements i.e. gates alone explains the performance of genetic circuits. Based on this assumption, the host chassis (the cell that receives a specific genetic construct) is generally ignored and the interplay of a genetic circuit with the host context is most often overlooked—an issue that has been identified essential for the predictability of synthetic biology devices [19]. Both the burden imposed by synthetic constructs on the host [21, 22] and the impact of context on genetic activity [20], have phenotypic implications that cannot be predicted from a gene-centric standpoint. Recently, the term host-awareness [48, 54] has been coined to bring attention to this problem, which is at the core of the lack of part interoperability [23] (i.e., parts that show similar performance in different host contexts).

While most synthetic biology efforts make use of only one host chassis to develop and characterize genetic constructs, potential applications may require the same genetic devices to work with different cell types [24]. For instance, circuit-constructs

optimised in *Escherichia coli* for rapid prototyping, might be implanted into *Pseudomonas putida* for a bioremediation application [15] or into *Geobacter sulfurreducens* for bioelectricity production [55]. However, circuit performance will likely differ in different chassis, gene dosages and vectors, highlighting the importance of context in host-circuit design. As a result, the performance of a given genetic logic device would not only be a consequence of its DNA sequence but also would be influenced by its context. Within this scenario, modifying the context could fine-tune the performance of logic gates, thus engineering reconfigurable genetic logic devices which share the same sequences but exhibit different behaviors [25]. In the work presented below we inspect these scenarios by analysing quantitatively the behaviour of a collection of genetic inverters in different strains of the same species, in other species and in either case with the same devices borne by low, medium and high-copy number vectors. The results expose that playing with these biological backgrounds expand the range of parameters that rule the behaviour of each construct. On this basis, we entertain that context-variability should be considered an advantage for circuit design rather than being seen as problematic.

3. Results

3.1. Generation of gate-context libraries.

To generate enough data on the contextual dependencies of genetic inverters we made use of 20 NOT logic gates assembled with a suite of promoters and repressors first developed as components of the CELLO platform for *E. coli* [4] and then re-cloned in broad host range vectors of different copy numbers for delivery to different types Gram-negative hosts [50]. The logic function (NOT or inverter) corresponds to a genetic device that expresses a target gene (output high) if it is not negatively regulated (input low) and vice-versa. The inverters used are pairs of a specific regulator (repressor) and its cognate promoter (**Figure 10A**). The characterized transfer functions measured the impact on promoter activity (output; captured by the expression level of an *ypf* reporter fused downstream) generated by specific concentration of regulator (input). In order to manipulate the expression level of the regulator, its coding sequence was placed under the control of a *lac* promoter, which was externally induced by IPTG. In order to characterise the gates, the corresponding were entered in the bacterial host, which was then used to measure the NOT function

(**Figure 10B**). Both the IPTG concentration and the output *yfp* fluorescence were converted to relative promoter units (RPUs) to standardise the characterisations. The reference dataset re the behavior of the 20 gates under inspection in *E. coli* NEB10 β —12 main gates plus 8 variants—was retrieved from Nielsen et al. [4] (**Table S5**).

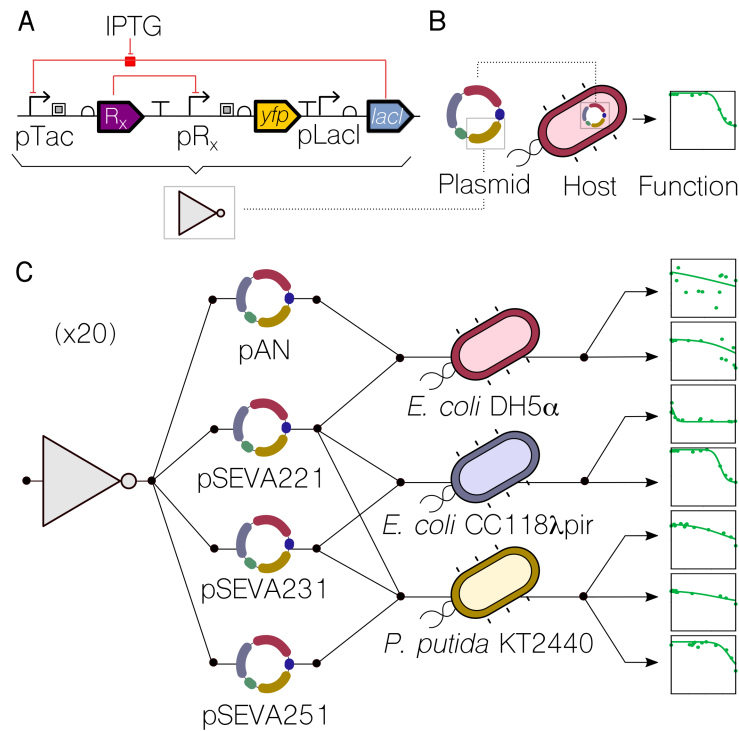


Figure 10: Generating a library of gate-context devices.

A. Genetic inverters (NOT logic gates) were placed in between the pTac/LacI system (the input) and the *yfp* gene (the output). Key components are a repressor (Rx) and its cognate promoter (pRx). **B.** For a genetic construct to be measured, it needs to be cloned into a plasmid which, in turn, is transformed into a host cell—thus using a single context. **C.** In this work, each genetic inverter (from an initial library of 20 gates) was measured in a number of context setups. These setups were based on combinations of 2 plasmid backbones (pAN and pSEVA), one of which with 3 different origins of replication—RK2 (221), pBBR1 (231), RFS1010 (251)—and 3 different hosts (*E. coli* DH5 α , *E. coli* CC118 λ pir, *P. putida* KT2440). The performance of the resulting 135 gate-context devices was characterized experimentally by using flow cytometry and analyzed computationally to find the impact of contextual dependencies on inverter’s behavior.

To assess the impact of the host context on gate performance, both the plasmid backbone and the cellular chassis were changed. As far as the carrier backbone is concerned, gates were cloned into the pAN and pSEVA [44] backbones, considering different origins of replication that led to low (RK2, pSEVA221), medium (pBBR1, pSEVA231) and high (RFS1010, pSEVA251) copy numbers. This contextual feature accounted for dynamics generated by circuit burden, since more copies of the same

gate would increase the cost (of running it) to the cellular machinery. Regarding the chassis, we used two *Escherichia coli* (DH5 α and CC118 λ pir) strains and one *Pseudomonas putida* (KT2440) strain. Combinations of these resulted in a library of gate-backbone-host devices (**Figure 10C**) where the final performance cannot be explained by the genetics of the NOT logic gate alone. That is, the DNA sequence of the constructs is not enough to predict the behavior of the gate—information about the context is then essential for understanding the genotype-to-phenotype dynamics. As shown in **Figure 10C**, each logic gate in this study can have up to 7 context-dependent dynamic behaviors, some of which differ significantly. Specifically plots shown in **Figure 10C** correspond to the characterizations of gate PhlF (one of the 20 gates of the initial library) in seven different contexts. While the performance changes abruptly in some cases (e.g., in contexts 3 and 4), it did not change significantly in others (e.g. in contexts 5 and 6), suggesting that contextual dependencies act as a hidden layer of parameters that must be carefully considered to achieve a predictable logic gate design—an issue which has been traditionally overlooked.

3.2. Effects of cross-context portability

When a genetic logic gate is either passed onto another organism, or carried by a different backbone, the interplay between itself and the context changes [56]. Contextual dependencies are adjusted. These modifications alter the expression levels of a gate, its dynamic range and (in some cases) its logic function. Moreover, context-dependent changes of qualitative behavior imply that the dynamics of the interplay between context and construct are nonlinear. That is, a given pair of gates may suffer similar modifications in one context but very different in another. For example, PsrA-R1 and PhlF-P2 show these effects (**Figure 11A**). When both gates are hosted by chassis *E. coli* DH5 α , the backbone (either pAN or pSEVA221) seems to play a key role in the logic outcome of PhlF-P2, which becomes more step-like with pSEVA221 (i.e. sharper transition from on to off). In contrast to this, gate PsrA-R1 does not follow that trend and remains qualitatively unchanged, although absolute expression values drop. Using the same backbone (pSEVA221) we then tested the context impact of varying the *E. coli* strain. Whilst the performance of PhlF-P2 is qualitatively the same (with smaller dynamic range), PsrA-R1 shows a qualitative change, becoming more step-like, thus showing more desirable behavior than in other contexts. These inconsistencies in changes of qualitative behavior of gates highlight the difficulty of

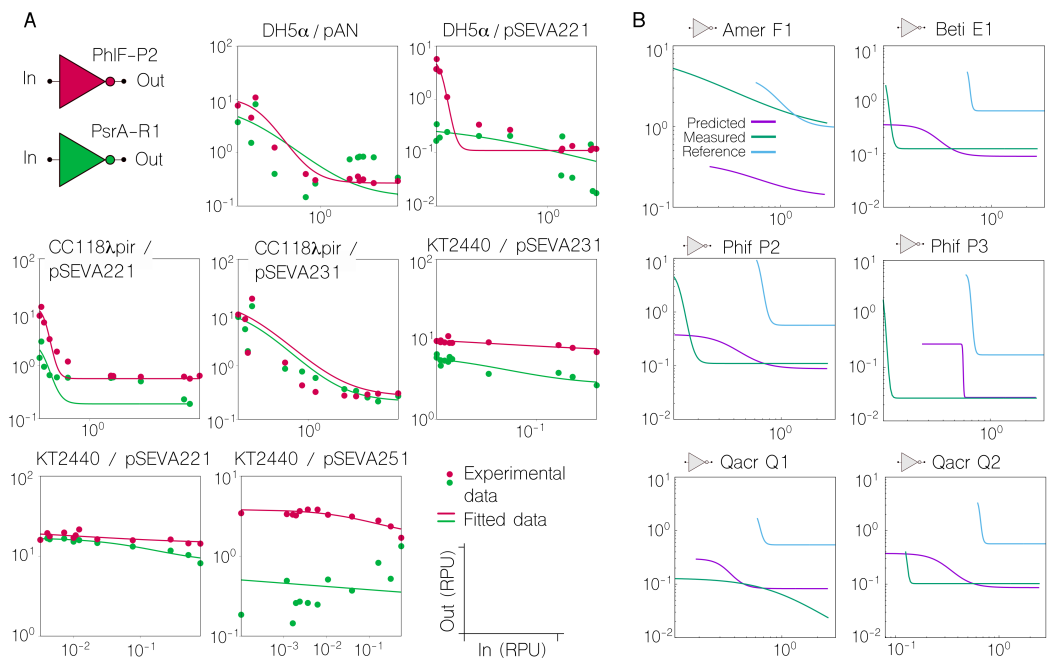


Figure 11: Non-linear effects in the cross-context portability of inverters.

A. Plots comparing the characterization of two gates, PsrA-R1 and PhIF-P2, in different contexts. As well as each individual characterization is differing across contexts, the relationship between the two characterizations also differs, depending on both strain and plasmid. That is, some contextual changes impact on a similar way on the performance of two inverters, while others impact on a different way—what we refer to as non-linear modifications. **B.** Non-linearities made the prediction of gate performance changes between contexts an overarching challenge. Predictions were made for gates in the DH5 α ::pSeva221 context ('Predicted' line), based on their characterizations in CC118 λ pir::pSeva221 ('Reference' line). The actual characterization of the gate is shown for comparison ('Measured' line). The gate upon which these predictions are based is AmtR-A1. It can be seen that the linear transformation used for prediction cannot accurately produce all the differences in behavior seen between contexts. For example, although translations in the Input and Output axis appear to be predicted well in some cases (see for example Qacr-Q2), more qualitative changes in the shape of the response curve cannot be addressed by this linear transformation (see for example Qacr-Q1).


compensating for such effects in order to engineer context-independent circuits [56]. However, there are also more predictable contextual changes in which that strategy may work well. For example, when both gates are hosted by *E. coli* CC118 λ pir, changing the backbone from pSEVA221 to pSEVA231 (that only differ in the origin of replication) generates almost the exact same phenotypic modification. Finally, a marked difference occurs when the gates are hosted by *P. putida* KT2240. In these contexts, the gates lose their NOT logic, regardless of choice of backbone (pSEVA221, pSEVA231 and pSEVA251). The characterization of the full library (20 gates) is shown in Annex **Table S6**.

As a result of this non-linear performance in cross-context experiments, the issue of inter-context predictions arose as a formidable challenge. For example, an attempt to predict the performance that a number of gates would display into the context *E. coli* DH5 α (pSEVA221)—host and plasmid—could not match all experimental data, revealing differences where the non-linear patterns were stronger (**Figure 11B**). For this prediction, we took as a reference the performance of the same gates into the context *E. coli* CC118 λ pir (pSEVA221) i.e. only change in host cell, and the measurement of only one of the gates (AmtR-A1) into the new context *E. coli* DH5 α (pSEVA221). With the information provided by the characterization of AmtR-A1 in both context setups, we applied linear transformations to predict the performance of the rest of the gates into the new scenario. As expected, some of the gates showed a relatively good prediction (good candidates for portability applications), but that was not the case for all of the constructs. Although predictable gate portability is then highlighted as an open problem, contextual dependencies offer a unique opportunity for fine-tuning gate performance, which we carefully analysed as explained next.

3.3. Enhanced gate compatibility by fine-tuning contextual dependencies.

Building a genetic circuit by coupling genetic logic gates requires an assessment of their compatibility, to determine which gates can be connected. In order to connect two gates, the output levels of one must match the input levels for the other. If not, it may result in failure of the overall circuit logic [4, 57, 58]. This is one of the major bottlenecks that restrict the depth of genetic logic circuits and limit scalability, since not every gate within a library will be compatible with another. The analysis of inter-gate compatibility is therefore fundamental for circuit design and is an integral part of current synthetic biology Computer Aided Design tools [4, 59]. However, knowledge about the effect of context on gate compatibility has until now been lacking.

In order to tackle this contextual issue we first scored the matching of all the gate pairs in the library according to their input and output thresholds (**Figure 12A**). The inclusion of the input thresholds in the output ones defines a pair as “compatible”; each pair of gates in the library was categorized as such (see Methods for details of this calculation). Moreover, the information provided by the compatibility was complemented by the introduction of a similarity score (**Figure 12B**). While the former relates two different gates, the latter relates the same gate to itself when



varying contextual dependencies. This score quantifies the impact of specific context variations on each gate.

In this analysis, constructs were considered as gate-context entities (e.g. *E. coli* DH5 α (pAN::PsrA-R1) or *E. coli* DH5 α (pSEVA221::PsrA-R1), rather than individual gates alone (e.g. PsrA-R1) so that results account for the performance of a gate in a given context. We consider that high numbers of compatible pairs in a library is a desirable trait, and examine the impact of the two contextual features we focus on (backbone and host) on this metric, both independently and together. **Figure 12C** (left) shows the results of the former, where the compatibility is assessed in the case of gates sharing backbone pAN and host *E. coli* DH5 α . In this scenario, the number of compatible pairs found within the library was 67. However, when we allowed the calculation algorithm to consider all possible backbone options (while keeping strain constant), the number of compatible pairs increases to 203 (a 303% increase, **Figure 12C** middle). We conclude that consideration of backbone as a design parameter results in a more flexible, and reconfigurable, library with the ability to include dynamics that are not captured by just DNA sequences e.g., the copy number of circuits (thus their burden to the cell).

Gate compatibility was enhanced further by allowing the algorithm to consider the host as a design parameter, as well as backbones (**Figure 12C** right). Although the number of compatible pairs relative to the total number of pairings being evaluated remained almost constant at roughly 2%, the absolute number increased to 697. Due to the consideration of contextual dependencies, the original gate library of only 20 genetic devices was increased to 135 different functions. Therefore, there were many more options to evaluate and more compatible pairs found. However, some of these pairs correspond to gates that are compatible only if they are inside different hosts. For example, the gates HIyIIR-H1 and AmeR-F1 can only be matched (i.e., their function is complementary) if the former is hosted by *E. coli* DH5 α and the latter by *E. coli* CC118 λ pir. This suggests that taking a multicellular (distributed) computing approach [60-62] may be a suitable strategy if coupling the functions of these two constructs. In multicellular computations, a predefined function is distributed across different engineered bacterial strains (or species), which are connected in such a way that the output of one cell is the input of another one. Therefore, considering the host

of a genetic construct within circuit design will allow for building both intra- and inter-cellular computations.

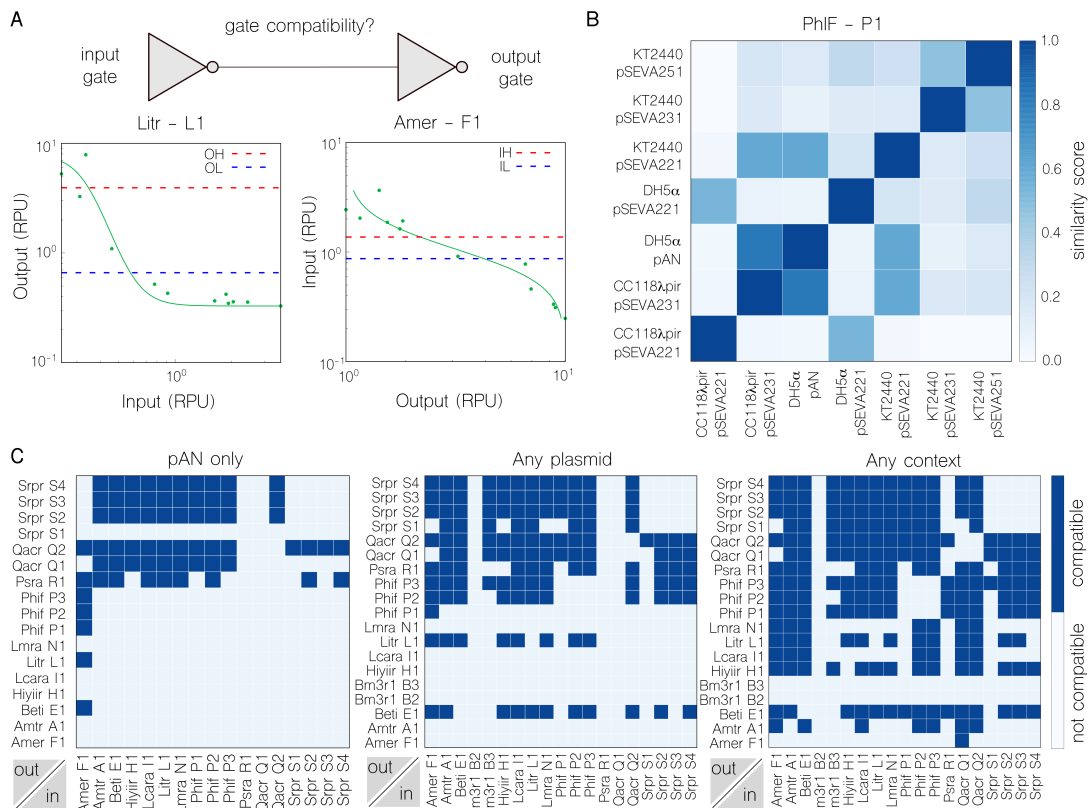


Figure 12: Comparing inverter compatibility and similarity across contexts.

A. Gate compatibility indicates if two gates can be sequentially assembled—the output of the first gate is compatible with the input of the second—or not. Since the IH (Input High) and IL (Input Low) thresholds of the output gate, Amer-F1, lie between the OH (Output High) and OL (Output Low) thresholds of the input gate, LitR-L1, this pairing is compatible. **B.** A heatmap of similarity scores (which refers to how similar the shape of both inverter’s transfer function is) calculated using discrete Fréchet distance between the characterization of PhIF-P1 in each of the seven contexts (darker is more similar). Most values within the score scale are covered, which highlights context contribution to final gate behavior. **C.** Maps of compatible pairs for the gates characterized in: the strain DH5 α with pAN as the only plasmid for all inverters (left), the strain DH5 α with any variation in plasmid type (middle) and in any context choice (right). Considerably more compatible pairs are found when freedom is given in the choice of backbone, rising from 67 (left) to 203 (middle) pairs. The freedom to use both backbone and strain (right) as a design parameter yields the most compatible pairs (697) and maximum utilisation of gate combinations in the library (~68%).

3.4. Context-aware design rules for layered logic gates

The design of synthetic genetic circuits typically overlooks contextual features by considering that phenotypic performance can be explained by the DNA sequence of the synthetic construct alone. However, this over-simplification has negative implications; for example, it requires considerable effort to adapt a genetic circuit to a

new host [63]. The fact that genetic constructs show different dynamics depending on their context is not necessarily a disadvantage for pre-defined circuit design—could we rationally use such variability? To begin to address this question, we carried out computations in order to identify the circuit depth (i.e., number of layers) that could be achieved by connecting gates within our library, and assessed the impact of contextual effects in such a chain (**Figure 13**). First, when considering all gates in the

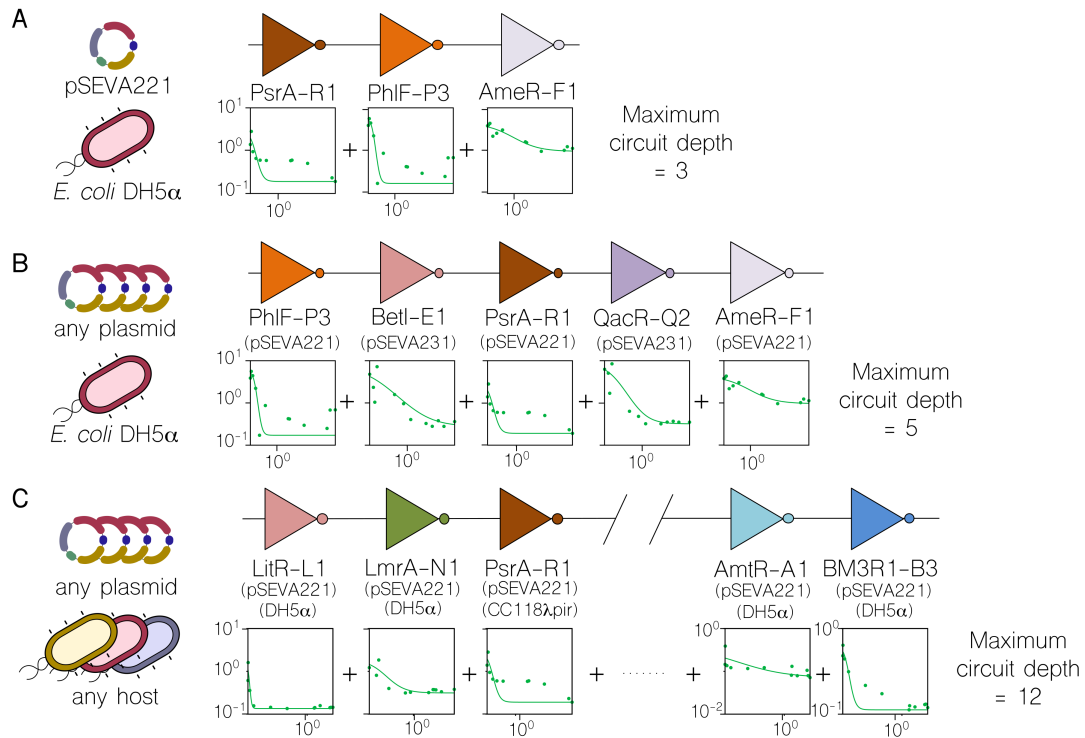


Figure 13: Calculation of maximum circuit depth as a result of layering inverters.

Based on the compatibility between gates, these were layered within the library in order to evaluate the impact of contextual dependencies on circuit size. **A.** The maximum depth calculated when the computational method is forced to consider all gates carried by the low copy-number plasmid pSEVA221 and hosted by *Escherichia coli* CC118 λ pir, is 3 gates-deep. **B.** If the algorithm is free to select any plasmid (but still forced to CC118 λ pir), the maximum depth increases to 5. In this scenario, two gates are carried by the medium copy-number plasmid pSEVA231. **C.** In the last analysis, the calculation used all contextual dependencies, including the variation in host chassis. The maximum number of gates layered increases to 12 (only 5 shown in figure—refer to Annex for more information). In the sketch shown in the figure, 4 out of the 5 gates were characterized in the strain *Escherichia coli* DH5 α . For all graphs: x-axis refers to the input and y-axis to the output (both RPU).


same context, with backbone pSEVA221 and hosted by *E. coli* CC118 λ pir, the maximum depth was 3 (**Figure 13A**). That is, there are 3 gates that can be connected consecutively while maintaining the correct logic output (i.e., logic values 0/1 are effectively transmitted from beginning to end). Every other valid configuration will result in fewer (or the same) number of layers. We find that increasing the number of

contexts available can significantly increase the maximum depth computed by the search algorithm. As shown in **Figure 13B**, allowing another context by including gates characterized with any backbone (but still hosted by *E. coli* CC118 λ pir) increases the maximum depth to 5. This can be further improved upon by allowing freedom in the choice of host, for a total of 7 contexts (**Figure 13C**). In this case, the computed maximum depth is 11 — a circuit depth that is far beyond the current state-of-the-art for synthetic circuitry [4].

