DETECTION AND QUANTITATIVE CHARACTERIZATION OF BIOMOLECULAR INTERACTIONS WITH MICROFLUIDIC MECHANICAL TRAPPING

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Steven R. Bates

November 2010

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Stephen R. Quake) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Markus W. Covert)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(James E. Ferrell, Jr.)

Approved for the University Committee on Graduate Studies

Abstract

In the era of rapid high-throughput DNA sequencing and increasing bioinformatic capabilities to analyze large amounts of biological data, it is especially important to continue to develop high-throughput experimental proteomic methods. Probing physical interactions among proteins and nucleic acids is a powerful approach to gain insight into their functional relationships. Microfluidic tools are of great potential value to this field because they make it possible to run hundreds or thousands of independent experiments in parallel and at a high spatial density.

Here I present and explain MITOMI (mechanically induced trapping of molecular interactions), an *in vitro* method that was initially created to measure the affinities of transcription factors binding to DNA. I contributed to the development of the system so it could be applied to mapping protein-protein interaction networks, and used it to study the interactions of *E. coli* RNA polymerase in order to better understand the regulation of bacterial transcription.

I also extended the utility of MITOMI by adapting it to be able to measure interaction kinetics and to more efficiently measure affinities. With the ability to discover interactions at a large scale and to quantitatively characterize them on the same platform, MITOMI constitutes a valuable contribution to proteomic methodology.

Acknowledgements

I want to thank all of the family, friends, and colleagues who supported me during this difficult process.

First and foremost, my parents, Bob and Nancy, whose unconditional love and pride in me prepared me to choose and follow this path for myself, and my sister Dianne, for whom I try to live up to being a role model. The Johnsons—my aunts, uncles, and cousins—have inspired me with their humor and warm sense of family.

Among friends, Dustin Ngo, for understanding me better than anyone, and Shor Bowman, for understanding me better than almost anyone. Also Nick Cizek and Jennifer Burns, as well as Jake Ditmore and everyone else who helped me take breaks from work and relax, particularly over the last year.

And finally, colleagues and collaborators who helped me through their advice and example, particularly Thomas Snyder, Paul Blainey, Doron Gerber, Aaron Streets, Polly Fordyce, Matthias Meier, Josh Weinstein, and Jenny Jiang.

Contents

List of Tables

List of Figures

Chapter 1

Background and Introduction to MITOMI

Molecular biology concerns how cells function in terms of their components, biological macromolecules. Although the molecular players also include the lipids, carbohydrates, metallic ions, and water molecules that constitute the cellular environment, the first-order view of the field is framed in terms of Francis Crick's central dogma, which focuses on the relationships among DNA, RNA, and proteins[1], with DNA and RNA being primarily information carriers, and the actual cellular operations being primarily performed by proteins.

The emergence of high-throughput sequencing techniques over the past two decades has made possible a comprehensive knowledge of organisms' gene sequences. As a result it is now relatively straightforward to get the full "molecular parts list" for any cell whose genome can be sequenced. The next challenge is developing high-throughput methods to determine how the parts work together functionally, and one major approach to this question is to map how the molecules interact physically.

There are two classes of high-throughput methods in widespread use for mapping protein-protein interactions. The first, the yeast two-hybrid system and its variants, is

based on the principle that transcription factors have distinct DNA-binding domains and transcription activation domains. In a two-hybrid assay, the domains are genetically separated: the binding domain is genetically fused to a potential interaction partner, designated the bait, and the activation domain is fused to a second potential interactor, the prey. In principle, the binding domain binds to the upstream activation sequence of a reporter gene, which is expressed only if bait and prey interact and bring the activation domain into proximity to the gene to activate transcription[2]. The technique is well-suited to adapation to high-throughput screens by generating libraries of bait and prey plasmid constructs, and combining them into a large number of potentially interacting pairs. With the right choice of reporter gene, colonies containing interacting pairs can be easily identified by color or antibiotic resistance. In practice, two-hybrid approaches suffer from a high false-positive rate[3], and false-negatives from bait or prey failing to localize to the nucleus are also a concern[4].

An alternate approach is affinity purification followed by mass spectrometry (AP-MS). Here, baits are designed with an affinity purification tag, each bait is expressed *in vivo* in a single clone, cells are lysed, and the baits are purified, with any physiological interaction complexes coeluting. Then the complexes are separated by liquid chromatography or gel electrophoresis, and the components are analyzed with standard MS methods to identify the bait's binding partners[5]. By creating a library of tagged baits, AP-MS can also easily be scaled to high-throughput screens. It is important to note that although the results of two-hybrid screens and AP-MS screens can be and are compared[6], they are searching subtly different interaction spaces. The two-hybrid approach probes for direct binary interactions, while AP-MS methods can detect multicomponent complexes, so identified interactors may not be directly contacting the bait. AP-MS methods also suffer from their own biases, since weak, transient interactions are less likely to survive the purification steps[7].

