Reconstructing and Discovering New σ^{54} Regulation Mechanisms in a High Throughput Manner

Lior Levy

Reconstructing and Discovering New σ^{54} Regulation Mechanisms in a High Throughput Manner

Research Thesis

In partial fulfillment of the requirements for the

Degree of Master of Science in Biotechnology & Food Engineering

Lior Levy

Submitted to the Senate of
the Technion - Israel Institute of Technology

Adar B month 5776, Haifa, March 2016

The research thesis was done under the supervision of

Asst. Prof. Roee Amit in the department of Biotechnology and Food Engineering

The generous financial help of the Technion- Israel Institute of Technology is gratefully acknowledged.

Table of contents

Table of figures

Table of Tables

Table of equations:

Abstract:

Bacterial enhancers are non-translated DNA sequences, which play a fundamental roll in gene regulation and functions as a type of molecular integrator that determines when, where, and how much of a particular gene is expressed. A bacterial enhancer is typically comprised of an upstream activating sequence (UAS) which binds oligomeric activators (also known as enhancer binding proteins- EBPs) that provide the necessary energy for the formation of an open complex on σ^{54} dependent promoters. Characterization of the relationship between the UAS and its cognate σ^{54}-promoter have been performed previously in low scale and for very specific promoters.

Bacteria can use a variety of mechanisms to regulate the expression of specific genes, a few examples are: (i) competition on the promoter site: in which the concentration change of an inducer or repressor can result in the activation or silencing of specific genes. (ii) Looping based regulation in σ^{54} dependent promoters: in which DNA binding proteins bound to the looping region can increase or lower the probability of the loop formation and by that control the activation of these kind of promoters. (iii) RNA level regulation: in which secondary structures in the RNA can cause the pausing of an mRNA translation, and (iv) gene regulation by roadblocks and RNA polymerase (RNAP) pausing: in which various DNA binding proteins (e.g. transcription factors or repressors) or even other RNAPs can block the transcription of a trailing RNAP.

I decided to perform a comprehensive characterization of the UAS affinity for EBP's effect on σ^{54}-promoter activation in enhancer systems, covering all the Ntr regulated promoters with native and synthetic UASs in a combinatorial manner. In addition, I also performed a high throughput experiment using an oligo-library of 12,000 different sequences in order to understand the mechanism behind a silencing phenomenon observed in my experiments and its prevalence in the genomes of *E.coli* and *V.cholera*.

My results showed that the σ^{54}-promoter's activation efficiency is dependent more on promoter's strength than on the UASs affinity for the EBP. Moreover, I was able to show that it is not possible to predict the UASs affinity for EBP only by the cumulative affinity of its comprising binding sites. In an interesting turn of events my work was also able to

show a surprising silencing phenomenon observed using inactivated glnKp promoter and two different upstream promoters (glnAp1 and pLac/Ara). A σ^{54}:RNAP holoenzyme roadblock regulation mechanism was ruled out using site-directed mutagenesis of the glnKp's sequence and by the comparison of expression using a ΔRpoN ($\Delta\sigma^{54}$) strain, showing that the silencing effect is related to the flanking sequences of the promoter and not to the core consensus sequence. Finally, I was able to show that this silencing phenomenon is widespread in the genomes of *E.coli* and *V.cholera*, showing prevalence of 25% in our tested sequences. Future work should be carried out in order to reveal the mechanism/s behind the silencing phenomenon, first focusing on a Shine Dalgarno sequestering mechanism.

Abbreviation:

AMP- Ampicillin

aTc- Anhydrotetracycline

BA- BioAssay

bEBP- bacterial enhancer binding protein

BS- Binding site

CIP- Calf Intestinal

DMSO - Dimethyl sulfoxide

DNA - Deoxyribonucleic acid

dntp- Deoxynucleotide

EBP- Enhancer binding protein

EDTA- Ethylenediaminetetra-acetic acid

IPTG- Isopropyl β-D-1-thiogalactopyranoside

KAN- Kanamycin

LB- Lysogeny broth

MBW- Molecular biology water

mRNA- Messenger ribonucleic acid

NtrC- Nitrogen regulatory protein C

O.D- Optical density

PBS- Phosphate Buffered Saline

PCR- Polymerase Chain Reaction

PNK- Polynucleotide kinase

Pol- Polymerase

RBP- Ribosome binding protein

RNA- Ribonucleic acid

RNAi- RNA interference

RNAP- Ribonucleic acid polymerase

TAE- Tris-acetate-EDTA

TE- Tris EDTA

TetR- Tetracycline repressor

TSS- Transcription start site

UAS- Upstream activating sequence

UPW- Ultra pure water

UTR- Untranslated region

°C- Celsius degree

µg- Microgram

µl- Microliter

µm- Micrometer

gr- Gram

hr- Hour

M- Molar

mg- Milligram

min Minutes

ml- Milliliter

mM- Millimolar

ng- Nanogram

rpm- Rounds per minute

sec- second

V- volts

v/v- Volume per volume

Δ- Gene deletion

σ- Sigma

1 Introduction:

1.1 Transcription initiation in bacteria

Transcription initiation in bacteria is highly regulated. It is facilitated by the binding of RNA polymerase (RNAP) and a modular subunit named sigma (σ) factor, which mutually forms a holoenzyme complex required for both directing the polymerase to a specific promoter and DNA melting [1,2]. In *E.coli*, there are two classes of σ factors (σ^{70} and σ^{54}), which differ in amino acid sequence, domain structure and in the open complex formation pathway [3,4]. Most of the *E.coli*'s housekeeping and growth related genes are transcribed by the σ^{70} family, whereas σ^{54} plays a major role in nitrogen limiting and stress conditions [5,6].

σ factors recognize and direct the RNAP to DNA binding determinants located in the promoter region. Members of the σ^{70} family recognize sites located at -10 and -35 with respect to the transcription start site (TSS) [4]. In contrast, σ^{54} recognizes the sequence located at -12 and -24 with respect to the TSS [7]. Moreover, The binding of σ^{70}:RNAP holoenzyme to the promoter is sufficient for transcription initiation, while σ^{54} forms a transcriptionally incompetent stable closed complex with the RNAP at the promoter site which requires energy in order to isomerase into transcriptionally competent open-complex [5]. The energy needed derives from ATP hydrolysis catalyzed by activator proteins known as bacterial enhancer binding proteins (bEBPs) which bind to cognate DNA domains (also named enhancer sites or upstream activating sequence-UAS) located upstream of the promoter and activate the process by making contact with the closed complex through DNA looping as shown in Figure 1 [7,8]. These UASs are also effective when put as far as 1000bp away from the promoter [9,10]. It is important to note that in some cases, DNA looping is facilitated by DNA bending proteins such as ArgR or the integration host factor (IHF) [11–13].

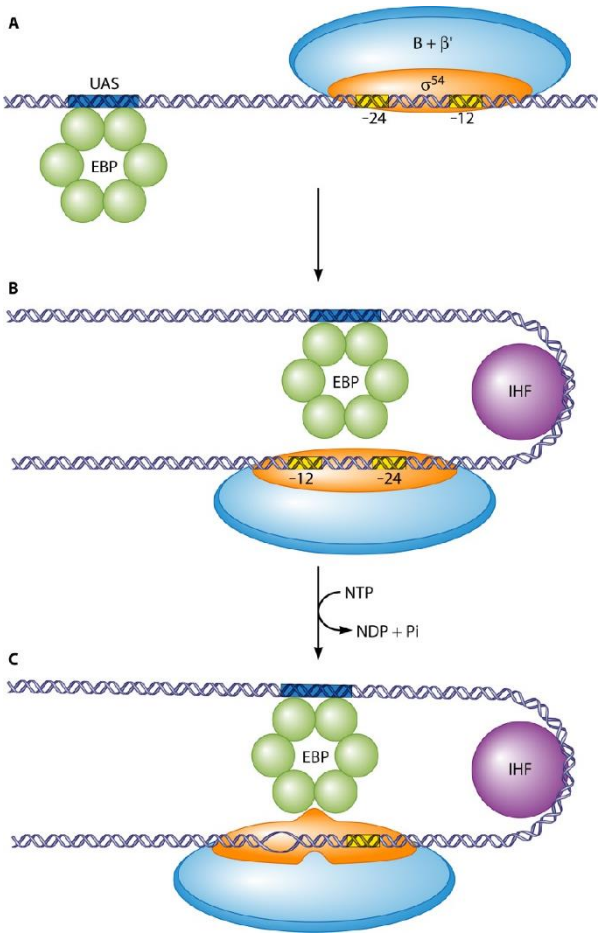

Figure 1: Activation of the σ^{54}:RNAP holoenzyme.

(A) RNAP is directed to the -12 and -24 promoter elements by the σ^{54} protein, activation of the σ^{54}:RNAP holoenzyme by the bacterial enhancer binding protein (bEBP, shown as a green hexamer) is dependent on the bEBP binding to an upstream activating sequence (UAS) upstream of the transcriptional start site. (B) DNA looping occurs, sometimes mediated by other proteins such as integration host factor (IHF). (C) The closed complex isomerization is promoted by ATP hydrolysis. Adopted from reference [1].

1.1.1 Bacterial enhancer binding proteins (bEBPs)

bEBPs mostly have similar structure: N-terminal regulatory domain responding to extrinsic stimulatory signal, a central ATPase domain accountable for ATPase activity and the interaction to the σ^{54}:RNAP holoenzyme, and a C-terminal DNA-binding sites which binds one or more binding sites [14]. Hexamerisation of the bEBP is required in order for the

ATPase function to be active and the energy derived from ATP hydrolysis is coupled to the σ^{54}:RNAP holoenzyme isomerization from closed to open complex [15,16].

Each bEBP in the cell is regulated by its own signal transduction pathway, allowing a very tight regulation on σ^{54} dependent gene expression as different environmental needs arises [17]. Some bEBPs need to be modified (i.e. phosphorylated) in order to promote their binding to their respective UAS. Such bEBP is the nitrogen regulatory protein C (NtrC, also named NRI) which has a role in regulating glutamine synthesis [8,15] and is phosphorylated/dephosphorylated by NtrB (also named NRII). NtrC-p binds to the UAS as dimers and recruits a third dimer from the cytoplasm in order to form the hexamer needed for ATP hydrolysis [1,7].

NtrC regulates a number of operons in *E.coli*: it activates the expression of glnK-amtB operon (an alternate PII and an ammonia transporter), the glnALG operon (glutamine synthetase and Ntr regulator), glnHPQ operon (glutamine transport), astCADBE operon (arginine catabolism) and nac (a σ^{70} dependent transcriptional activator). It is also known to repress glnAp1 and glnLp of the in the glnALG operon [18].

1.1.2 Ntr regulated σ^{54} dependent promoters

The best nitrogen donors in *E.coli* are glutamine and glutamate, therefore in the case of nitrogen limiting conditions, these molecules need to be synthetized and their expression regulated. In *E.coli*, there are four NtrC regulated operons that are responsible for the levels of glutamine and glutamate. These operons are regulated by the activation of their respective σ^{54} promoters by NtrC [18–20]. The promoters are: (i) glnAp2- part of the glnALG operon. It has five upstream binding sites for NtrC and it can be activated in low NtrC levels due to its high affinity binding sites (#1 and #2). (ii) glnKp- part of the glnK-amtB operon has two low affinity NtrC binding sites upstream, therefore this promoter needs higher levels of NtrC in the cell in order to be activated [21]. (iii) nacp- regulates the expression of the Nac gene. Known to have two NtrC binding sites with varying affinities. (iv) glnHp2- part of the glnHPQ operon is thought to have four NtrC binding sites distributed around the promoter. In this case, DNA looping is associated with a DNA binding protein called IHF. The effect of IHF can be enhancing or diminishing, depending on its binding site location relative to the active UAS. It has been shown that if the UAS is

relatively close to the promoter (~120bp from the transcription start site) IHF binding enhances transcription, while for UASs located a few hundred bases away, the effect of IHF can be repressing [22]. Another nitrogen donor in the cell is arginine, in nitrogen limiting conditions and when arginine is present in the solution- the astCADBE operon is activated. The promoter regulating the operon is astCp2 which is known to work in high concentrations of NtrC in the cell and its activation is also mediated by a DNA binding protein called ArgR [12,23].

1.2 Repressive regulatory mechanisms in bacteria

Bacteria uses a variety of mechanisms and allocate many resources to control how much of and when a gene would be expressed. Regulation is typically carried out at the transcriptional level (e.g. during transcription initiation or elongation), and can be either inhibitory of promoter activity or enhancing the formation of a valid transcript. In addition, regulation also takes place at the translational level by controlling the ultimate levels of mRNA and the ability of the ribosome to translate. In this part, we will go over some of the main inhibitory transcriptional mechanisms used in bacteria.

1.2.1 Competition based regulation

Competition based regulations are the most common form of repression in gene expression. The mechanism, identified originally by Jacob and Monod [24] for the Lac repressor typically involves a competition for binding between a protein (repressor) and the RNA polymerase at a σ^{70} promoter site, typically between the -10 and -35 elements. In this mechanism, if the repressor is bound to the promoter, the RNAP is unable to form a stable holoenzyme, thus preventing transcription. There are numerous documented examples for this mechanism in *E.coli* and other bacteria, which include the pAra promoter with the AraC protein [25], pLac promoter with the LacI protein [26] and the pR promoter with the lambda repressor [24], to name a few. This kind of competition is frequently used in synthetic systems as one can induce expression of a desired gene by changing the balance of the protein-RNAP competition by adding an inducer, which lowers the number of the repressor protein bound to the promoter.

Other lesser prevalent competition mechanisms are also known, and include two different RNAPs competing for binding to a sequence containing overlapping promoters. In this

case, the RNAPs can inhibit the occupancy of the adjacent promoter. For example, the Crl gene's promoter sequence in *E.coli* was found to contain two types of promoters- a σ^{70} promoter and a σ^{54} promoter, the two RNAPs compete with each other for binding the sequence [27]. Another competition-based mechanism includes the formation of repressive DNA loops, which make the promoter inaccessible to the RNAP if enclosed inside the loop [28].

1.2.2 Looping based regulation

In bacterial enhancer architectures, isomerization of the promoter's open complex is facilitated by DNA looping (as described before). Thereby, by interfering with loop formation, inhibition of transcription can be facilitated. Looping formation interference can be made by the binding of transcription factors (TFs) in the looped DNA sequence. Native examples for such regulation is the IHF protein that binds its site in the glnHPQ operon looping region. As described before, IHF has a dual effect on gene regulation- it can enhance loop formation but it can also repress its formation, depending on the UAS location [22], a second example for such regulation is the #3 and #4 NtrC binding sites in the glnAp2 looping region which were found to repress expression in high NtrC concentration [29]. Looping based regulation was also tested via a synthetic biology approach as a way of controlling expression in synthetic enhancer systems. It was shown that TFs binding to the enhancer's looping region can repress loop formation and by that the target gene expression [30]. The number of TF binding sites, protein size and relative position of the binding sites within the loop was also shown to play an important role in this form of regulation [31].

1.2.3 RNA based regulation

Regulatory phenomenon also occurs at the post-transcriptional level, but for the most part it is poorly understood. RNA based regulation include, transcriptional termination via hairpins or Rho-based mechanism, anti-sense RNA or RNAi (RNA interference), RBP (ribosome binding protein) binding to inhibit ribosome, 5' UTR (untranslated region) secondary structure, etc.

One of the main regulatory mechanisms employed at the RNA level is the use of sequences located on the mRNA's 5' UTR in order to control gene regulation in a *cis* manner. One example of this kind of *cis* regulation is the riboswitch. Riboswitches are complex folded

domains located on the non-coding region of the mRNA which bind a specific metabolite and can then control various aspects of gene regulation such as translational initiation, transcript elongation etc. by creating changes in the RNA structure [32,33]. A second example of *cis*-regulatory elements that can modulate transcription elongation or translational initiation are the attenuators. Attenuators are RNA segments in the non-translated region that can form different secondary structures that can cause either a premature termination of transcription, or a hairpin structure which sequesters the Shine-Dalgarno sequence thus affecting the initiation of translation [34–36].