4. Discussion

A fundamental driving force for synthetic biology [64, 65] is the clarification of mechanistic assumptions as our understanding of molecular processes increases, which allows scientists to add novel tools to the catalogue for engineering living systems. Although the cellular environment consists of much more than DNA, circuit design [4, 59] typically revolves around genetic elements (promoters, terminators, RBSs...) in order to link genotype to phenotype—an over-simplified reductionist approach. The comfortable, yet error-prone, assumption that engineered parts alone can ultimately explain phenotypic performance needs to be expanded upon [19]. This leads us to consider what has been termed genetic background [20] and host-aware [48] dynamics: cellular features and constraints that have an impact on circuit performance but are not captured by the DNA sequences of the construct. In recent years, several of these features have been analyzed: the impact of having limited cellular resources [52, 66] (e.g. ribosomes) to “spend” on synthetic constructs, the effects of placing DNA parts in different genomic locations [67, 68], the role played by metabolism in genetic control [69, 70], or even genetic stability [71] due to evolution over time. All these effects turn the portability of genetic circuits into an overarching challenge—the fine-tuning of a circuit to work inside a different host (to the one it was originally built-in) is still a major task [56]. Furthermore, it limits the scope of biological circuits by solely using a DNA-insert toolbox for designing circuits.

Here, we use the word “context” to refer to the molecular background of the cell beyond genes and analyze how such context can be used for improving biocircuit design. By “dependencies”, we mean the constraints imposed by the context on a given genetic construct. Therefore, genetic logic gates are exposed to contextual



dependencies that influence their phenotypic behavior. Although synthetic biology is a field full of metaphors [72] already, we entertain here a new one that we consider to provide a useful conceptual frame: the use of contextual dependencies as in a software engineering problem. Any piece of software, or program, must run inside a specific environment (e.g. operating system) and software engineers usually face the problem of adapting it to the particular dependencies of the environment/context at stake. Under this metaphor, genetic circuits are considered software (instead of hardware [73]) whose performance is deeply linked to context-specific dependencies, which can allow designers to access functions that could not be coded otherwise. In this paper, we propose that contextual dependencies are important parameters for circuit design, and focus on [i] backbone carrying the construct, and [ii] cellular host in which the construct performs.

In this work we exploited a library of 20 genetic inverters (NOT logic gates), which are combined with 4 different backbones and 3 cellular strains to give a total of 135 gate-context constructs. In this regard, the number of functions exposed by the library increases by 675% due to the addition of these two contextual dependencies. With this new library we carried out experiments in order to assess the implications of adding context to the context-free initial collection of NOT gates. First, the characterization of the constructs showed how gate behavior changed across contexts in a nonlinear fashion. That is, the phenotypic modifications in the performance of one gate across two contexts may not match those of another gate under the same contextual transformations. This has major implications for the portability of genetic devices, since not all genetic components may be affected in the same way upon host change—thus building complex portable devices will become difficult (if not entirely impossible). Second, our experiments suggested that the compatibility of gates (so that they could be composed; the output of the first being the input of the second) does not only depend on selected genetic inserts, but also on their context. While only 67 compatible pairs were found in the original library of 20 inverters, the number increased to 697 in the new library. For instance, by allowing gates to be carried by 4 different backbones, the computational algorithm was able to evaluate the compatibility of 4 functions instead of 1 and return not only the name of the compatible gate but also the name of the backbone to use for carrying it. This allows reconfiguration of genetic constructs, since the same piece of DNA-insert can have

different behaviors depending on rationally selected contextual dependencies. Finally, the use of the cellular host as a separate design parameter allowed identification of gate pairs that were only compatible if connected gates were located in different strains/species. This suggests that multicellular computing approaches [60-62, 74] may be best suited for a given set of functions, and establishes rational criteria for the selection of cellular chassis in such distributed consortia from a bottom-up design.

In a similar way to living systems that use a number of mechanisms to go from genotype to phenotype, we advocate for the development of genetic circuits by considering whole-cell dynamics—including contextual dependencies. This will result in the design of biological circuits that are closer to the internal workings of natural systems—therefore more robust, reliable, predictable and reproducible.


5. Materials and Methods

5.1. DNA and strain construction.

All cloning steps were done in *E. coli* CC118 λ pir. Primers are ordered from Merck Sigma Aldrich, Inc. The repressible and inducible systems were previously described by Voigt Lab and acquired by the courtesy of Voigt Laboratory in MIT (USA). Description of the 20 different NOT gates moved into broad host range pSEVA backbones are described in [50]. Components of the original inverters, like terminators, RBSs, insulators, etc. were kept the same during the SEVA conversion. SEVA backbones have two terminators, T0 and T1 which are important to lower potential leakages. Required oligo list can be found in the Annex (**Table S3**). The pAN backbone [4] has a kanamycin resistance gene with a p15A origin of replication which is ~ 15 copy number in *E. coli* NEB10 β strain.

5.2. Medium and experimental protocols.

In all experiments (unless stated otherwise) M9 minimal medium for *E. coli* and M9 medium for *P. putida* were used. The ingredients of the M9 medium used are as following: for 250 ml of liquid medium, 25 ml 10X M9 salts, 500 μ l of 1M MgSO₄, 2.2 ml of 20% carbon source (glucose for *E. coli* and citrate for *P. putida*), 125 μ l of 1% Thiamine, 2.5 ml of 1% Casaminoacids and milliQ-H₂O up to 250 ml. Concentration of kanamycin used is 50 μ g ml⁻¹ in the experimentation procedures.



IPTG was used as inducer for pTac/LacI inducible system in 12 different concentrations diluted from 1M stock concentration that are 0, 5, 10, 20, 30, 40, 50, 70, 100, 150, 200, 500 and 1000 μ M. For synchronizing the cells in the experimental procedure, cultures are started from a single colony picked from LB agar plate which is each time freshly prepared from -80C glycerol stock by inoculating it O/N in 1 ml M9 minimal medium. O/N cultures after saturation were diluted by \sim 666 times to inoculate 200 μ l M9 minimal medium in 96 well plate for 24h, which is enough to reach to 0.2 - 0.3 OD in 96 well plate after which for halting the growth cells were kept on cold platform during the measurements.

5.3. Flow cytometry analysis.

Miltenyi Biotec MACS flow cytometer at channel B1 with an excitation of 488 nm and emission of 525/50 nm was used for measuring YFP fluorescence distribution of each sample. 30000 events were defined as the statistically sufficient amount under singlet gating for each sample. Calibration of the flow cytometer was done daily by using MACSQuant Calibration Beads. Throughout flow cytometer measurements samples were always kept on cold 96 well plate platforms. For the analysis of the data FlowJo software was used. In the analysis, gating was done via usage of auto-option and allowing to cover at least 50% of the whole events run while Forward and Side scatters were plotted, and the same gating conditions were kept for all samples in the same group.

5.4. Fluorescence data pre-filtered by cell size.

In order to unify fluorescence measures between and within flow cytometry experiments, we analyzed fluorescence and scattering values. Variation in cell size across experiments showed that median fluorescence values were decisively affected, therefore inaccurate for the sake of comparison. To compare between experiments, we took the distribution of fluorescence for single scattering values. A full description of this process is detailed in Annex Figure S 1.

5.5. Standard fluorescence measurements.

Two extra plasmids were used for measurements, the autofluorescence plasmid (Backbone::1201), and the reference standard plasmid (Backbone::1717) that triggers

yfp expression under the pLacI constitutive promoter. In order to derive reference promoter units (RPU), the following equation was applied:

Equation 3:

$$RPU = \langle YFP \rangle - \frac{\langle YFP \rangle_{\text{autofluorescence}}}{\langle YFP \rangle_{\text{standardization}} - \langle YFP \rangle_{\text{autofluorescence}}}$$

$\langle YPF \rangle$ stands for the median fluorescence value of the inverter that is to be standardized into RPU, $\langle YPF \rangle_{\text{autofluorescence}}$ is the median fluorescence value of the auto-fluorescence plasmid, $\langle YPF \rangle_{\text{standardization}}$ indicates the median fluorescence value from the standardization plasmid. RPU values were calculated in transfer function plots for different data points using at least 6 inducer levels covering the range of induction up to saturation.

5.6. Data fitting.

The pre-filtered experimental data were fitted to a 4-parameter hill equation of the form:

Equation 4:

$$h(x) = y_{\min} + \frac{(y_{\max} - y_{\min}) k^n}{k^n + x^n}$$


The parameter values for y_{\min} and y_{\max} were set to the minimum and maximum of the corrected experimental data. The values for k and n were then fitted using the least squares method from the `scipy.optimize` Python package [75], with logarithmic residuals.

5.7. Calculating compatibility.

Thresholds OL, OH, IL and IH were computed from the parameters of the fitted hill curves according to the definitions given in [1]. OL and OH are twice y_{\min} and half of y_{\max} , respectively. IL and IH are the values of x for which the output of the fitted hill function is equal to OL and OH, respectively. Accordingly, the values of IL and IH were calculated with the following formulae:

Equation 5:

$$IL = \left(\frac{k^n y_{\max}}{(y_{\max} - 2y_{\min})} \right)^{1/n}$$



$$IH = \left(\frac{k^n (y_{max} - 2y_{min})}{y_{min}} \right)^{1/n}$$

Inverters were considered operational under the condition that $OH > OL$ and $IH > IL$ for their fitted hill curve. For a pair of operational inverters, A and B, their compatibility score was defined as:

Equation 6

$$\min \left(\ln \left(\frac{IL_B}{OL_A} \right), \ln \left(\frac{OH_A}{IH_B} \right) \right)$$

with the implication that inverter A can be connected as input to inverter B, if and only if their compatibility score is positive.

5.8. Computation of inverter chains.

Chains of compatible inverters were found by creating a table of compatibility between available inverters, for which the entry for a compatible pair was 1, and all other entries were 0. This table was then treated as the adjacency matrix of the graph of all possible connections, and the longest paths were enumerated using a depth-first search of the graph. Paths in which the same repressor was used more than once were excluded from the results, thus imposing an upper bound of 12 on path length.

5.9. Similarity measure.

The discrete Frechet distance [76] was used to measure similarity of the shapes of two experimental curves, after first log transforming and min-max normalization of the data along both axes. The Frechet distance was then subtracted from 1 in order to produce a metric that increases as the shape of the curves becomes more similar. The discrete Frechet distance was computed using the ‘similarity measures’ Python package [77].

5.10. Prediction.

The goal of the prediction is to transform the characterization of gates in a source context, to a characterization in the target context. A single operable gate was selected arbitrarily upon which to base the prediction. The ‘scipy.optimize’ Python package [75] is used to compute a linear transformation matrix, which when applied to the

source characterization, minimizes the L1Loss between the transformed characterization and the target's true characterization. Predictions for other gates in the library are then made by applying the same transformation to their characterization in the source context.

VI. CHAPTER 4

TOOLS FOR STANDARDIZATION OF
PSEUDOMONAS PUTIDA

VI. CHAPTER 4. Tools for standardization of *Pseudomonas putida*

1. Abstract

Standardization is a key issue in synthetic biology that has the potential to promote reproducibility to communicate experimental outputs. A standardization chassis (e.g. of *Pseudomonas putida*) could be a useful start. In this study we have considered two directions to endorse standardization of *P. putida*: I- Reference promoter, II- RNA Polymerase immunocapture method. A reference promoter means standardized genetic measure, a milestone according to which quantitative outcomes can be reported in gene expression. To this end, we have identified a promoter out of a pool of chromosomally integrated promoters with a unique capacity of least disturbance towards environmental changes e.g. growth conditions. In order to complement this reference promoter, we have developed another technique that allows RNAP immunocapture via epitope tagging of the β' subunit of RNAP. As is the case with many other bacteria, the $\alpha_2\beta\beta'\omega$ core RNA polymerase (RNAP) of *P. putida* interacts *in vivo* at any given time with a suite of sigma factors, transcriptional regulators and auxiliary proteins with a relative composition ruled by the physiological state of cells. In this work we have adopted an *in vivo* immunocapture approach to inspect the effect of typical environmental stresses undergone by *P. putida* in the configuration of the RNAP interactome. To this end the genome of the KT2440 strain of this species was alternatively inserted with a His-tag, a Myc-tag and an E-tag in the 3' end of the *rpoC* gene, which corresponds to a permissive and protruding site of the β' subunit of RNAP. MALDI-TOF analyses and LC-ESI-MS/MS of the thereby seized complexes recurrently revealed the association of the core components of RNAP (RpoA, RpoB, RpoC) with two major sigma RpoS (σ^S) and RpoD (σ^{70}) along with chaperones DnaK, GroEL and HtpG, the ratio of which changed under stress conditions. In contrast, association of RNAP with the product of RpoZ (the so called ω subunit) was weakly detected only under exponential state, but not other conditions tested. Despite the limited resolution of the technology, the data suggests a highly dynamic interplay of the RNAP with a variety of partner proteins that varies with growth conditions and environmental cues.



2. Introduction

In synthetic biology standardization is central, given to its nature embracing engineering principles for biology. The challenge of standardization [78] and the need for common standards is taken into the agenda of researchers on illuminating the necessity and focusing on the importance of standards [79-84]. Standards hold importance in synthetic biology as much as other places where engineering is applied e.g. industry [14]. In the design and reproducibility standardized approaches help to convey information in a package of specified version that is understandable and comparable for others, making the knowledge diffusion efficient. There are several forms of standardization efforts among which we believe to start with emphasizing basics of cellular machineries i.e. gene expression and transcriptional tools. In this study, we have introduced two tools – a standard promoter and transcriptional tool to work in combination with standard promoter – that are to promote standardization efforts in synthetic biology and in *P. putida*.


An *in vivo* standard promoter was developed towards achieving comparable research outcomes. Polymerase per second (PoPS) [27] and the Kelly standard [26] are two examples of such an *in vivo* standard. Our method is to initiate the basis to combine these two approaches within the chromosome of a chassis organism by using a countable promoter expression in PoPS. The proposition was to identify a reference promoter that has stable expression even under changing environmental conditions to embrace wider experimental compositions. In order for identification of such a standard promoter that is orthogonal to cellular practices, a pool of synthetic constitutive promoters is identified [85]. These promoters are covering a wide range of promoter strengths and have single copies in Tn-7 transposon site of chromosome. Characterization of promoters from Zobel et al 2015 in *P. putida* was completed in 3 growth conditions. Meanwhile, the performance of the selected reference promoter was observed under changing pH conditions. Identification of the reference promoter was completed with the following step being RNA Polymerase (RNAP) modification of the corresponding *P. putida* strain in order to facilitate an all-in-one chassis with a standard promoter for PoPS calculations. We have developed a technique to reliably pull-down RNAP for quantification purposes and tested the capacity of this technique

by identify protein-protein interactions, and the protein-DNA interactions would be the next level.

The soil bacterium *Pseudomonas putida* is evolutionary endowed with unique characteristics towards environmental stresses (such as high resistance to solvents and the ability to execute harsh biochemical reactions) [15, 29, 31]. It thus offers a great chance to exploit its distinctive traits for many environmental and industrial applications missing in other microbial cell platforms. Since *P. putida* has become a model *chassis* in microbial biotechnology, it has been comprehensively investigated from different aspects [28-35]. However, such a high potential asset is missing for inclusive RNA Polymerase studies.

RNA Polymerase (RNAP) is responsible for transcription that is copying the DNA sequence into RNA. During transcription process it makes several interactions e.g. DNA, RNA and protein interactions. Among protein interactions, sigma factors are primarily important for RNAP activity. Structurally, RNAP is composed of 5 subunit proteins in the core enzyme and turns into active DNA directed holoenzyme with addition of a sigma factor. Sigma factors are not only important for basic understanding of cellular gene transcription, but also may be important especially in synthetic biology applications i.e. re-wiring cellular rna transcription pathways.

RNAP interactome mapping was shown for several organisms [33, 39] yet, sampling of RNAP interacting proteins under industrially related conditions in *P. putida* was not shown to this point. Among emerging new chassis in synbio *P. putida* is one of them [28, 40]. There is still very limited information regarding its transcription sigma factors [86]. Hitherto, detailed testing of interactome dynamics in response to environmental stimuli and identification of RNAP interacting proteins is a missing piece. This matter is addressed by developing an *in vivo* approach that is modifying a permissible site of one of the subunits of core enzyme (β' subunit) by addition of an immunoprecipitation tag (His, Myc or E-Tag separately). The use of this technique is shown by associating some of the predicted sigma factors with their environmental stimulus. Physical interactions with RNA Polymerase (RNAP) and its various interactome proteins are mapped for *P. putida*.



Here, the impact of various environmental conditions over RNAP interactome dynamics is studied for the model organism *Pseudomonas putida*. This should seize attention for industrial and environmental applications. Tested and shown results here is a strong candidate to do further analysis on relating impacts i.e. growth conditions with transcriptionally related protein.

Heat-shock and solvent stress conditions are tested in order to identify alterations in the interactome of RNAP under different stress conditions. The profile of protein-protein interactions under a given stress condition helps to identify regulators that are expressed as a response. In this study, it was possible to identify many predicted proteins of RNAP interactome under two above-mentioned stress conditions. This is an indicator that by trying more conditions it may be possible to affiliate new proteins with new conditions.

Here, sampling of RNAP and interacting proteins under a variety of environmental conditions is achieved here and a proof of concept is shown. This is useful for direct studies of RNAP interactions and identification of conditionally transcribed regulatory proteins in *P. putida*. The pull-down of RNAP showed direct verifications of some sigma factors and RNAP interactions under predefined stress and growth conditions. A quantitative fold change study is done and some interactions predicted for homologues in the literature are experimentally shown for *P. putida* and mapped too. These findings should contribute into understanding RNAP interactions of *P. putida* to further degrees.

3. Results and Discussion

3.1. Reference promoter

A pool of constitutive and orthogonal synthetic promoters was characterized throughout growth. This characterization took place on 10 promoters that were previously identified out of 30 promoters at Zobel et al. [85]. Chromosomally integrated 10 promoters are tested for their expression pattern at exponential growth under different media (**Figure 14**). LB, M9 with Citrate and M9 with Succinate as carbon source were used. Results are reported as GFP values normalized with OD. **Figure 14a.** and **b.** show two negative controls as a. being the WT strain that is lack

of any chromosomal integrations and b. containing the integrated cassette without a promoter sequence. Outcomes show that there is a fluctuation over time at expression rates of promoters. KT_BG28 (**Figure 14g**) was excluded due to having very close expression values to negative control. Among the rest, KT_BG35 and KT_BG37 strains were selected as showing the best correlation between growth conditions for their promoter expression levels (Materials and Methods). These two promoters later

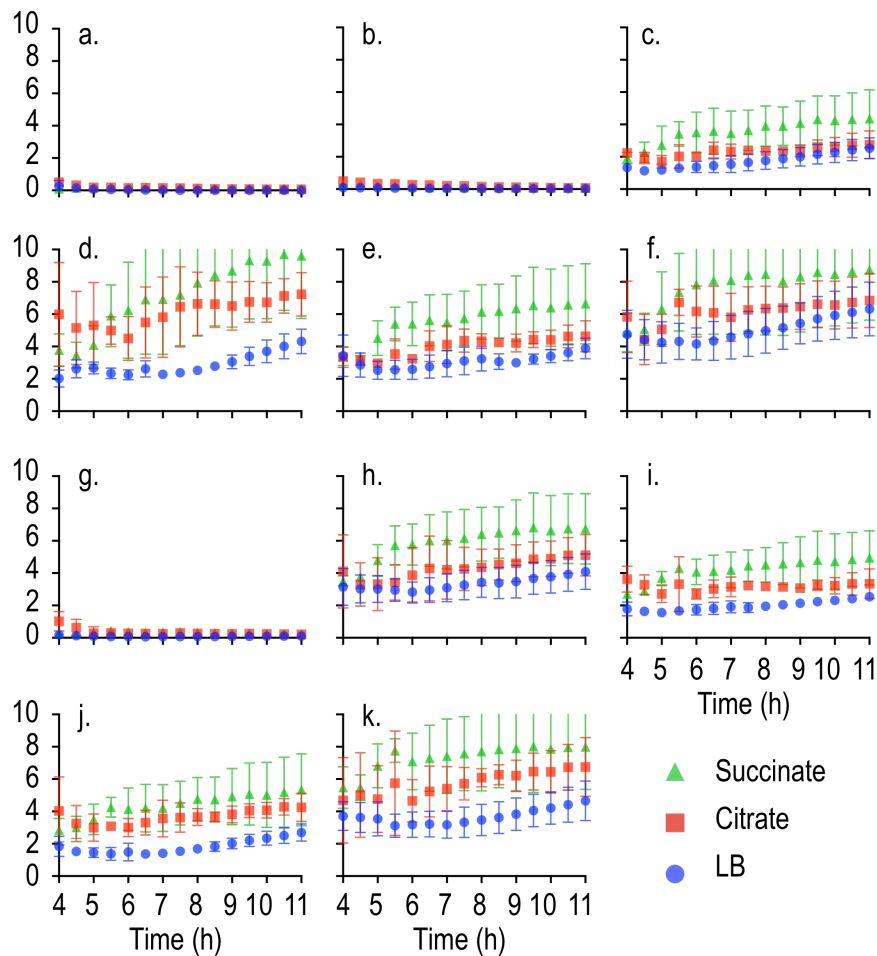


Figure 14: Promoter characterization under 96 well plate growth condition.

Circle (blue) is LB medium, rectangle (red) is M9 minimal medium with 0.2% citrate as carbon source and triangle (green) is M9 minimal medium with 0.2% succinate as carbon source. y – axes stand for GFP/OD values ($\times 10^3$) and x – axes stand for time (h). Promoter strains are as following a. KT2440 b. KT_BG c. KT_BG13 d. KT_BG17 e. KT_BG19 f. KT_BG25 g. KT_BG28 h. KT_BG34 i. KT_BG35 j. KT_BG37 k. KT_BG42.

were tested under the same growth conditions as before, but in flask this time (**Figure 15**). KT_BG37 (**Figure 15B**) showed the least disturbance among 3 growth media used when compared to KT_BG35 (**Figure 15A**), suggesting as a standard KT_BG37 would be more resilient. In addition, this promoter showed higher endurance towards

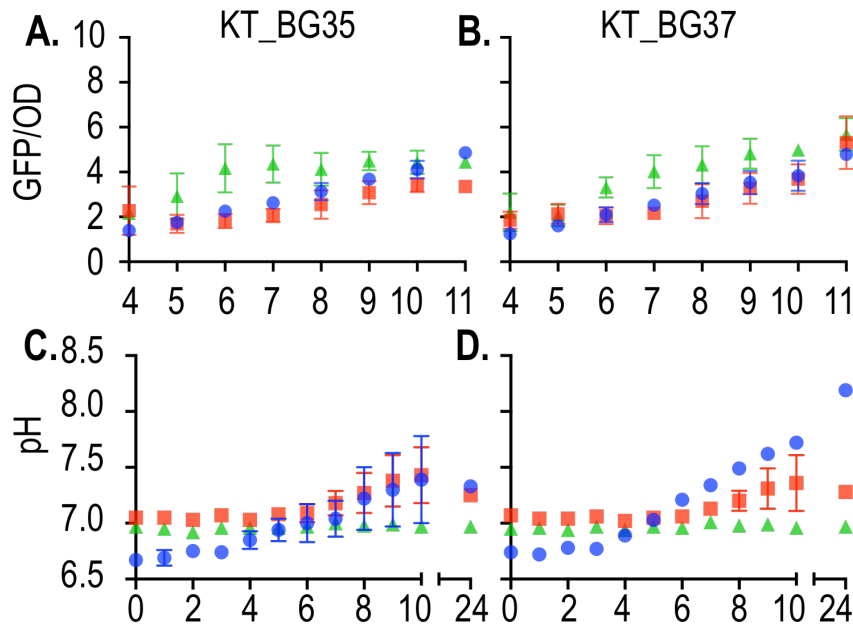


Figure 15: Promoter characterization under flask growth.

Circle (blue) is LB medium, rectangle (red) is M9 minimal medium with 0.2% citrate as carbon source and triangle (green) is M9 minimal medium with 0.2% succinate as carbon source. For A. KT_BG35 and B. KT_BG37 y – axes stand for GFP/OD values (x10³) and x – axes stand for time (h). For C. KT_BG35 and D. KT_BG37 y – axes stand for pH values and x – axes stand for time (h).

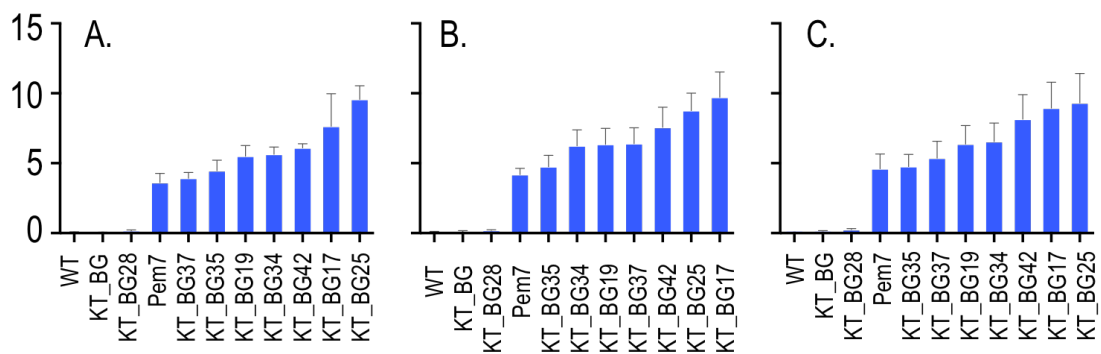


Figure 16: Promoters comparative strengths under 3 growth conditions.

y – axes stand for GFP/OD values (x10³) and x – axes stand for promoter carrying strains. **A.** LB medium **B.** M9 minimal medium with 0.2% citrate as carbon source **C.** M9 minimal medium with 0.2% succinate as carbon source.

pH alterations (**Figure 15C** and **D**). These results suggest the promoter in KT_BG37 strain as a standard promoter that is more orthogonal, durable to environmental changes and has a correlated constitutive character throughout the exponential growth at different media. Its promoter strength (after O/N growth) in comparison to the rest of the tested promoters and Pem7 can be found in **Figure 16**, together with their sequences in **Figure 17**. Defining KT_BG37 strain as the ideal standard promoter

carrying strain, a genomic manipulation took place to modify its RNAP subunit RpoC by addition of immunoprecipitation tags (His tag and Myc tag) – that is, to prepare this strain for future quantitative analysis such as defining actual RNA Polymerase correspondence at a given time period (e.g. PoPS).