In this thesis I describe my work on developing a third major strategy for protein-protein interactions based on microfluidics. The multi-layer soft lithography approach to microfluidics developed in the Quake Lab uses the viscoelastic polymer polydimethylsiloxane (PDMS) as the chip material. Advantages of using PDMS include its gas permeability and nontoxicity, which allows for biocompatible devices, and its transparency, which allows for optical readout. The single most important benefit, however, is that its mechanical flexibility allows for the creation of monolithic microfluidic valves[8](Figure 1.1). These chips have at least two layers, a flow layer and a control layer, with the channels in each layer having a height on the order of tens of μm. The flow layer contains the medium and reagents for the experiment of interest, while the control layer contains a fluid (usually oil, water, or air) that acts as a control medium. Both flow lines and control lines connect to ports that couple the chip to the outside world. In the case of the control lines, this coupling is a connection

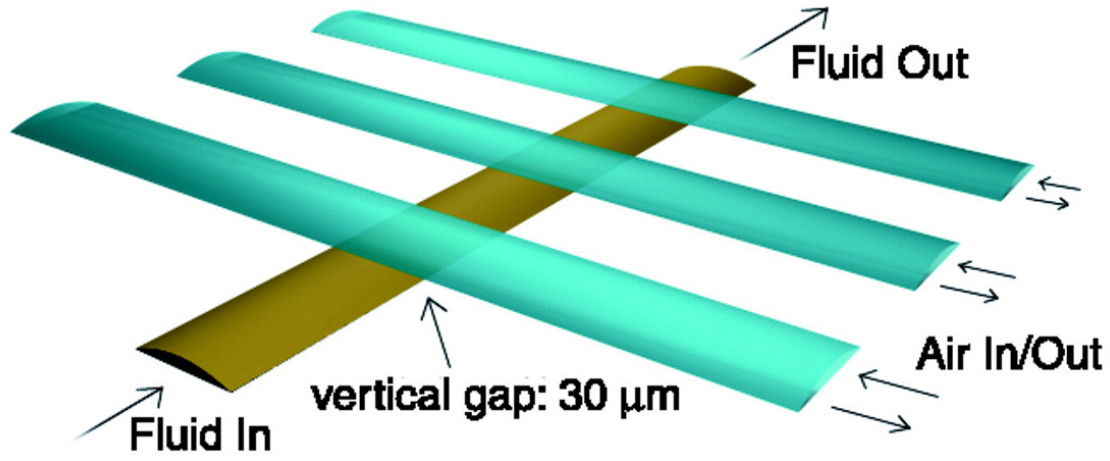

Figure 1.1: Cartoon of valve design in a PDMS chip. A control line of sufficient width crossing a flow line creates a valve which can be closed by pressurizing the control line. Figure reproduced from Unger *et al.*, 2000[8]. Instead of air, the microfluidic experiments described in this thesis used water as a control medium except where otherwise indicated.

to solenoid valves that can switch between different pressures applied to the line. If a control line is pressurized, it will swell, and in places where it crosses a flow line at a high enough aspect ratio (crossing width divided by flow line height) it will expand into the flow line enough to completely block flow. The valve closing process can be thought of as similar to stepping on a garden hose.

This type of valve plays a role in microfluidic chip design roughly analogous to the role of transistors in integrated circuits. Valves as a basic design element are combined in various ways to create complex devices for specific biochemical[9] or biological[10] applications.