1.2.4 Roadblocks and polymerase pausing in bacteria

Another form of regulation that may take place at the transcriptional level is "road-blocking". During transcription elongation, a processing RNAP's action may be interrupted by many kinds of roadblocks along the DNA helix such as DNA-bound proteins (transcription factors, repressors, nucleoid associated factors, etc.), other RNA and DNA polymerases [37,38], or the replication machinery. These encounters may cause the elongating RNAP to be (i) stalled- can be resolved by the recruitment of the Mutation frequency decline factor (Mfd) which clears stalled complexes from a DNA template [39,40], (ii) backtracked- can be resolved by the recruitment of the gene regulator factors GreA and GreB which induce internal cleavage of the transcript [37,39,41] or (iii) knocked off by the roadblock [37,39]. As for a collision between two RNAPs, promoter arrangements effects the type of the collision. It has been shown that in a tandem promoter configuration: a rear-end collision can cause a trailing RNAP to aid the leading RNAP to escape transient pausing [38,39,42,43], front-end collision was shown using phage RNAPs, indicating that two opposing phage RNAPs may pass each other and retain their activity [43]. Transcription by RNAP can also be interrupted by pauses which play a diverse regulatory role [44]. RNAP pauses can facilitate RNA folding, factor recruitment, transcription termination and also a way for transcription to be synchronized with translation in prokaryotes. It has been shown that a pause site may comprise of a 16 nucleotides consensus sequence which has distinct features (e.g. GG at the upstream edge) [44]. These examples of regulatory mechanisms indicate that transcriptional regulation can take place not only before initiation, but also during the actual processive production of the mRNA molecules.

1.3 Synthetic biology as a novel basic research approach for regulation

Synthetic biology forces us to test what we think we understand about Biology, by allowing us to take "characterized" genomic elements and rewiring them into new contexts. Therefore, as opposed to the traditional approach for the engineering of proteins and other regulatory elements to obtain the desired behavior, synthetic biology relies primarily on the manipulation of existing gene network architecture. The field is inspired by electrical engineering, computer science and information theory, using fundamental elements from these fields to help guide our designs.

Synthetic biology approaches rely on "biological parts" (e.g. promoters, RBSs etc.) in order to construct composite biological objects that will be used to build full genetic circuitry. The use of the knowledge and elements from different doctrines has enabled researchers to create various logic gates (e.g. AND, OR, XOR etc.) [45] demonstrating the potential to harness the molecular biology parts that evolution produced to form the back-bone of a new hard-wired programming language.

A few examples of these new functions are, (i) modular counter circuit that can count inducible events according to programmed input [46]- can be used in cells which needs accurate count accuracy of tightly controlled processes. (ii) Toggle switch circuit which can switch between two states in a fast and easy manner [47] and (iii) artificial clock which shows oscillation behavior that may lead to engineering new biological functions in cells [48].

Besides from creating computer based circuitry, the synthetic biology approach allows us to study regulatory mechanisms in a modular orderly fashion. A recent paper from our lab [31] demonstrated the utility of this approach, by allowing us to characterize looping repression in σ^{54} enhancer architecture in a systematic fashion. Here, we were able to characterize the effect of TF size, orientation and the length of the loop on gene expression using both a synthetic biology experimental approach and a supporting thermodynamic model.

In this thesis, I use a synthetic biology approach to search for additional regulatory mechanisms that can be distilled via a mixed and match approach by massively engineering novel regulatory elements from naturally occurring parts such as σ^{54} promoters and NtrC binding sites. Using this approach, I found that one particular σ^{54} promoter glnKp is able to substantially inhibit upstream transcribing promoters leading to reduction in gene expression by an order of magnitude. As a result, I constructed a large scale library to search for this effect in known σ^{54} promoters from two different bacteria, and found that about 25% of known or putative σ^{54} promoter are capable of silencing expression from an upstream promoter. Further analysis and experiments showed that the silencing effect can be localized to a conserved pyrimidine 5-mer oligos in the flanking sequences of the core σ^{54} promoter, and as a result the effect is in all likelihood a form of post-transcriptional regulation.

2 Research goals:

2.1 The connection between σ^{54} promoters and the affinity of UASs to EBPs was discussed previously in a small scale for specific variants. Therefore:

2.1.1 In this thesis, I would like to perform a comprehensive characterization of the connection between UAS's affinity for NtrC and all Ntr regulated σ^{54} promoters in *E.coli*.

2.1.2 Second, I would like to characterize new synthetic tandems of NtrC binding sites and test their affinity with the Ntr regulated σ^{54} promotes.

2.1.3 Third, I would like to test the effect of a specific UAS found in *E.coli* with an embedded σ^{70} promoter on activation and expression from σ^{54} promotes.

2.2 During the first experiment I came across a silencing phenomenon originated from one of the tested σ^{54} promoters. That opened up new objectives for this research:

2.2.1 Characterization of the silencing phenomenon from glnKp promoter- testing the effect of promoter orientation, the effect of different σ^{70} promoter and trying to unveil the mechanism behind the silencing effect using a high-throughput assay.

2.2.2 Test the prevalence of the silencing effect in the genomes of *E.coli* and *V.cholera*, and other bacterial strains.

3 Materials and Methods

3.1 Reagents and Chemicals

Agilent

Herculase II Fusion DNA Pol.

Becton Dickinson

Bacto™ Tryptone, Bacto™ Yeast Extract, Bacto™ Agar.

Bio-lab

Ethanol, MBW, TAE (X10).

Biological Industries (Beth-Haemek, Israel):

PBS, UPW.

Biologix

Cell lifter.

Biorad:

electroporation cuvettes.

Cayman Chemicals:

aTc.

Gadot (Israel):

Glycerol.

Hy-Labs (Israel):

Taq-Ready-Mix.

IDT:

Primers.

Invitrogen

TE.

Lucigen:

CloneDirect® Rapid ligation kit.

Merck:

NaCl, $MgSO_4$, DMSO.

New-England-Biolab (NEB):

Restriction Enzymes, Ligases, Q5 Pol, Tac Pol, CIP, PNK.

SeaKem:

LE Agarose.

Sigma-Aldrich:

Kanamycin, Ampicillin, NaOAc, Primers, IPTG.

ThermoFisher Scientific:

Glycogen.

3.2 Bacterial Strains:

- *Escherichia coli* Top10 cells (Genotype: F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G), was used for cloning purposes.
- *Escherichia coli* Δ*RpoN* Top10 cells (Genotype: F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G, Δ*RpoN*), was used for the Δ*RpoN* silencing experiment.
- *Escherichia coli* 3.300LG cells (Genotype: Δ*Gln*L:Δ*Gln*G) used in the Amit lab for testing synthetic enhancers.
- *E.cloni*® 10G (Genotype: F - *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *end*A1 *rec*A1 Φ80d*lac*Z ΔM15 Δ*lac*X74 *ara*D139 Δ(*ara*,*leu*)7697 *gal*U *gal*K *rps*L (Str^R) *nup*G λ-*ton*A). From Lucigen. Cat# LC-60117-2. Was used for transformation of the oligo-library.

3.3 Vectors:

- pACT-Tet: a high copy number vector, containing a selection marker of ampicillin resistance and expressing the repressors TetR and LacI and a dephosphorization defective mutant NRII enzyme [30].
- pLP-RbsK-RA62S-TET-combinatorial-bb: a low copy number vector, containing a selection marker of kanamycin and 3 TetR binding sites in the enhancer loop region. The plasmid was used for the combinatorial experiments.
- pPROLar A122: a low copy number vector, containing a selection marker of kanamycin. The plasmid was used, after modification, for the transcription silencing experiments.
- pUC19: a high copy number vector, containing a selection marker of ampicillin resistance and was used for selection of the unstained variant of the transcription silencing experiments.

3.3.1 Vectors design:

pLP-RbsK-RA62S-TET-combinatorial-bb: the vector (Figure 2B) was created based on the pLP-RbsK-RA62S-TET (Figure 2A) that is used extensively in the lab [30]. The vector's purpose was to act as a backbone for the combinatorial experiments. For that purpose, the mCherry gene was removed from pLP-RbsK-RA62S-TET plasmid using reverse PCR, and a circuit (gBlock®-IDT) consisting of two restriction enzymes (NdeI/KpnI), RBS, mCherry and terminator was added using the Gibson assembly method (Gibson *et al*. 2009). The NdeI/KpnI restriction sites were supposed to become the landing pad for the combinatorial minigenes.

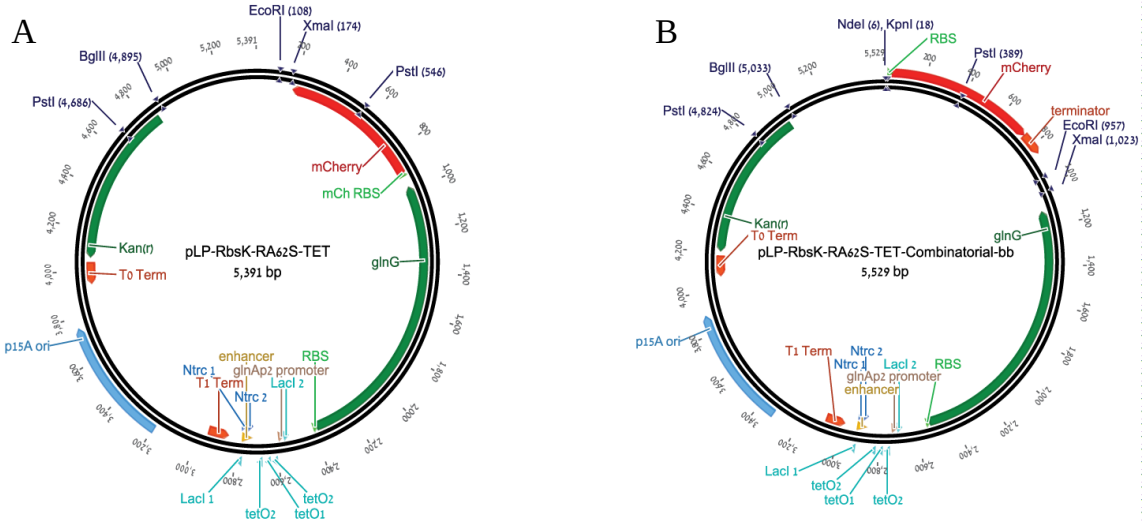

Figure 2: Combinatorial experiment vectors.

(A) pLP-RbsK-RA62S-TET map. The plasmid consists of a synthetic enhancer, glnG gene encoding for NtrC protein and mCherry as a reporter for the synthetic enhancer circuit [30]. (B) pLP-RbsK-RA62S-TET-combinatorial-bb map. RBS, mCherry and terminator were added downstream of an added NdeI/KpnI sites.

pPROLar-A122-eyfp-Ara-glnKp-silencing: the vector (Figure 3B) was created by modifying pPROLar-A122-eyfp (Figure 3A) which was present in the lab. The vector's purpose was to test transcription silencing effect of glnKp. For that purpose, the pPROLar-A122-eyfp plasmid was linearized using double digestion by EagI/EcoRI (EagI removes 39 bp from the 3' end of the KanR gene). "Ara-glnKp" circuit (gBlock® -IDT) consisting

of the end part of KanR gene, two Lac-ara promoters, glnK promoter, mCherry gene and a double terminator (Figure 3C) was added using the Gibson assembly method. Positives clones were verified by growing the transformed cells on Kanamycin agar plates.

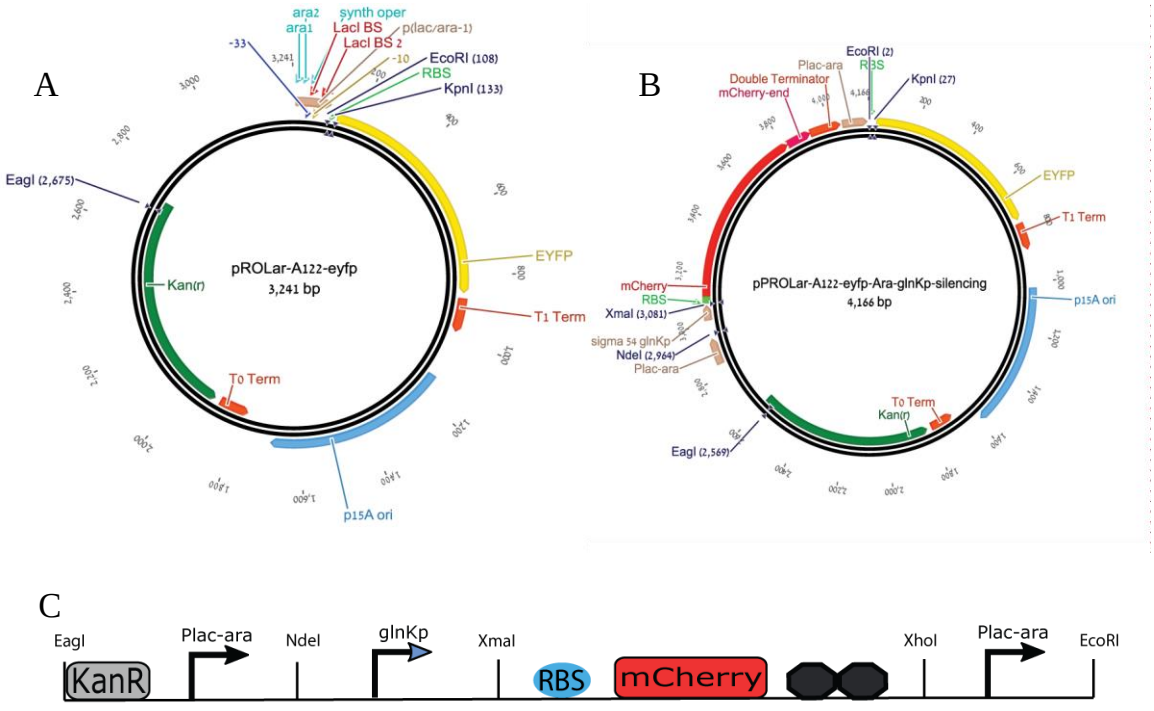

Figure 3: transcription silencing vectors and gBlock® design.

(A) pPROLar-A122-eyfp map. The plasmid consists of EYFP gene regulated by pLac/Ara promoter. It was used as a base plasmid to create pPROLar-A122-eyfp-Ara-glnKp-silencing (B) pPROLar-A122-eyfp-Ara-glnKp-silencing map. pLac/Ara, glnKp, mCherry and double terminator was added. (C) Transcription silencing Ara-glnKp gBlock® circuit design.

pPROLar-A122-eyfp-No-54-Ara-silencing: the vector was created by removing glnK promoter from the pPROLar-A122-eyfp-Ara-glnKp-silencing vector by using reverse PCR. The vector had two purposes: 1. positive control for the glnKp-silencing experiments. 2. Backbone for the oligo-pool transcription silencing experiment.

pPROLar-A122-eyfp-No-70-Ara-silencing: the vector was created by removing the first lac-ara promoter from the pPROLar-A122-eyfp-Ara-glnKp-silencing vector by using

reverse PCR. The vector's purpose was to be a negative control for the glnKp-silencing experiments.

3.4 Methods:

3.4.1 Cloning:

Cloning was done using one of the two following methods:

1. Restriction enzyme- carried out as recommended in NEB's protocols.
2. Gibson assembly- carried out as described in reference [49].

3.4.2 Plasmid production and purification:

Plasmids were produced and purified using NucleoSpin Plasmid Easy Pure Kit (Macherey-Nagel) for plasmidial DNA extraction and purification according to the manufacturer's protocol. In brief, bacteria transformed with the proper plasmids were grown over night (maximum 16 hr) in 5 ml of LB (10mg/ml NaCl, 10mg/ml Tryptone, 5mg/ml Yeast extract) added with 100µg/ml Ampicillin or 25µg/ml Kanamycin. Following centrifugation (Thermo Scientific, Heraeus Megafuge 16R, 5000 rpm for 5 min), purification continued according to the manufacturer's protocol. Following purification, DNA amounts were quantified using nano-drop (Thermo Scientific NanoDrop 2000 Spectrophotometer).