Synthetic Promoters Sequence	
Design Format	TTAATTAANNNNNTTGACANNNNNNNNNNNNNNNNNNNTATAATNNNNNNACCTAGG
BG28	TTAATTAAGTGGTTGACA-----TGGATATAATGTATGTACCTAGG
BG37	TTAATTAAGTGAATTGACATGTCAATTTTATGTTGTATAATATAACTACCTAGG
BG35	TTAATTAATTATTTGACATGCGTGATGTTTAGAATATAATTTGGGGACCTAGG
BG19	TTAATTAACCAATTGACAATCAACAGGAAACACTTTATAATACGAGAACCTAGG
BG34	TTAATTAATAAATTGACATCGAAATCGAACACATGTATAATCGCTTAACCTAGG
BG42	TTAATTAAGCCCATGACAAGGCTCTCGCGCCAGGTATAATTGCACGACCTAGG
BG17	TTAATTAACGGAGTTGACAACACTCGAAAAGCCGAGTATAATCAGATGACCTAGG
BG25	TTAATTAAGCCCGTTGACATGACATGGTTTGTAGGGTATAATGTGGCGACCTAGG

Figure 17: Synthetic promoters sequences.

Blue sites show the PacI (TTAATTA) and AvrII (CCTAGG) restriction regions. Red sites show the conserved regions of sigma70 promoter -35 (TTGACA) and -10 (TATAAT) [85]. Sequences are ordered based on activities in LB medium.

This could be done by taking advantage of protein-DNA interactions. However, in *P. putida* there were not any methods known to pull-down RNAP to our knowledge. This made us to develop a second tool for standardization of *P. putida* – an RNAP pull-down technique.

3.2. RNA Polymerase immunocapture method

As RNA Polymerase is an enzyme that realizes crucial workload for the cell, manipulating genes for RNAP subunit is not trivial. Yet, a permissible site (based on our previous experience) was modified in *rpoC* gene. This gene encodes for DNA-directed RNAP subunit β' that is a part of the core enzyme. The modification took place in the upstream of TAA stop codon (**Figure 18A**). In the genome of *P. putida* KT2440, *rpoC* is genetically tagged with three different immunoprecipitation peptide-encoding sequences separately. These options give the flexibility for using a tag of interest i.e. His tag for cost effective RNAP capturing or Myc and E tags for high specificity of protein identification. In our case, Myc tag was picked for the

experimental procedure. These three strains (Tas108, Tas110 and Tas114) and two strains of KT_BG37 (Tas103 and Tas106) are ready to be distributed on request of researchers (**Table S7**). All modifications were done chromosomal level via usage of Martinez et al., genome manipulation technique [87]. Glycine amino acids are added at the 5' of the tags in order to allow flexibility for exposure of the tags [88]. The size of the tags is ~1 kDa and RpoC is ~155 kDa, a minute size in comparison. Verifications took place in dna and protein levels. In gene level, for the set of primers one primer always targeted direct sequence of the tag and PCR verifications showed correct sizes ~0.7 kB (**Figure 18B**) and the locations. A list of the primers used in this study can be found in **Table S3**. In peptide level Western Blot (**Figure 18C**) showed the tags attached to RpoC are at the correct size (~155 kDa) and accessible after crosslinking.

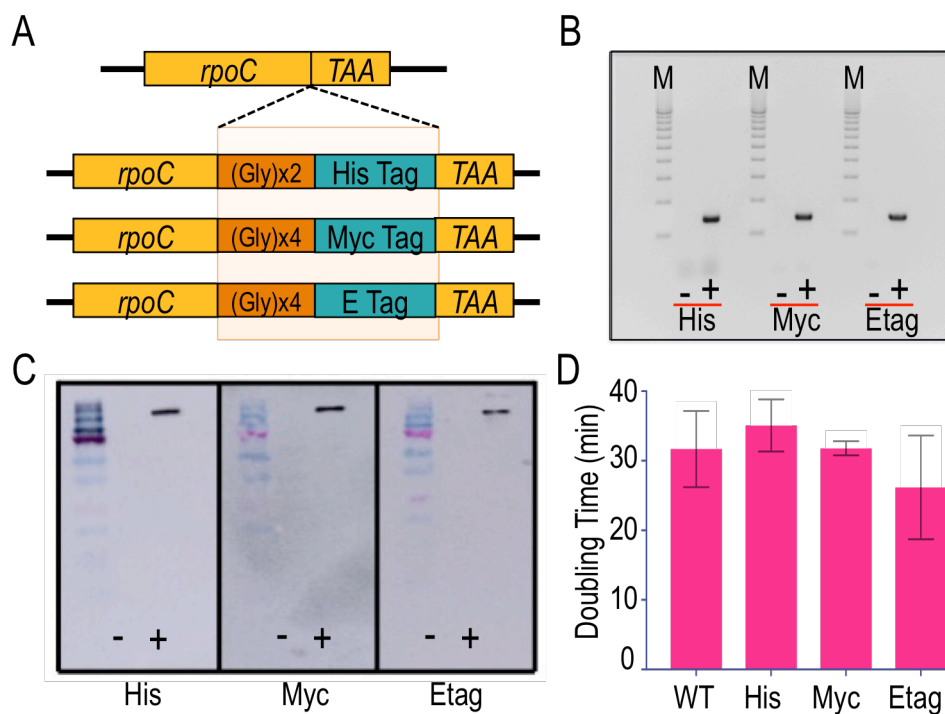


Figure 18: Genetic tagging verifications of *rpoC*.

A. The genome of wild type strain is modified. Three different tags are added in the downstream of the *rpoC* gene before the stop codon. These tags are His tag, Myc tag and Etag. Modifications are done in chromosomal level. Glycine amino acids are added between the RpoC and the tag in order to give flexibility and exposure to the tag. The schemes are not scaled, RpoC is ~155 kDa and each tag is ~1 kDa **B.** PCR verification of the tags by using tag specific primer sets. (M) stands for marker, (-) for WT control, (+) for modified strain. Each tag is presented with WT control PCR and tagged strains show right size (~0.7 kB). **C.** Western Blot analysis for verifying each strain with corresponding tags. WB analysis shows that the band size (~155 kDa) corresponds to RpoC size. Each analysis is run separately. All three tags show correct sizes. **D.** Doubling times are measured as an indication of whether

growth rate is affected by *rpoC* manipulation. Tagged strains show close proximity to WT strain in doubling times.

WB is done on the cell extract after lysis as described under Materials and Methods. Additionally, visual molecular dynamics program is used for identifying tag's possible position on 3D structure of core enzyme (**Figure 19**). Doubling time was observed to see if immediate growth defects present due to the modifications. Results (**Figure 18D**) show that tagged *P. putida* strains have a very close proximity with WT strain in doubling times. To eliminate concerns over possibility of morphological impacts, colony formations were checked. **Figure 20** indicates no major morphologic changes in tagged strains over 3 days growth on M9 agar plates.

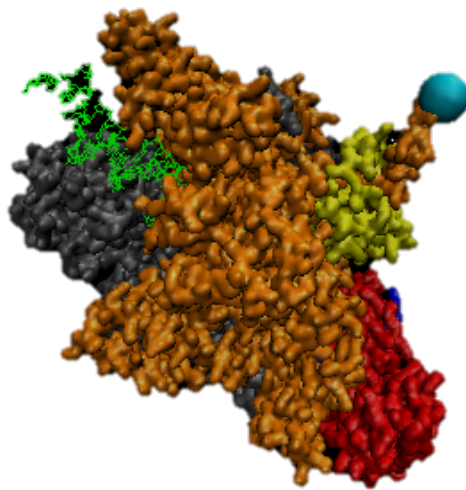


Figure 19: Representation of an RNA Polymerase core enzyme.

RNA Polymerase core enzyme with subunits RpoA (red), RpoB (Gray), RpoC (Orange) and RpoZ (Yellow). Blue ball is spatial coordinate of final amino acid of RpoC after which tag sequence is inserted. DNA position is shown with green color.

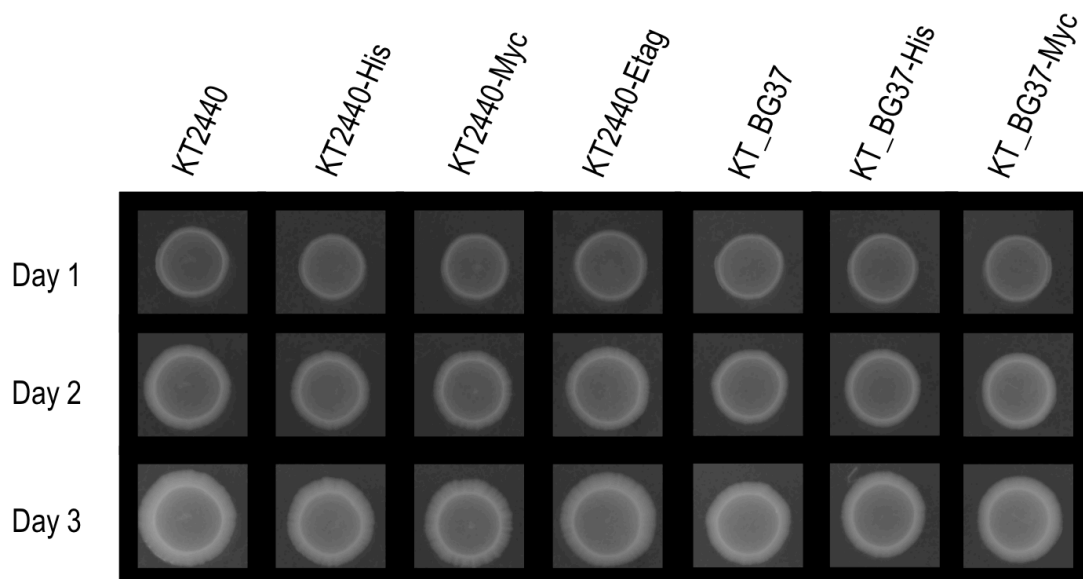


Figure 20: Colony morphology.

For 5 modified strains, colony formation on a M9 minimal media 1% agar plates is controlled over 3 days of period in comparison to wild type *P. putida* and negative control KT_BG37.

An experimental procedure is established (**Figure 21**), for immunocapture of RpoC interacting proteins (see Materials and Methods). RpoC interacting proteins were identified by using MALDI-TOF and LC-ESI-MS/MS proteomics techniques. Immunocapture of RNAP was facilitated in total of 6 common growth and stress conditions (**Figure 22A**). These are I- Heat-shock II- Exponential III- Overnight IV- Sheer Stress V- 5days Starvation VI- Solvent Stress. SDS-Page is run and designated bands from 5 out of 6 conditions were cut out for identification with MALDI-TOF analysis; and out of 6 conditions exponential growth, heat-shock & solvent stress conditions were also analyzed with LC-ESI-MS/MS technique. In MALDI-TOF the bands are picked considering the variations among each condition. Our results show a list of protein-protein interactions that were predicted (according to String-DB, low to high confidence) and also some nascent interactions (**Figure 22B**). Schematic representation of protein interactions is shown in **Figure 22C**.

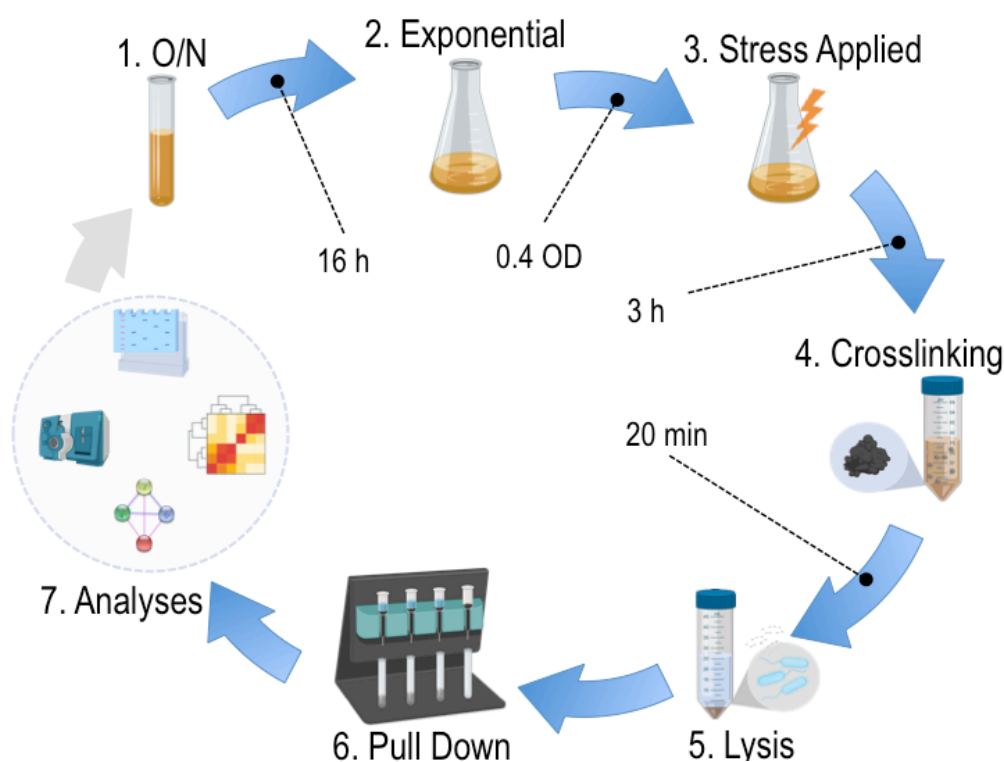


Figure 21: Experimental procedure.

Cultures are equilibrated by 0.02 OD inoculation from an overnight culture. Exponential state is reached once 0.4 OD is obtained. Stress condition is applied over 3 hours period which is followed by a crosslinking step. Lysis is applied before immunomagnetic separation with Myc tag. Proteomics analysis, SDS-Page or other downstream bioinformatics analyses are done.

Some of the new interactions were not predicted (based on homology analysis in other hosts) to have primary degree protein-protein interactions with RNAP i.e. chaperon ClpB. That is, ClpB has a first-degree protein-protein interaction with chaperon DnaK which has primary interaction with RNAP [89], supporting the idea that RpoC sampling may also be useful for capturing second-degree interactions of RNAP. Raw data of MALDI-TOF, the identified proteins and their corresponding scores are included in Supplementary Data Raw.

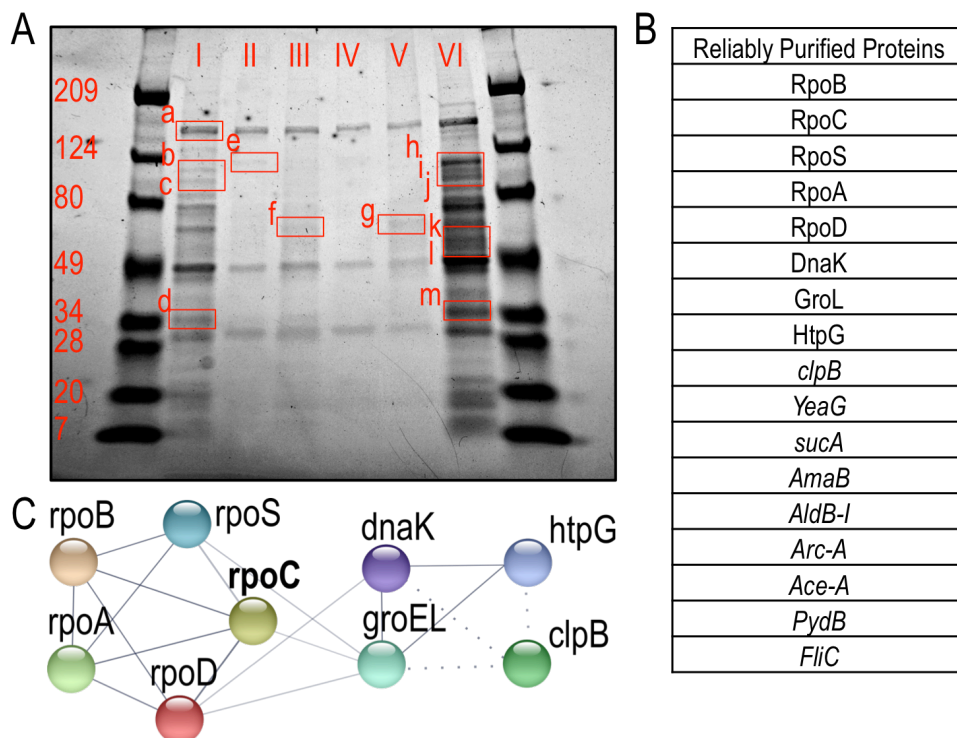



Figure 22: Pull-down of RpoC interacting proteins under various environmental conditions.

A. SDS-Page gel shows 5 different stress conditions and pulled down proteins with RpoC under these conditions. These conditions are I- Heat-shock II- Exponential III- Overnight IV- Sheer Stress V- 5days Starvation VI- Solvent Stress. Bands that are analyzed are shown with a red rectangle. (Note: For SDS-Gel Contrast and color code is applied for printing purposes, original version can be found in supplementary) **B.** The list shows all reliably pulled down proteins. In the list proteins predicted to be interacting with RpoC in other hosts (not italic) and new interactions (*italic*) are shown. **C.** The interaction network of these proteins is shown. Each node in the network shows all the proteins produced by a single protein-coding gene locus. Color code is for representative purposes.

Under heat-shock stress, 4 bands were identified: RpoB and RpoC (RNAP subunits), SucA (a decarboxylase from TCA cycle), ClpB and HtpG (chaperons). RpoB and RpoC (a) bands are seen at ~155 kDa band size. These are two RNAP core enzyme subunits with very close sizes, causing co-identification from the same band. SucA



(b) is another protein identified in heat shock stress and ClpB (c) too. ClpB is a chaperon protein. HtpG (d) is another chaperon protein that is active in heat-shock damages [90]. The band size HtpG is isolated is around half size of HtpG protein (~72 kDa) which may speculate to possible protein breaks. Most of the identified proteins are seen in their expected band sizes in SDS-Page gel, however not all. This may cause from involuntary breaks or formed complexes due to using protein denaturalizing agent and crosslinking agent. Under exponential growth condition, gel extracted band is (e) corresponding to GroL (chaperone) that promotes refolding to prevent misfolding. Its size is 60kDa, however GroL dimerizes e.g. in *Mycobacterium* [91, 92]. This may be the case in *P. putida* too given to the band twice bigger band size it is extracted from. ArcA (f), pulled-down under overnight stress condition, is an amino acid degradation enzyme, arginine deiminase. It is stimulated by the accumulation of sigma S factor during entry into stationary phase [93, 94].

This may explain its abundance in the gel. Under 5 days starvation condition, the isolated sample is identified as RpoA (g), RNAP core enzyme subunit. It dimerizes for molecular activity [95-98] which may explain its presence in a band size around double of its own. Under solvent stress (1-3-PD) condition, isolated bands (h, i, j, k, l, m) have shown less consistency in MALDI-TOL analysis. Bands (i) and (k) could not be identified and in (l) and (m) traces of RpoB and RpoC were seen. (h) is identified as RpoA and (j) is identified as FliC, a flagellin protein. Seeing an extracellular flagellin protein may be a contamination introduced due to the stress condition over inducing its expression. Exponential state, heat shock and solvent stress conditions are expanded with LC-ESI-MS/MS analysis for a quantitative approach. Purification fold change is calculated by comparison of purified and unpurified (whole cell extract) samples at exponential growth condition (Materials and Methods) and fold changes are calculated (based on PSMs values) among proteins that are predicted in String-db as interacting with RNAP. On average purification factor is seen to be around ~4 fold. Note that here the main purpose is the proof of concept for showing that RNAP sampling and its interactome could be inspected under a variety of conditions.

Heat-shock and solvent (1-3-PD) stress conditions (**Figure 23**) were used for sampling RNAP interactions. Fold change values are presented with whisker plots in reference to exponential state. Whisker plots are drawn based on RpoC interacting

proteins. The proteins in the top 10% for fold change increase are depicted in **Figure 23A** (heat-shock) and in **Figure 23B** (solvent stress) (full list in Supplementary Data Raw). 10 proteins with highest fold change values for both stress conditions can be seen on **Figure 23C** and **Figure 23D**. Among these proteins 5 of them are shared for both stresses. These 5 proteins are ClpB (chaperon), SdhA (succinate dehydrogenase), AceF (a component of pyruvate dehydrogenase complex), KgdA (aldolase activity) and EtfB (electron transfer flavoprotein subunit β). The KEGG Pathways distribution (of the 10 proteins) is given as following: in heat-shock condition carbon metabolism (5 proteins), citrate cycle (3 proteins), biosynthesis of antibiotics (5 proteins), microbial metabolism in diverse environments (5 proteins); for solvent stress condition whereas, biosynthesis of secondary metabolites (6 proteins), citrate cycle (3 proteins), biosynthesis of antibiotics (5 proteins) and microbial metabolism in diverse environments (5 proteins).

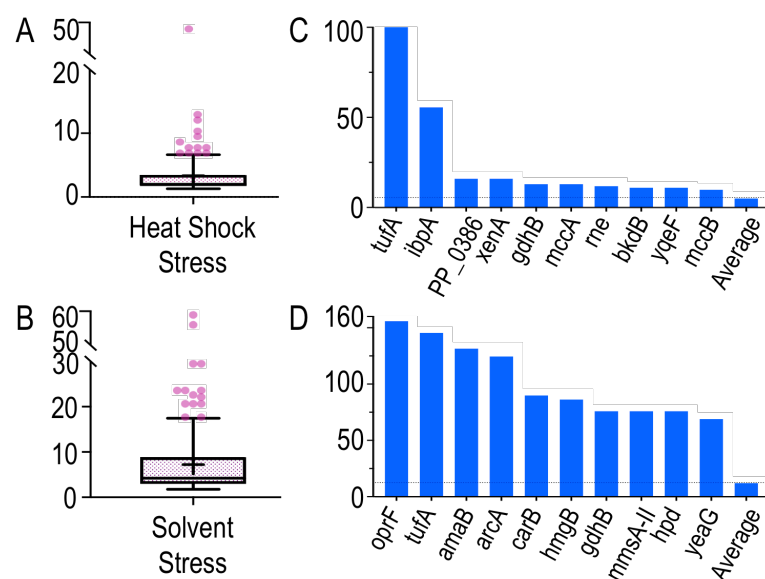



Figure 23: Fold changes for heat shock and solvent stress conditions.

Fold changes for heat shock and solvent stress conditions as exponential state being reference. Whisker plots show fold changes for proteins exist in exponential growth and the stress condition. Bar graphs show proteins only exist in the stress conditions. **A. & B.** Pulled-down protein distribution under heat-shock stress and solvent stress conditions. Proteins over 90% top are shown in filled circles. **C. & D.** Top 10 proteins pulled-down only in heat-shock or solvent stress conditions and not in exponential growth condition. Average shows the average of all pulled-down proteins in the corresponding stress condition.