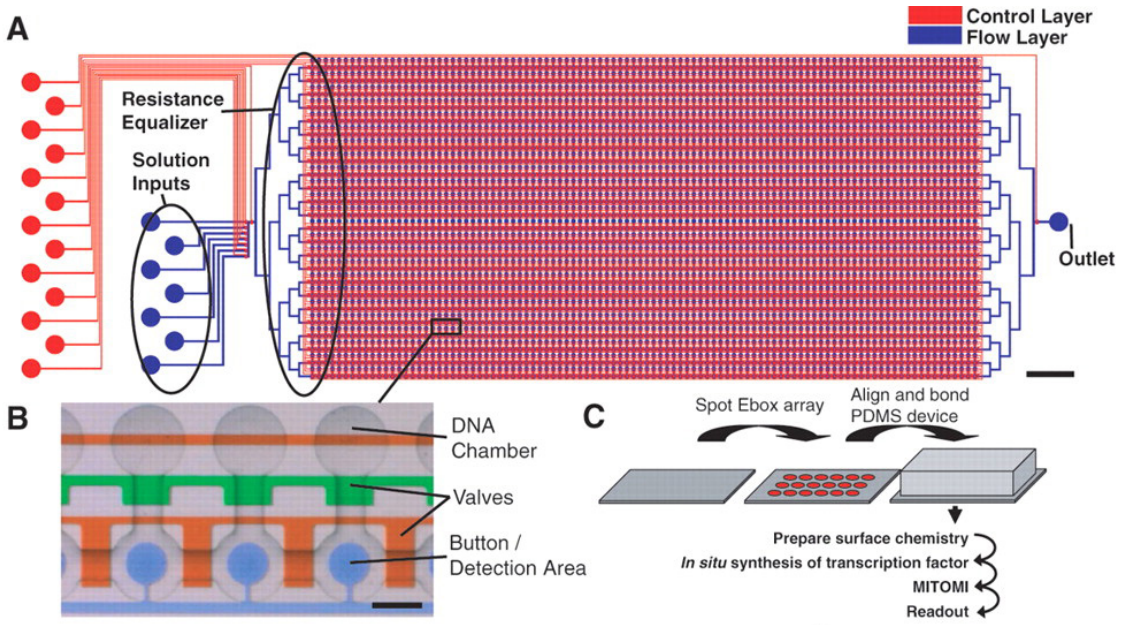

Figure 1.2: MITOMI schematic as reproduced from Maerkl and Quake, 2007[11]. **A.** Overall chip layout. This chip comprises 24 rows, each with 100 unit cells. Scale bar indicates $2mm$. **B.** Three sequential unit cells. Neck valves are colored green, sandwich valves orange, and button valves blue. As is typical, one external control port actuates all neck valves. The same is true of all sandwich valves, and of all button valves. Scale bar indicates $150\mu m$. **C.** The DNA template arraying strategy, and the workflow of a MITOMI experiment.

One such class of devices is MITOMI (mechanically induced trapping of molecular interactions) chips (Figure 1.2). These chips were developed to study biological macromolecules and, to take advantage of microfluidics' throughput-increasing capabilities, they are designed to work with an array of nucleic acid sequences, which may either participate in interesting interactions themselves or be expression templates for proteins that participate in such interactions. Nucleic acid arrays are created by using a microarray printer to deposit an array of sequences onto a glass epoxy slide, itself exactly the sort used in microarray experiments (Figure 1.2C).

Then the MITOMI chip is aligned over the array, and thermally bonded to the glass slide. The bonding protocol I used was usually $\sim 53^{\circ}C$ overnight, though higher temperatures and correspondingly shorter bonding times can be used. The floor of the flow channels in the bonded MITOMI chip is the slide surface, with PDMS on all other sides. The center-center distance between each MITOMI unit cell in a single row is 300-400μm depending on chip design (Figure 1.2B), with rows spaced by 600-800μm, allowing for a unit cell density of 300-600cm^{-2}. Adjacent unit cells can be functionally separated by actuating the chip's "sandwich valves" (Figure 1.2B) so that each unit cell corresponds in a sense to an independent experiment. The "neck valves" isolate the DNA chambers from the reaction chambers during the surface chemistry steps (described below). Each unit cell's reaction chamber has a "button" valve which effects the mechanical trapping that gives the technique its name. When actuated, a button valve contacts the surface underneath but does not block flow in the channel.

A MITOMI experiment begins with a series of sequential surface chemistry steps. Successive reagent steps are separated with a $10min.$ wash step during which buffer is flowed. The buffer used in all the experiments described in this thesis is phosphate buffered saline (PBS) at pH 7.4 unless otherwise indicated. The very first reagent introduced to the chip is biotinylated bovine serum albumin (BSA) (flowed for $30min.$),

which binds directly to the epoxy glass surface. Then neutravidin is flowed for $20min.$ and creates another chemical layer by interacting with the biotin groups. The buttons are closed halfway through the wash step that follows, and another $30min.$ biotinylated BSA step blocks the flow channel surface that isn't protected by the buttons. Next a biotinylated antibody is flowed for $22min.$, during the last 20 of which the buttons are reopened. In this way the antibody is deposited only in a defined spot under the button.

There are several options at this point. If the DNA templates printed were an expression array, an *in vitro* transcription and translation (ITT) mix can be introduced, and the neck valves opened to allow the mix to dead-end fill DNA chambers (since PDMS is gas permeable, dead-end filling by introducing liquid at sufficient pressure to displace the air in the channels is a standard technique). Then an ITT incubation follows, typically for $2hrs.$ The Roche RTS-500 ITT mix I used calls for a $\sim 30°C$ incubation temperature; I used a hot plate set to $\sim 32°C$ to compensate for the temperature drop across the epoxy slide. The ITT reaction is generally run with sandwich valves closed to segregate the unit cells, and neck and button valves open. Although the expression templates in each DNA chamber can include more than one sequence, one of the expression products will be immobilized by the antibody under the button. Borrowing and adapting yeast two-hybrid terminology, this product is designated the bait, with prey being molecules that can only bind to the spot indirectly through the bait.