3.4.3 PCR product purification:

Following PCR, salts and proteins were removed using Wizard® SV Gel and PCR Clean-Up system (Promega) for DNA purification from gels and in-vitro enzymatic reactions. The procedure was done according to the manufacturer's protocol.

3.4.4 DNA extraction and purification from gel:

DNA extraction and purification from gel was done mostly according to the manufacturer's protocol of Wizard® SV Gel and PCR Clean-Up system (Promega) for DNA purification from gels and *in-vitro* enzymatic reactions. In the transcription silencing experiments, DNA extraction and purification from gel was done using MIDI GeBA Flex Tube Dialysis Kit (Gene-Bio-Application L.T.D) for DNA purification from gels. In brief, gel slice

containing the desired DNA size was inserted together with 700µl of UPW into the GeBA Flex Tube. The tube was placed in a designated tray and was immersed fully with fresh TAE in an electrophoresis apparatus. The two membranes of the tube were in parallel to the electric field in order to permit the electric current to pass through the tube. Current of 125V was applied for 1hr and then for 3 min in the reverse current. Following DNA elution, the solution was pipetted and transferred to a 1.5ml microcentrifuge tube which was then centrifuged for 1 min at maximum speed. DNA-containing solution was transferred to a clean 1.5ml microcentrifuged tube. DNA was later precipitated by adding 1:10 µl of 3M NaOAc, 1:200 µl of Glycogen and 1:1 µl of Isopropanol. Following overnight incubation in -20°C, washes using 1ml of 70% ethanol were applied. Eventually the pellet was re-suspended in 20µl of MBW.

3.4.5 Combinatorial experiment cassette design:

Combinatorial cassettes were ordered as dsDNA minigenes from gen9 Inc. each minigene we ordered was ~500 bp long, and contained the following parts (Figure 4): NdeI restriction site, variable sequences of NtrC tandem UAS and σ^{54} promoter (shown in Table 1 and Table 2) and KpnI restriction site at the 3' end. The NtrC tandem UAS and σ^{54} promoter were separated by a looping segment of 70 bp.

Insertion of minigene cassettes to the combinatorial plasmid was done by double digesting both cassettes and plasmids by NdeI/KpnI, followed by ligation to pLP-RbsK-RA62S-TET-combinatorial-bb and transformation to 3.300LG *E.coli* cells containing pACT-Tet vector.

Figure 4: Combinatorial experiments cassette design.

NtrC tandem UAS binding sites and σ^{54} promoter sequences were changed according to table 1 and table 2. Each promoter was ordered with each of the NtrC tandem UAS binding sites. NdeI and KpnI restriction sites were used for cloning.

#	Tandem UAS Name	UAS affinity to NtrC	Native promoter	Distance from native TSS	UAS sequence	Source	Site affinity to NtrC
1	KK	+	GlnKp	-115	TGGTGC	References: [21,50]	+
			GlnKp	-87	TGCACTGTCATAGTGCG	References: [21,51]	+
2	glnAp1	+++	glnAp2	-140	TGCACCAACATGGTGCT	References: [18,50]	+++
			glnAp2	-108	AGCACTATATTGGTGCA	References: [18,50]	+++
3	HH	++	glnHp2	-135	TGCACAATTTTAGCGCA	References: [22,50]	+
			glnHp2	-109	TGCCCCAGAATGGTGCA	References: [22,50]	+++
4	CC	++	astCp2	-275	ATGTCAACGATGGCGCA	References: [18,50]	++
			astCp2	-253	TGCCCGCTTTTGGTGCG	References: [18,50]	++
5	AA	+	glnAp2	-68	TTTTGCACGATGGTGCG	References: [18,50]	+
			glnAp2	-45	AACGCCTTTTAGGGGCA	References: [18,50]	+
6	KH	+	GlnKp	-87	TGCACTGTCATAGTGCG	References: [21,50]	+
			glnHp2	-79	GCCCTATAAATCGTGCA	References: [22,50]	+
7	AH	++	glnAp2	-89	ATTCACATCGTGGTGCA	References: [18,50]	+
			glnHp2	-122	CGCACCAGATTGGTGCC	References: [22,50]	+++
8	CP	+++	astCp2	-233	TGCGTCAGAATGGCGCA	References: [18,50]	+++
			nacp	-152	TGAACCATCGTGGTGCA	References: [20,50]	+++
9	HA	++	glnHp2	-79	GCCCTATAAATCGTGCA	References: [22,50]	+
			glnAp2	-140	TGCACCAACATGGTGCT	References: [18,50]	+++
10	AC	++	glnAp2	-68	TTTTGCACGATGGTGCG	References: [18,50]	+
			astCp2	-275	ATGTCAACGATGGCGCA	References: [18,50]	+++

Table 1: NtrC UAS binding sites used in the combinatorial experiment.

Sequence origin, native promoter, location and the affinity of the individual and tandem UAS to NtrC is detailed. Affinity is represented as plus signs. One plus- weak affinity, two pluses- medium affinity, three pluses- strong affinity.

Promoter name	promoter sequence
glnKp	TTAACTTCCTGCTCTCTTTCTCGTTTTTCATTTCTGGCACACCGCTTGCAATACCTTCTT
glnHp2	GCCGCATCTCGAAAAATCAAGGAGTTGCAAAACTGGCACGATTTTTTCATATATGTGAAT
astCp2	TAGCCTCCGCCGTTTATGCACTTTTATCACTGGCTGGCACGAACCCTGCAATCTACATTT
nacp	TTGGTTAGCTTGTACATCAACACCAAAATAAAACTGGCAAGCATCTTGCAATCTGGTTGT
glnAp2	GCATGATAACGCCTTTTAGGGGCAATTTAAAAGTTGGCACAGATTTCGCTTTATCTTTTT

Table 2: σ^{54} promoters used in the combinatorial experiment.

Promoter sequences used in the combinatorial experiment. Reference: [50].

3.4.6 Combinatorial experiment expression assay:

Expression assay for all combinatorial experiments was carried out as described by Brunwasser-Meirom *et al.* [31]. Briefly, the combinatorial experiment strains were grown overnight (not more than 16 hr) in fresh LB with the appropriate antibiotics (100µg/ml Amp and 25µg/ml Kan) followed by a 1:100 dilution with fresh LB and antibiotics (Amp/Kan). Cells were grown to mid-log phase (O.D600 of ~0.6) as measured by a spectrophotometer (Novaspec III, Amersham Biosciences) followed by resuspension with a low-growth/low-autofluorescence BA buffer (for 1L: 0.5 g Tryptone, 0.3 ml Glycerol, 5.8 g NaCl, 50 ml 1M MgSO4, 1 ml 10xPBS buffer at pH7.4, and 950 ml DDW) with the appropriate antibiotics (Amp/Kan). 1mM IPTG was added at this point to hinder the lacI which is expressed from the pACT-TET plasmid and represses the glnAp2 promoter in the NtrC level module. 2ml of the resuspended culture with IPTG and antibiotics were dispensed in duplicates to a 48-well plate. Appropriate concentration out of 24 levels of aTc were dispensed to each well (2, 1.6, 0.85, 0.55, 0.36, 0.23, 0.15, 0.1, 0.06, 0.04, 0.03, 0.02, 0.01, 0.007, 0.005, 0.003, 0.002, 0.001, 0.0008, 0.0005, 0.0004, 0.0002, 0.0001, 6.7e-6) ng/µl, covering four to six orders of magnitude. The plates were then incubated at 37°C shaker for 3hr until cultures reached steady-state growth and fluorescent protein reached maturity. Measurements of mCherry fluorescence and O.D600 were done by dispensing 200µl of culture into each well of a 96-well plate and were carried out by a plate-reader (Tecan F200).

3.4.7 Robot measurement:

High resolution combinatorial experiments were carried out as described by Brunwasser-Meirom *et al.* [31]. Briefly, the experiments were performed on a Tecan EVO 100 MCA 96 multichannel liquid handling system. Experiments were done as described above with slight changes: cells were grown to O.D600 of ~0.1 in a 96-well plate, than centrifuged and resuspended in BA. aTc was added manually to an inducer plate, and were distributed automatically in different concentrations to the 96-well plate. The plate was incubated at 37°C shaker for 2 hr and both mCherry fluorescence and O.D600 were measured every 20 min.

3.4.8 glnKp transcription silencing expression assay:

glnKp transcriptional silencing expression assay was performed as follows: the glnKp silencing strains were grown overnight (not more than 16 hr) in fresh LB with the appropriate antibiotic (25µg/ml Kan) followed by a 1:100 dilution with fresh LB and antibiotic (Kan). Cells were grown to mid-log phase (O.D600 of ~0.6) as measured by a spectrophotometer followed by resuspension with a low-growth/low-autofluorescence BA buffer with the appropriate antibiotics (Kan). Following incubation in a shaker for 3 hr in 37°C, mCherry to EYFP fluorescence ratio was measured by a FACSAria (Becton-Dickinson).

3.4.9 glnKp transcription silencing in ΔRpoN strain expression assay

glnKp transcriptional silencing expression assay was performed as described in 3.4.8, in a Top10 ΔRpoN strain.

3.4.10 Oligo-library experiment cassette design:

Oligo-library was ordered as ssDNA oligos from two different companies (Twist Biosciences and CustomArray Inc.). Each oligo we ordered was ~150 bp long and contained the following parts: 5' primer binding sequence, NdeI restriction site, specific 10 bp barcode, variable tested sequence, XmaI restriction site and 3' primer binding sequence. The barcode and the promoter sequence were separated by a spacer segment of 23 bp (cassette design is shown in Figure 5). Tested sequences were grouped into 5 groups as described in Table 3.

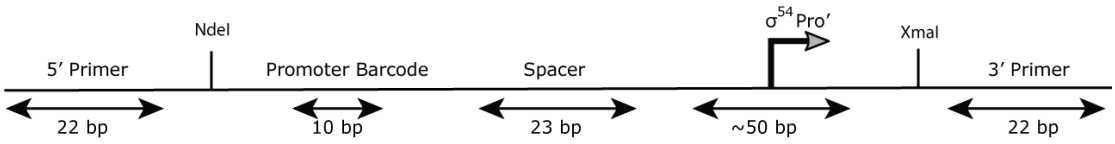

Figure 5: Oligo-library "silencing" cassette design.

Group name	#sequences*	Source	comments
glnKp perturbations	268	Mutations done in glnKp sequence.	Understand the mechanism of -glnKp silencing effect
Known σ^{54} promoters	456	References: [50,52]	Test known σ^{54} promoters for transcription silencing effect
Genome wide consensus σ^{54} binding scan	11,430	Ecoli K12 genome. GenBank: U00096.2 V.cholera genome. GenBank: CP003069.1	Test for silencing effect of consensus σ^{54} binding sequences across the genome
No promoter (control)	260	Ecoli K12 genome. GenBank: U00096.2	Low σ^{54} binding consensus score
σ^{70} promoter (control)	250	Reference: [50]	Test silencing effect for σ^{70} promoters

Table 3: Oligo-library sequences groups.

* The number includes both forward and reverse promoter orientation (1:1)

3.4.10.1 σ^{54} consensus binding site scoring

The consensus probability matrix for σ^{54} binding (Appendix 1) was based on the compilation of 186 σ^{54} promoters (Table 3 in [52]). The genomes of *E.coli* and *V.cholera* were scanned using a Matlab script that assigns a σ^{54} probability score to all possible 16 bp-long sequences, based on similarity to the consensus site. In detail, each base in the 16 bp sequence is given a value of 0-1 according to the table in Appendix 1. The values of all 16 bases are summed, and the total is normalized by first subtracting the lowest possible total (1.679) and then dividing by the difference between the highest possible total (11.2) minus the lowest possible total (1.679), resulting in a final score in the range [0,1] (shown in Equation 1). Genomic sequences with scores in the range [0.765, 1] were chosen as candidates for σ^{54} binding. Genomic sequences with scores in the range [0, 0.5] were chosen as candidates that were highly unlikely to bind σ^{54} and were taken as the "no promoter" group.

$$SCORE = \frac{\sum(matrix\,values) - 1.679}{11.2 - 1.679}$$

Equation 1: σ^{54} consensus probability score.

3.4.10.2 glnKp mutations design

Single nucleotide mutations in the core glnKp promoter TGGCACACCGCTTGCA were chosen based on the score of the mutated sequence. For each position, mutations were chosen to reflect all possible mutated scores attainable for that position. If different mutations at a given position resulted in the same score, only one mutation was chosen for that position and score. There were some exceptions to this rule: all possible single-nt mutations were chosen for positions 7 and 8, and no mutations for the CGC at positions 9-11 were chosen since the consensus promoter is apparently insensitive to the sequence of these bases. We avoided choosing mutations that contained known transcription-factor binding sites listed in the RegulonDB database [50].

Mutations in the glnKp flanking regions (outside of the 16-nt core promoter) were designed using Matlab by generating random replacement sequences that differed from the glnKp promoter flanking regions at all positions. An equal probability was given to all 3 possible replacement nucleotides at each position. We checked that mutated sequences contained no known TF binding sites, based on the regulonDB database [50]. The mutated flanking sizes ranged from 0 (the original promoter) to 24 nt (entire flanking region mutated), in size increments of 1 nt. Four different mutated sequences were chosen for all flanking sizes. Since the 16-nt core promoter is not located at the center of the full 50-nt glnKp promoter sequence, mutated flanking regions upstream and downstream of the core promoter were not necessarily of equal length. To accommodate for this asymmetry, for each choice of flanking size we left equally-sized flanking regions adjacent to the core promoter unchanged, both upstream and downstream of the promoter. The size of the mutated flanking region therefore corresponds to the larger of the two mutated flanking regions (upstream and downstream). The mutated flanking sequences for the reverse orientation of the glnKp promoter were chosen by running the same Matlab script on the reverse complement of the glnKp promoter, and therefore do NOT correspond to the reverse complement sequences of the mutated direct promoters.

3.4.11 Oligo library cloning

High resolution oligo library cloning was based on cloning protocol developed by the Segal group [53]. Briefly, ssDNA library in a specific concentration was added to a PCR reaction

mix (970 µl polymerase buffer, 243 µl 10mM dNTPs, 970 µl Forward-primer, 970 µl Reverse-Primer, 97µl Herculase polymerase, up to 4800µl UPW), the mix was than dispensed into a 96-well plate. Final DNA concentrations varied as a factor of the library source (Twist Bioscience-0.025 ng/µl, CustomArray- 0.004 ng/ µl). Oligo library was amplified using PCR machine Mastercycler proS (Eppendorf) with the following program: step1: 95°C 1min, Step2: 95°C 20seconds, Step3: 50°C 20 sec, Step4: 68°C 1min, Step5: 68°C 4min. the number of cycles for steps 2-4 varied as a factor of the library source (Twist Bioscience-8 cycles, CustomArray- 25 cycles). Following PCR in a 96 well plate, DNA content was concentrated using Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-30 membrane (EMD Millipore) for DNA purification and concentration- content of wells was joined and divided into 5 Amicon tubes. Concentration continued according to the manufacturer's protocol. All tubes were joined eventually into one 1.5ml microcentrifuge tube. Purification of the concentrated PCR product was performed as described above. Following purification, dsDNA was cut overnight with XmaI and NdeI at 37°C and run in electrophoresis gel for separation of the desired size. Gel purification of the desired band was performed as described above. The cut and cleaned DNA fragments were ligated to 150 ng of the cut plasmid "pPROLar-A122-eyfp-No-54-Ara-silencing" using clone direct ligase (30°C for 30min followed by heat inactivation for 15 min at 70°C). Ratio between inserts and plasmid was 1:1 in order to reduce multiple inserts in ligation reactions. Ligated plasmids were transformed electrochemically (1600V, 5τ) to *E.cloni*® 10G electrocompotent cells (Lucigen Corporation) using Multiporator (Eppendorf) and plated on 28 Kan 14cm agar plates in order to conserve complexity. In the following morning, each plate was treated as follows: 10ml of LB with antibiotic (Kan) was poured into the plate and the colonies were scrapped using cell lifter (Biologix), the culture was transferred to an Erlenmeyer for growth. Transcription silencing expression assay will be elaborated below.