A network of predicted interactome of RpoC based on homologs in other bacterial species is given in **Figure 24**. 19 out of 25 predicted proteins were identified under



two stress conditions. By diversifying environmental conditions more proteins may be identified. Exponential state is the reference for fold change calculations. For both heat-shock and solvent stress conditions, PykA (pyruvate kinase), PyrG (CTP synthase) and NusA (transcription termination and antitermination protein – transcription factor) are shared proteins that are overexpressed in addition to Pnp that is only under solvent stress. Pnp is polyribonucleotide nucleotidyltransferase that is involved in mRNA degradation. Its overexpression supports the idea that under the solvent stress cell prefers shutting down activities other than survival reflex. Observed PSMs means that those proteins are captured under one or two of the defined conditions (RpoC is given as a reference point). Ndk (nucleoside diphosphate kinase) and RpsH (ribosomal protein) are overexpressed in solvent stress and RpsQ (ribosomal protein) is not detected in heat-shock condition. Ndk is nucleoside diphosphate kinase that regulates synthesis of nucleoside triphosphates in general. Nucleoside diphosphates are used as RNA synthesis substrates and also as energy sources for the cell. RpsH is a 30S ribosomal protein. RpsQ as being one of the primary rRNA binding proteins that interacts with 16S ribosomal RNA [99] is an important piece of transcription/translation mechanism. RelA (GTP diphosphokinase activity) and DksA (RNAP binding transcription factor) are captured in trace amounts. RpoZ shows no appearance under stress conditions and shows only a limited abundance under exponential condition. This may be explained with small size of this protein and the dynamic interactions of RNAP subunits. Despite RpoZ, rest of the

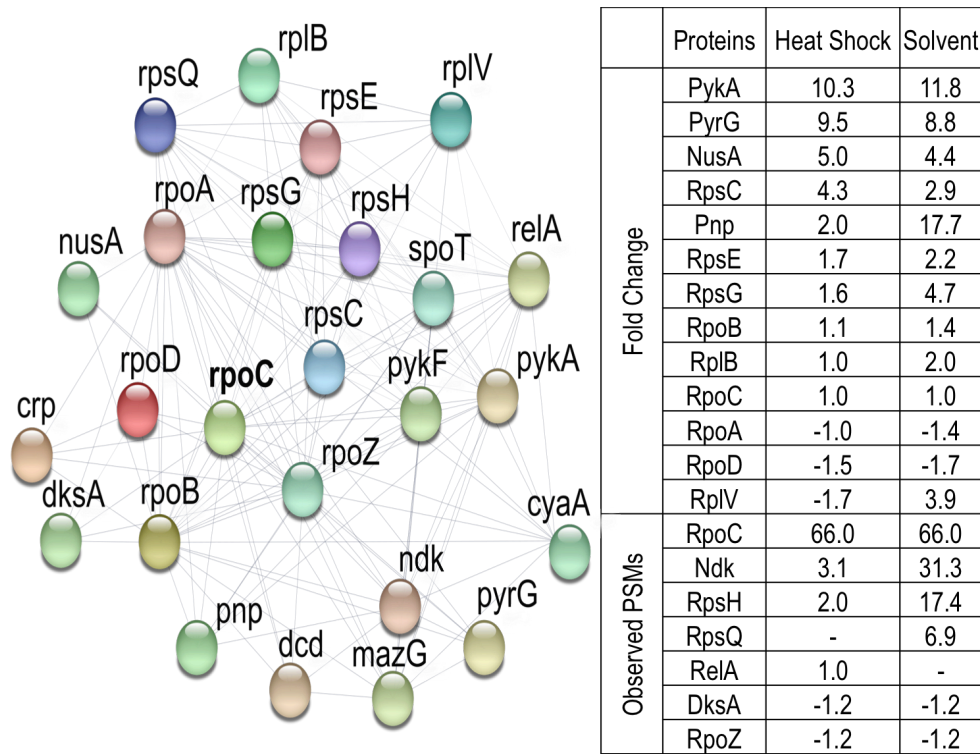


Figure 24: Quantitative comparison of Exponential state, Heat Shock and Solvent Stress conditions.

A. Mapping of predicted RpoC interactions for *P. putida* based on homologs in other species (String-db). **B.** The fold change (fc) values show the fc variation for Heat Shock and Solvent Stress in reference to exponential state, meaning proteins that are present in all three states at the same time could be calculated for fc values. Negative values show fold change decrease. Observed PSMs means that the proteins are only captured in either one or two of the conditions. Negative values mean the difference from exponential state. Dash (-) means absence of the protein at a given condition. Each node in the network shows all the proteins produced by a single protein-coding gene locus. Color code is for representative purposes.

core enzyme subunits are captured in abundance. The data responds to sampling of core transcriptional machinery and its interacting proteins under different environmental conditions.

Here, immunocapture of RNAP is tested through epitope tagging RpoC subunit. By using the epitope tagging, the interactome is examined under various conditions in *Pseudomonas putida*. This is shown on five different stress conditions and growth state. As an outcome of this study, characterization of Formaldehyde and 1-3-Propanediol is also realized for *P. putida*. In brief, two valuable tools are developed in this study, a standard promoter and an RNAP pull-down technique. That is, towards standardization efforts in synthetic biology *P. putida* should serve as a key candidate.

4. Conclusion

To promote standardization efforts in synthetic biology, two angles were exploited here, discovering a standard promoter and engineering a pull-down method for RNAP of *P. putida*. A universal quantification approach was targeted by characterizing a reference promoter in combination with RNAP pull down technique to work within the same promoter carrying strain. KT_BG37 strain was described as the reference promoter strain and two versions of it with immunoprecipitation tags (His tag and Myc tag) on RpoC subunit of RNAP were prepared. In *P. putida* an RNAP based pull-down technique was missing up to day, therefore we took the liberty to develop and to establish the limits of the developed technique by investigating more on RNAP interacting proteins. In this study, it is shown that RNA polymerase and its interacting proteins can be sampled *in vivo* in *P. putida*. C-terminus genetic tagging of one of the subunits of RNA Polymerase core enzyme, RpoC, is realized. The tagging is done in the genome. Several scale verifications for genomic integrations are done i.e. genomic scale (PCR), protein scale (WB and Proteomics) and biologically relevant (Doubling Time) verifications. By sampling RNAP interactome, an important field in environmental bacterium *P. putida* is addressed. Configuration of RNAP interactome is analyzed quantitatively by calculation of fold changes at given conditions and robustness of the sampling was checked by implementation under various growth and stress conditions. Common databases (e.g. String-db, uniprot) were used to consult homologs in other species to do comparisons and confirm pulled-down proteins. Enrichment in the concentrations of RNAP interacting proteins was also shown in the purified samples. Moreover, FA and 1-3-PD chemicals were characterized for *P. putida*. Sampling under diverse growth and stress conditions was succeeded for characterization of sigma factors and identification of new RNAP interactions, together with quantification of RNAP participating pathways. KT_BG37 reference promoter in combination with pull-down technique can be helpful for further standardization and quantification purposes.

5. Materials and Methods

5.1. String-DB

String-DB [86] is used as the reference database for protein-protein interaction information. Unless otherwise is stated, highest confidence configuration (>0.90) is

used for defining minimum required interaction score. This means only known interactions with a high confidence are considered. Active interaction sources are picked as "Experiments, Databases and Co-expression".

5.2. Visual Molecular Dynamics analysis

To represent spatial location of inserted tag sequence in the 3D structure of RNAP, a similarity assay is used (as complete crystal structure of *P. putida* RNAP core enzyme was not found in the protein database - PDB). 3D structure of *Thermus thermophilus* RNAP (with pdb code 4WQS.pdb) was available. Basing on this RNAP structure, a 3D possible representation is created to visualize tag representation. The NCBI blast results of RNAP protein sequence similarity between *Thermus thermophilus* and *Pseudomonas putida* can be seen as following (sequentially the query cover and percentage of identity): RpoA: 93%, 43%; RpoB: 95%, 52%; RpoC: 77%, 50%, (RpoZ: low coverage). **Figure 19** shows an approximation, yet based on Visual Molecular Dynamics analysis tags spatial representation corresponds to outwards. VMD version used is Version 1.9.1.

5.3. Functional Analysis

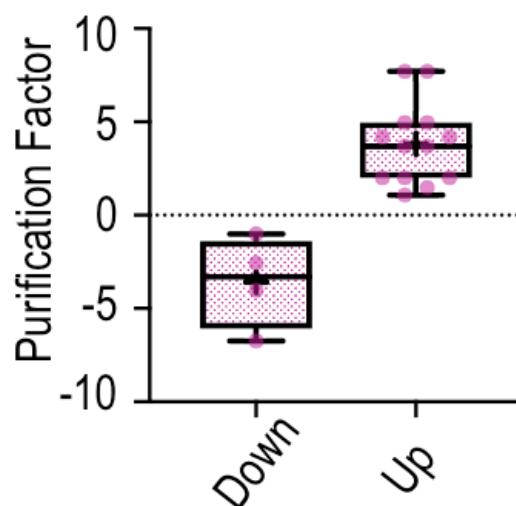



Figure 25: Purification factor.

Exponential state RNA Polymerase subunit β' interacting proteins in highest confidence String-db list. Fold changes are shown for changed PSM values between before and after purification process. Down (mean= 3.6) shows proteins lowered in concentration after purification process and Up (mean=3.8) shows proteins that are concentrated after purification.



Fold changes are calculated based on PSMs values obtained at experimental purified condition vs either heat-shock PSMs values or solvent stress PSMs values obtained as a result of LS-ESI-MS/MS technique. In **Figure 25** the comparison is done between unpurified whole cell extract protein PSMs versus purification applied proteins PSMs in order to understand enrichment obtained under defined conditions. Normalization is done based on RpoC PSMs values for purified samples and based on total number of PSMs for the enrichment analysis. The purification factor intuitively can be increased by increased volume of immunoprecipitation beads used for the same amount of elution buffer volume. This can be optimized for users preferences. Down values in **Figure 25** shows the proteins that show a concentration decrease after sampling. This may be due to the weaker interaction of those proteins with RpoC. This might mean that as stronger interactions are preserved with increased purification factor, weaker interactions may be lost. For fold changes in **Figure 24**, the calculations are based on PSMs values unless otherwise mentioned. When a protein is not found in a condition, instead of fold change the observed PSMs are reported.

5.4. Growth and Media

Luria-Bertani (LB) medium is used throughout the experimental procedure. Optimum temperature for growth of *P. putida* KT2440 is 30C degree. For flask experiments air shaker is used and for 96 well plate experiments 96-well microtiter: SpectraMax M2 microplate reader (Molecular Devices, Sunnyvale, CA, USA) is used. For heat-shock stress condition, the experimental setup is preserved as in **Figure 21**, and 42C degree temperature is applied accordingly at an air shaker. Doubling time analyses depend on 3 separate experiments where 20 ml LB is used in a 100 ml flask at 30C degree air shaker. Cultures are inoculated with 0.02 OD initial optical densities from overnight cultures.

5.5. Formaldehyde Characterization

Formaldehyde (FA) is used as the crosslinking agent. The application is developed based on Sendy et al. crosslinking procedure [100]. According to results seen in **Figure 26**, Formaldehyde Crosslinking Calibration, a final concentration of 1% Formaldehyde is used with 20 minutes of incubation time at 30C. This characterization is done via targeting GFP protein migration with WB under various

FA concentrations. In addition, possible effect of FA concentrations on protein functionality is observed in the sense of fluorescence.

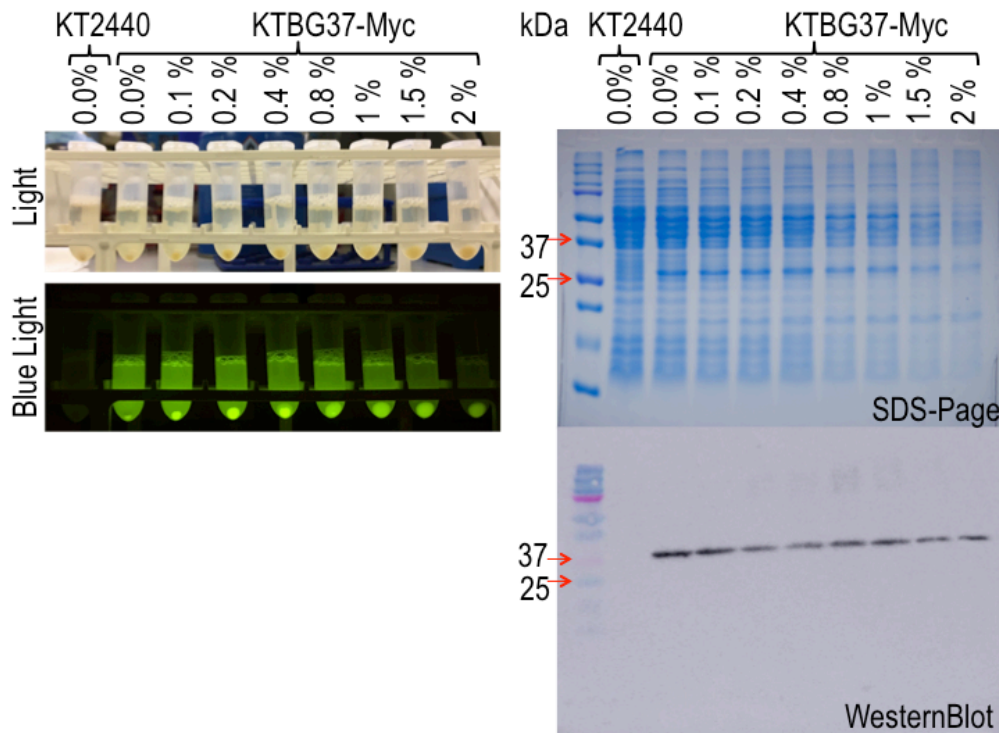


Figure 26: Formaldehyde characterization in *Pseudomonas putida*.

A detailed analysis on the concentration of FA is tested for *Pseudomonas putida* KT2440.

5.6. Lysis

Lysis is done via using BugBuster protein extraction reagent and Lysonase bioprocessing reagent according to their defined protocols.

5.7. Western Blotting

For doing Western Blot analysis in this study the following procedure is followed. After running SDS-Page gels, the stacking gel part (4% Acrylamide) is removed. The running gel is accommodated in between filter sponges (soaked in Transfer buffer) and the activated membrane. Membrane is activated in Methanol for a minute before used. The sandwich is transferred into Transfer Buffer to soak even more for electroconductivity. The transfer is done by using 0.1 Ampers for 30 minutes. After transfer membrane is blocked for unspecific bindings with 3% milk in Buffer A for 1 hour at room temperature. Antibody with POD specific to the condition defined is used 1:10000 dilution. This process is completed with addition of chemiluminescence

for peroxidase and after 1 minute incubation samples are measured with Amersham™ Imager 600.

5.8. RNA Polymerase Pull-down

To equilibrate the starting culture an O/N 2 ml LB culture is started from -80C stock. 0.02 OD culture is started from the O/N culture in a 20 ml LB medium. Once the culture reached to exponential state (~0.4 OD) the stress is applied. The procedure continues with crosslinking with formaldehyde and lysis steps. Later, immunoprecipitation is done by using MACS molecular's μMACS Epitope Tag Protein Isolation Kits protocol and magnetic apparatus. After elution is completed analysis can be done as a downstream application such as SDS-Page and proteomics.

5.9. Sequencing

Sequencing is done as a Sanger sequencing. The aligned data can be found in Supplementary Data Raw. **Figure 27** shows sequencing result alignment as close-ups to the tag sequence. Primers used for sequencing can be found at **Table S3**.

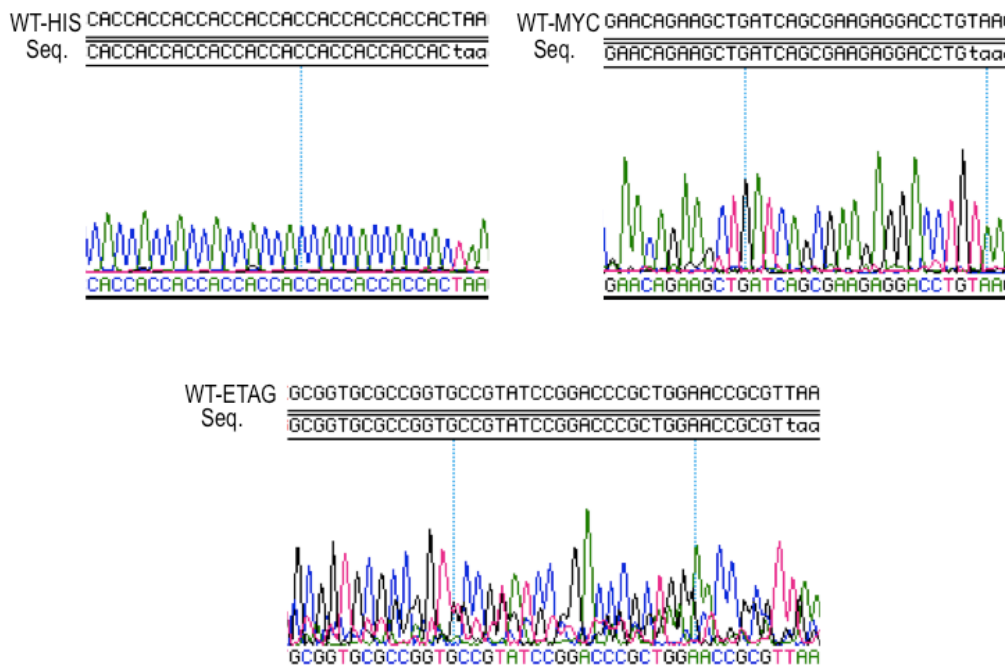


Figure 27: Sanger sequences of *in vivo* tags. Complete set of tag sequences are seen in each figure at the upstream of stop codon TAA.

5.10. MALDI-TOF and LC-ESI-MS/MS

MALDI-TOF and LC-ESI-MS/MS (ultra short gradient) protocols are followed by CNB-CSIC proteomics department as a paid service. The protocols are used as in Arcos et al. and Bodega et al. [101, 102]. All affiliated LC-ESI-MS/MS data can be found in Supplementary Data Raw.

5.11. 1-3 Propanediol Characterization

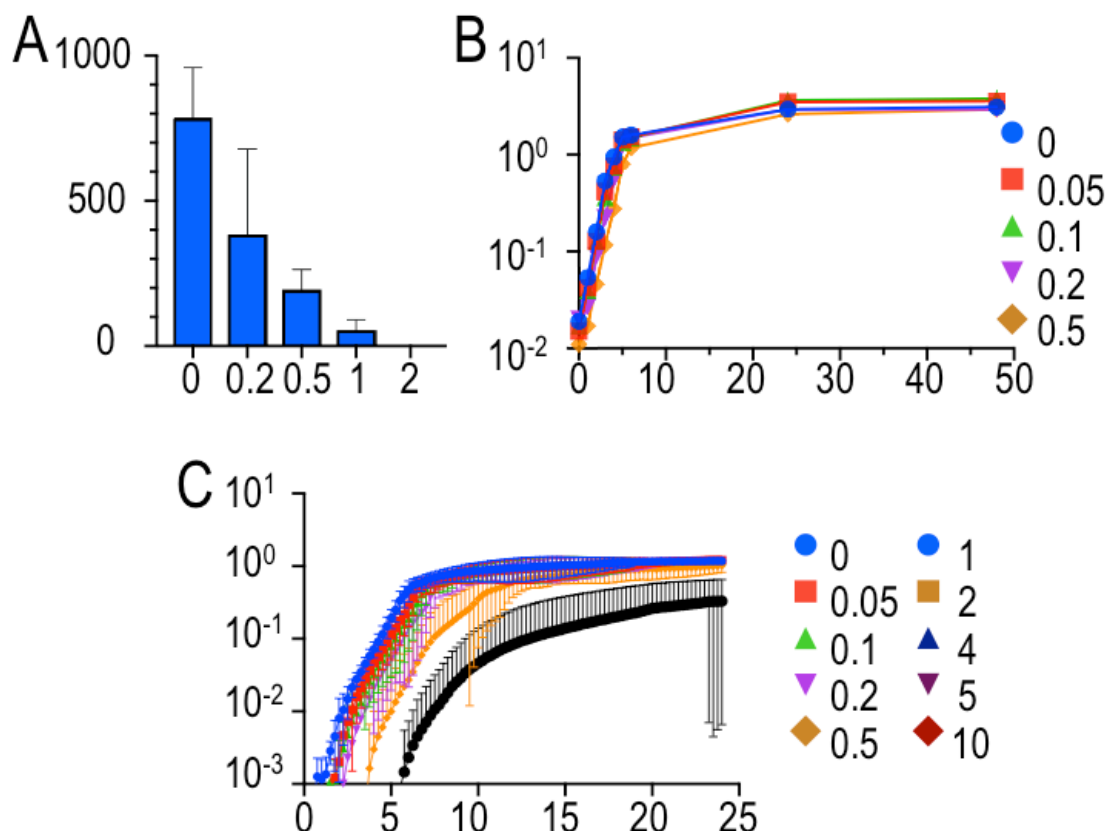



Figure 28: 1,3-Propanediol characterization in *Pseudomonas putida*.

A. Plate analysis y-axis Number of Colonies vs x-axis Concentration (in Molar). Showing the number of colonies formed after seeding same amount of cultures from each concentration of 1-3 PD applied samples. Cultures are exposed to 1-3-PD at 0.4 OD for 3 hours. The effect of 1-3-PD on halting growth after 3 hours exposure is seen. **B.** Flask analysis y-axis OD(600) vs x-axis Time (hours). Four different concentrations of 1-3-PD is controlled for its effects on growth. Flasks are inoculated at 0.02 OD with co-presence of corresponding 1-3-PD levels. The OD measurements are for 48 hours period. **C.** 96 well plate analysis y-axis OD(600) vs x-axis Time (hours) is shown. 1-3-PD effect is measured in higher concentrations in 96 well plate. Cultures are started from 0.002 OD and measurements continued for 24 hours in thermo controlled plate reader.

The effect of 1,3-Propanediol concentrations is measured with 3 different approaches (**Figure 28**). In each of these approaches cultures are started from an equilibrated O/N LB cultures. In the first approach the survival analysis is done with exponential phase



growing cells which are exposed to 200 mM, 500 mM, 1 M and 2 M 1-3-PD for 3 hours with a starting concentration of 0.4 OD exponential phase cells where 0 mM is used as control. Seeding took place in the end of 3 hours exposure time of solvent applications and then colony seeding of 10 μ l of twice 10^{-2} serial dilution is applied from each flask on LB agar plate for overnight. And results are reported as the comparison of survival colony numbers averaged from 4 experimental repeats with 2 technical replicates. In the second approach 0, 50, 100, 200, 500 mM 1-3-PD are used in flask cultures. Starting OD as 0.02 OD with co-presence of corresponding solvent concentrations. The OD is tracked up to 48 hours, and for first 12 hours an OD measurement is done approximately every 1 hour. In the third approach solvent's effect on 24-hour exposure is checked for higher concentrations of 1-3-PD. The cultures are inoculated with a starting 0.002 OD of 2 experimental repeats with 8 technical replicates in 96 well plate in order to test the effect of 1-3-PD over 24 hours growth. All data related to 1-3-PD characterization can be found in Supplementary Data Raw.

5.12. Correlation and Outlier Analyses

Correlation analyses were done based on GFP/OD values at exponential growth state (between 4th to 11th hours) by using default correl function of Excel 2011 Version 14.6.6. Outlier analysis was done on a data point showed significant deviation. The formula used for this analysis is the common outlier test that searches for Quartile 1 and 3 and 1.5 times of Interquartile Range to define Lower and Upper bounds of the dataset. The same excel version was used with default quartile function for this analysis.

5.13. Genetic Tagging in *rpoC* gene

Three tags used are sequentially His, Myc and E-tag which are the common tags in laboratory practices. The genomic manipulations are done to *rpoC* gene's downstream by adding these tags right before the stop codon to form three individual strains with one of the tags in each. The protocol used for this process is the one explained by Martinez et al. [87]. It is pEMG plasmid based on chromosomal integration with homologous recombination technique (**Figure 29**).

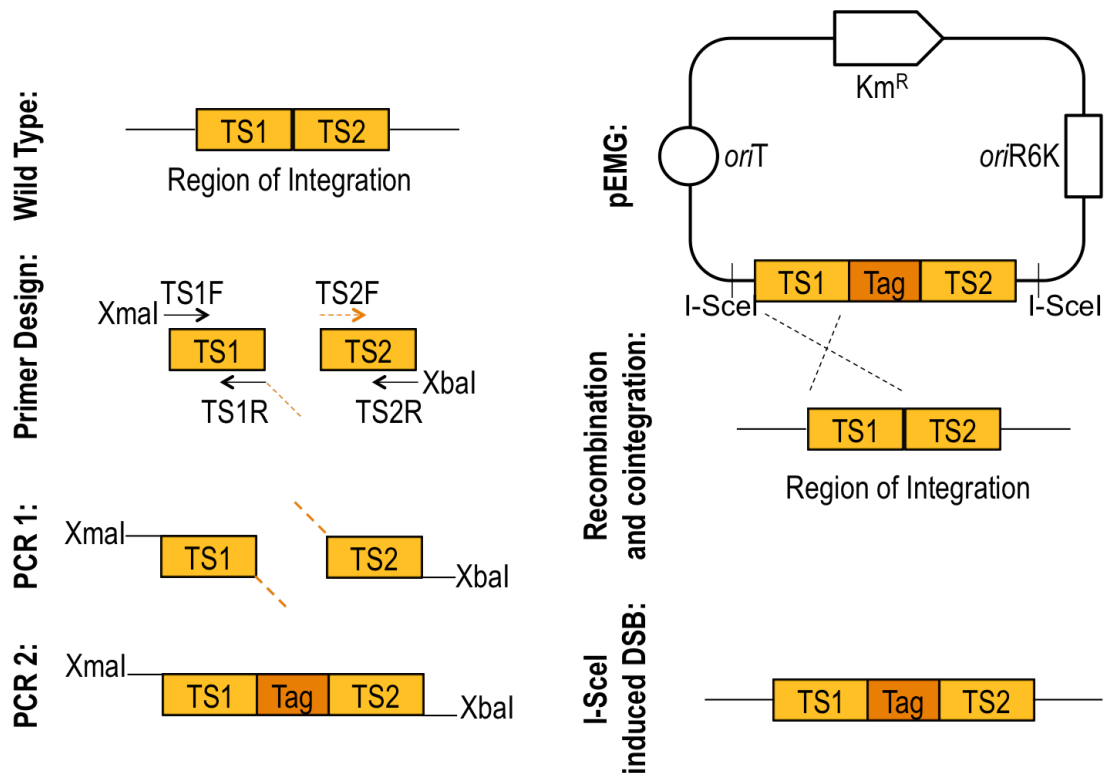


Figure 29: Genetic Tagging in *rpoC* Gene.