Alternatively, the bait can be introduced from off-chip in the same way as the surface chemistry reagents were. This may make sense if there is a single bait to be deposited in all chip spots, or a relatively small number of baits—MITOMI chips optionally include address valves that can set the flow path to only pass through a subset of the chip rows. If each row is individually addressable, the absolute limit to the number of different baits introduced in this manner is equal to the number of

rows, though in practice the necessity of introducing the baits serially creates a lower, "soft" limit.

Assuming one bait and one prey per unit cell, there are really four possibilities: coexpression from arrayed bait-prey template pairs, expression from arrayed bait templates with prey introduced from off-chip, introduction of bait from off-chip and expression from arrayed prey templates, and introduction of bait then prey from off-chip.

Once the ITT step, if there is one, is complete, the button and neck valves are closed and sandwich valves are opened, the ITT mix is washed out, and fluorescently labeled antibody targeted to the prey is introduced. The sandwich valves are closed

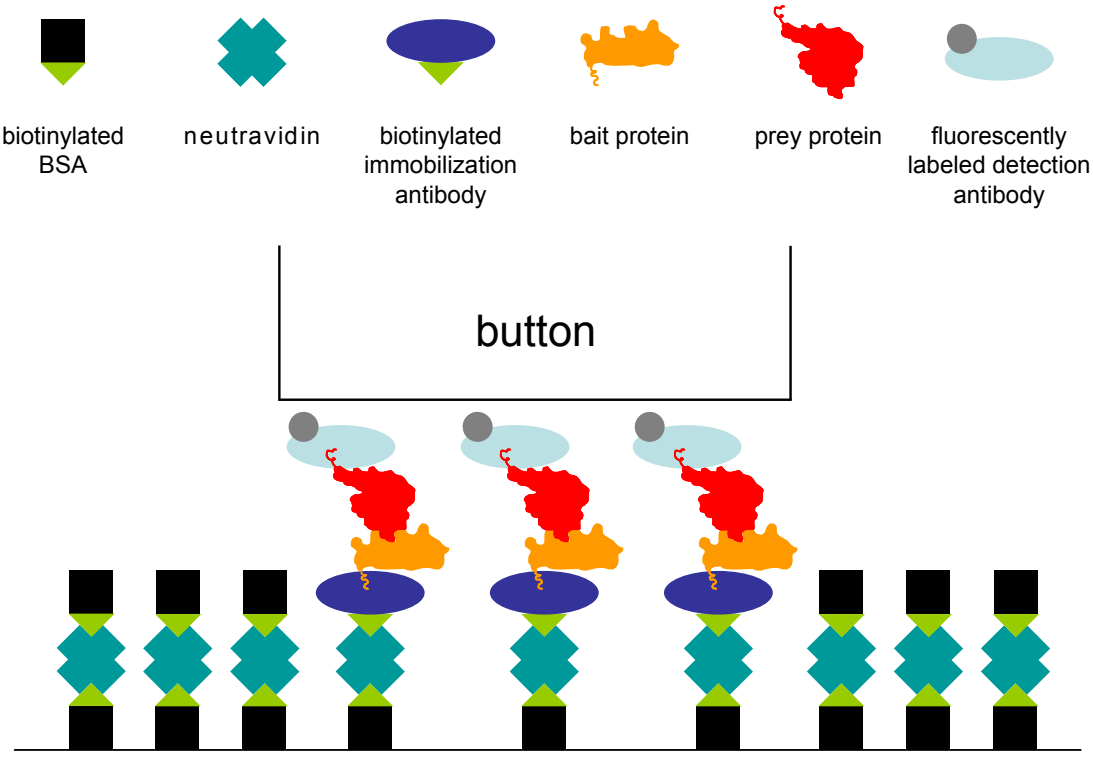

Figure 1.3: Cartoon of a mechanically-trapped interacting bait-prey pair after labeling, showing the layers of surface chemistry holding the bait in place.

and the button and neck valves are opened, the chambers are allowed to reequilibrate at room temperature—$1hr.$ being the typical incubation time for this step—and the button and neck valves are opened and the sandwich valve closed one more time. Finally excess label is washed out and the chip is fluorescently imaged. Fluorescent detection of the labeling antibody under a button establishes the presence of prey, and consequently the existence of a physical interaction between bait and prey. Mechanical trapping during the steps following expression is central to the method. The button holds in place any prey that is pulled down by immobilized bait and minimizes any loss of the probed interaction as reagents are exchanged.

The question of labeling strategy presents even more options than the question of how to introduce bait and prey. The first use of MITOMI was to map the binding energy landscape of transcription factors interacting with DNA sequences[11], and if the prey is DNA, labeling is relatively simple because of the wide availability of commercially produced, fluorescently labeled oligonucleotides. My graduate work began with adapting the method to protein-protein interactions, and we eventually settled on generalizing the strategy used for bait immobilization: add epitope tag sequences to the proteins, and target the epitope tags with fluorescently labeled antibodies.