3.4.12 Oligo-library transcription silencing expression assay

Oligo-library transcription silencing expression assay for the transformed oligo-pool library was developed based on [53] and was carried out as follows:

3.4.12.1 Culture growth

Library containing bacteria were grown with fresh LB and antibiotic (Kan). Cells were grown to mid-log phase (O.D600 of ~0.6) as measured by a spectrophotometer (Novaspec III, Amersham Biosciences) followed by resuspension with BA buffer and the appropriate antibiotic (Kan). Culture was grown in BA for 3 hours prior to sorting by FACSAria cell sorter (Becton-Dickinson).

3.4.12.2 FACS sorting

Sorting was done at low sample flow rate and sorting speed of ~20,000 cells per sec. Cells were sorted into 16 bins (100,000 cells per bin) according to the mCherry to EYFP ratio in two groups: (i) bins 1-8: high resolution on low ratio bins (30% scale), (ii) bins 9-16: full resolution bins (3% scale).

3.4.12.3 Sequencing preparation

Sorted cells were grown over night in 5ml LB and appropriate antibiotic (Kan). In the next morning, cells were lysed (TritonX100 0.1% +TAE 1%:15ul, culture: 5ul. 99°C for 5min and 30°C for 5min) and the DNA from each bin was subjected to PCR with different 5' primer containing a specific bin barcode. PCR products were purified using Wizard® SV Gel and PCR Clean-Up system for DNA purification from gels and in-vitro enzymatic reactions (Promega). Equal amount of DNA (20 ng) from each bin were joined to one 1.5 microcentrifuge tube for further analysis.

3.4.12.4 Sequencing

Sample was sequenced using Illumina MiSeq using "MiSeq Reagent Kit V3" 150SR, 20% PhiX were added as a control. From each read the bin barcode and the sequence of the strain were extracted using a custom python script: which fixes the read's orientation for all the reads to in the same orientation, identification of the constant sequences in the read and extracting the variables: bin barcode, sequence barcode and the variable tested sequence and eventually mapped all the reads to combinations of tested sequence and expression bin. This resulted in ~5,000,000 uniquely mapped reads each containing a sequence and expression bin barcode pair.

3.4.12.5 Deriving expression ratio profiles

We first removed all reads mapped to bin number 16 from the analysis to eliminate biases originating from out of range fluorescence measurements. Next we filtered out sequences with low read counts keeping only those with at least 30 reads in one of the sets of bins (1-8, 9-15). We than generated a single profile by replacing bin 9 with bins 1-8 and redistributing the reads in bin 9 over bins 1-9 (while correcting for the relative width of bin 9 to bins 1-8). Next, we estimated for each sequence the corresponding fraction of cells in each bin based on the number of sequence reads from that bin that mapped to that strain (the reads of each bin were first normalized to match the fraction of the bin in the entire population). This procedure resulted in an expression ratio profiles over 14 bins for ~1400 strains.

3.4.12.6 Deriving mean expression ratio

For each of the ~1400 sequences, we defined the mean expression ratio as the weighted average of the ratios at the geometric centers of the bin, where the weight of each bin is the fraction of the strain in that bin.

3.4.12.7 Minimal HG enrichment tests

We sorted the sequences according to their mean expression ratio values and calculated minimal hyper geometric enrichment [54] scores for each of the strain groups (known σ^{70} promoters, known σ^{54} promoters, glnKp perturbations, etc.) for both the top and the bottom of the sorted list. The test was preformed using the xlmhg python library [55] version 1.1rc3.

3.4.12.8 Motif detection

Sequence motifs were identified in the mean expression ratio sorted list using DRIMust [56], a tool for discovering sequence motif enrichment in sorted lists of sequences.

4 Results:

4.1 Combinatorial experiment design:

σ^{54}-promoters can be classified according to their activating EBP. Typical promoter architectures include multiple potential UASs, each consisting of a tandem of binding sites with varying affinities to the EBP [21,51]. In order to understand the UAS affinity-promoter connection I designed and carried out a combinatorial experiment in which I measured the expression level of mCherry created by various UAS-promoter combinations. To address this question I used a two module system: a measurement module and an NtrC induction module. In the measurement module I tested 5 different, well studied, σ^{54}-promoters (glnAp2, glnKp, glnHp2, nacp and astCp2) with 10 different UAS sequences each consisting of a tandem of NtrC binding site with different affinities (Figure 6A). In particular, I chose: the 5 natural UASs of the tested promoters, four UAS sequences were synthetic tandems (created by a mix of individual binding sites from the natural UAS sequences) and one UAS had a σ^{70} promoter embedded within it (will be referred to as its promoter's name: glnAp1) as found in *E.coli* (materials and methods- Table 1). The measurement module included a synthetic enhancer consisting of one of the ten UAS sequences, a 70 bp spacing sequence (designed so to not bind any known *E.coli* TF), one of the five promoters, and an mCherry reporter protein. Given the different possible combinations, I constructed 50 different measurement modules.

In order to properly control my experiment, and to ensure that fluorescent output will only be due to our engineered synthetic enhancer circuit, I used ΔglnG (ΔNtrC) strain and another home-built synthetic enhancer circuit expressing NtrC (the NtrC induction module), which can be induced by anhydrous tetracycline (aTc) (Figure 6B). In brief, the circuit is based on the lab's past synthetic enhancer designs, in which it was shown that TetR can substantially repress NtrC expression when bound to the looping. Upon induction with aTc, TetR is removed from the looping region enabling the formation of DNA loops and resultant gene expression (for the mechanism refer to [30]). The circuit can thus be used as an inducible promoter with substantially less leaky profile as compared with a more standard tool like pTet.

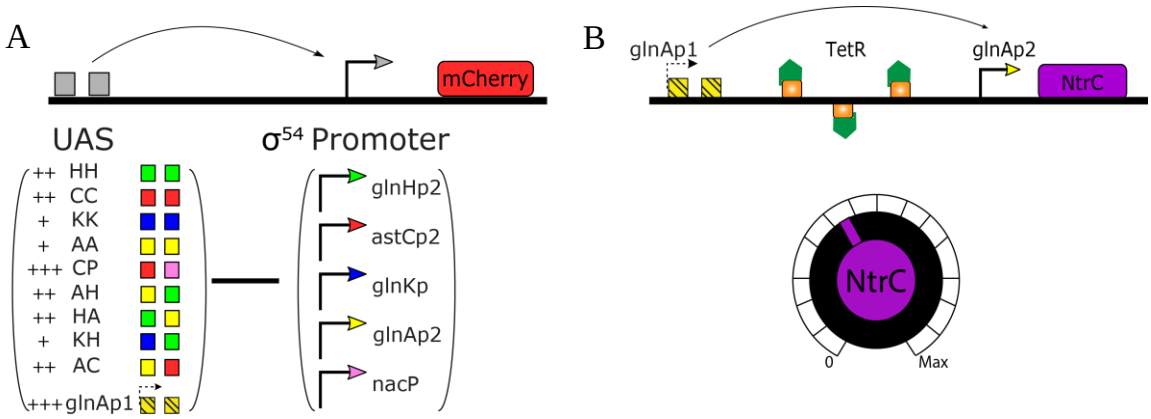

Figure 6: Combinatorial experiment circuits design.

(A) The measurement module: five σ^{54} promoters differentiated by an arrow color, color of the different NtrC binding site in the UAS tested are as the native promoter, one UAS contained the glnAp1 promoter. Affinities of UASs to NtrC are described as plus signs- one plus: weak binding, two pluses- medium binding, three pluses- strong binding (B) The NtrC induction module: Basal level of NtrC was created using glnAp1 promoter, three TetR sites were located in the looping region. NtrC levels were controlled using the addition of an aTc inducer.

4.1.1 Combinatorial experiment-control results:

It was shown that UASs can activate transcription from long distances [9]. In the experiment's design the two modules were both based on σ^{54} architecture (thus both containing a UAS and σ^{54} promoter) and located on the same plasmid. Therefore, I initially wanted to test whether the NtrC-inducing module's UAS can cross-activate the tested σ^{54}-promoter in the measurement module by DNA looping. To achieve this goal, I removed the UAS from the measurement module and measured mCherry fluorescence while rising concentrations of aTc were added (leading to rising levels of NtrC, Figure 7A). I showed that 3 out of 5 promoters (glnAp2, glnKp and nacp) presented rising mCherry fluorescence as aTc concentrations rose, as opposed to glnHp2 and astCp2 that showed the same levels as were without any σ^{54}-promoter in the measurement module (Figure 7A- light blue), indicating that cross activation within my system is possible, depending on the tested promoter. Given that the distance on the plasmid between the NtrC induction module's UAS to the measurement module's σ^{54}-promoter is 3000 bp, this is the farthest activation

with NtrC that has been recorded to date. Interestingly, the two promoters, which did not exhibit this effect (glnHp2 and astCp2) are known to be weaker σ^{54}-promoters, and thus it may be possible that the largest possible separation between a UAS and a σ^{54}-promoter will be a direct function of promoter strength.

In order to verify that the only expression of mCherry seen in the experiment results originated from the tested σ^{54}-promoter and not by other component in the plasmid, I removed the promoter from the measurement module and measured mCherry fluorescence as described before. As shown in Figure 7B, glnAp1 UAS, which contains a σ^{70} promoter, was the only UAS that showed mCherry expression. Fluorescence results for glnAp1 shows up to two fold repression when aTc concentration rose to the maximum, consistent with the fact that this particular UAS acts as a promoter and as a binding site for NtrC. Here, as more NtrC is created, it binds the NtrC binding sites and therefore represses glnAp1 promoter activity.

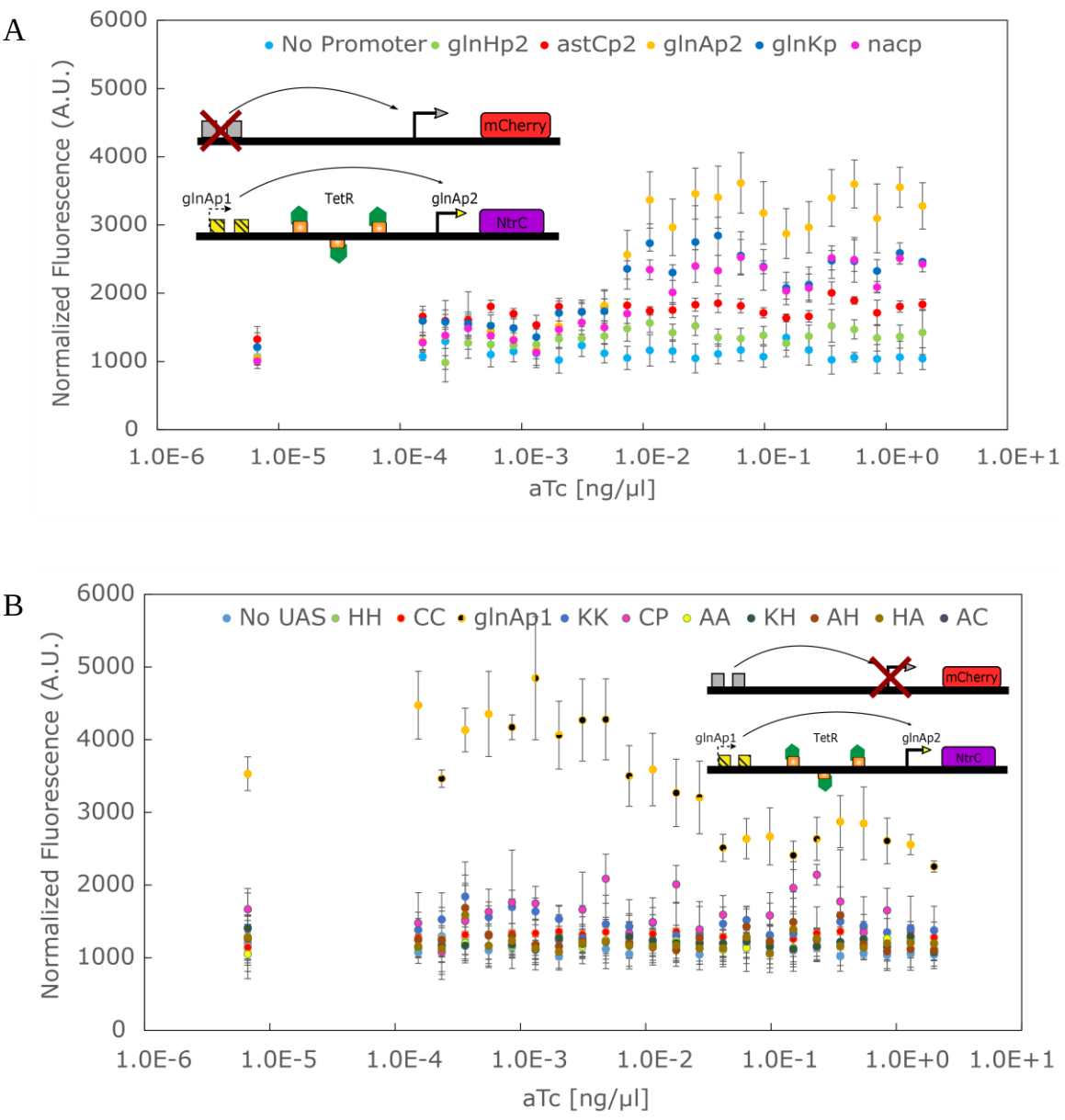

Figure 7: Combinatorial experiment-control results.

(A) Comparing different σ^{54} promoters' activity without UAS in the measurement module. glnAp2, nacp and glnKp showed higher mCherry expression as aTc concentration rose, while glnHp2 and astCp2 did not. (B) Comparing different UASs transcription ability without σ^{54} promoter in the measurement module. Most UASs showed no mCherry expression. glnAp1 showed repression effect as aTc concentrations rose.

4.1.2 Combinatorial experiment- results:

In order to quantify the measurement modules gene expression's output, mCherry fluorescence was measured using 24 different aTc concentration for each of the 50 UAS-promoter combination produced. Figure 8 shows an example of two promoter activity results for every single UAS combo in the design (i.e. ten data sets per panel). In Figure 8A, I plot the mCherry fluorescence measurements for nacp, which shows rising fluorescence values as with aTc concentration, even without UAS (light blue -indicating cross activity as showed in Figure 7A). In addition, the glnAp1 UAS showed a minor decrease in fluorescence as aTc concentration rose consistent with the control measurements shown in Figure 7B. In Figure 8B, I plot the mCherry fluorescence measurements for the ten synthetic UAS combos for astCp2 promoter. Here, the data show similar behavior to the one seen in Figure 7B in which there was no promoter in the measurement module (i.e. no mCherry expression for most UAS and glnAp1 showed repression effect as aTc concentration rose), indicating that astCp2 promoter did not act as a promoter in our experiment, or had a very weak effect.

In order to be able to compare between the different variants in an easy manner, fluorescence values with the minimum and maximum aTc concentration were used. Figure 9 shows the fluorescence results for the maximum (Figure 9A) and minimum (Figure 9B) NtrC levels (defined by the aTc concentration). As shown in Figure 9A, when NtrC levels were at the maximum- all promoters except astCp2 showed a minimum two fold change in expression compared to the "no promoter" variants. Moreover, when comparing between fluorescence derived from different UASs for a specific promoter- no significant change was seen for most promoters (glnHp2, astCp2, glnKp and nacp) and the mean normalized fluorescence value for each promoter was ~3000 A.U. The only promoter that did present a change in fluorescence while changing the UAS was glnAp2 which showed statistically different results than the other promoters (F test), glnAp2 results show that some UASs cause very high expression (CC,KH) and some very low expression (AH). It is problematic to address the expression yielded from combinations containing the glnAp1 UAS because mCherry expression levels originate from both tested promoter and glnAp1 promoter.