Region of integration on wild type sequence is depicted. Design of primers are made according to the sequence of interest, in our case the *rpoC* gene. Primer sequences used for this study can be found at the Annex. The procedure follows two PCRs, one for creating TS blocks with introduction of flanking sites (the cloning sites and the tag to be inserted) and second one for making the final double stranded DNA for cloning purposes. The TSs block is cloned into pEMG plasmid. This is a suicide plasmid and used for triggering homologous recombineering in order to have a cointegration of the whole plasmid into the genome. This cointegration is induced with double stranded breaks (DSB) by I-SceI sites which conclude into final integration of the immunocapture tag. This process is given as 50% efficiency in Martinez et al., 2012 [87].


VII. DISCUSSION

VII. DISCUSSION

This Thesis advocates the importance of revealing new frontiers for genetic devices and establishing fundamental methods to upgrade *Pseudomonas putida* as a synthetic biology chassis.

Automation is an accelerating endeavor of contemporary synthetic biology. Several well-established setups have been introduced. Genetic parts optimization is a common application in circuit designs, yet cumbersome and costly. It is promoted here that having libraries of inverters originally optimized for *Escherichia coli* NEB10 β (borrowed from Cello study) helps to have a quick start for new organisms. The inverters set was reconstructed in SEVA collection backbones (in low, medium and high copy numbers) without changing the operation modules. Availability of choice among 3 separate ORIs enables minimizing toxicity effect on the design. Obtained results show that in Gram-negative bacteria these inverter libraries can be characterized rapidly. This approach is expected to have an impact on faster diffusion of logic gate based designs and in relation to this diffusion of CelloCAD in new organisms.

Using standard libraries for characterization purposes emerge chances to make comparisons between hosts. As it is discussed in Chapter 3 of this Thesis, context effect is a parameter to consider for circuit designs. Standard libraries are cut out for this purpose. This may expand our understanding on host effect and interoperability concept for parts and circuits. Hence, by preparing SEVA circuit design libraries, initial step is taken towards standardization of logic based circuit designs, and also opened doors for intensifying efforts on host and context influence in circuit design. *Pseudomonas putida* KT2440 was used for the proof of concept testing of our libraries. This showed that the use of automation tool CelloCAD can be expanded towards new organisms in a fast fashion. We argue that standardization step taken on logic-based operations here can serve towards further momentum. In this regard, having this standardized, broad host range and ready to use inverter package in SEVA format would help to automatize circuitry designs.




Having a standardized automation library for biocircuit design may help to shed light on host related effects on performing circuits. Circuitry parts are mainly reduced to DNA elements and mostly circuit design process takes into account solely DNA, although living matter is composed of a more complex intracellular environment [4, 59]. This surely shapes how to think about genotype (circuit) to phenotype (circuit function) connections. A central aspect of synthetic biology is engineering the living matter. As intending so, conversion of molecular systems into already known mechanistic pathways helps to simplify biological concepts. That is, new parts are introduced into the engineering assemblage for the use of synthetic biologists. Up to now revealed borders of circuit design is tested with growing new applications. Over-reductionist approach that focuses on solely (even though standardized) DNA parts in order to portray circuitry performance ultimately is obliged to elaborate on [19]. Due to the nature of biological matter it is essential to consider genetic background in which the circuit performs [20]. This is also known as context dependencies or recently coined term 'host-aware' [48] design. In host-aware design, the endeavor considers phenotypic impacts, which may not be explained with simplification of the phenomenon into merely DNA sequence information, by including the context in which circuit is introduced.

Context effect has been explored in several angles. The effect on genetic control imposed by metabolism [69, 70], diffusion rate [103] and genomic location accessibility of integrated circuitry elements displayed by genomic position [67, 68], resource competition [52, 66] (e.g. transcriptional and translational machineries) in between indigenous and exogenous elements showed the indisputable presence of context as a parameter for circuit design. That is to say, accomplishing interoperability is a key challenge considering these impacts. Still, re-tuning parts of synthetic constructs in a new host (other than what it is constructed for) is a cumbersome effort [56] which is disregarding the plausible advantageous context driven phenotypes (by sole consideration of DNA genetic parts) for design of biocircuits.

Intracellular effects (e.g. cellular mechanisms) on synthetic parts are here discussed as context effect. Our results depicted a way to take advantage of context effect to improve synthetic construct designs. The way constructs are prone to context effect

defines the contextual dependencies. Hence, contextual dependencies impact biocircuit function. As an accelerating field, synthetic biology is exposed to blossom of analogies. Yet, here we dared to borrow metaphors from computer science to establish the outline of contextual dependencies. Consideration of genetic circuits as a software overlays better with its biological representation when compared to presenting them as a hardware [73]. Biocircuits perform in an environment (cellular context) like software function in computer hardware environment. Contextual dependencies differentiate the performance of software/biocircuit that would not be possible disregarding them. In this study two aspects of contextual dependencies are investigated: the backbone (carrying biocircuit) and the host (biocircuit is performing at).

135 gate-context is created by using 3 bacterial hosts, 4 backbones and 20 NOT gates (inverter logic operator). Analysis showed almost 7-fold increase in achieved gate phenotypes with addition of these two contextual dependencies as compared to disregarding context effect. Having a direct comparison between context-less and context-added analysis was enabled with the library prepared. Repercussions of context addition to the initial NOT library depicted influences of context in circuit design. Another important outcome is that having complex devices completely interoperable may be of a challenge otherwise non-possible, as we have observed that transition between contexts is not a kept function for two different gates. That is, context effect alters its impact based on the genetic elements, hence there is not a linear correlation between two inverters for instance that are transitioning from context A to context B. Another outcome of our experiments with the new library is that compatible gates (two gates connected within threshold levels) is noticeably increased little over 10-fold. That is, there are more available part combinations to design biocircuit when considered context effect. To illustrate, when a gate is analyzed with the algorithm in for example 3 backbones, analysis return with 3 phenotypes carrying the context information of backbone. Therefore, re-configuration can be done for biocircuit elements to achieve more phenotypes from the very same DNA-insert (i.e. NOT gate) under chosen contextual dependencies. Our results also showed that some compatible pairs are obtained only when two gates performed in different hosts. That is, via considering host as a design parameter with a bottom-up approach multi-cellular computation can be achieved for selected functions in a



consortium of chosen chassis. In order to achieve robust and predictable genetic circuits, this study advocates biocircuits that are designed to perform functions whilst considering cellular environment – that is, genetic circuit design with reconfigurable genetic inverters using contextual dependencies.


Engineering principles start with reproducible outcomes. Having proper tools is a must to address standardization of *P. putida*. To standardize *P. putida* and to prepare it as a standard chassis for synthetic biology, two aspects may be prioritized. That is, characterization of a reference promoter and a tool for its quantification. The reference promoter is observed through changing environmental conditions to assign most reproducible promoter. The promoter is improved with addition of a translational coupler and being single copy in chromosome (integrated at the Tn-7 transposon site) is another advantage, avoiding copy number variations. The reference promoter is picked among 10 promoters that were previously selected out of 30 synthetic promoters [85]. The identified reference promoter is complemented with an RNAP immunocapture technique by implementing one more genetic tagging of *rpoC* gene at the same strain. The combination of two tools within the same strain is to create a standardization chassis in which PoPS calculations could be done over a standard promoter. The strain is prepared in two versions of immunoprecipitation tags: a Histidine tag and a Myc tag. The pull-down technique due to its potential could be multi-use, as far as we are aware there is not presence of a second tool specifically designed to pull RNAP in *P. putida*. Therefore, after preparation of the standard strain, very same tool was extensively exploited on defining RNAP protein-protein interactions i.e. sigma factors.

Extending the use of genetic devices may benefit from exploring essentials of *P. putida*. Identification of transcriptional regulators is a key point for discovery of natural switches, logic operations and circuits. RNA Polymerase is a good target to study these regulation mechanisms, as it is in the center of transcription. Our study converted a subunit (RpoC) of RNA Polymerase of *Pseudomonas putida* KT2440 into a pull-able agent and hence created basis for immunocapture feature to work on it. To increase the options 3 different immuno-peptide tags were genetically added (into wild type *P. putida*) and presented for the disposal of researchers. Modification took place on the genome to keep the natural workings of cell. Verification steps were

realized in order to observe if any effects were created over cellular machinery: genomic verification with PCR, protein level verification with SDS-Page and Western Blot, growth verification with doubling time analysis and colony morphology analysis. Within the covered angles, no inconsistencies were found. Tag sizes together with addition of flexibility enabling glycines are relatively minuscule (~1 kDa) compared to RpoC (~155 kDa).

Epitope tagged RNAP subunit *rpoC* is developed to work on RNAP interacting proteins of *P. putida*. The experimental procedure was tested for common growth conditions and several stress conditions. The outcome is verified by using common databases (e.g. string-db, uniprot). For given conditions immunocaptured proteins showed that predicted interactions can be re-shown and moreover new interactions can be captured. SDS-Page band extractions were analyzed with MALDI-TOF which showed e.g. ClpB interaction with RNAP. ClpB is known to be interacting with DnaK [89] that is an RNAP interacting protein. In this case the interaction seen supports the idea that retrieving secondary interactions of RNAP is also possible. In addition, a high-throughout protein identification technique (e.g. LC-ESI-MS/MS) was used for quantitative analysis over the three of the conditions defined. The results were converted into quantitative information to analyze proteomics data computationally. Fold change values between selected stress conditions (heat-shock vs. solvent stress) showed that interactome flow could be quantitatively tracked. That is, expressed proteins and possible affiliations –based on the environmental stress or growth condition– can be matched. Increasing the number of tested conditions would help to create sufficient background for experimental identification of RNAP interacting putative proteins i.e. transcription factors, sigma factors etc.

Immunocapture of epitope tagging of RNAP may be useful for understanding cellular machineries. Evaluation of 42C heat-shock vs. 1-3-PD solvent stress gave insights about bacteria anti-stress endeavor. 5 proteins having largest fold changes are shared after heat-shock or solvent stress application. KEGG pathways such as citrate cycle, biosynthesis of antibiotics and microbial metabolism in diverse environments for 10 highest fold change values were shared between two stresses.



Sampling of RNA Polymerase interactome of *Pseudomonas putida* is accomplished under various conditions (i.e. for growth, for stress) that are considered common for research applications. Here, it is advocated that this would be handy to shed light on transcriptional regulation unknowns of *P. putida* and to accelerate the upgrade of *P. putida* as a synthetic biology chassis.

Pseudomonas putida, with genetic circuitry design advancements and tools developed for exploration of cellular functions related to transcriptional machinery, was upgraded for the predictability and quantitative studies and therefore increased its chances as a model organism for deep engineering. *Pseudomonas putida* may be advocated as an upgraded chassis for synthetic biology due to the progresses realized and novelties added into the bacterium as a result of this Thesis.

VIII. CONCLUSIONS

VIII. CONCLUSIONS

This work has led us to the following conclusions:

1. The Broad Host Range (BHR) libraries introduced in SEVA format expands accessibility of available NOT logic gates for cellular computing and the applicability of circuit design automation processes with CelloCAD in Gram-negative bacteria.
2. A NOR gate designed via using CelloCAD and is predicted to not comply with *Escherichia coli* was effectively implemented in *Pseudomonas putida* and showed a clear NOR gate performance.
3. Considering contextual-variability as a parameter for genetic circuit design increases the number of possible combinations per genetic construct and the available configuration of gates to obtain novel response functions.
4. BG37 synthetic promoter performs as an orthogonal constitutive promoter in *Pseudomonas putida* under tested typical media conditions and *P. putida* KT_BG37 strain is introduced as a reference strain for comparative gene expression analyses.
5. An *in vivo* immunocapture approach samples the effect of distinctive environmental stresses undergone by *P. putida* in the configuration of the RNAP interactome and yields exploration of a suite of sigma factors, transcriptional regulators and auxiliary proteins with a relative composition ruled by the physiological state of cells.

IX. REFERENCES

IX. REFERENCES

1. Xin, F., et al., *Chapter 9 - Biosynthetic Technology and Bioprocess Engineering*, in *Current Developments in Biotechnology and Bioengineering*, S.P. Singh, et al., Editors. 2019, Elsevier. 207-232.
2. Kumar, J., L.K. Narnoliya, and A. Alok, *Chapter 6 - A CRISPR Technology and Biomolecule Production by Synthetic Biology Approach*, in *Current Developments in Biotechnology and Bioengineering*, S.P. Singh, et al., Editors. 2019, Elsevier. 143-161.
3. Stanton, B.C., et al., *Genomic mining of prokaryotic repressors for orthogonal logic gates*. *Nat Chem Biol*, 2014. **10**: 99-105.
4. Nielsen, A.A., et al., *Genetic circuit design automation*. *Science*, 2016. **352**: aac7341.
5. Miyamoto, T., et al., *Synthesizing biomolecule-based Boolean logic gates*. *ACS Synth Biol*, 2013. **2**: 72-82.
6. Xiang, Y., N. Dalchau, and B. Wang, *Scaling up genetic circuit design for cellular computing: advances and prospects*. *Nat Comput*, 2018. **17**: 833-853.
7. Watanabe, L., et al., *iBioSim 3: A Tool for Model-Based Genetic Circuit Design*. *ACS Synth Biol*, 2019. **8**: 1560-1563.
8. Mohammadi, P., N. Beerenwinkel, and Y. Benenson, *Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy*. *Cell Syst*, 2017. **4**: 207-218.
9. Hillson, N.J., R.D. Rosengarten, and J.D. Keasling, *j5 DNA assembly design automation software*. *ACS Synth Biol*, 2012. **1**: 14-21.
10. Taketani, M., et al., *Genetic circuit design automation for the gut resident species *Bacteroides thetaiotaomicron**. *Nat Biotechnol*, 2020. **38**: 962-969.
11. Wang, B., et al., *Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology*. *Nat Commun*, 2011. **2**: 1-9.
12. Amos, M. and A. Goñi-Moreno, *Cellular Computing and Synthetic Biology*, in *Computational Matter*, S. Stepney, S. Rasmussen, and M. Amos, Editors. 2018, Cham: Springer International Publishing. 93-110.
13. Slomovic, S., K. Pardee, and J.J. Collins, *Synthetic biology devices for in vitro and in vivo diagnostics*. *Proc Natl Acad Sci USA*, 2015. **112**: 14429-14435.
14. de Lorenzo, V., et al., *The power of synthetic biology for bioproduction, remediation and pollution control*. *EMBO Rep*, 2018. **19**: 45658
15. Dvořák, P., et al., *Bioremediation 3.0: engineering pollutant-removing bacteria in the times of systemic biology*. *Biotechnol Adv*, 2017. **35**: 845-866.
16. Conde-Pueyo, N., et al., *Synthetic Biology for Terraformation Lessons from Mars, Earth, and the Microbiome*. *Life*, 2020. **10**: 14.
17. Inniss, M.C. and P.A. Silver, *Building synthetic memory*. *Curr Biol*, 2013. **23**: 812-816.

18. Grozinger, L., et al., *Pathways to cellular supremacy in biocomputing*. Nat Commun, 2019. **10**: 1-11.
19. Cardinale, S. and A.P. Arkin, *Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems*. Biotechnol J, 2012. **7**: 856-866.
20. Phillips, K.N.O., et al., *Diversity in lac Operon Regulation among Diverse Escherichia coli Isolates Depends on the Broader Genetic Background but Is Not Explained by Genetic Relatedness*. mBio, 2019. **10**: 2232-22319.
21. Gyorgy, A., et al., *Isocost lines describe the cellular economy of genetic circuits*. Biophys J, 2015. **109**: 639-646.
22. Ceroni, F., et al., *Quantifying cellular capacity identifies gene expression designs with reduced burden*. Nat Methods, 2015. **12**: 415.
23. Institute of, M., *The Science and Applications of Synthetic and Systems Biology: Workshop Summary*, R.C. Eileen, A.R. David, and P. Leslie, Editors. 2011, Washington, DC: The National Academies Press.
24. Kim, J., et al., *Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis*. Essays Biochem, 2016. **60**: 303-313.
25. Goñi-Moreno, A. and M. Amos, *A reconfigurable NAND/NOR genetic logic gate*. BMC Syst Biol, 2012. **6**: 126.
26. Kelly, J.R., et al., *Measuring the activity of BioBrick promoters using an in vivo reference standard*. J Biol Eng, 2009. **3**: 4.
27. Canton, B., A. Labno, and D. Endy, *Refinement and standardization of synthetic biological parts and devices*. Nat Biotechnol, 2008. **26**: 787-93.
28. Nikel, P.I. and V. de Lorenzo, *Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism*. Metab Eng, 2018. **50**: 142-155.
29. Martinez-Garcia, E. and V. de Lorenzo, *Pseudomonas putida in the quest of programmable chemistry*. Curr Opin Biotechnol, 2019. **59**: 111-121.
30. Sanchez-Pascuala, A., et al., *Functional implementation of a linear glycolysis for sugar catabolism in Pseudomonas putida*. Metab Eng, 2019. **54**: 200-211.
31. Nikel, P.I. and V. de Lorenzo, *Assessing Carbon Source-Dependent Phenotypic Variability in Pseudomonas putida*. Methods Mol Biol, 2018. **1745**: 287-301.
32. Espeso, D.R., et al., *Dynamics of Pseudomonas putida biofilms in an upscale experimental framework*. J Ind Microbiol Biotechnol, 2018. **45**: 899-911.
33. Belda, E., et al., *The revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis*. Environ Microbiol, 2016. **18**: 3403-3424.
34. Aparicio, T., V. de Lorenzo, and E. Martinez-Garcia, *CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in Pseudomonas putida*. Biotechnol J, 2018. **13**: 1700161.

35. Akkaya, O., et al., *The Metabolic Redox Regime of Pseudomonas putida Tunes Its Evolvability toward Novel Xenobiotic Substrates*. mBio, 2018. **9**: 1512-1518.
36. Volke, D.C., et al., *Physical decoupling of XylS/Pm regulatory elements and conditional proteolysis enable precise control of gene expression in Pseudomonas putida*. Microb Biotechnol, 2020. **13**: 222-232.
37. Makart, S., M. Heinemann, and S. Panke, *Characterization of the AlkS/P(alkB)-expression system as an efficient tool for the production of recombinant proteins in Escherichia coli fed-batch fermentations*. Biotechnol Bioeng, 2007. **96**: 326-36.
38. Park, J. and H.H. Wang, *Systematic and synthetic approaches to rewire regulatory networks*. Curr Opin Syst Biol, 2018. **8**: 90-96.
39. Bates, S.R. and S.R. Quake, *Mapping of protein-protein interactions of E. coli RNA polymerase with microfluidic mechanical trapping*. PLoS One, 2014. **9**: 91542.
40. Calero, P. and P.I. Nikel, *Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms*. Microb Biotechnol, 2019. **12**: 98-124.
41. Szafranski, P., C.L. Smith, and C.R. Cantor, *Principal transcription sigma factors of Pseudomonas putida strains mt-2 and G1 are significantly different*. Gene, 1997. **204**: 133-8.
42. Slusarczyk, A.L., A. Lin, and R. Weiss, *Foundations for the design and implementation of synthetic genetic circuits*. Nat Rev Genet, 2012. **13**: 406-20.
43. Shin, J., et al., *Programming Escherichia coli to function as a digital display*. Mol Syst Biol, 2020. **16**: 9401.
44. Martinez-Garcia, E., et al., *SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts*. Nucleic Acids Res, 2020. **48**: 1164-1170.
45. Martinez-Garcia, E., et al., *SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities*. Nucleic Acids Res, 2015. **43**: 1183-1189.
46. Nikel, P.I., E. Martinez-Garcia, and V. de Lorenzo, *Biotechnological domestication of pseudomonads using synthetic biology*. Nat Rev Microbiol, 2014. **12**: 368-379.
47. Khan, N., et al., *A broad-host-range event detector: expanding and quantifying performance between Escherichia coli and Pseudomonas species*. Synthetic Biology, 2020. **5**. ysaa002.
48. Boo, A., T. Ellis, and G.B. Stan, *Host-aware synthetic biology*. Curr Opin Syst Biol, 2019. **14**: 66-72.
49. Beal, J., et al., *Reproducibility of Fluorescent Expression from Engineered Biological Constructs in E. coli*. PLoS One, 2016. **11**: 0150182.
50. Tas, H., Á. Goñi-Moreno, and V. de Lorenzo, *A standardized broad host range inverter package for genetic circuitry design in Gram-negative bacteria*. bioRxiv, 2020: 2020.07.14.202754.

51. Tas, H., et al., *Contextual dependencies expand the re-usability of genetic inverters*. bioRxiv, 2020: 2020.07.15.204651.
52. Darlington, A.P., et al., *Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes*. Nat Commun, 2018. **9**: 1-12.
53. Stoof, R., et al., *FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering*. bioRxiv, 2020: 2020.07.23.217869.
54. Darlington, A. and D.G. Bates. *Host-aware modelling of a synthetic genetic oscillator*. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2016.
55. Ueki, T., et al., *Genetic switches and related tools for controlling gene expression and electrical outputs of Geobacter sulfurreducens*. J Ind Microbiol Biotechnol, 2016. **43**: 1561-1575.
56. Kushwaha, M. and H.M. Salis, *A portable expression resource for engineering cross-species genetic circuits and pathways*. Nat Commun, 2015. **6**: 1-11.
57. Prindle, A., et al., *Rapid and tunable post-translational coupling of genetic circuits*. Nature, 2014. **508**: 387-91.
58. Gorochoowski, T.E., et al., *Genetic circuit characterization and debugging using RNA-seq*. Mol Syst Biol, 2017. **13**: 952.
59. Appleton, E., et al., *Design automation in synthetic biology*. Cold Spring Harb Perspect in Biol, 2017. **9**: a023978.
60. Regot, S., et al., *Distributed biological computation with multicellular engineered networks*. Nature, 2011. **469**: 207-11.
61. Kylilis, N., et al., *Tools for engineering coordinated system behaviour in synthetic microbial consortia*. Nat Commun, 2018. **9**: 2677.
62. Goñi-Moreno, A., et al., *Biocircuit design through engineering bacterial logic gates*. Nat Comput, 2011. **10**: 119-127.
63. Kittleson, J.T., G.C. Wu, and J.C. Anderson, *Successes and failures in modular genetic engineering*. Curr Opin Chem Biol, 2012. **16**: 329-336.
64. Ausländer, S., D. Ausländer, and M. Fussenegger, *Synthetic biology—the synthesis of biology*. Angew Chem Int Ed, 2017. **56**: 6396-6419.
65. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline*. Mol Syst Biol, 2006. **2**: 28.
66. Nikolados, E.-M., et al., *Growth defects and loss-of-function in synthetic gene circuits*. ACS Synth Biol, 2019. **8**: 1231-1240.
67. Stoof, R., A. Wood, and A. Goni-Moreno, *A Model for the Spatiotemporal Design of Gene Regulatory Circuits*. ACS Synth Biol, 2019. **8**: 2007-2016.
68. Goñi-Moreno, A.n., et al., *Deconvolution of gene expression noise into spatial dynamics of transcription factor–promoter interplay*. ACS Synth Biol, 2017. **6**: 1359-1369.

69. Oyarzún, D.A. and G.-B.V. Stan, *Synthetic gene circuits for metabolic control: design trade-offs and constraints*. J R Soc Interface, 2013. **10**: 20120671.
70. Goñi-Moreno, A. and P.I. Nikel, *High-performance biocomputing in synthetic biology—integrated transcriptional and metabolic circuits*. Front Bioeng Biotechnol, 2019. **7**: 40.
71. Couto, J.M., et al., *The effect of metabolic stress on genome stability of a synthetic biology chassis Escherichia coli K12 strain*. Microb Cell Fact, 2018. **17**: 8.
72. de Lorenzo, V., *Beware of metaphors: chassis and orthogonality in synthetic biology*. Bioeng Bugs, 2011. **2**: 3-7.
73. Danchin, A., *Bacteria as computers making computers*. FEMS Microb Rev, 2008. **33**: 3-26.
74. Gorochowski, T.E., et al., *Toward Engineering Biosystems With Emergent Collective Functions*. Front Bioeng Biotechnol, 2020. **8**: 705.
75. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python*. Nat Methods, 2020. **17**: 261-272.
76. Fréchet, M.M., *Sur quelques points du calcul fonctionnel*. Rend Circ Mat Palermo, 1906. **22**: 1-72.
77. Jekel, C.F., et al., *Similarity measures for identifying material parameters from hysteresis loops using inverse analysis*. Int J Mater Form, 2019. **12**: 355-378.
78. Tas, H., et al., *The synthetic microbiology caucus: are synthetic biology standards applicable in everyday research practice?* Microb Biotechnol, 2020. **13**: 1304-1308.
79. Muller, K.M. and K.M. Arndt, *Standardization in synthetic biology*. Methods Mol Biol, 2012. **813**: 23-43.
80. Hillson, N.J., et al., *Improving Synthetic Biology Communication: Recommended Practices for Visual Depiction and Digital Submission of Genetic Designs*. ACS Synth Biol, 2016. **5**: 449-51.
81. Hecht, A., et al., *A minimum information standard for reproducing bench-scale bacterial cell growth and productivity*. Commun Biol, 2018. **1**: 219.
82. de Lorenzo, V. and M. Schmidt, *Biological standards for the Knowledge-Based BioEconomy: What is at stake*. N Biotechnol, 2018. **40**: 170-180.
83. Beal, J., et al., *The long journey towards standards for engineering biosystems: Are the Molecular Biology and the Biotech communities ready to standardise?* EMBO Rep, 2020. **21**: 50521.
84. Arkin, A., *Setting the standard in synthetic biology*. Nat Biotechnol, 2008. **26**: 771-774.
85. Zobel, S., et al., *Tn7-Based Device for Calibrated Heterologous Gene Expression in Pseudomonas putida*. ACS Synth Biol, 2015. **4**: 1341-1351.

86. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. Nucleic Acids Res, 2019. **47**: 607-613.
87. Martinez-Garcia, E. and V. de Lorenzo, *Transposon-based and plasmid-based genetic tools for editing genomes of gram-negative bacteria*. Methods Mol Biol, 2012. **813**: 267-83.
88. Tas, H., et al., *An Integrated System for Precise Genome Modification in Escherichia coli*. PLoS One, 2015. **10**: 0136963.
89. Seyffer, F., et al., *Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces*. Nat Struct Mol Biol, 2012. **19**: 1347-1355.
90. Chuang, S.E. and F.R. Blattner, *Characterization of twenty-six new heat shock genes of Escherichia coli*. J Bacteriol, 1993. **175**: 5242-52.
91. Qamra, R., V. Srinivas, and S.C. Mande, *Mycobacterium tuberculosis GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins*. J Mol Biol, 2004. **342**: 605-17.
92. Kumar, C.M., et al., *Facilitated oligomerization of mycobacterial GroEL: evidence for phosphorylation-mediated oligomerization*. J Bacteriol, 2009. **191**: 6525-38.
93. Mika, F. and R. Hengge, *A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of sigmaS (RpoS) in E. coli*. Genes Dev, 2005. **19**: 2770-81.
94. Chang, D.E., D.J. Smalley, and T. Conway, *Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model*. Mol Microbiol, 2002. **45**: 289-306.
95. Zhang, G. and S.A. Darst, *Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain*. Science, 1998. **281**: 262-6.
96. Igarashi, K. and A. Ishihama, *Bipartite functional map of the E. coli RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP*. Cell, 1991. **65**: 1015-22.
97. Hu, Y., et al., *Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the sigma A-containing RNA polymerase holoenzyme*. Nucleic Acids Res, 2012. **40**: 6547-57.
98. Degen, D., et al., *Transcription inhibition by the depsipeptide antibiotic salinamide A*. Elife, 2014. **3**: 02451.
99. Held, W.A., et al., *Assembly mapping of 30 S ribosomal proteins from Escherichia coli. Further studies*. J Biol Chem, 1974. **249**: 3103-11.
100. Sendy, B., et al., *RNA polymerase supply and flux through the lac operon in Escherichia coli*. Philos Trans R Soc Lond B Biol Sci, 2016. **371**: 20160080.
101. Bodega, G., et al., *Young and Especially Senescent Endothelial Microvesicles Produce NADPH: The Fuel for Their Antioxidant Machinery*. Oxid Med Cell Longev, 2018. **2018**: 3183794.

102. Arcos, S.C., et al., *Proteomic profiling and characterization of differential allergens in the nematodes Anisakis simplex sensu stricto and A. pegreffii*. *Proteomics*, 2014. **14**: 1547-1568.
103. Kuhlman, T.E. and E.C. Cox, *DNA-binding-protein inhomogeneity in E. coli modeled as biphasic facilitated diffusion*. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2013. **88**: 022701.

X. ANNEXES

X. ANNEXES

I. Decomposing fluorescence and scattering

Fluorescence values are decomposed in two components: a part that is experiment dependent and a part that is due to volume variation and disagreement.

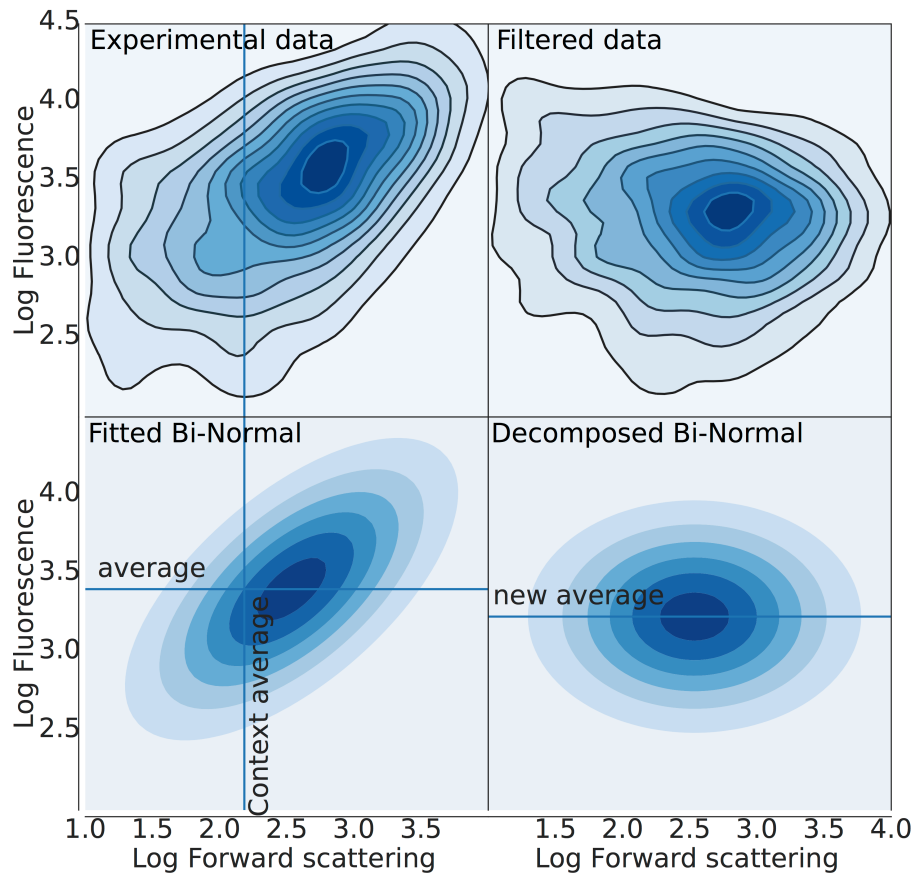


Figure S 1: Example of filtering of flow cytometry data.

Top left, gated Experimental data. Bottom left, experimental data is fitted to a bi-normal function. These fits are done for all experiments. Vertical line indicates mean scattering with the same context. Bottom right, scattering decomposition is calculated on and applied fit. Top right, this transformation is applied as a filter to the experimental data.

II. Compatibility tables for individual strains

In order that the genetic circuits may be treated as logic NOT gates, the continuous variable (experimentally obtained standardized fluorescence) must be interpreted as a discrete variable (representing logic 1 or 0). Thresholds are therefore required to partition the input and output fluorescence values into groups that are to be interpreted as a logic value, or rejected as ambiguous. Compatibility between two gates is a qualitative measure of the agreement of these thresholds. In particular, the output thresholds of the ‘input gate’ must not lie in the group of inputs that would be rejected as ambiguous by the ‘output gate’. Smaller ambiguous regions will increase the numbers of compatible pairs *in silico*, but circuits built from such pairs may behave unpredictably in the presence of noise or measurement error of a real system. Larger ambiguous regions will guard against the effect of noise, at the cost of flexibility in design. In this study a thresholding scheme is used that has been used previously for the same library [4]. Further, two gates are considered that they may only be connected if they are compatible, and a library with many compatible gates is desirable.

Here the compatibility is shown between pairs of gates in individual strains, illustrating what can be achieved by incorporating backbone as a design parameter, without changing host. The superior performance found in the DH5 α and CC118 λ *pir* *E. coli* strains, in comparison to the *P. putida* strain KT2440, may reflect the fact that the library components were initially selected for use in an *E. coli* host, and that choice of host significantly impacts behavior of genetic parts.

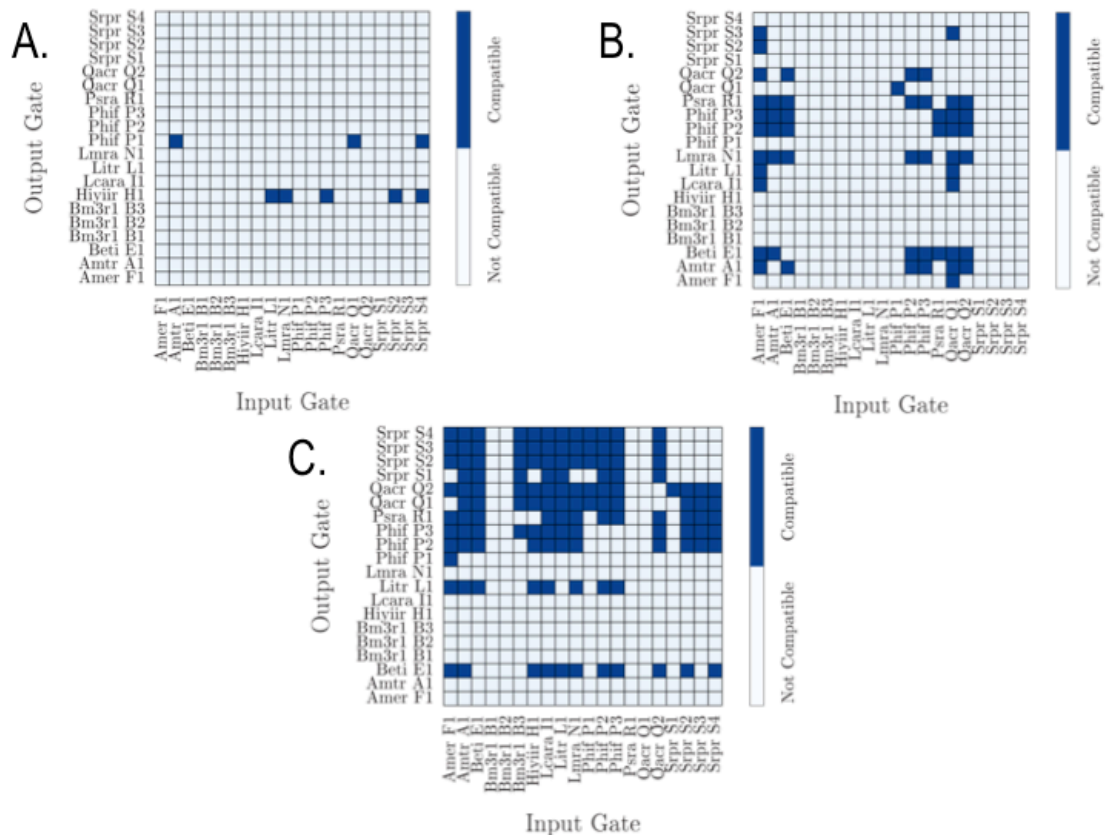


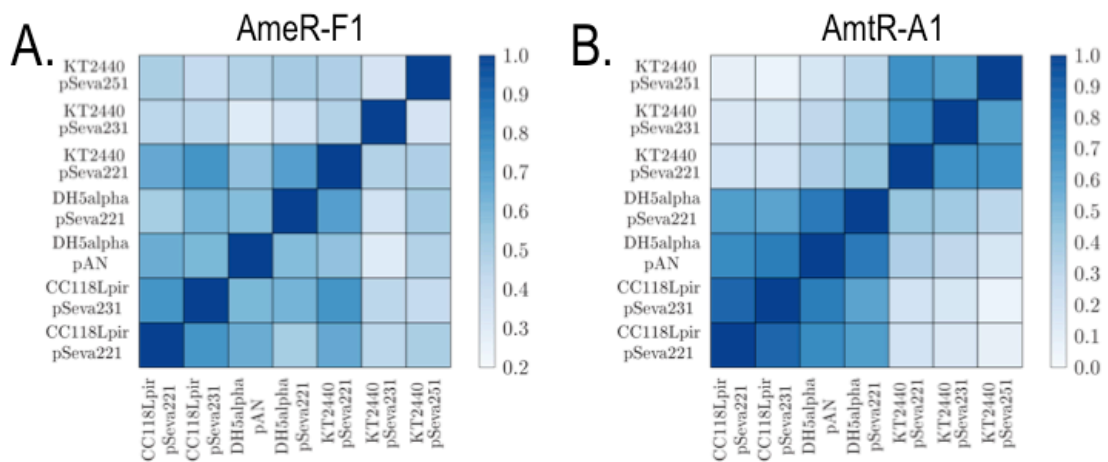
Figure S 2: Compatibility tables for the library in different hosts.

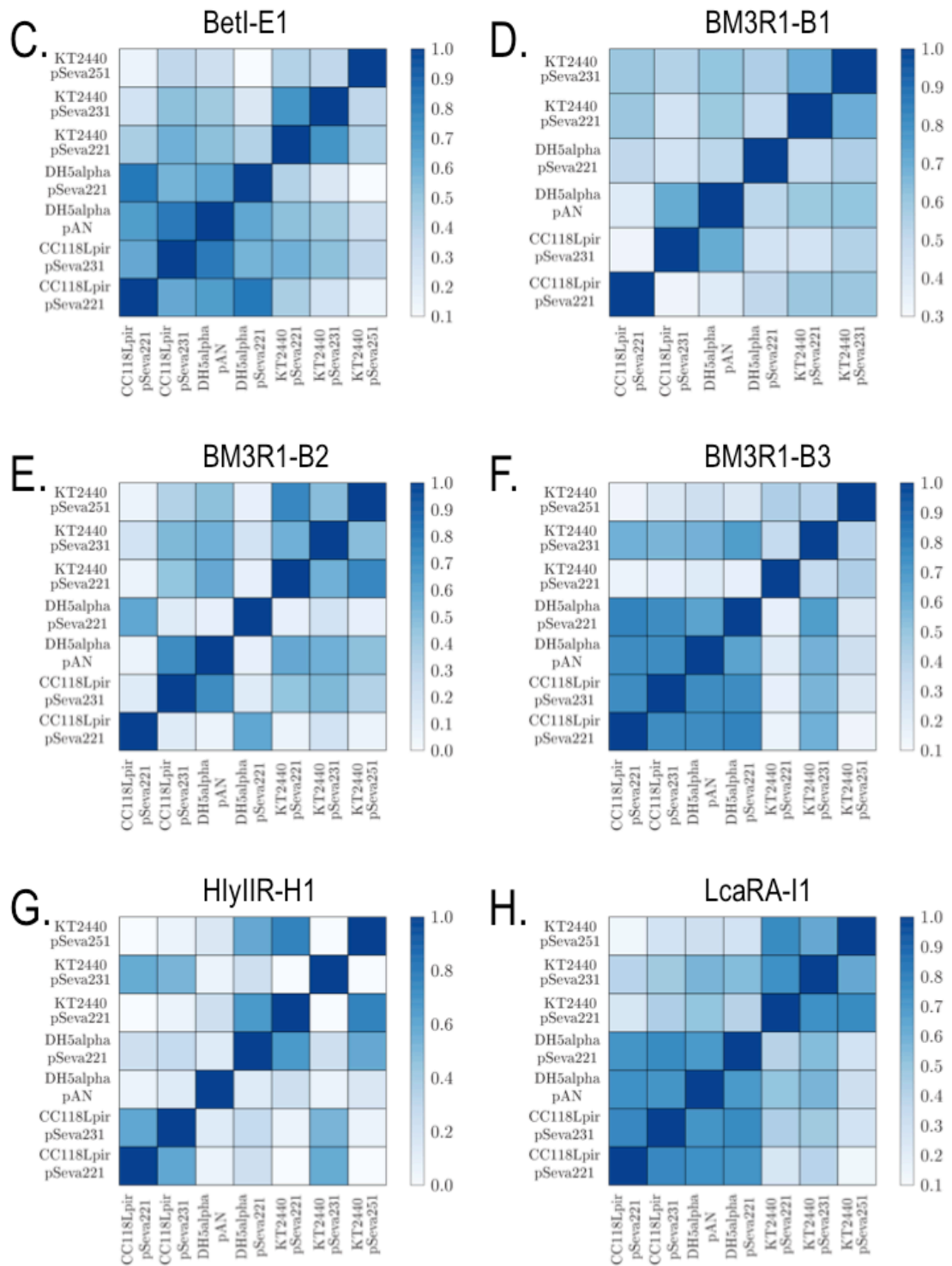
Input gate on the x-axis is the first gate, whose output provides the input for ‘Output gate’ on the y-axis. Two gates are compatible if their thresholds agree, and they do not use the same repressor molecule. Compatibility table for gates characterized in **A.** KT2440 **B.** CC118 λ pir **C.** DH5 α .

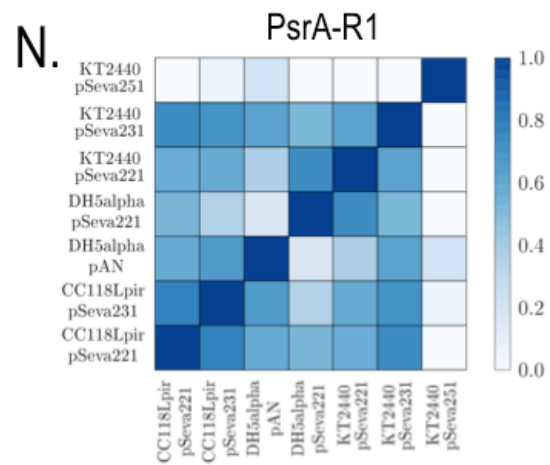
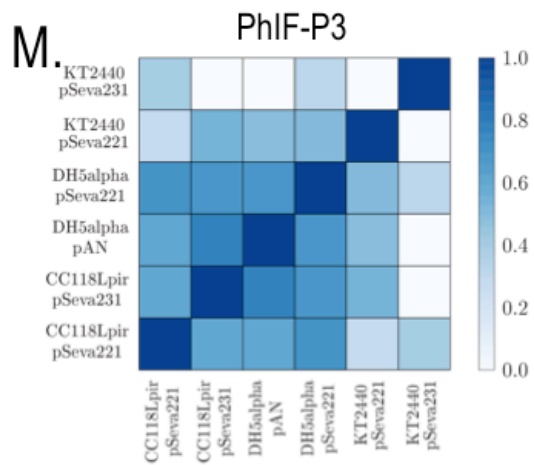
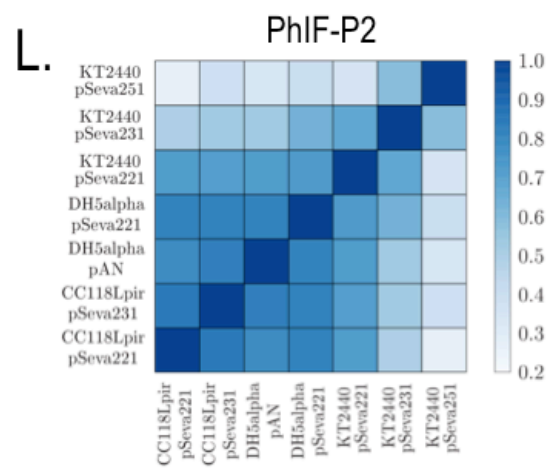
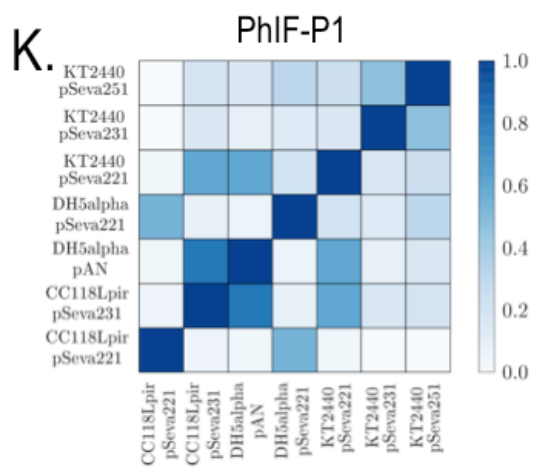
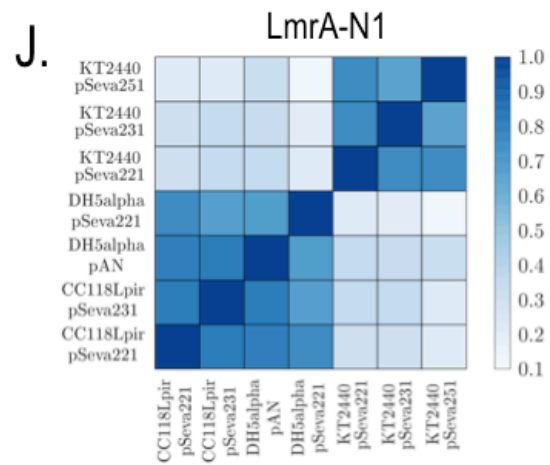
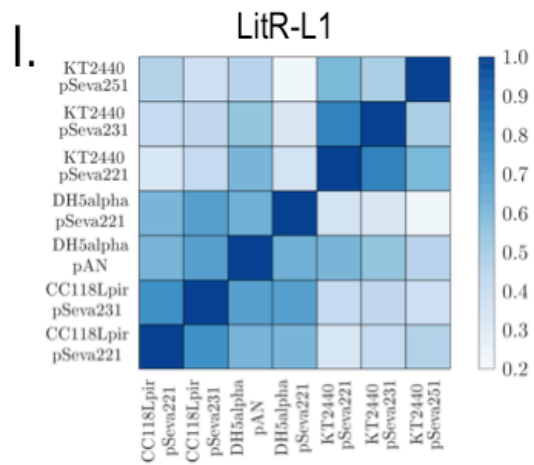
III. Context Similarity

This study found that the same genetic logic gate can exhibit differing behavior depending upon the context in which it was situated. The characterization of the genetic logic gate may be both quantitatively and qualitatively different. Qualitative changes, such as to the shape of the response curve, are particularly interesting when considering the effects of context-circuit interplay, because they suggest these interactions are nonlinear phenomena.

Changes are attempted to be quantified in curve shape using a similarity measure as described in V. CHAPTER . Log transformation of the curve ensures that deviations in the upper regions of input and output are not disproportionately penalized. The min-max normalization in both input and output dimensions captures the shape information of the curve. Methods based on comparison of the gradient of the curves were also considered, and produce similar results.







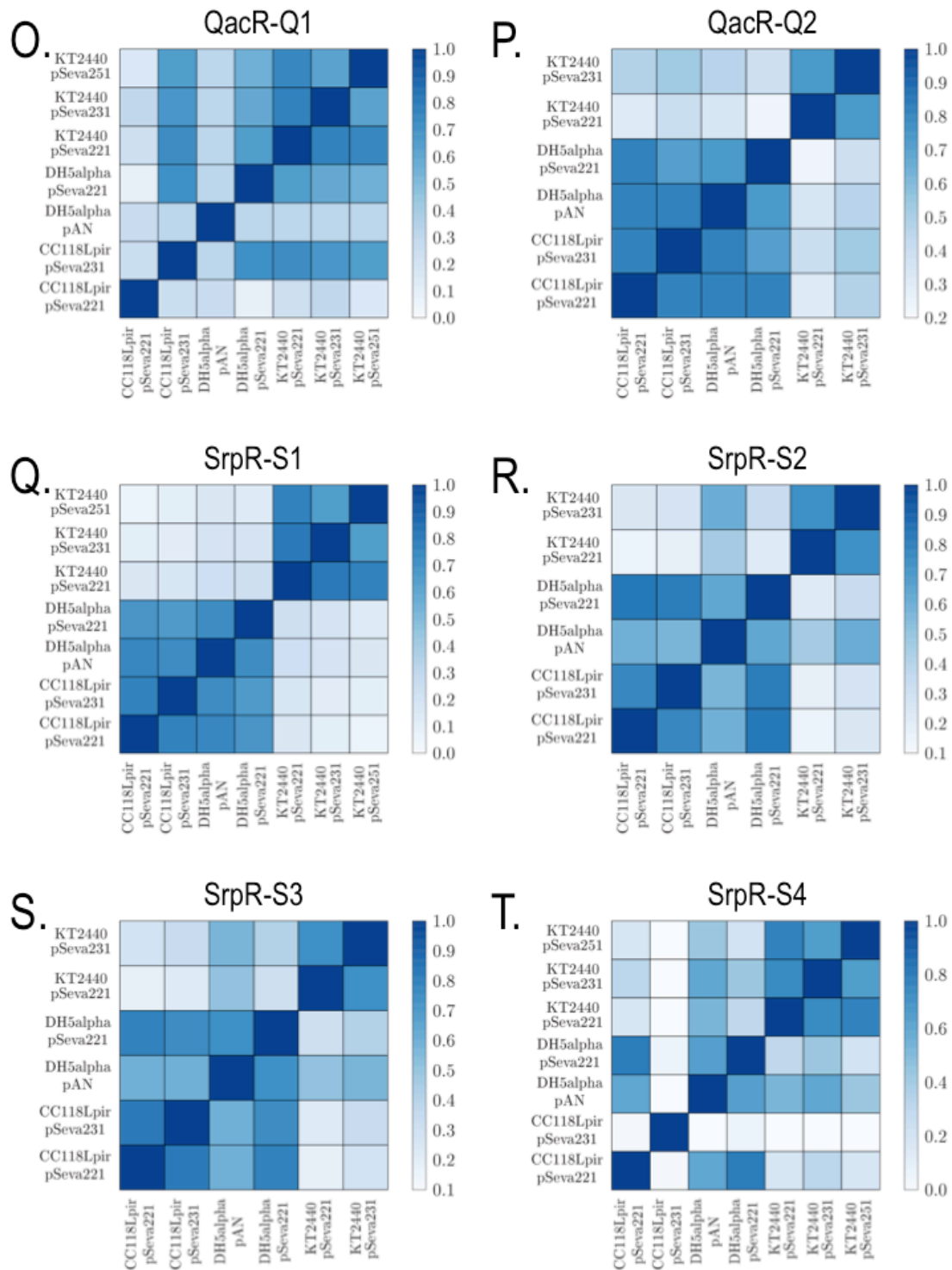


Figure S 3: Similarity scores heatmap for 20 gates in 7 contexts.