The expression templates for bait and prey are created by a multistep PCR method (Figure 1.4). Starting with the gene of interest cloned in a plasmid vector, the first PCR (30 cycles) adds coding sequences for any desired epitope tags to the open reading frame, as well as start and stop codons. The second PCR (10 cycles), the primers of which we refer to as extension primers, adds untranslated region to both sides. For experiments using the RTS-500 expression system, which includes T7 RNA polymerase for transcription, the extension primers include T7 promoter and terminator sequences. The third PCR could also be considered the second stage of the second PCR; after the 10 cycles of extension are complete, final primers are added and another 30 cycles are immediately initiated. The final primers add no additional

sequence and are merely used to effect full amplification to the desired final DNA product concentration (typically $0.5\mu M$).

We experimented with several epitope tags in the early stages of developing the system, but settled on three that worked reliably: a 6-His tag that was recognized by a penta-His antibody (Qiagen), a T7 epitope tag (MASMTGGQQMG; antibody from Novagen) and a c-Myc epitope tag (EQKLISEEDL; antibody from Sigma). Adding the tag to the C-terminus is preferred, to avoid recognition of partially translated peptide sequences. However, we typically added an N-terminal c-Myc tag to the baits as a way to measure bait expression levels, since the immobilization epitope tag, either 6-His or T7, already occupied the C-terminus. The labeling mixture would then include both an expression labeling antibody and an interaction labeling antibody, to be imaged in separate fluorescent channels.

There are disadvantages to this labeling method, and other possibilities exist. All

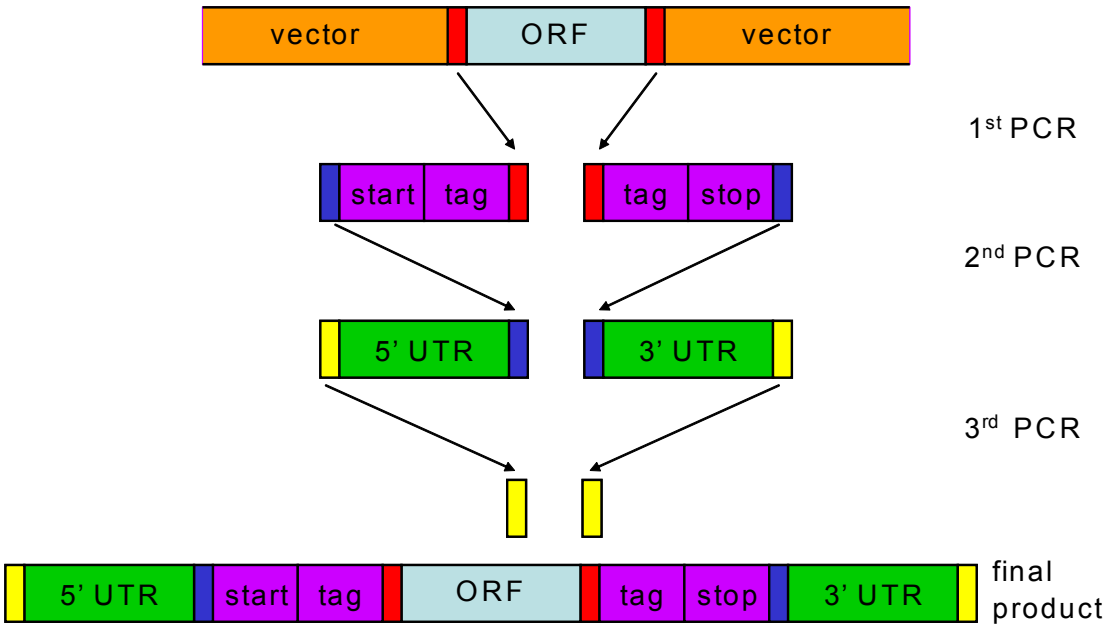

Figure 1.4: Multistep PCR method for generating linear expression templates.

other things being equal, a covalent label would be preferable, to avoid loss of signal if the labeling antibody is not at a high enough concentration to create an equilibrium with prey molecules fully labeled. Accordingly, I did try in some cases to use preys that were designed as fusions with green fluorescent protein (GFP; see Chapters 3 and 4). However, this approach would be much more difficult to adapt to high-throughput experiments than PCR amplification from a library of prey genes.

There is also the problem of using a label that is comparable in size to the species being labeled. This is even more of an issue with antibody labeling and with another method we considered, quantum dot labeling, than it is with GFP labeling. There is a concern that the interactions we are trying to detect would be sterically disrupted. An ideal solution may be to use a labeling method that covalently attaches a small-molecule fluorescence tag to a genetically encoded amino acid sequence[12][13]. I did not have the chance to adapt this kind of strategy to my experiments, but I am convinced that developing labeling tools is the single most important area of research needed for MITOMI to reach its full potential.