As for when there NtrC levels in the cell were at the minimum (Figure 9B), no mCherry fluorescence was seen for most of the UAS-promoter combinations. As can be seen in Figure 9B, one specific UAS, glnAp1 UAS, had shown high mCherry expression although NtrC was absent in the system. This can be explained by the fact that glnAp1 UAS has a σ^{70} promoter function that does not need activation from NtrC. Interestingly, mCherry expression from glnAp1 UAS was seen for all of the promoters tested except for glnKp in the absence for of NtrC. Thus, the 50 bp sequence encoding the glnKp sequence (including a core promoter and flanking sequences which do not encode a known binding site for a TF) was somehow able to silence mCherry expression from glnAp1. The remaining experiments in the thesis are designed to explore this silencing effect, its prevalence in other σ^{54} promoters, its over-all occurrence rate in the genomes, and a possible regulatory mechanism.

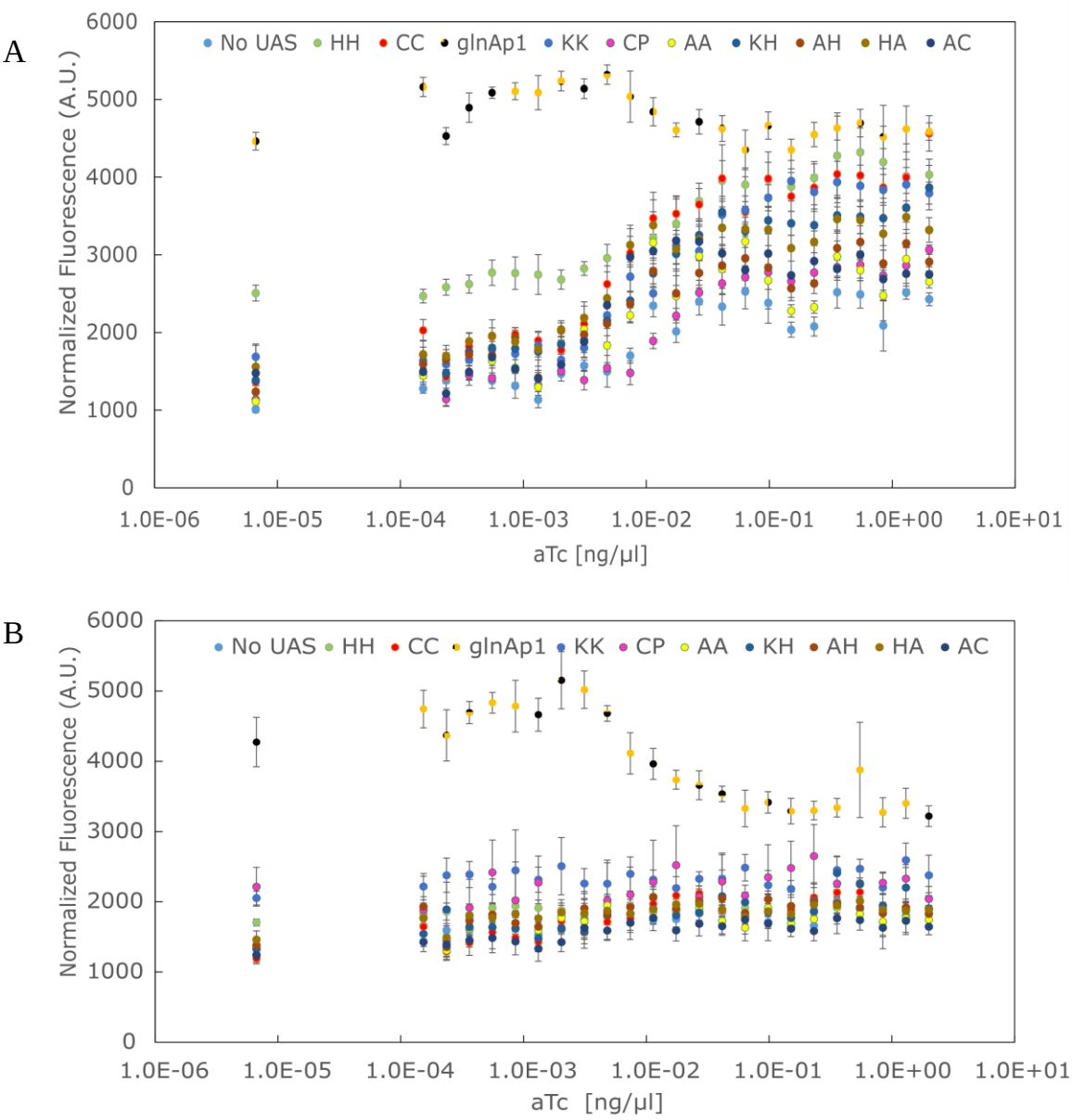

Figure 8: nacp and astCp2 activity with different UASs in rising aTc concentrations.

(A) nacp results- showed an increase in fluorescence as aTc concentration rose, variant without UAS showed same pattern. glnAp1 UAS showed a minor decrease in fluorescence. (B) astCp2 results- showed no change in fluorescence as aTc concentration rose in all UASs tested except for glnAp1, glnAp1 showed repression effect as aTc concentrations rose.

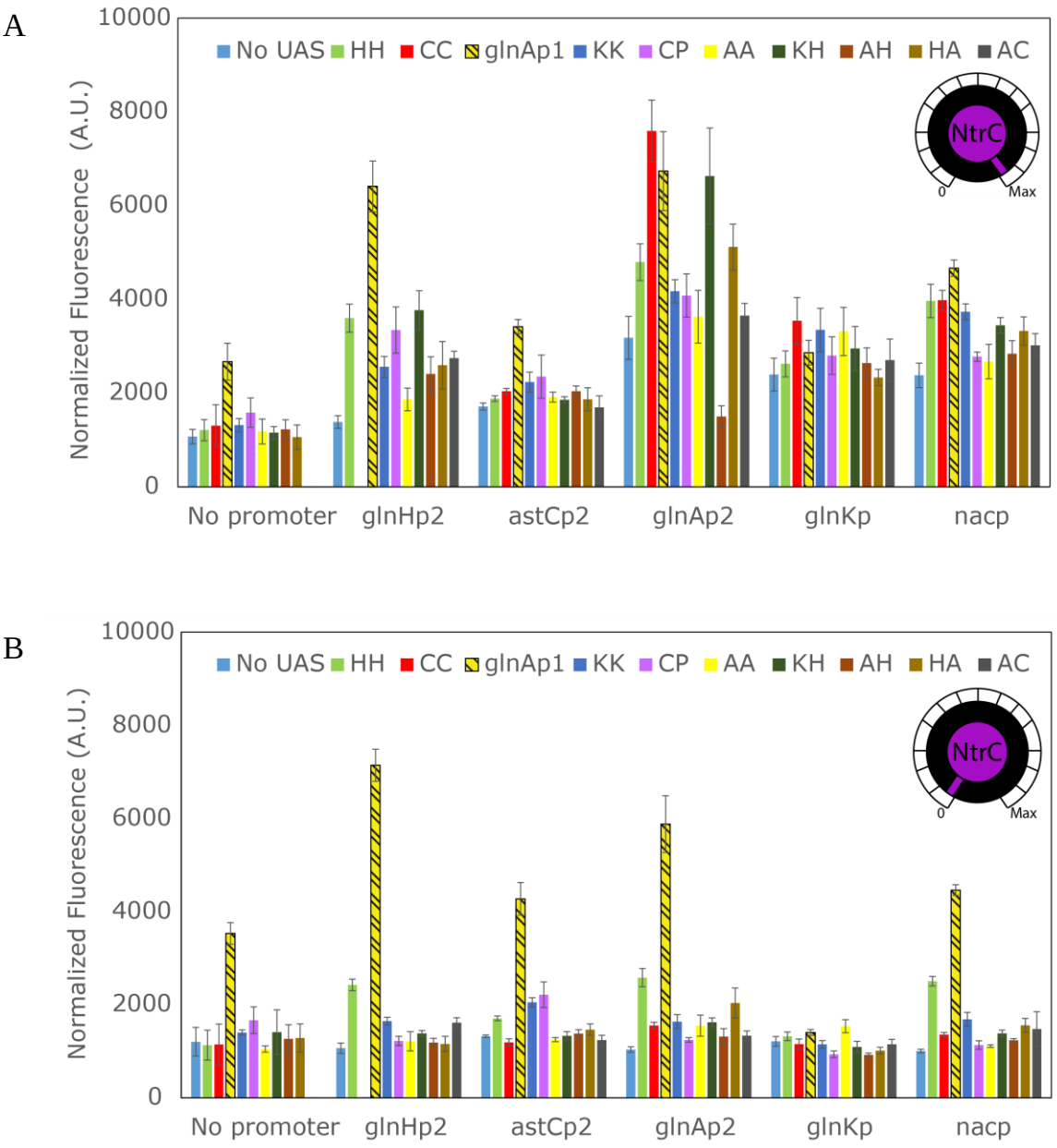

Figure 9: Combinatorial experiment results.

(A) Normalized fluorescence for all promoter-UAS combinations with maximum NtrC in the system. Most promoters didn't show change in expression with the UAS change except for glnAp2. (B) Normalized fluorescence for all promoter-UAS combinations without NtrC in the system. High expression of mCherry was seen for glnAp1 variant for all promoters tested except for glnKp.

4.1.3 glnKp's silencing effect is unidirectional

In order to further understand the mechanism behind glnKp's silencing effect (seen in Figure 9B) - glnKp's sequence orientation effect was tested. UAS-promoter combination: "glnAp1 UAS-glnKp" was tested again without NtrC in the system, in this case, glnKp's sequence was inserted to the measurement module in the reverse orientation. The results in Figure 10 show that while glnKp placed in the forward orientation can silence expression form glnAp1 (as shown by the middle bar), the promoter in the reverse orientation did not show a silencing effect at all and had a similar effect on mCherry expression as if there was no promoter at all (Figure 10– right and left bars respectively), thereby allowing me to conclude that glnKp's silencing effect is unidirectional.

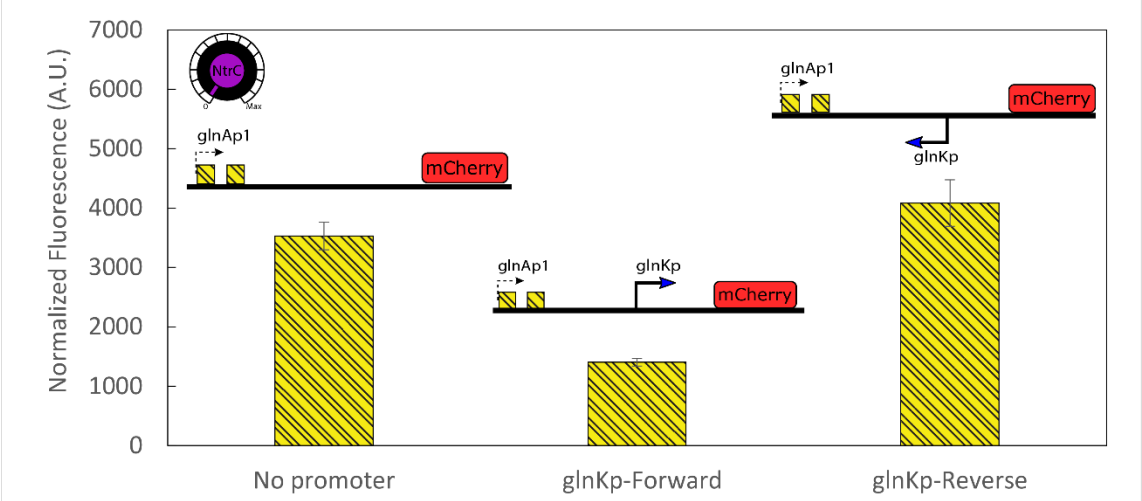

Figure 10: Transcription silencing orientation effect in glnKp.

Testing glnKp orientation effect on expression without NtrC in the system. glnKp in the forward orientation silences expression from glnAp1 UAS while placing glnKp in the reverse orientation eliminates the silencing effect and values are as in the no-promoter variant.

4.2 glnKp's transcription silencing:

The results shown in Figure 9B showed that glnKp's sequence could silence the expression of mCherry originated from glnAp1. Trying to explain this result, I postulated an hypothesis in which a stalled σ^{54}:RNAP holoenzyme (a state which occurred when there was no NtrC) can block a trailing elongating RNAP.

In order to validate this hypothesis, I constructed a new circuit (Figure 11A) containing the glnKp sequence upstream from a reporter mCherry gene and downstream from the known pLac/Ara promoter. A second pLac/Ara promoter was placed upstream of an eyfp reporter gene in order to normalize the results. In order for σ^{54}:RNAP holoenzyme to be bound to the glnKp promoter, no UAS was added to this design, thus allowing me to treat the poised polymerase as a very large transcription factor or DNA binding protein.

In order to quantify the silencing effect I carried out my measurement at the single cell level on a flow cytometer. The ratio of mCherry/eyfp fluorescence was the indicator for the silencing effect (low ratio means silencing). Three circuits were tested in this assay (Figure 11B): 1. "no pLac/Ara": a circuit without pLac/Ara upstream from glnKp and mCherry (Figure 11B-blue). 2. "no glnKp": a circuit without the glnKp sequence between pLac/Ara and mCherry (Figure 11B-purple). 3. The full circuit containing both promoters upstream of mCherry (Figure 11B-orange). "no pLac/Ara" circuit showed very low mCherry/eyfp ratio, indicating that no mCherry was expressed (specific fluorescent protein expression for all variants can be seen in Figure 11C). "no glnKp" circuit showed a nine fold higher ratio then the "no pLac/Ara" variant, indicating high mCherry expression originated from pLac/ara. The full circuit variant showed a fivefold decrease in ratio compared to the "no glnKp" variant and a slight increase in ratio as compared with the "no pLac/Ara" variant, indicating low mCherry expression levels. These results show that the presence of the glnKp sequence in mCherry's 5' UTR region is consistent with the ability to silence transcription of a trailing elongating RNAP. However, one can imagine other regulatory mechanisms that are unrelated to transcription and take place at the post-transcriptional level.

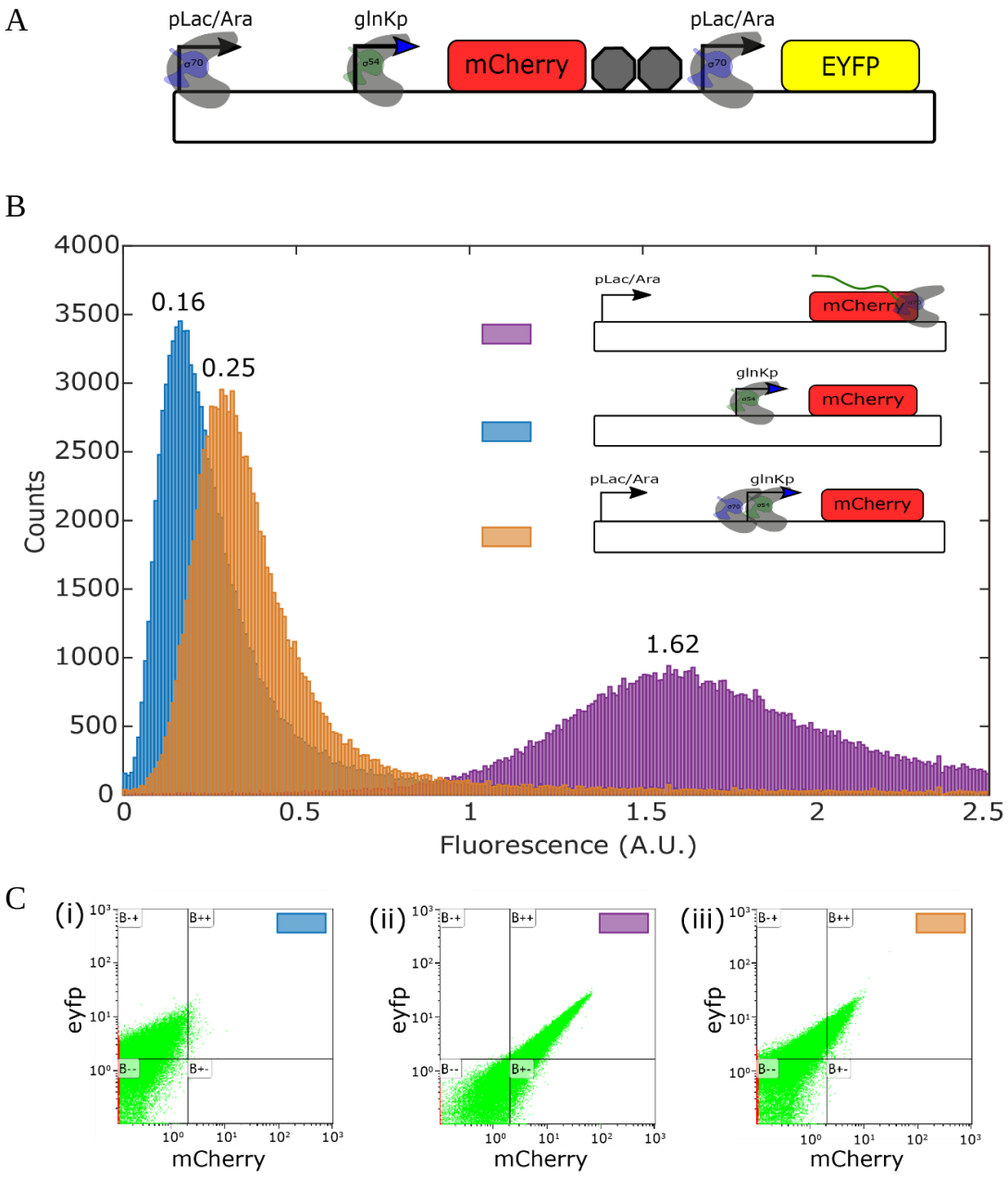

Figure 11: glnKp transcription silencing results.