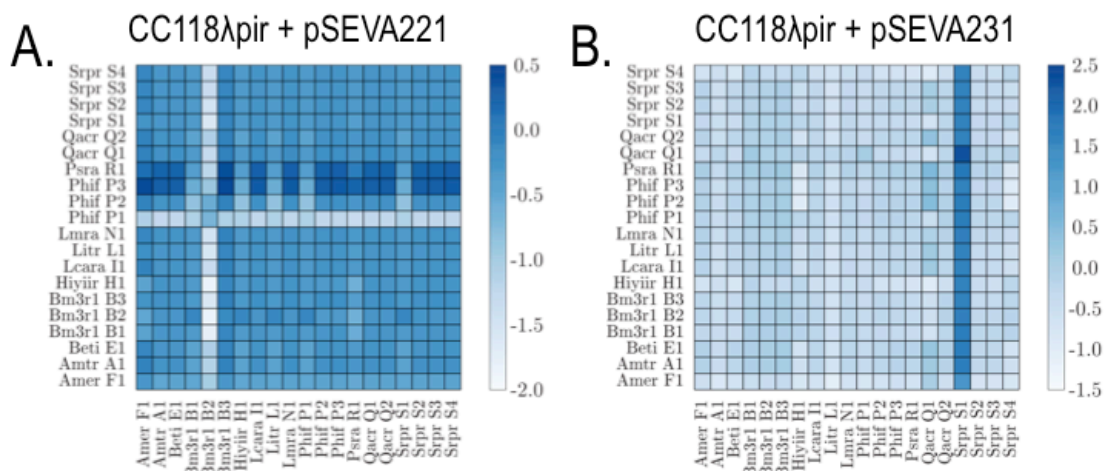
Similarity scores shown for all gates in all 7 contexts. A high similarity score (darker squares) is best. The minimum and maximum scores are 0 and 1, respectively.

IV. Compatibility Scoring

As compatibility tables indicate which pairs of gates may be connected, they offer no indication as to which pairs are most or least compatible. It may be desirable, from an optimization perspective, to select pairs of gates for which the first's output thresholds lie as far away from the second's ambiguous region as possible whose. Doing so will improve performance in spite of noise and provide a greater margin for error.

Conversely, the library may be optimized for a specific context by redesign of the parts. In this case, it would be a desire to know which pairs would be compatible with only small changes to their existing behavior, such that the reward for the optimization efforts are maximized.

A compatibility score was computed to measure these characteristics, as defined in V. CHAPTER . A positive/negative score for a pair of gates indicated the pair is compatible/incompatible. Further, a positive score represents the minimum of the maximum perturbations to the thresholds that could be tolerated, while retaining the compatibility of the pair. A negative score represents the maximum of the minimum perturbations to the thresholds that would be required to make the pair compatible that could be allowed to any of thresholds. Thus, pairs of gates can be ranked in terms of compatibility.



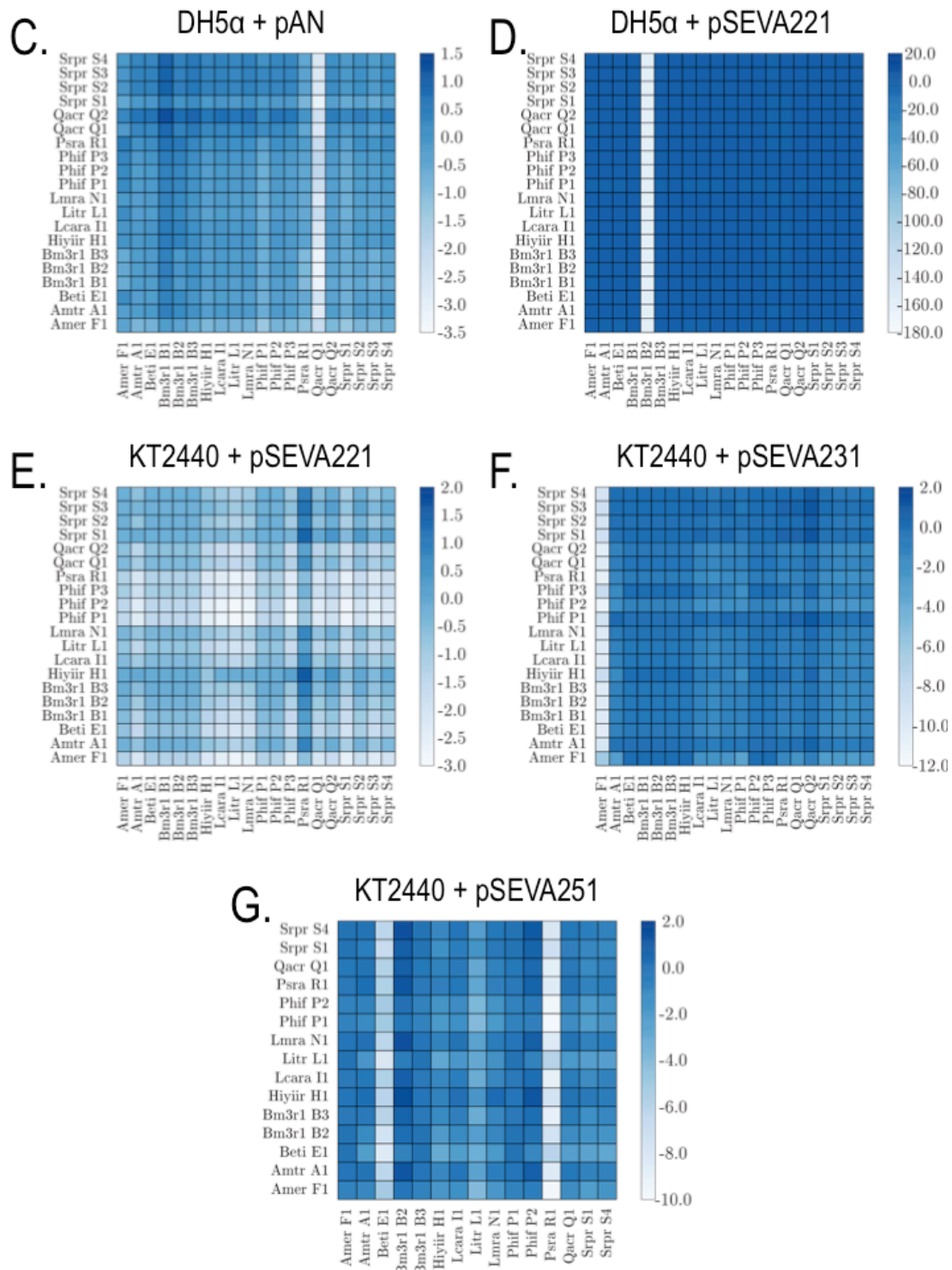


Figure S 4: Compatibility score heatmaps for gates in the 7 contexts.

Compatibility scores between pairs of gates for all contexts. Higher scores are better and indicate more compatible pairs. Negative scores indicate incompatible pairs.

V. Provided codes and data

Supplementary Data Raw and Supplementary Data Scripts (with the codes that implement the *in silico* methods) from this Thesis can be found and downloaded at:

https://fairdomhub.org/data_files/3823?version=1

In order to obtain the results from this particular study, only MATLAB_R2012b is needed (III. CHAPTER 1) and Python3.6 or greater is needed (V. CHAPTER 3). However, to generate the figures used in V. CHAPTER 3 further requires gnuplot and L^AT_EX. Once these dependencies are installed, the command:

```
python3 produce_figures.py full-update
```

once run from the project root directory will perform the analysis and place figures in the appropriate subdirectories. Docker users may find it more convenient to use the Dockerfile associated with the project.

It should also be possible to run the same analysis on a different dataset, if the data is provided as a csv file with the expected fields and format. The processed data from this study that is provided with the codes can act as a guideline.

In addition, all the corresponding scripts and data are saved in a CD format and deposited with this Thesis.

VI. List of Supplementary Tables

1. Table for Function Portability

Table S 1: Calculation of Sensor OFF/ON activities.

Promoter	State	<YFP>	<YFP>RPU	<YFP>0	RPUs
Ptac	OFF	0.63	8.44	0.61	0.003
	ON	14.8	8.44	0.61	1.812
Ptet	OFF	0.62	9.6	0.6	0.002
	ON	54.35	9.6	0.6	5.972
Pbad	OFF	1.56	9.38	0.6	0.109
	ON	113	9.38	0.6	12.802
Pbad (Glucose)	OFF	1.24	9.15	0.61	0.066
	ON	263	9.15	0.61	27.36
Pm	OFF	0.71	4.89	0.7	0.002
	ON	3.12	4.89	0.7	0.578
PalkB	OFF	5.98	4.84	0.69	1.275
	ON	9.043	4.84	0.69	2.013

Table S 2: Characterization and measurement.

Stock Name	Library	Work	Definition
Tas425	KT2440 + pSeva221::1718	This study	Inducer Characterization - Ptac/LacIq - IPTG
Tas426	KT2440 + pSeva221::1719	This study	Inducer Characterization - Ptet/TetR - aTc
Tas427	KT2440 + pSeva221::1720	This study	Inducer Characterization - Pbad/AraC - L-arabinose
Tas436	KT2440 + pSeva227YFP	This study	Autofluorescence
Tas437	KT2440 + pSeva227_Pem7_YFP	This study	RPU standard
Tas438	KT2440 + pSeva228YFP	This study	Promoter Efficiency -Pm/XylS
Tas439	KT2440 + pSeva229YFP	This study	Promoter Efficiency -PalkB/AlkS
Tas440	KT2440 + pSeva228YFP::AmeR-F1	This study	Repressor Efficiency - AmeR-F1 - with Pm/XylS
Tas441	KT2440 + pSeva229YFP::AmeR-F1	This study	Repressor Efficiency - AmeR-F1 - with PalkB/AlkS
Tas382	KT2440 + pSeva221::NOR-SrpR-S1	This study	NOR Gate - Low Copy
Tas383	KT2440 + pSeva231::NOR-SrpR-S1	This study	NOR Gate - Medium Copy

Note that Table S 2 should be consider with some of the constructs of Table S 6.

2. Table of primers

Table S 3: Complete set of primers used.

Primers	Sequence (5'→3')	Note
Gate-Ctrl- Unvrsl-1	TCTAGGGCGGCGGATTTG	The position right before the MCS. Same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl- Unvrsl-3	TGGTGAGCAAGGGCGAG	The position right after the multiplicity region of the gates This is same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl- Unvrsl-4	ACCTTAGCTACCAGTCCGC	This position is same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl- Unvrsl-5	ACAATCTTCTCGCGCAACG	This position is same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl- Unvrsl-6	CGGTGAAGGGCAATCAGCT	This position is same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl- Unvrsl-7	GAATGCATTATATGCACTCAGCGC	The last one for covering the integrated region and same for pSEVA221, pSEVA231, pSEVA251
Gate-Ctrl-2- AmeR-F1	TTTTAGCAGCAAAAACGCACTGG	The 500th position for the AmeR-F1
Gate-Ctrl-2- AmtR-A1	CCTGCTGAAAAGCACCGTTG	The 500th position for the AmtR-A1
Gate-Ctrl-2- BetI-E1	GTCTGCATGCACTGCCG	The 500th position for the BetI-E1
Gate-Ctrl-2- BM3R1- B123	CGGTCTGGCAAATGAACGTG	The 500th position for BM3R1-B1 BM3R1-B2 and BM3R1-B3 gates and can be used with these gates
Gate-Ctrl-2- HIyRII-H1	TGCCGAACTTTCTGGAAAAAACC	The 500th position for HIyRII-H1
Gate-Ctrl-2- lcaRA-I1	ACAAAAGCAACTATAGCATCGATGC	The 500th position for lcaRA-I1
Gate-Ctrl-2- litR-L1	CTGAACAAAGTGAAAAACGAGTTTCAC	The 500th position for LitR-L1 gate
Gate-Ctrl-2- LmrA-N1	AAGAATATATCCGCCAGAAAATCGC	The 500th position for LmrA-N1
Gate-Ctrl-2- PhIF-P123	AAAATGAAAGCGAACAGGTGCG	The 500th position for PhIF-P1 PhIF-P2 and PhIF-P3
Gate-Ctrl-2-	GGAACGTGAACTGGAACGTC	The 500th position for PsrA-R1

PsrA-R1		
Gate-Ctrl-2- QacR-Q123	CAAATCAAATGCAAAACCAACCGC	The 500th position for QacR-Q1 and QacR-Q2
Gate-Ctrl-2- SrpR-S1234	TCCGCTGGAAGTGGATTTTACAC	The 500th position for the SrpR-S1 SrpR-S2 SrpR-S3 and SrpR-S4 gates
PS1	AGGGCGGCGGATTTGTCC	SEVA Internal Control Primers
PS2	GCGGCAACCGAGCGTTC	SEVA Internal Control Primers
PS3	GAACGCTCGGTTGCCGC	SEVA Internal Control Primers
PS4	CCAGCCTCGCAGAGCAGG	SEVA Internal Control Primers
PS5	CCCTGCTTCGGGGTCATT	SEVA Internal Control Primers
PS6	GGACAAATCCGCCGCCCT	SEVA Internal Control Primers
pAN_pSeva F	CCTAGATTAATTAACACCCCTTGTATT ACTGTTTATGTAAGC	Forward primer to pop out the gate
pAN_pSeva R	GTCTAAACTAGTCGTCGGCGTAGAGGA TC	Reverse primer to pop out the gate
TS1F-XmaI	ATTACCCGGGCGGTAGCGAACCGTACGA AGAG	General TS1F
TS2R-XbaI	CGCGTCTAGAGACCCCTGCCGACCTTA	General TS2R
TS2F-His	GCGGCCACCACCACCACCACCACCACC ACCACCACTAAGTACAGGGCAAGGCC TTAGTGGTGGTGGTGGTGGTGGTGGTGG	His Tag specific TS2F
TS1R-His	TGGTGGCCGCCATTACCGCTGGAATTC G	His Tag specific TS2R
TS1R-Myc	CAGGTCCTCTTCGCTGATCAGCTTCTGTT CGCCGCCGCCGCCATTACCGCTGGAATTC AGCGCTTC	Myc Tag specific TS1R
TS2F-Myc	GCGGCCGCGGCGAACAGAAGCTGATCA GCGAAGAGGACCTGTAAGTACAGGGCAA GGCCCC	Myc Tag specific TS2F
TS1R-Etag	ACGCGGTTCCAGCGGGTCCGGATACGGC ACCGGCGCACCGCCGCCGCCATTAC CGCTGGAATTCAGCGCTTC	E Tag specific TS1R
TS2F-Etag	GCGGCCGCGGCGGTGCGCCGGTGCCTG ATCCGGACCCGCTGGAACCGCGTTAAGT ACAGGGCAAGGCCCC	E Tag specific TS2F
QCekF	GTAAAGAGACCAAGGGCAAGCGTCGCCT GGTGATCACTCCGACCGACGGTAGCG	General Forward primer to check insertions
QCekR-Myc	CAGGTCCTCTTCGCTGATCAGCTTCTGCT CGCCGCC	Myc Tag specific Reverse primer to check insertions
QCekR-His	GTGGTGGTGGTGGTGGTGGTGGTGGTGG	His Tag specific Reverse primer to

	TGGCCGCC	check insertions
QCek-Myc-F	GAACAGAAGCTGATCAGCGAAGAGGACC TG	Myc Tag specific Forward primer
QCek-Myc-R	CAGGTCCTCTTCGCTGATCAGCTTCTGTT C	Myc Tag specific Reverse primer
QCek-Etag-R	ACGCGGTTCCAGCGGGTCCGGATACGGC ACCGGCGCACC	E Tag specific Reverse primer
QCek-Etag-F	GGTGCGCCGGTGCCGTATCCGGACCCGC TGGAACCGCGT	E Tag specific Forward primer
QCek-F	CGTCGTCCGAAAGAAGCCTCGATT	Insertion check
QCek-R	TACTTCGGATCGTCAAGGATCTCACGTTT TGCT	Insertion check
pSW-F	GGACGCTTCGCTGAAAATA	pSW plasmids transformation and curation check
pSW-R	AACGTCGTGACTGGGAAAAC	pSW plasmids transformation and curation check
M13-F	GTAAAACGACGGCCAGT	Cloning target sequences into pEMG check
M13-R	GGAAACAGCTATGACCATG	Cloning target sequences into pEMG check
FinSeq-F	AAGTTCACCCGAGTGCTGCTGGGT	Final sequencings
FinSeq-R	CACGCCACGACGCTGCGG	Final sequencings

Note: The positions are written based of pSeva221::QacR-Q2 plasmid.

Note: PS primers, M13 primers and pSW primers were used from the literature

3. Table of Insulated gate response function parameters *P. putida* vs *E. coli*

E. coli (E) values (taken from cellocad.org) are values originally calculated in Cello study with pAN backbone. *P. putida* (P) values are calculated for this study with pSEVA221 backbone.

Table S 4: Insulated gate response function parameters *E. coli* vs *P. putida*.

Repressor	n (E)	n (P)	K (E)	K (P)	Ymax (E)	Ymax (P)	Ymin (E)	Ymin (P)	Scores (E)	Scores (P)
pSeva221_AmeR_F1	1.6039	1.3979	0.212	0.0256	13.99	13.76	0.33	8.74	7.91E-04	4.79E-04
pSeva221_AmtR_A1	3.2122	1.1279	0.1769	0.0532	2.13	1.79	0.08	0.18	5.15E-05	2.45E-04
pSeva221_BetI_E1	2.2714	1	0.5841	0.0744	2.29	2.77	0.07	1.35	1.94E-04	3.58E-04
pSeva221_BM3R1_B1	2.3822	1.3901	0.4403	0.3345	2.29	1.4	0.13	1.23	1.67E-04	1.88E-05
pSeva221_BM3R1_B2	2.4073	1.3947	0.8582	0.0778	11.62	0.76	0.29	0.31	2.04E-04	3.55E-04
pSeva221_BM3R1_B3	1.6956	1.4	0.2238	0.09	12.62	0.28	1.15	0.23	7.03E-04	3.83E-05
pSeva221_HIYIR_H1	2.8351	1.4003	0.4909	0.031	8.15	0.33	0.36	0.03	8.40E-05	6.24E-05
pSeva221_lcaRA_I1	1.686	1.0302	0.2209	0.0464	9.93	5.7	0.41	0.48	6.41E-04	0.0015
pSeva221_LitR_L1	2.0769	1.1814	0.1299	0.061	15.55	3.62	0.38	0.33	0.0019	6.09E-04
pSeva221_LmrA_N1	2.2534	1.0197	0.3919	0.0429	7.45	3.63	0.79	0.29	1.30E-04	0.001
pSeva221_PhIF_P1	4.5298	1.3983	0.141	0.0199	17.76	17.73	0.1	12.31	1.75E-04	2.70E-04
pSeva221_PhIF_P2	3.4206	1.3985	0.538	0.0176	18.84	20.72	0.12	11.6	8.06E-04	6.29E-04
pSeva221_PhIF_P3	3.6435	1	0.4543	0.0516	23.94	28.02	0.13	3.92	0.0012	0.0065
pSeva221_PsrA_R1	2.5288	1	0.2157	0.1284	11.31	20.16	0.11	8.62	1.64E-04	0.0015
pSeva221_QacR_Q1	2.3025	1.0669	0.4532	0.0714	12.88	4.12	0.15	0.22	2.43E-04	0.0017
pSeva221_QacR_Q2	1.9083	1	0.4048	0.0665	22.73	4.99	0.77	1.18	3.08E-04	0.0018
pSeva221_SrpR_S1	7.8788	1.3967	0.0923	0.0452	5.77	1.13	0.07	0.03	5.24E-04	3.92E-04
pSeva221_SrpR_S2	2.8736	1	0.2554	0.0836	8.89	1.3	0.07	0.27	1.57E-04	2.57E-04
pSeva221_SrpR_S3	2.865	1.0239	0.1851	0.1064	20.29	1.13	0.19	0.07	2.29E-04	4.11E-04
pSeva221_SrpR_S4	2.5123	1	0.3868	0.0906	9.34	1.16	0.08	0.12	2.31E-04	7.72E-04

4. Tables of constructs.

Supplementary tables that are not included in the main text can be found here.

4.1. List of gates from the Cello work.

Table S 5: 12 main gates and 8 variants that were obtained from the Cello work.

Inverters		Study
pAN::AmeR-F1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::AmtR-A1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::BetI-E1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::BM3R1-B1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::BM3R1-B2	NOT Gate	Nielsen <i>et al</i> ¹
pAN::BM3R1-B3	NOT Gate	Nielsen <i>et al</i> ¹
pAN::HlyIIR-H1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::lcaRA-I1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::LitR-L1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::LmrA-N1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::PhIF-P1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::PhIF-P2	NOT Gate	Nielsen <i>et al</i> ¹
pAN::PhIF-P3	NOT Gate	Nielsen <i>et al</i> ¹
pAN::PsrA-R1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::QacR-Q1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::QacR-Q2	NOT Gate	Nielsen <i>et al</i> ¹
pAN::SrpR-S1	NOT Gate	Nielsen <i>et al</i> ¹
pAN::SrpR-S2	NOT Gate	Nielsen <i>et al</i> ¹
pAN::SrpR-S3	NOT Gate	Nielsen <i>et al</i> ¹
pAN::SrpR-S4	NOT Gate	Nielsen <i>et al</i> ¹
pAN::1201	Autofluorescence	Nielsen <i>et al</i> ¹
pAN::1717	Standardization	Nielsen <i>et al</i> ¹
pAN::1818	Promoter Activity	Nielsen <i>et al</i> ¹

4.2. List of context dependent inverters

Table S 6: List of all libraries for context dependent inverters.