Whichever labeling method is used, the resulting fluorescent image needs to be carefully interpreted. We used Axon GenePix Pro software, which is designed for analyzing microarray images (Figure 1.5). The images of the MITOMI chip in the various fluorescence channels are uploaded into Genepix as a multi-image TIF file and aligned if necessary. Genepix alllows for the creation of arrays of circular spots that define regions for analysis. Each of these analysis spots has an associated "background" region consisting of the annulus with inner diameter equal to the spot's diameter and outer diameter equal to twice the spot's diameter. To analyze our experiments, we align the analysis spots over the spots in the image defined by the button, and measure the median fluorescence of the pixels in the analysis spot minus the median fluorescence in the background region (background-subtracted signal). The "raw" signal without background subtracted may be preferable depending on the situation.

This background-subtracted or raw median signal assigns a single numerical value to each MITOMI spot for each channel and is used for all analysis after this point, for instance statistical hypothesis testing to determine if each bait-prey combination in an interaction screening experiment gives a significantly higher signal than negative control MITOMI spots.

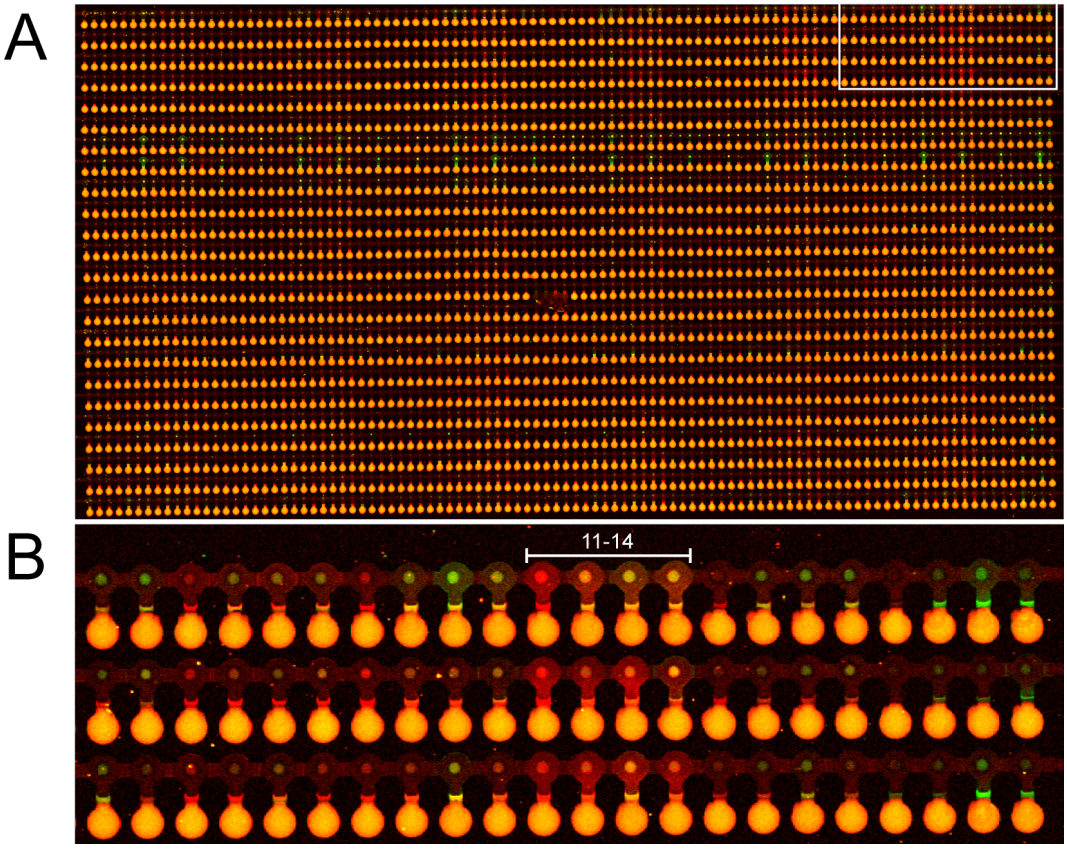

Figure 1.5: **A.** Genepix image of MITOMI experiment, showing relative intensity of interaction channel (red) signal and bait expression channel (green) signal. The largest circles are the DNA chambers, still filled with fluorescent label in solution. The MITOMI spots are smaller and less obvious. **B.** Blowup of top right corner of chip. The center-center distance between unit cells is $700\mu m$ vertically and $320\mu m$ horizontally. The 11th through 14th chambers from the left in each row contain the same prey (but different baits) and illustrate how one protein can consistently give a higher background signal than others, because of nonspecific binding outside the MITOMI spot.

The use of proper controls for these experiments is extremely important, because of the possibility for some proteins to bind nonspecifically (Figure 1.5B). In the context of MITOMI, certain proteins appear to be "stickier" than others. This problem of stickiness may extend to the antibodies themselves. Fluorescence signal in a MITOMI spot cannot be assumed to be entirely due to the labeling antibody binding to the epitope tags of prey molecules interacting with bait molecules which are bound to the immobilization antibody. Any step in that chain could be replaced by nonspecific binding, such as between labeling antibody and bait, between prey and immobilization antibody, or even between the labeling and immobilization antibodies.