(A) glnKp transcription silencing circuit design. σ^{54}: RNAP holoenzyme bound to glnKp, pLac/Ara was placed in two locations: upstream of glnKp and mCherry, and upstream of eyfp reporter gene. Double terminator separates the two parts. (B) FACS results of the transcription silencing effect for glnKp, fluorescence is shown for the ratio of mCherry to eyfp. Purple: no glnKp in between pLac/Ara and mCherry- high mCherry/eyfp

fluorescence ratio. Blue: no pLac/Ara upstream of mCherry- low mCherry/eyfp fluorescence ratio. Orange: glnKp between pLac/Ara and mCherry-mCherry/eyfp fluorescence ratio was low again. Median values on top of each histogram. (C) mCherry and eyfp fluorescence FACS results for the different glnKp silencing circuits, (i) no pLac/Ara upstream of mCherry, (ii) no glnKp in between pLac/Ara and mCherry and (iii) glnKp between pLac/Ara and mCherry .Colors as in B.

4.3 glnKp's silencing does not change in a ΔRpoN strain

In order to provide further support for the transcriptional blocking mechanism, I tested my glnKp transcription silencing variants (elaborated in results 4.2) in a ΔRpoN ($\Delta\sigma^{54}$) *E.coli* strain created in the lab by measuring the mCherry to eyfp ratio as previously described. In a ΔRpoN strain the σ^{54}:RNAP holoenzyme should not bind the σ^{54} promoter (due to the lack of σ^{54}). Therefore, in the variant with both promoters (pLac/Ara and glnKp), I expected to see a recovery in the fluorescence when the road-blocking is lifted. However, FACS histogram median results (Figure 12) reveal that the absence of σ^{54} protein in the cells did not affect the silencing effect seen in the WT strain. These results indicate that the silencing effect seen with glnKp promoter sequence is likely not related to the binding or unbinding of the σ^{54}:RNAP holoenzyme.

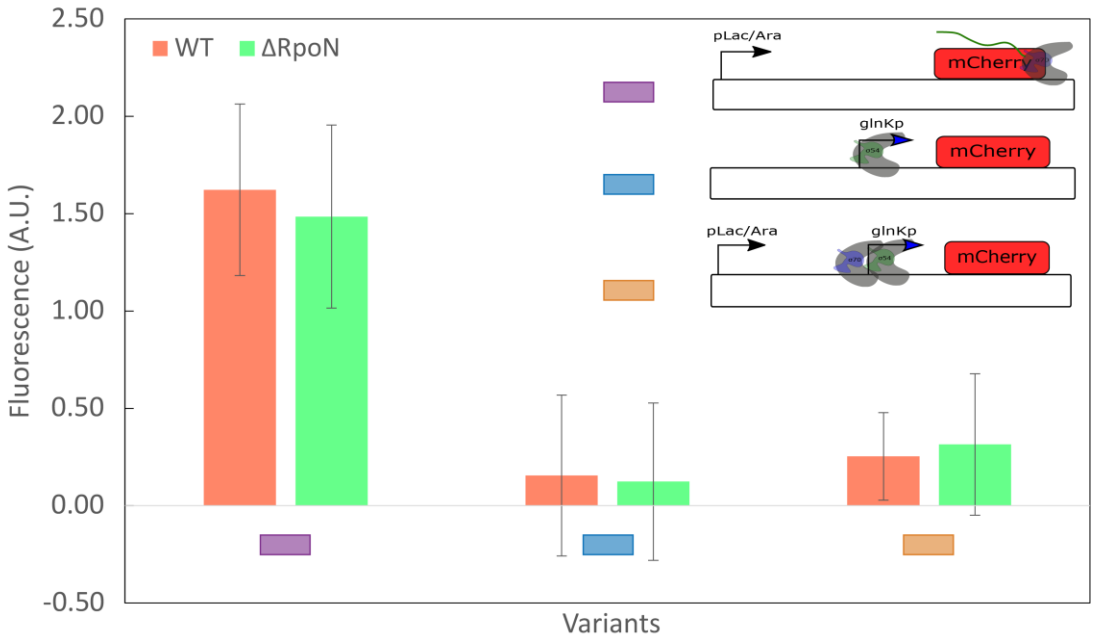

Figure 12: glnKp silencing median fluorescence in ΔRpoN and WT strains.

Purple: no glnKp in between pLac/Ara and mCherry- high mCherry/eyfp fluorescence ratio. Blue: no pLac/Ara upstream of mCherry- low mCherry/eyfp fluorescence ratio. Orange: glnKp between pLac/Ara and mCherry- no effect on transcriptional silencing in the ΔRpoN strain.

4.4 Oligo-library transcription silencing experiment

In order to search for the silencing mechanism responsible for the glnKp promoter inhibitory effect, I designed an oligo-pool library with the following four goals in mind:

1. Distilling the transcription silencing mechanism.
2. Expanding the testing for transcription silencing effect in known σ^{54} promoters.
3. Scanning for non-promoter transcription silencing sequences, which contains putative σ^{54} binding consensus sequence.
4. Determining the effect of the promoter orientation on gene expression.

To achieve these goals I ordered an oligo-library of ~12,000 barcoded sequences and cloned it into the transcription silencing circuit (Figure 13). Plasmids were transformed

into E.cloni® cells and were sorted in a FACS-sorter (FACS-Aria II) into 14 expression bins according to the ratio of mCherry/eyfp fluorescence. Next, bin barcodes were added to each sequence by PCR and the sequences were run in Illumina MiSeq next generation sequencer in order to obtain the mean fluorescence ratio for each tested sequence based on the distribution of its sequencing reads across the sorted expression bins (For further detail on the protocol see materials and methods). In this thesis, I present the analysis of results from ~1400/12,000 sequences that were characterized by this high-throughput approach.

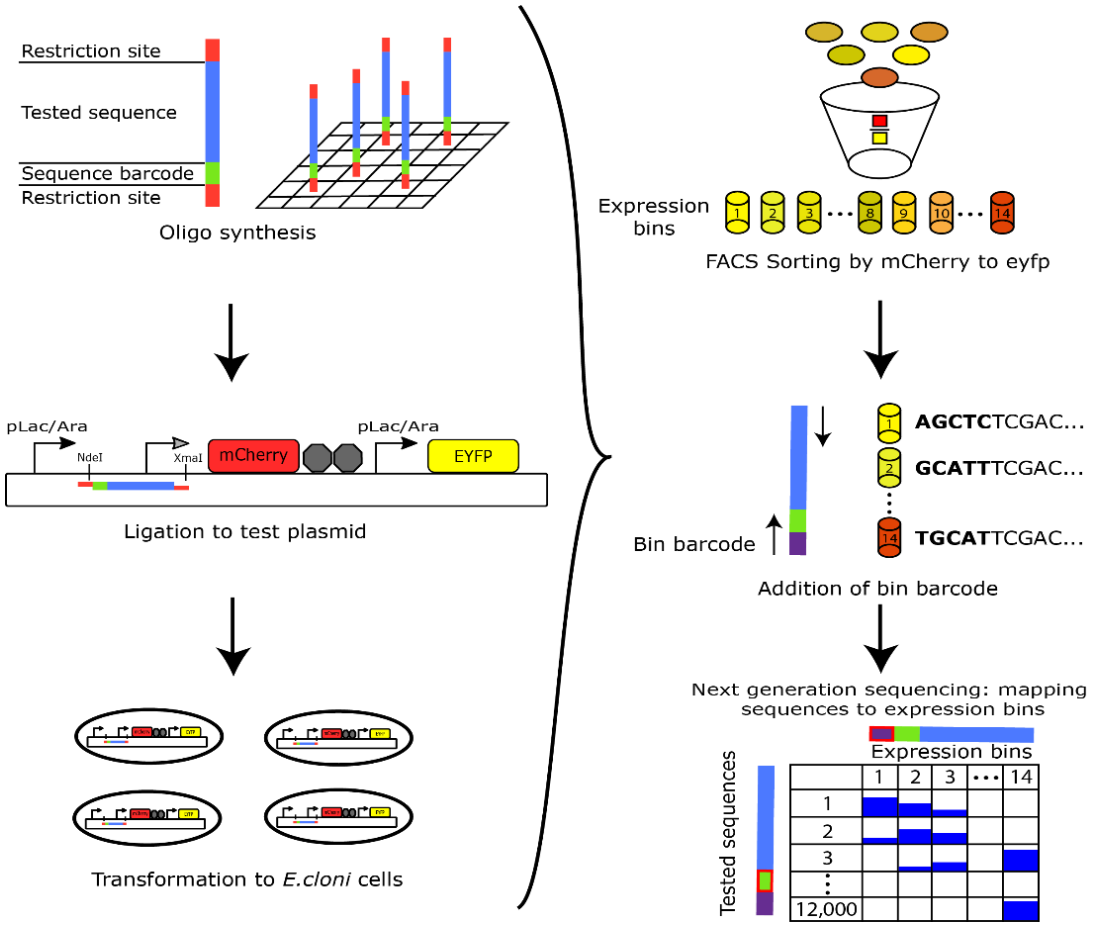

Figure 13: Illustration of the high-throughput transcription silencing experimental method.

4.4.1 Non σ^{54} consensus sites and σ^{70} promoters do not silence transcription

As part of my experiment I was interested in determining the transcription silencing ability of sequences that were not supposed to silence transcription according to my hypothesis (negative controls). I used two groups of sequences: (i) Sequences with low σ^{54} consensus score (will be referred to as "no promoter", scoring method elaborated in materials and methods). (ii) Various σ^{70} promoters. The first group should not present transcription silencing due to the fact that the sequence should not bind σ^{54}:RNAP holoenzyme and the latter should present high mCherry expression, attributable to the active promoter that is present in the sequence. The negative control groups showed (Figure 14) sequence enrichment (according to minimal hyper geometric distribution- materials and methods) with low p.value ($p<0.05$, 0.01 respectively) in the higher mean fluorescence ratio scoring, indicating that, as expected, sequences from these groups do not silence transcription and that my method can identify non-silencing sequences.

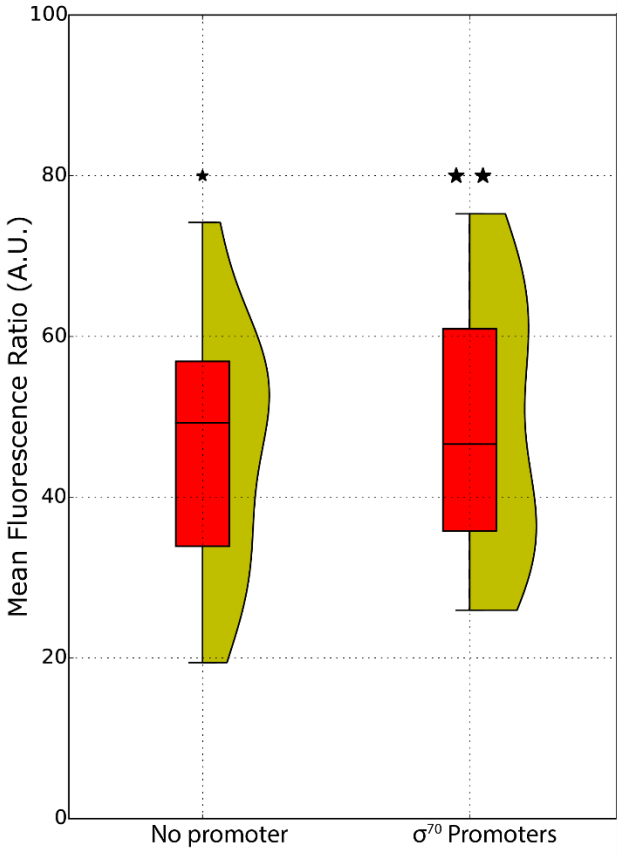

Figure 14: Oligo-library negative control results.

Distribution of the mean fluorescence ratio for the sequences from "no promoter" and "σ^{70} promoters" groups. Enrichment of high mean fluorescence ratio scoring is seen in both groups. Stars represents p.value: one star <0.05, two stars <0.01.

4.4.2 Flanking region plays an important role in transcription silencing

I next used the results from my library to get a better clue as to the mechanism for transcriptional silencing. I mutated glnKp's sequence at the core promoter (16 bp) and the flanking regions (34 bp) for the purpose trying to understand the mutations' effect on transcriptional silencing (Figure 15). Single mutations in the core region were carried out based on the less probable base in the consensus sequence (materials and methods) and the flanking region was mutated in an orderly fashion, mutating one base pair at a time and randomizing the previously mutated base pairs (materials and methods). The results in Figure 15 show that most of the forward orientated glnKp sequences were silencing transcription, showing mean fluorescence ratio of about 25 while the reverse orientated glnKp sequences were mostly non-silencing with mean fluorescence ratio of about 40 (calculation of mean fluorescence ratio is explained in materials and methods). Moreover, I noticed that mutations in the flanking region had a dramatic effect on transcriptional silencing. In particular, altering the mean fluorescence ratio of the forward variants from 15 (forward-wild type) to a maximum of 55, while for the reverse variants some flanking mutations raised the mean fluorescence ratio (with respect to the reverse-wild type variant) and some lowered it down by 2.2 fold in other cases. Interestingly, core region mutations did not show much effect on transcriptional silencing. Consequently, this led me to suspect that the silencing effect was not encoded within the core-promoter region, but rather within the core's flanking regions.