Stock Name	Library	Work	Definition
Tas74	<i>E.coli</i> DH5α pAN::Amer-F1	This study	NOT Gate
Tas75	<i>E.coli</i> DH5α pAN::AmtR-A1	This study	NOT Gate
Tas76	<i>E.coli</i> DH5α pAN::BetI-E1	This study	NOT Gate
Tas77	<i>E.coli</i> DH5α pAN::BM3R1-B1	This study	NOT Gate
Tas78	<i>E.coli</i> DH5α pAN::BM3R1-B2	This study	NOT Gate
Tas79	<i>E.coli</i> DH5α pAN::BM3R1-B3	This study	NOT Gate
Tas364	<i>E.coli</i> DH5α pAN::HIyIIR-H1	This study	NOT Gate
Tas81	<i>E.coli</i> DH5α pAN::lcaRA-I1	This study	NOT Gate
Tas82	<i>E.coli</i> DH5α pAN::LitR-L1	This study	NOT Gate
Tas83	<i>E.coli</i> DH5α pAN::LmrA-N1	This study	NOT Gate
Tas84	<i>E.coli</i> DH5α pAN::PhIF-P1	This study	NOT Gate
Tas85	<i>E.coli</i> DH5α pAN::PhIF-P2	This study	NOT Gate
Tas86	<i>E.coli</i> DH5α pAN::PhIF-P3	This study	NOT Gate
Tas87	<i>E.coli</i> DH5α pAN::PsrA-R1	This study	NOT Gate
Tas88	<i>E.coli</i> DH5α pAN::QacR-Q1	This study	NOT Gate
Tas89	<i>E.coli</i> DH5α pAN::QacR-Q2	This study	NOT Gate
Tas90	<i>E.coli</i> DH5α pAN::SrpR-S1	This study	NOT Gate
Tas91	<i>E.coli</i> DH5α pAN::SrpR-S2	This study	NOT Gate
Tas365	<i>E.coli</i> DH5α pAN::SrpR-S3	This study	NOT Gate
Tas93	<i>E.coli</i> DH5α pAN::SrpR-S4	This study	NOT Gate
Tas385	<i>E.coli</i> DH5α pSEVA221::Amer-F1	This study	NOT Gate
Tas386	<i>E.coli</i> DH5α pSEVA221::AmtR-A1	This study	NOT Gate
Tas387	<i>E.coli</i> DH5α pSEVA221::BetI-E1	This study	NOT Gate
Tas388	<i>E.coli</i> DH5α pSEVA221::BM3R1-B1	This study	NOT Gate
Tas389	<i>E.coli</i> DH5α pSeva221::BM3R1-B2	This study	NOT Gate
Tas390	<i>E.coli</i> DH5α pSeva221::BM3R1-B3	This study	NOT Gate
Tas391	<i>E.coli</i> DH5α pSeva221::HIyIIR-H1	This study	NOT Gate
Tas392	<i>E.coli</i> DH5α pSeva221::lcaRA-I1	This study	NOT Gate
Tas393	<i>E.coli</i> DH5α pSeva221::LitR-L1	This study	NOT Gate
Tas394	<i>E.coli</i> DH5α pSeva221::LmrA-N1	This study	NOT Gate
Tas395	<i>E.coli</i> DH5α pSeva221::PhIF-P1	This study	NOT Gate
Tas396	<i>E.coli</i> DH5α pSeva221::PhIF-P2	This study	NOT Gate
Tas397	<i>E.coli</i> DH5α pSeva221::PhIF-P3	This study	NOT Gate
Tas398	<i>E.coli</i> DH5α pSeva221::PsrA-R1	This study	NOT Gate
Tas399	<i>E.coli</i> DH5α pSeva221::QacR-Q1	This study	NOT Gate
Tas400	<i>E.coli</i> DH5α pSeva221::QacR-Q2	This study	NOT Gate
Tas401	<i>E.coli</i> DH5α pSeva221::SrpR-S1	This study	NOT Gate
Tas402	<i>E.coli</i> DH5α pSeva221::SrpR-S2	This study	NOT Gate
Tas403	<i>E.coli</i> DH5α pSeva221::SrpR-S3	This study	NOT Gate
Tas404	<i>E.coli</i> DH5α pSeva221::SrpR-S4	This study	NOT Gate
Tas1	<i>E.coli</i> CC118λpir pSeva221::Amer-F1	This study	NOT Gate
Tas2	<i>E.coli</i> CC118λpir pSeva221::AmtR-A1	This study	NOT Gate

Tas3	<i>E.coli</i> CC118λpir pSeva221::BetI-E1	This study	NOT Gate
Tas4	<i>E.coli</i> CC118λpir pSeva221::BM3R1-B1	This study	NOT Gate
Tas5	<i>E.coli</i> CC118λpir pSeva221::BM3R1-B2	This study	NOT Gate
Tas6	<i>E.coli</i> CC118λpir pSeva221::BM3R1-B3	This study	NOT Gate
Tas7	<i>E.coli</i> CC118λpir pSeva221::HIyIIR-H1	This study	NOT Gate
Tas8	<i>E.coli</i> CC118λpir pSeva221::lcaRA-I1	This study	NOT Gate
Tas9	<i>E.coli</i> CC118λpir pSeva221::LitR-L1	This study	NOT Gate
Tas10	<i>E.coli</i> CC118λpir pSeva221::LmrA-N1	This study	NOT Gate
Tas11	<i>E.coli</i> CC118λpir pSeva221::PhIF-P1	This study	NOT Gate
Tas12	<i>E.coli</i> CC118λpir pSeva221::PhIF-P2	This study	NOT Gate
Tas13	<i>E.coli</i> CC118λpir pSeva221::PhIF-P3	This study	NOT Gate
Tas14	<i>E.coli</i> CC118λpir pSeva221::PsrA-R1	This study	NOT Gate
Tas15	<i>E.coli</i> CC118λpir pSeva221::QacR-Q1	This study	NOT Gate
Tas16	<i>E.coli</i> CC118λpir pSeva221::QacR-Q2	This study	NOT Gate
Tas17	<i>E.coli</i> CC118λpir pSeva221::SrpR-S1	This study	NOT Gate
Tas18	<i>E.coli</i> CC118λpir pSeva221::SrpR-S2	This study	NOT Gate
Tas19	<i>E.coli</i> CC118λpir pSeva221::SrpR-S3	This study	NOT Gate
Tas20	<i>E.coli</i> CC118λpir pSeva221::SrpR-S4	This study	NOT Gate
Tas367	<i>E.coli</i> CC118λpir pSeva231::Amer-F1	This study	NOT Gate
Tas368	<i>E.coli</i> CC118λpir pSeva231::AmtR-A1	This study	NOT Gate
Tas23	<i>E.coli</i> CC118λpir pSeva231::BetI-E1	This study	NOT Gate
Tas24	<i>E.coli</i> CC118λpir pSeva231::BM3R1-B1	This study	NOT Gate
Tas25	<i>E.coli</i> CC118λpir pSeva231::BM3R1-B2	This study	NOT Gate
Tas26	<i>E.coli</i> CC118λpir pSeva231::BM3R1-B3	This study	NOT Gate
Tas27	<i>E.coli</i> CC118λpir pSeva231::HIyIIR-H1	This study	NOT Gate
Tas28	<i>E.coli</i> CC118λpir pSeva231::lcaRA-I1	This study	NOT Gate
Tas29	<i>E.coli</i> CC118λpir pSeva231::LitR-L1	This study	NOT Gate
Tas30	<i>E.coli</i> CC118λpir pSeva231::LmrA-N1	This study	NOT Gate
Tas31	<i>E.coli</i> CC118λpir pSeva231::PhIF-P1	This study	NOT Gate
Tas32	<i>E.coli</i> CC118λpir pSeva231::PhIF-P2	This study	NOT Gate
Tas33	<i>E.coli</i> CC118λpir pSeva231::PhIF-P3	This study	NOT Gate
Tas34	<i>E.coli</i> CC118λpir pSeva231::PsrA-R1	This study	NOT Gate
Tas369	<i>E.coli</i> CC118λpir pSeva231::QacR-Q1	This study	NOT Gate
Tas36	<i>E.coli</i> CC118λpir pSeva231::QacR-Q2	This study	NOT Gate
Tas37	<i>E.coli</i> CC118λpir pSeva231::SrpR-S1	This study	NOT Gate
Tas38	<i>E.coli</i> CC118λpir pSeva231::SrpR-S2	This study	NOT Gate
Tas39	<i>E.coli</i> CC118λpir pSeva231::SrpR-S3	This study	NOT Gate
Tas40	<i>E.coli</i> CC118λpir pSeva231::SrpR-S4	This study	NOT Gate
Tas193	<i>P.putida</i> KT2440 pSeva221::Amer-F1	This study	NOT Gate
Tas194	<i>P.putida</i> KT2440 pSeva221::AmtR-A1	This study	NOT Gate
Tas195	<i>P.putida</i> KT2440 pSeva221::BetI-E1	This study	NOT Gate
Tas196	<i>P.putida</i> KT2440 pSeva221::BM3R1-B1	This study	NOT Gate
Tas197	<i>P.putida</i> KT2440 pSeva221::BM3R1-B2	This study	NOT Gate
Tas198	<i>P.putida</i> KT2440 pSeva221::BM3R1-B3	This study	NOT Gate
Tas199	<i>P.putida</i> KT2440 pSeva221::HIyIIR-H1	This study	NOT Gate
Tas200	<i>P.putida</i> KT2440 pSeva221::lcaRA-I1	This study	NOT Gate

Tas201	<i>P.putida</i> KT2440 pSeva221::LitR-L1	This study	NOT Gate
Tas202	<i>P.putida</i> KT2440 pSeva221::LmrA-N1	This study	NOT Gate
Tas203	<i>P.putida</i> KT2440 pSeva221::PhIF-P1	This study	NOT Gate
Tas204	<i>P.putida</i> KT2440 pSeva221::PhIF-P2	This study	NOT Gate
Tas205	<i>P.putida</i> KT2440 pSeva221::PhIF-P3	This study	NOT Gate
Tas206	<i>P.putida</i> KT2440 pSeva221::PsrA-R1	This study	NOT Gate
Tas207	<i>P.putida</i> KT2440 pSeva221::QacR-Q1	This study	NOT Gate
Tas208	<i>P.putida</i> KT2440 pSeva221::QacR-Q2	This study	NOT Gate
Tas209	<i>P.putida</i> KT2440 pSeva221::SrpR-S1	This study	NOT Gate
Tas210	<i>P.putida</i> KT2440 pSeva221::SrpR-S2	This study	NOT Gate
Tas211	<i>P.putida</i> KT2440 pSeva221::SrpR-S3	This study	NOT Gate
Tas212	<i>P.putida</i> KT2440 pSeva221::SrpR-S4	This study	NOT Gate
Tas213	<i>P.putida</i> KT2440 pSeva231::Amer-F1	This study	NOT Gate
Tas214	<i>P.putida</i> KT2440 pSeva231::AmtR-A1	This study	NOT Gate
Tas215	<i>P.putida</i> KT2440 pSeva231::BetI-E1	This study	NOT Gate
Tas216	<i>P.putida</i> KT2440 pSeva231::BM3R1-B1	This study	NOT Gate
Tas217	<i>P.putida</i> KT2440 pSeva231::BM3R1-B2	This study	NOT Gate
Tas218	<i>P.putida</i> KT2440 pSeva231::BM3R1-B3	This study	NOT Gate
Tas219	<i>P.putida</i> KT2440 pSeva231::HIyIIR-H1	This study	NOT Gate
Tas220	<i>P.putida</i> KT2440 pSeva231::lcaRA-I1	This study	NOT Gate
Tas221	<i>P.putida</i> KT2440 pSeva231::LitR-L1	This study	NOT Gate
Tas222	<i>P.putida</i> KT2440 pSeva231::LmrA-N1	This study	NOT Gate
Tas223	<i>P.putida</i> KT2440 pSeva231::PhIF-P1	This study	NOT Gate
Tas224	<i>P.putida</i> KT2440 pSeva231::PhIF-P2	This study	NOT Gate
Tas225	<i>P.putida</i> KT2440 pSeva231::PhIF-P3	This study	NOT Gate
Tas226	<i>P.putida</i> KT2440 pSeva231::PsrA-R1	This study	NOT Gate
Tas227	<i>P.putida</i> KT2440 pSeva231::QacR-Q1	This study	NOT Gate
Tas228	<i>P.putida</i> KT2440 pSeva231::QacR-Q2	This study	NOT Gate
Tas229	<i>P.putida</i> KT2440 pSeva231::SrpR-S1	This study	NOT Gate
Tas230	<i>P.putida</i> KT2440 pSeva231::SrpR-S2	This study	NOT Gate
Tas231	<i>P.putida</i> KT2440 pSeva231::SrpR-S3	This study	NOT Gate
Tas232	<i>P.putida</i> KT2440 pSeva231::SrpR-S4	This study	NOT Gate
Tas233	<i>P.putida</i> KT2440 pSeva251::Amer-F1	This study	NOT Gate
Tas234	<i>P.putida</i> KT2440 pSeva251::AmtR-A1	This study	NOT Gate
Tas235	<i>P.putida</i> KT2440 pSeva251::BetI-E1	This study	NOT Gate
Tas237	<i>P.putida</i> KT2440 pSeva251::BM3R1-B2	This study	NOT Gate
Tas238	<i>P.putida</i> KT2440 pSeva251::BM3R1-B3	This study	NOT Gate
Tas239	<i>P.putida</i> KT2440 pSeva251::HIyIIR-H1	This study	NOT Gate
Tas240	<i>P.putida</i> KT2440 pSeva251::lcaRA-I1	This study	NOT Gate
Tas241	<i>P.putida</i> KT2440 pSeva251::LitR-L1	This study	NOT Gate
Tas242	<i>P.putida</i> KT2440 pSeva251::LmrA-N1	This study	NOT Gate
Tas243	<i>P.putida</i> KT2440 pSeva251::PhIF-P1	This study	NOT Gate
Tas244	<i>P.putida</i> KT2440 pSeva251::PhIF-P2	This study	NOT Gate
Tas246	<i>P.putida</i> KT2440 pSeva251::PsrA-R1	This study	NOT Gate
Tas247	<i>P.putida</i> KT2440 pSeva251::QacR-Q1	This study	NOT Gate
Tas249	<i>P.putida</i> KT2440 pSeva251::SrpR-S1	This study	NOT Gate

Tas252	<i>P.putida</i> KT2440 pSeva251::SrpR-S4	This study	NOT Gate
Tas65	<i>E.coli</i> CC118λpir pSeva221::1201	This study	Empty plasmid for autofluorescence
Tas68	<i>E.coli</i> CC118λpir pSeva221::1717	This study	RPU standard plasmid for standardization
Tas317	<i>E.coli</i> CC118λpir pSeva221::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas66	<i>E.coli</i> CC118λpir pSeva231::1201	This study	Empty plasmid for autofluorescence
Tas69	<i>E.coli</i> CC118λpir pSeva231::1717	This study	RPU standard plasmid for standardization
Tas370	<i>E.coli</i> CC118λpir pSeva231::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas94	<i>E.coli</i> DH5α pAN::1201	This study	Empty plasmid for autofluorescence
Tas95	<i>E.coli</i> DH5α pAN::1717	This study	RPU standard plasmid for standardization
Tas366	<i>E.coli</i> DH5α pAN::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas361	<i>E.coli</i> DH5α pSeva221::1201	This study	Empty plasmid for autofluorescence
Tas362	<i>E.coli</i> DH5α pSeva221::1717	This study	RPU standard plasmid for standardization
Tas363	<i>E.coli</i> DH5α pSeva221::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas257	<i>P.putida</i> KT2440 pSeva221::1201	This study	Empty plasmid for autofluorescence
Tas260	<i>P.putida</i> KT2440 pSeva221::1717	This study	RPU standard plasmid for standardization
Tas266	<i>P.putida</i> KT2440 pSeva221::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas258	<i>P.putida</i> KT2440 pSeva231::1201	This study	Empty plasmid for autofluorescence
Tas261	<i>P.putida</i> KT2440 pSeva231::1717	This study	RPU standard plasmid for standardization
Tas267	<i>P.putida</i> KT2440 pSeva231::1818	This study	P _{Tac} activity plasmid for promoter activity
Tas259	<i>P.putida</i> KT2440 pSeva251::1201	This study	Empty plasmid for autofluorescence
Tas262	<i>P.putida</i> KT2440 pSeva251::1717	This study	RPU standard plasmid for standardization
Tas285	<i>P.putida</i> KT2440 pSeva251::1818	This study	P _{Tac} activity plasmid for promoter activity

4.3. List of standardization and RNAP immuno-capture strains.

Table S 7: Standardization and RNAP immuno-capture Strains

Stock Name	Library	Work	Definition
Tas103	<i>P.putida</i> KT_BG37-RpoC::HisTag	This study	Standard promoter strain with RpoC tagged with His tag
Tas106	<i>P.putida</i> KT_BG37-RpoC::MycTag	This study	Standard promoter strain with RpoC tagged with Myc tag
Tas108	<i>P.putida</i> KT2440-RpoC::HisTag	This study	RPIT application strain with RpoC tagged with His tag
Tas110	<i>P.putida</i> KT2440-RpoC::MycTag	This study	RPIT application strain with RpoC tagged with Myc tag
Tas114	<i>P.putida</i> KT2440-RpoC::Etag	This study	RPIT application strain with RpoC tagged with Etag

VII. Summaries in other languages¹

1. Spanish Version

1.1. Presentación

La biología sintética actual es conocida por los esfuerzos en el diseño de biocircuitos y la implementación de los circuitos genéticos. La metodología típica consiste en asignar contrapartes biológicas de los estados de entrada y salida a través de una analogía extraída de las ciencias computacionales, es decir, la informática y la ingeniería electrónica. Las entradas son entidades bio- o fisico- químicas (v. g. temperatura, radiación, un compuesto o cualquier tipo de inductor) o maquinarias operativas como los efectores celulares (v. g. proteínas, secuencias de ARN o ADN, metabolitos celulares) que contienen un diseño inteligente detrás con una capa de cálculo y resultados que se traducen en la acción / producto de interés (v. g. inducción de taxis específicas, liberación de moléculas, inducción de expresión génica, muerte celular programada, reubicación de información de ADN, etc.) [1, 2].

Aunque muchos CAD (Diseño Asistido por Computador) de biocircuitos conocidos se construyen sobre la idea de partes optimizadas para el huésped [4, 10], podría ser un enfoque competente hacer uso de las partes ya desarrolladas y reutilizarlas en otros huéspedes (portabilidad del dispositivo). Después de los intentos contemporáneos informados sobre las interacciones entre el anfitrión y los dispositivos genéticos, parece razonable pensar que una forma de lograr la portabilidad del dispositivo podría ser tener en consideración el efecto contextual. En consecuencia, el efecto contextual podría formalizarse para abordar la idea de reconfigurabilidad.

Una de las tareas centrales de la biología sintética es la reproducibilidad de los resultados de investigación, es decir, la construcción de estándares para compartir resultados además del desarrollo de herramientas y chasis para propósitos específicos. La expresión génica es un tema en el que los esfuerzos de estandarización son visibles [26, 27], aunque todavía son necesarias algunas mejoras.

En este estudio se han desarrollado y probado los conceptos dichos arriba y mostrado los resultados afiliados.

¹ *Disclaimer: Please note that the original language of this Thesis is English. Spanish version is

1.2. Objetivos

El **objetivo general** de esta Tesis es:

Actualización de *Pseudomonas putida* como chasis de biología sintética para diseños de circuitos genéticos, investigando el concepto de interoperabilidad así como desarrollando herramientas de estandarización y de metodologías cuantitativas para la ingeniería de mecanismos reguladores de la transcripción.

Este objetivo general incluye los siguientes **objetivos** parciales **específicos**:

1. Diseñar un conjunto de herramientas de biocircuitos de amplio rango de huéspedes *ad hoc* para la implementación de la automatización del diseño en bacterias Gram-negativas distintas de *Escherichia coli*.
2. Investigar el diseño de circuitos automatizados en *Pseudomonas putida*.
3. Establecer la variabilidad contextual como influencia en la portabilidad de dispositivos genéticos y su uso para mejorar el concepto de interoperatividad.
4. Diseñar un chasis reforzado con herramientas de estandarización incorporadas dentro del genoma.
5. Desarrollar un método de inmuno-captura para revelar la maquinaria de regulación de la transcripción y para el mapeo cuantitativo del interactoma de *Pseudomonas putida*.

1.3. Conclusiones

Este trabajo nos ha llevado a las siguientes conclusiones:

1. Las bibliotecas Broad Host Range (BHR) introducidas en formato SEVA amplían la accesibilidad de las puertas lógicas NOT disponibles para la computación celular y la aplicabilidad de los procesos de automatización de diseño de circuitos con CelloCAD en bacterias Gram-negativas.
2. Una puerta NOR diseñada mediante el uso de CelloCAD y con un mal funcionamiento predicho para con *Escherichia coli* se implementó eficazmente en *Pseudomonas putida* y mostró un rendimiento claro.
3. Considerar la variabilidad contextual como un parámetro para el diseño de circuitos genéticos aumenta el número de posibles combinaciones por construcción genética y la configuración disponible de puertas para obtener nuevas funciones de respuesta.
4. El promotor BG37 actúa de forma constitutiva y ortogonal en *Pseudomonas putida* en las condiciones de los medios típicos ensayados y mientras que la cepa de *P. putida* KT_BG37 se introduce como cepa de referencia para análisis comparativos de expresión génica.
5. El enfoque de inmunocaptura in vivo muestrea el efecto de las tensiones ambientales distintivas sufridas por *P. putida* en la configuración del interactoma RNAP y produce la exploración de un conjunto de factores sigma, reguladores transcripcionales y proteínas auxiliares con una composición relativa regida por el estado fisiológico de células.

2. Turkish Version²

2.1. Sunum

Mevcut sentetik biyoloji, biyo-devre tasarımı ve genetik devrelerin uygulanmasındaki çabalarla bilinir. Tipik metodoloji, hesaplamalı bilimlerin, yani bilgisayar bilimi ve elektronik mühendisliğinden yapılan bir analogi aracılığıyla girdi (input) ve çıktı (output) durumlarının biyolojik karşılıklarının atanmasından oluşur. Girdiler, akıllı bir tasarım içeren biyo- veya fiziko-kimyasal varlıklar (örneğin sıcaklık, radyasyon, toksin veya herhangi bir tür dış etken), hücrel efektörler (örneğin proteinler, tasarlanmış RNA veya DNA dizileri, hücrel metabolitler) gibi çalışan makineler, bir hesaplama katmanı ve ilgili eylem / ürün gibi sonuçlar (örneğin, spesifik toksinlerin indüksiyonu, moleküllerin salınması, gen ifadesinin uyarılması, programlanmış hücre ölümü, DNA bilgisinin yeniden konumlandırılması, vb.) olabilirler [1, 2].

Bilinen birçok biyo-devre CAD (Bilgisayar Destekli Dizayn)'ları, konak için optimize edilmiş parçalar [4, 10] fikri üzerine inşa edilmiş olsa da, bu noktaya kadar geliştirilen parçalardan yararlanmak için, konağa özgü tasarlanmış genetik parçaların diğer konaklarda yeniden kullanılması (örneğin, aygıt taşınabilirliği) etkin bir yaklaşım olabilir. Konak ve genetik aygıtlar arasındaki etkileşimlerle alakalı olarak rapor edilmiş olan mevcut girişimler takip edildiğinde, cihaz taşınabilirliğini sağlamanın bir yolunun bağlamsal (hücre içi) etkiyi dikkate almak olabileceğini düşünmek yerinde olabilir. Sonuç olarak, bağlamsal etki yeniden yapılandırılabilirlik fikrini ele almak için önem teşkil edebilir.

Araştırma sonuçlarının tekrarlanabilir olması, sentetik biyolojinin temel sorularından biridir ki bu da sonuçları paylaşmak için standartların oluşturulmasını ve belirli amaçlar için standartlaştırılmış araç gereçleri ve şasileri gerekli kılar. Gen ifadesinin rapor edilmesi şekli standardizasyon çabalarının görünür olduğu bir alandır [26,27], ancak bu alan bazı iyileştirmelerden fayda sağlayabilir.

Bu çalışmada yukarıda bahsi geçen fikirler geliştirilmiş, denenmiş ve ilgili sonuçlar gösterilmiştir.

² *Disclaimer: Please note that the original language of this Thesis is English. Spanish version is shared as a reference for cataloging purposes. Turkish version is shared on the preference of the author and revised only by the dissertation owner.

2.2. Hedefler

Bu Tezin **genel amacı**:

Pseudomonas putida'nın genetik devre tasarımları için sentetik biyoloji şasisi olarak güncellenmesi, bağıntısız çalışabilirlik (interoperability) kavramının incelenmesi, transkripsiyonel düzenleyici mekanizmaların mühendisliği için standardizasyon araçları ve niceliksel metotlar geliştirilmesidir.

Bu genel hedef, aşağıdaki **belirli alt hedefleri** içerir:

1. *Escherichia coli* dışındaki Gram-negatif bakterilerde tasarım otomasyonunun uygulanması için geniş konak kapsamlı (broad-host-range) biyo-devre araç tasarımı.
2. *Pseudomonas putida*'da otomatikleştirilmiş devre tasarımının araştırılması.
3. Bağlamsal değişkenliğin (contextual variability), genetik aygıtların taşınabilirliği (portability) üzerinde ve bağıntısız çalışabilirlik kavramını geliştirmek için kullanımında bir parametre olarak belirlenmesi.
4. Genoma entegre standardizasyon araçlarıyla güçlendirilmiş bir şasinin mühendisliği.
5. Transkripsiyon düzenleme mekanizmasını ortaya çıkarmak ve *Pseudomonas putida* interaktomunun nicel haritalaması için bir immüno-yakalama yöntemi geliştirilmesi.

2.3. Sonuç

Bu çalışma sonucunda aşağıdaki sonuçlara ulaşılmıştır:

1. SEVA formatında hazırlanan Geniş Konak Aralığı (Broad-Host-Range) kütüphaneleri, hücresel hesaplama için mevcut DEĞİL (NOT) mantık kapılarının erişilebilirliğini ve CelloCAD ile devre tasarımı otomasyon işlemlerinin uygulanabilirliğini Gram-negatif bakterilerde arttırmıştır.
2. CelloCAD kullanılarak tasarlanan ve *Escherichia coli* ile uyumlu olmadığı tahmin edilen bir NOR devresi, *Pseudomonas putida*'da etkin bir şekilde uygulanmıştır ve sarıh bir NOR devre performansı elde edilmiştir.
3. Konak hücre bağlamsal değişkenliğinin (cellular context) genetik devre tasarımı için bir parametre olarak kullanılması, yeni yanıt fonksiyonları (response function) elde ederek genetik kapıların olası kombinasyonlarının sayısını ve kapıların mevcut konfigürasyonunu artırır.
4. Test edilen tipik ortam koşulları altında *Pseudomonas putida*'da BG37 promotörü ortogonal ve devamlı ifade edilen (constitutive) bir promotör olarak işlev göstermektedir ve *P. putida* KT_BG37 bakterisi, karşılaştırmalı gen ekspresyon analizleri için referans şasi olarak sunulmuştur.
5. *In vivo* immüno-yakalama tekniği *Pseudomonas putida* RNAP'ının interaktom konfigürasyonunda farklı çevresel streslerin etkisi ile oluşmuş değişiklikleri ortaya koyması sayesinde fizyolojik durum tarafından indüklenen sigma faktörlerin, transkripsiyon düzenleyicilerinin ve yardımcı proteinlerin keşfini sağlar.