Figure 1.6 shows some results from an experiment to evaluate the effects of nonspecific binding. One of the baits, marB, is well-behaved, with a high expression signal only in the spots with the correct immobilization antibody and with the marB actually expressing. However, lrp and rpoS, when expressed, give a significant expression signal even in the absence of penta-His antibody. This indicates nonspecific binding of these baits both to T7 immobilization antibody and to the unblocked neutravidin that is presented in the absence of antibody. The slight increase in signal when marB is expressed with the wrong antibody compared to nonexpressing marB spots indicates that marB exhibits a small but detectable amount of nonspecific binding itself.

In experiments where I was probing potential interactions to create maps of interaction networks, I tried to adhere to the philosophy that the difference between experimental and control should be a change of one variable. In the context of detecting an interaction between a given bait and a given prey, this meant using as negative controls chambers that were expressing the same prey, but no bait. The question is one of differential binding: does the antibody bind significantly more to a MITOMI spot in the presence of the bait than in its absence? This controls for nonspecific binding of both prey and labeling antibody to the spot. Of course, prey-less spots expressing the same bait would also help detect the effects of sticky baits,

and I made some use of this type of control as well. Arguably, even control chambers expressing both bait and prey but lacking immobilization antibody might be worthwhile, although I did not attempt this last type of control because of the difficulty of implementation in high-throughput experiments.

With this framework in place, MITOMI has the potential to make great contributions to proteomics and the understanding of cellular function. In the remainder of this thesis, I will report on how I used MITOMI to study a biologically interesting

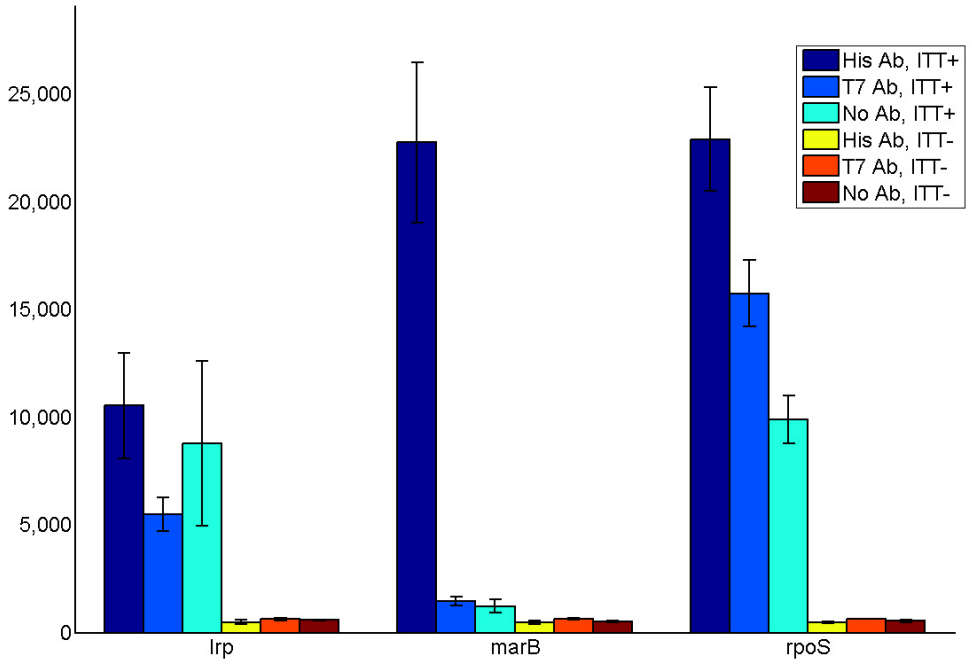

Figure 1.6: Nonspecific binding in MITOMI experiments. Example of MITOMI expression signals for three baits, averaged across replicates, ± standard deviation. All baits were designed with N-terminal c-Myc tags and C-terminal 6-His tags. A penta-His or T7 immobilization antibody, or no antibody, was deposited by the standard surface chemistry procedures. ITT mix was withheld from some MITOMI chambers before the expression step. All chambers were labeled with Cy3-conjugated c-Myc antibody.

protein-protein interaction network as well as discuss modifications I made to the basic MITOMI experimental setup to extend the applications of the method.

Chapter 2

Mapping Interactions of *E. coli* RNA Polymerase

I initiated a project to use the MITOMI methods described in Chapter 1 to map out a small network of interactions involving the subunits of *E. coli* RNA polymerase. The goals of the project were two-pronged. My interests centered around making MITOMI more quantitative, by measuring kinetics and affinities of biomolecular interactions. In practice, however, an ideal MITOMI proteomics project would begin by mapping out a network of interactions in high-throughput, then continue by focusing on a smaller set of interactions of interest for quantitative characterization. I wanted a single project that would illustrate both steps of the process. With that in mind, it was important to choose a biologically interesting network in order to give the best presentation of the methodology.