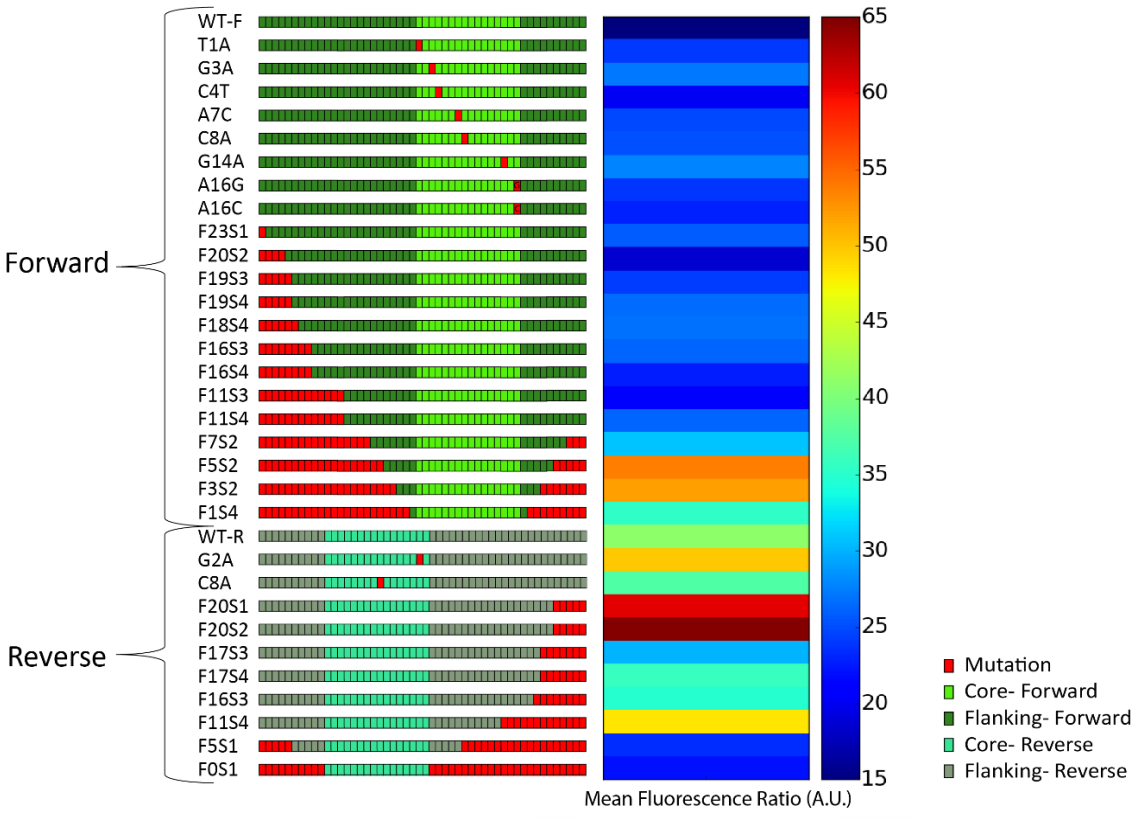

Figure 15: glnKp perturbations mean fluorescence ratio results.

Mean fluorescence ratio for the different mutations in glnKp's sequence. Bases location and orientation represented in colors, mutated bases colored as red. Flanking region mutations seems to have greater impact on mean fluorescence ratio and thereby on transcription silencing effect.

4.4.3 *E.coli*'s and *V.cholera*'s genomes contains silencing sequences

I also used the library to screen additional σ^{54} sequences in the genomes of *E.coli* and *V.cholera* which are able to silence transcription. The genomes were scanned for the σ^{54} consensus sequence (TGGCACACCGCTTGCA) and the highest scoring sequences were synthesized and tested (materials and methods). I was interested in discovering intragenic-in-phase transcriptional silencing sequences with the thought that these sequences may be able to downregulate gene expression in certain conditions. Distribution of these sequences vs mean fluorescence ratio is shown in Figure 16 and as can be seen, the sequences effect on transcription range from silencing (low ratio, <30) to not-silencing (high ratio, >50)

where the mean is at 39. When locating the sequences original position on the genome I noticed that most of the sequences tested were intragenic, in correlation with the gene percentage of *E.coli* and *V.cholera*'s genomes. In the intragenic group, the in-phase-with-transcription sequences are 48%, some of them show low mean fluorescence ratio, indicating that these sequences may be able to downregulate their respective genes. The out-of-phase sequence may have a different regulatory role, such as limiting opposite strand transcription or anti-phase sequences, but these were outside the scope for the present analysis. Either way, a significant percentage (~20-25%) of the sequences screened in my library exhibited some form of silencing (ratio <30) indicating that the phenomenon uncovered in the initial glnKp experiments are wide-spread and may be pervasive throughout bacterial genomes.

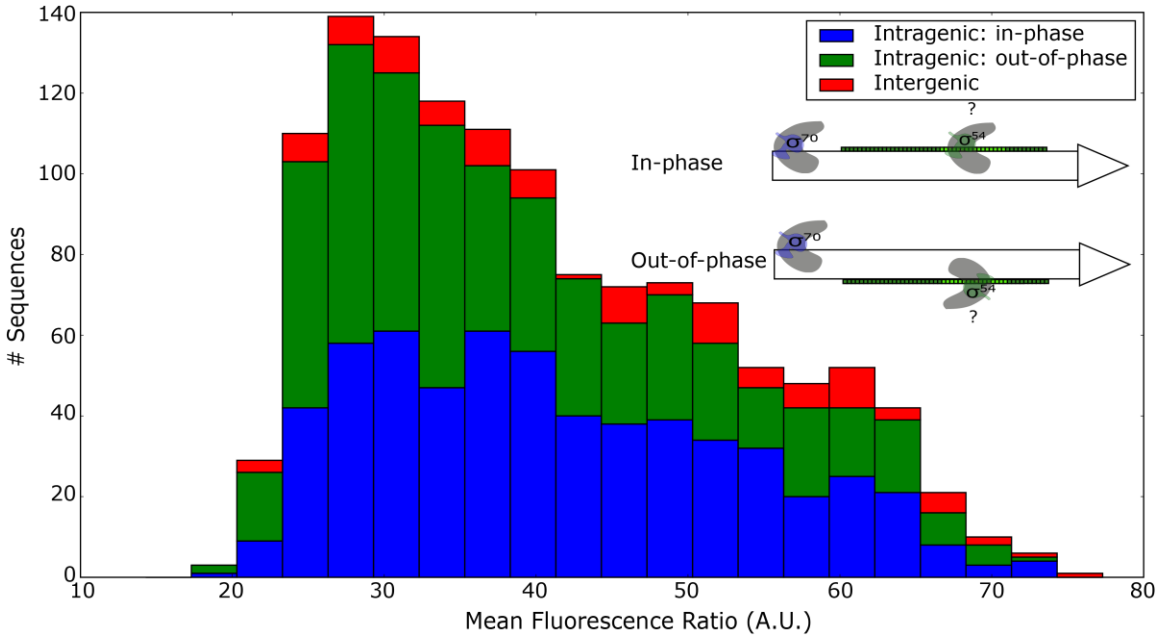

Figure 16: Genome-scan sequences mean fluorescence ratio distribution.

Genome-scan sequences show a wide distribution ranging from below 20 fluorescence ratio to above 70 (mean is 39). Each bar shows the distribution within the specific fluorescence ratio range (fixed range) for the sequences location and orientation with respect to its gene's transcription: in-phase (blue), out-of-phase (green) and intergenic (red). Most sequences are intragenic and some intragenic-in-phase sequences show transcription silencing effect (mean fluorescence ratio below 30).

5 Discussion

Although bacterial enhancers have been studied for quite some time and different enhancer regulation mechanisms were characterized, the relation between the upstream activating sequence's (UAS) EBP affinity and the enhancer's σ^{54} promoter is not well understood [30,57,31]. In this thesis, I provide a comprehensive insight on the relation between these two components, showing that the UAS's affinity for EBP (NtrC in this case) does not play the most critical role in activation of most of the promoters tested. Moreover, due to surprising results obtained, this study was also able to dive into and try to understand an undocumented silencing phenomenon by using a method of scanning thousands of sequences in a high throughput manner [53].

Activation of σ^{54} dependent promoters requires the contact between an oligomeric activator (EBP) bound to an upstream sequence (UAS) to a σ^{54}:RNAP holoenzyme bound to a promoter in a close complex. The contact is possible by forming a loop in the DNA, bringing the two components close to each other to facilitate a direct interaction [20,58]. Thereby, any factor that could alter the probability of interaction between the oligomeric activator and the promoter will affect the open-complex creation frequency. The factors which can affect such a collision are for example: the supercoiling of the DNA, the affinity of the UAS to the EBP, the promoter's dissociation constant (K_d) for the σ^{54}:RNAP holoenzyme [20,29] and the physical proteins bound in the looping region [30,31].

Based on a circuit design adapted from [51] I have been able to create a 'plug and play' system in which the efficiency of a UAS-σ^{54} promoter combination can be assessed in an *in-vivo*, easy, reproducible and high-throughput assay. My results imply that, as for Ntr regulated promoters, the σ^{54}-promoter's K_d for the σ^{54}:RNAP holoenzyme is a more dominant factor in the creation of the open complex then the affinity of the UAS for NtrC where it could be seen that UASs affinity's order was not in correlation with the activation efficiency for each tested promoter. Moreover, based on the results obtained from the synthetic UASs assembled in this experiment I can say that, as for the affinity of the UAS to NtrC, there is no additive effect when combining different affinity binding sites for NtrC into one UAS, meaning that High affinity synthetic UAS constructs did not present higher than average activation (e.g. CP UAS) while low affinity synthetic UAS constructs did

show high promoter activation (e.g. KH UAS). This implies that synthetic UASs are a viable option for researchers to clone in order to achieve new activation levels, with the limitation of not being able to predict the activation without proper testing. It must be noted that the location of the UAS relative to the promoter is also a major factor that should be taken into consideration when designing such a system due to the fact that it was shown that small changes in the position of the UAS can be critical for the activation [20], in my system the distance between the UAS and the promoter was fixed at 70bp for all variants.

Moreover, I was also able to demonstrate the farthest activation of a σ^{54} promoter with NtrC ever recorded to date (3000 bp, Figure 7A). This activation suggests that σ^{54} promoters on the genome of bacteria can be activated by a UAS located even a few thousands base pairs away. This, therefore, implies that our current understanding of these kind of promoters may be severely limited, and researchers should expand their scope to try to find and test UASs not only in the promoter vicinity, but also up to several thousand base-pairs away. Additionally, it is plausible to assume that researchers can evaluate promoter's strength by testing the largest possible separation between a UAS and a σ^{54} promoter. In the combinatorial experiment the weakest promoters were glnHp2 and astCp2 [12,22] and indeed I could show that these are the only promoters which did not show any activation originated from a distally positioned UAS. This kind of behavior is, again, relevant for synthetic biologists trying to create an enhancer system without any cross activation.

Bacterial enhancers architecture is a field studied intensively in the last few years [30,30,31] and it is only logical that the architecture itself will be a dominant factor in the activation of the promoter. One promoter in the combinatorial experiment, astCp2, was different than the other promoters, showing no activation throughout the whole experiment (can be seen in the Figure 8B) - This may be due to the fact that the native architecture of the astCADBE enhancer is much different than the one used in my experiment: in the native architecture the UAS is located at -233 and -255 from the transcriptional start site (in my experiment it was placed in ~ -140) and activation is mediated by a DNA bending protein ArgR [12] which binding sites were not present in my experiment.

The first promoter to be activated in nitrogen limiting conditions is glnAp2 [18,29]. This can be achieved by high affinity UAS which also contains a σ^{70} promoter embedded within it (glnAp1). A σ^{70} promoter embedded in a UAS in not a typical form of regulatory mechanism found in bacteria, but in the case of the glnALG operon this design enables the cell to initiate the expression from the glnALG operon and create large amounts of its products quickly (e.g. glnG which codes for NtrC) which are necessary for the cell's survival in a nitrogen limiting conditions. My results confirm previous findings, which suggested that NtrC bound to the UAS represses glnAp1 activity [59], while simultaneously turning on glnAp2. This is a simple attenuating mechanism, which can control the levels of NtrC. On the one hand, keep a steady low-level that is primed for activation, while on the other controlling that those same levels won't exceed a certain amount.

As a result of the "glnAp1" UAS's dual function, it was not surprising to witness expression without any NtrC in the cell, originated from glnAp1 promoter (Figure 9B). Interestingly, when glnKp's sequence was tested and placed downstream from the UAS, the expression from glnAp1 was suppressed. This silencing phenomenon witnessed only with glnKp in the 5'-3' direction (the phenomenon did not repeat when glnKp's sequence was placed in the reverse orientation- Figure 10) resulted in the postulation of a new hypothesis that "a stalled σ^{54}:RNAP holoenzyme can block a trailing elongating RNAP". This hypothesis was based on the assumption that the silencing effect resulted from a roadblock caused by the σ^{54}:RNAP holoenzyme bound to the promoter site. However, results from a ΔRpoN ($\Delta\sigma^{54}$) strain revealed that the silencing phenomenon did not originate from the holoenzyme itself (Figure 12), enabling new and exciting speculations on what could be the mechanism for such robust silencing.

Silencing/repression of expression can originate from transcriptional or translational repression mechanisms, or perhaps in this case it is a combination of the two. The different mechanisms are elaborated in the introduction, a few examples are: pausing of the RNAP, creation of a secondary structure in the RNA which sequesters the RBS, RNA interference etc. In order to understand the mechanism/s involved, high throughput method for scanning thousands of variants was carried out (Figure 13). Results from site-directed mutagenesis of the glnKp's sequence implied that the silencing effect may not be

related the core promoter sequence (which contained the consensus σ^{54} sequence), but is probably related to the flanking region of the promoter, which in my case covered up to 24 bp upstream and 16 bp downstream from the core (Figure 15), further proving that the σ^{54}:RNAP holoenzyme does not underlie the silencing effect. The dramatic change in the silencing effect caused by the mutations confirms my suspicions that there is a sequence related mechanism/s that can be localized to the glnKp σ^{54} promoter, but the particulars of this mechanism is yet to be discovered.

The silencing phenomenon's prevalence was also tested in the genomes of *E.coli* and *V.cholera*, where thousands of genome originated sequences containing variations of the σ^{54} consensus sequence were scanned using my high-throughput method. The results obtained (Figure 16) first indicate that 25% of the tested sequences from the genome are able to silence the expression of the tested target gene to some extent, suggesting that the phenomenon is wide-spread and may be pervasive throughout bacterial genomes. In addition, since ~50% of the "σ^{54}-promoter-liker" sequences that were selected for the library are intragenic, a significant silencing effect may be at work in these intragenic sequences as well. A potential correlation between intragenic σ^{54}-promoter-like sequences has been proposed in the literature, but no definitive relation had been established. In particular, genome-scale mapping of σ^{54}-promoters in *E.coli* did not reveal an enrichment in upregulation of genes in a ΔRpoN strains containing σ^{54}-promoter-like sequences [3]. However, another work provided evidence that σ^{54}:RNAP holoenzyme can be repressive for some genes [60]. My results provide further evidence that intragenic σ^{54}-promoter-like sequences may possess an unknown regulatory mechanism that is unrelated to the actual σ^{54}-RNAP complex itself. My conclusions provide support for both observation adding clarity to what seemed to be conflicting observations.

With the data obtained from my high-throughput experiment I was able to perform a bio-informatical analysis in search for a motif present in all the silencing sequences. An enrichment of a particular sequence motif can be the key to understand the mechanism behind the silencing phenomenon. This analysis resulted in the motif cCTT which was abundant in our most silencing sequences. Given that our pyrimidine-rich motif looks to be nearly the reverse complement of the ribosome binding site (RBS) used in my plasmids

(GGAGAA motif) mechanism, I hypothesized that this motif can generate a translational repression complex by forming a hairpin loop with the Shine-Dalgarno (SD) or RBS used in my system. This kind of translational regulatory mechanism was shown in previous work, which provide evidence for the existence of translation suppression mechanism by a 5' UTR stem loop which sequestered the SD sequence [61]. Consequently, in order to validate this mechanism I propose that the RBS sequences in my plasmid will be altered for some of the silencing strains. The mutated RBS should then be matched with a unique plasmid encoding for an altered 16s ribosomal RNA which encodes for the anti-SD sequence of the mutated RBS [62]. The use of an orthogonal expression system with an RBS that should not bind to our common motif will provide the necessary proof in order to validate the SD:anti-SD theory.

In conclusion, this work first and foremost presents a new approach to biological and synthetic biology research: starting with the simplification of a question to a simple design and continuing to a broad high throughput experiment. These experiments can be done using our ability to synthesize immense amounts of DNA in a rapid and cost-effective manner which can provide large amounts of data in a single experiment. This data can be analyzed by bio-informatical tools and give us a broader-than-ever prospective on biological systems. On the micro scale, my thesis was able to provide new insights into the (thought to be understood) σ^{54} enhancer systems by conducting a 'mix and match' type of experiment. Additionally, this work was able to identify and test the prevalence of a whole new regulation mechanism, first suspected to be a transcriptional regulatory mechanism, in a high-throughput manner. Additional work should be carried out in order to tease out the mechanism and implications of this new and exciting regulatory phenomenon.

6 Bibliography:

1. Bush, M. & Dixon, R. The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of σ54-Dependent Transcription. *Microbiol. Mol. Biol. Rev.* **76,** 497–529 (2012).
2. Murakami, K. S., Masuda, S. & Darst, S. A. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. *Science* **296,** 1280–1284 (2002).
3. Bonocora, R. P., Smith, C., Lapierre, P. & Wade, J. T. Genome-Scale Mapping of Escherichia coli σ54 Reveals Widespread, Conserved Intragenic Binding. *PLoS Genet.* **11,** (2015).
4. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.* **57,** 441–466 (2003).
5. Buck, M., Gallegos, M.-T., Studholme, D. J., Guo, Y. & Gralla, J. D. The Bacterial Enhancer-Dependent ς54(ςN) Transcription Factor. *J. Bacteriol.* **182,** 4129–4136 (2000).
6. Maeda, H., Fujita, N. & Ishihama, A. Competition among seven Escherichia coli σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* **28,** 3497–3503 (2000).
7. Buck, M. & Cannon, W. Specific binding of the transcription factor sigma-54 to promoter DNA. *Nature* **358,** 422–422 (1992).
8. Rappas, M., Bose, D. & Zhang, X. Bacterial enhancer-binding proteins: unlocking σ54-dependent gene transcription. *Curr. Opin. Struct. Biol.* **17,** 110–116 (2007).
9. Magasanik, B. Gene regulation from sites near and far. *New Biol.* **1,** 247–251 (1989).

10. Ninfa, A. J., Reitzer, L. J. & Magasanik, B. Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers. *Cell* **50,** 1039–1046 (1987).
11. Hoover, T. R., Santero, E., Porter, S. & Kustu, S. The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. *Cell* **63,** 11–22 (1990).
12. Kiupakis, A. K. & Reitzer, L. ArgR-Independent Induction and ArgR-Dependent Superinduction of the astCADBE Operon in Escherichia coli. *J. Bacteriol.* **184,** 2940–2950 (2002).
13. Santero, E. *et al.* Role of integration host factor in stimulating transcription from the σ54-dependent nifH promoter. *J. Mol. Biol.* **227,** 602–620 (1992).
14. Ghosh, T., Bose, D. & Zhang, X. Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol. Rev.* **34,** 611–627 (2010).
15. Bush, M. *et al.* The structural basis for enhancer-dependent assembly and activation of the AAA transcriptional activator NorR. *Mol. Microbiol.* **95,** 17–30 (2015).
16. Schumacher, J., Zhang, X., Jones, S., Bordes, P. & Buck, M. ATP-dependent Transcriptional Activation by Bacterial PspF AAA+Protein. *J. Mol. Biol.* **338,** 863–875 (2004).
17. Xu, H. & Hoover, T. R. Transcriptional regulation at a distance in bacteria. *Curr. Opin. Microbiol.* **4,** 138–144 (2001).
18. Reitzer, L. & Schneider, B. L. Metabolic context and possible physiological themes of sigma(54)-dependent genes in Escherichia coli. *Microbiol. Mol. Biol. Rev. MMBR* **65,** 422–444, table of contents (2001).

19. Blauwkamp, T. A. & Ninfa, A. J. Physiological role of the GlnK signal transduction protein of Escherichia coli: survival of nitrogen starvation. *Mol. Microbiol.* **46,** 203–214 (2002).
20. Feng, J., Goss, T. J., Bender, R. A. & Ninfa, A. J. Activation of transcription initiation from the nac promoter of Klebsiella aerogenes. *J. Bacteriol.* **177,** 5523–5534 (1995).
21. Atkinson, M. R., Blauwkamp, T. A. & Ninfa, A. J. Context-Dependent Functions of the PII and GlnK Signal Transduction Proteins in Escherichia coli. *J. Bacteriol.* **184,** 5364–5375 (2002).
22. Claverie-Martin, F. & Magasanik, B. Role of integration host factor in the regulation of the glnHp2 promoter of Escherichia coli. *Proc. Natl. Acad. Sci.* **88,** 1631–1635 (1991).
23. Schneider, B. L., Kiupakis, A. K. & Reitzer, L. J. Arginine Catabolism and the Arginine Succinyltransferase Pathway in Escherichia coli. *J. Bacteriol.* **180,** 4278–4286 (1998).
24. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).
25. Lee, N. L., Gielow, W. O. & Wallace, R. G. Mechanism of araC autoregulation and the domains of two overlapping promoters, Pc and PBAD, in the L-arabinose regulatory region of Escherichia coli. *Proc. Natl. Acad. Sci.* **78,** 752–756 (1981).
26. Kennell, D. & Riezman, H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *J. Mol. Biol.* **114,** 1–21 (1977).

27. Zafar, M. A., Carabetta, V. J., Mandel, M. J. & Silhavy, T. J. Transcriptional occlusion caused by overlapping promoters. *Proc. Natl. Acad. Sci.* **111,** 1557–1561 (2014).
28. Becker, N. A., Peters, J. P., Lionberger, T. A. & Maher, L. J. Mechanism of promoter repression by Lac repressor–DNA loops. *Nucleic Acids Res.* **41,** 156–166 (2013).
29. Atkinson, M. R., Pattaramanon, N. & Ninfa, A. J. Governor of the glnAp2 promoter of Escherichia coli. *Mol. Microbiol.* **46,** 1247–1257 (2002).
30. Amit, R., Garcia, H. G., Phillips, R. & Fraser, S. E. Building Enhancers from the Ground Up: A Synthetic Biology Approach. *Cell* **146,** 105–118 (2011).
31. Brunwasser-Meirom, M. *et al.* Using synthetic bacterial enhancers to reveal a looping-based mechanism for quenching-like repression. *Nat. Commun.* **7,** 10407 (2016).
32. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
33. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5,** 451–463 (2004).
34. Gama-Castro, S. *et al.* RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* **36,** D120–D124 (2008).
35. Henkin, T. M. & Yanofsky, C. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays* **24,** 700–707 (2002).

36. Yanofsky, C. Transcription Attenuation: Once Viewed as a Novel Regulatory Strategy. *J. Bacteriol.* **182,** 1–8 (2000).
37. Hao, N. *et al.* Road rules for traffic on DNA—systematic analysis of transcriptional roadblocking in vivo. *Nucleic Acids Res.* **42,** 8861–8872 (2014).
38. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference – a crash course. *Trends Genet.* **21,** 339–345 (2005).
39. Crampton, N., Bonass, W. A., Kirkham, J., Rivetti, C. & Thomson, N. H. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res.* **34,** 5416–5425 (2006).
40. Roberts, J. & Park, J.-S. Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr. Opin. Microbiol.* **7,** 120–125 (2004).
41. Komissarova, N. & Kashlev, M. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 1755–1760 (1997).
42. Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.* **22,** 4719–4727 (2003).
43. Ma, N. & McAllister, W. T. In a Head-on Collision, Two RNA Polymerases Approaching One Another on the Same DNA May Pass by One Another. *J. Mol. Biol.* **391,** 808–812 (2009).
44. Larson, M. H. *et al.* A Pause Sequence Enriched at Translation Start Sites Drives Transcription Dynamics In Vivo. *Science* **344,** 1042–1047 (2014).

45. Miyamoto, T., Razavi, S., DeRose, R. & Inoue, T. Synthesizing Biomolecule-based Boolean Logic Gates. *ACS Synth. Biol.* **2,** 72–82 (2013).

46. Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324,** 1199–1202 (2009).

47. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

48. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

49. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).

50. Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41,** D605–D612 (2013).

51. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli. *Cell* **113,** 597–607 (2003).

52. Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of σ54-dependent promoter sequences. *Nucleic Acids Res.* **27,** 4305–4313 (1999).

53. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30,** 521–530 (2012).

54. Eden, E., Lipson, D., Yogev, S. & Yakhini, Z. Discovering Motifs in Ranked Lists of DNA Sequences. *PLOS Comput Biol* **3,** e39 (2007).

55. Wagner, F. GO-PCA: An Unsupervised Method to Explore Biological Heterogeneity Based on Gene Expression and Prior Knowledge. *bioRxiv* 018705 (2015). doi:10.1101/018705

56. Leibovich, L., Paz, I., Yakhini, Z. & Mandel-Gutfreund, Y. DRIMust: a web server for discovering rank imbalanced motifs using suffix trees. *Nucleic Acids Res.* **41,** W174–W179 (2013).

57. Zhang, N. & Buck, M. A Perspective on the Enhancer Dependent Bacterial RNA Polymerase. *Biomolecules* **5,** 1012–1019 (2015).

58. Zhang, N. & Buck, M. A Perspective on the Enhancer Dependent Bacterial RNA Polymerase. *Biomolecules* **5,** 1012–1019 (2015).

59. Reitzer, L. J. & Magasanik, B. Expression of glnA in Escherichia coli is regulated at tandem promoters. *Proc. Natl. Acad. Sci.* **82,** 1979–1983 (1985).

60. J. Schaefer, C. Engl, N. Zhang, E. Lawton & M. Buck. Genome wide interactions of wild-type and activator bypass forms of 54. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv597

61. Shokeen, S., Patel, S., Greenfield, T. J., Brinkman, C. & Weaver, K. E. Translational Regulation by an Intramolecular Stem-Loop Is Required for Intermolecular RNA Regulation of the par Addiction Module. *J. Bacteriol.* **190,** 6076–6083 (2008).

62. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484,** 538–541 (2012).

7 Appendixes

7.1 Appendix 1: Consensus sequence probability for σ^{54} binding

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
P(A)	0.057	0.003	0	0.087	0.720	0.117	0.150	0.670	0.250	0.250	0.25	0.090	0.070	0.003	0.027	0.870
P(C)	0.057	0.003	0	0.740	0.093	0.650	0.150	0.165	0.250	0.250	0.25	0.090	0.070	0.003	0.920	0.065
P(G)	0.057	0.990	1	0.087	0.093	0.117	0.550	0.670	0.250	0.250	0.25	0.090	0.070	0.990	0.027	0.065
P(T)	0.830	0.003	0	0.087	0.093	0.117	0.150	0.165	0.250	0.250	0.25	0.730	0.790	0.003	0.027	0.870

Specific nucleotide probability for each position in the σ^{54} consensus sequence. as obtained from [52].

בנייה מחדש וגילוי של מנגנוני בקרה המבוססים על σ^{54} בתפוקה גבוהה

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר מגיסטר למדעים בהנדסת ביוטכנולוגיה ומזון

ליאור לוי

הוגש לסנט הטכניון - מכון טכנולוגי לישראל

אדר ב' תשע"ו חיפה מרץ 2016

<u>תקציר</u>:

Enhancers חיידקיים הם רצפי דנ"א לא מתורגמים, אשר להם תפקיד מרכזי בביטוי הגנים ומתפקדים כסוג של מתאם מולקולרי הקובע מתי, איפה וכמה ביטוי יהיה מגן מסויים. Enhancers חיידקיים בדרך כלל בנויים מרצף הפעלה היושב במעלה הזרם (UAS) הקושר אקטיבטורים אוליגומרים (אשר גם ידועים בשם חלבונים קושרי enhancer- EBP) אשר מספקים את האנרגיה הדרושה ליצירה של הקומפלקס הפתוח עבור פרומוטורים המבוססים על σ^{54}. אפיון של הקשר בין ה-UAS לפרומוטור σ^{54} אותו הוא מפעיל התבצע בעבר בסדר גודל קטן ועבור מספר מועט ביותר של פרומוטורים.

הבקרה על ביטוי גנים בחיידקים הינה רחבה מאוד וכוללת מנגנונים שונים כגון: (א) תחרות על אתר הפרומוטור- אשר מביאה לכך ששינוי בריכוז משרן או מעכב מביא להפעלה או השתקה של ביטוי גנים מסוימים, (ב) בקרה על יצירת הלולאה החיונית להפעלת פרומוטורים מבוססי σ^{54} - אשר בה חלבונים קושרי דנ"א הנקשרים ברצף הלולאה יכולים למנוע או להגביר את ההסתברות להיווצרותה ובכך לשלוט על אופי ביטוי הגנים הרלוונטיים, (ג) בקרה ברמת הרנ"א- בה מבנים שניוניים שונים ברמת הרנ"א יכולים להביא לעצירה או עיכוב של תרגום הרנ"א השליח, בקרה אחרונה אותה אציג כאן הינה (ד) בקרה על שעתוק הגנים על ידי עצירה או חסימה של הרנ"א פולימרז על ידי חסמים הנקשרים לרצף אותו הוא משעתק- חסמים אלה יכולים להיות חלבונים קושרי דנ"א (כגון פקטורי שעתוק או מעכבים) או אפילו רנ"א פולימרזות אחרים אשר נחסמו במהלך פעולתם.

החלטתי לבצע אפיון מקיף של האפקט הנובע מזיקת ה-UAS ל-EBP על ההפעלה של פרומוטורים המבוססים על σ^{54} במערכות של enhancers, האפיון יכלול את כל הפרומוטורים העוברים בקרה על ידי Ntr עם UASs טבעיים וסינטטיים בצורה קומבינטורית ובכך יכסה את הקבוצה העיקרית והנלמדת ביותר של פרומוטורים המבוססים על σ^{54}. בנוסף, ביצעתי ניסוי בתפוקה גבוהה על ידי שימוש בספריית תחלים (oligo-library) של כ-12,000 רצפים שונים לשם הבנת המנגנון הפועל מאחורי תופעת השתקה אשר נצפתה בניסויי ואת שכיחות התופעה הזו בגנומים של *E.coli* ושל *V.cholera*.

תוצאותיי מראות כי יעילות ההפעלה של פרומוטורים מבוססי σ^{54} תלויה יותר בחוזק הפרומוטור מאשר על זיקת ה-UAS ל-EBP. מעבר לכך, הצלחתי להראות שלא ניתן לחזות את הזיקה של UASs עבור EBP רק על ידי סכימת הזיקה של אתרי הקישור המרכיבים אותו אלא שיש לבחון כל UAS בפני עצמו עם הפרומוטור המתאים. בנוסף, היה באפשרותי להראות את ההפעלה הרחוקה ביותר שתועדה עד עתה על ידי UAS עבור פרומוטורים המבוססים על σ^{54}. השתלשלות אירועים מעניינת הביאה לכך שעבודתי הצליחה להראות תופעת השתקה מפתיעה באמצעות פרומוטור glnKp אינו משופעל אשר במעלה הזרם ממנו נבחנו שני פרומוטורים שונים (glnAp1 ו-pLac/Ara), הימצאות רצף הפרומוטור הצליחה להשתיק את הביטוי אשר

התקבל מאותם פרומוטורים אשר הוצבו במעלה הזרם ממנו ומכאן שהיה ברצוני לחקור תופעה זו. מנגנון בקרה שבו ההולואנזים σ^{54}:RNAP משמש כמחסום נשלל על ידי בדיקת ההשפעה של מוטציות נקודתיות ברצף הפרומוטור glnKp ועל ידי השוואה לביטוי המתקבל בזן ($\Delta\sigma^{54}$) ΔRpoN, ניסויים אלה הראו כי אפקט ההשתקה משוייך לרצפים מצידי ליבת הפרומוטור ולא לרצף הליבה עצמו. לבסוף, הצלחתי להראות כי תופעת ההשתקה הינה רחבה בגנומים של *E.coli* ושל *V.cholera*, המראים שכיחות של 25% של תופעה זו עבור הרצפים שנבדקו. עבודה עתידית צריכה להתבצע על מנת לחשוף את המנגנון/המנגנונים מאחורי תופעת ההשתקה, כאשר התמקדות ראשונית צריכה להיעשות על מנגנון בידוד ה-Shine Dalgarno.

בנייה מחדש וגילוי של מנגנוני בקרה המבוססים על σ^{54} בתפוקה גבוהה

ליאור לוי