To establish MITOMI as a viable method for protein-protein interaction detection, it would be important to use it on more systems, establish a comparison to previously established methods, and show that it could be used to discover interactions not accessible by those methods. I realized I'd discovered a suitable system when I came across a paper describing a proteome-scale study using affinity purification followed

by mass spectrometry (AP-MS) to map *E. coli* protein-protein interactions[14].

The paper spent a couple paragraphs of the discussion on interactions with RNA polymerase, and noted a failure to detect interactions of RNA polymerase with "transcriptional factors, possibly due to their presence in low concentration of transient interactions with RNAP, or high affinity to target DNA." I considered this an excellent opening for MITOMI. I would choose a pool of proteins and map their interactions with RNA polymerase. The pool would be selected partly to sample known interactions, and partly to sample interactions that weren't detected in previously high-throughput studies, but would be expected to occur, such as interactions with transcription factors. The interactions in the first category would provide a basis for comparison of MITOMI to AP-MS methods, while the second category would provide the opportunity to discover novel interactions. Building on the results of the initial mapping, I could then focus on quantitatively characterizing the identified interactions.

If Arifuzzaman *et al.*'s hypothesis about their failure to detect transcription factor interactions was correct, MITOMI would be ideal to capture them. Holding transient interactions in place is the point of MITOMI's eponymous mechanical trapping, the *in vitro* transcription and translation system is designed to boost expression of the proteins of interest as high as possible so that low abundance in the cell isn't an issue, and the lack of DNA or RNA other than the linear expression templates mitigates the interference of the interactions of interest by protein-nucleic acid interactions. (The absence of proteins at high concentration other than the bait and prey also has the effect of focusing the detection on binary interactions instead of interactions mediated by other proteins). Finally, the advantages of working on a system from the most-studied organism in biology shouldn't be underestimated: a large body of previous work would be available for contextualization of any new discoveries, and gene libraries would be readily accessible once the pool of interaction partners was

selected.

In *E. coli*, RNA polymerase is a multimeric protein complex consisting of two α subunits, one each of β, β' and ω subunits, and a σ factor[15]. The σ factor is necessary for transcription initiation, but after the first $\sim$10 nucleotides are transcribed, the initiation complex undergoes a conformational change to become an elongation complex, which includes dissociation of the σ factor, leaving the remaining subunits to continue. *E. coli* has multiple σ factors specific to different sets of promoters and environmental conditions, but most transcription makes use of a a general σ factor, the $70kDa$ major σ factor (σ^{70}). The ω subunit is not strictly necessary for transcription, and is understood to be mainly important to aid assembly of the complex[16][17].

The α ($37kDa$), β ($151kDa$), β' ($155kDa$), and σ^{70} subunits are expressed from the rpoA, rpoB, rpoC, and rpoD genes, respectively. I decided that the project would focus on the interactions of each of these four subunits individually against each member of the target pool of potential interaction partners: $4\times N$ interactions in total, with N the size of the target pool.

The next step was to select the pool. The J. Craig Venter Institute had a library of Gateway clones covering the entire *E. coli* genome, and if I ordered a selected subset, the clones would come in 96-well plates. So a single plate, $N = 96$, was a reasonable pool size. Coincidentally, Arifuzzaman *et al.* reported 96 proteins as interacting with at least one of the four subunits, but although it would have been tempting to consider exactly those proteins, I didn't want to focus all my attention on the results of one study.

So, to get a better idea of which interactions were previously established, I consulted two protein-protein interaction databases, the Database of Interacting Proteins (DIP) and SwissProt. The majority of rpoA, rpoB, rpoC, and rpoD interactions in the databases come from another large-scale study, by Butland *et al.*, also making

use of AP-MS[18], though there are also a number of interactions established by low-throughput methods, such as identifying contacts in crystal structures. The pool I selected was designed to sample different states of knowledge of RNAP interactions, as well as different classes of biological function.

Including the four RNAP subunits themselves, there are 26 proteins that were identified by both AP-MS studies as interacting with at least one of the subunits, of which my pool contained all but three: the ribosomal proteins rplD, rplE, and rpsD. I also selected an additional 33 that were reported by Arifuzzaman *et al.* but not Butland *et al.*, and an additional 10 that were reported by Butland *et al.* but not Arifuzzaman *et al.*. The 30 remaining slots were filled by six proteins with RNAP interactions reported in the databases by low-throughput methods only, and 24 with no RNAP interactions reported (Table 2.1).

Table 2.1: List of genes ordered from Venter Institute, giving gene common name; Genbank locus tag; previously reported interactions with RNAP subunits α (A), β (B), β' (C), or σ^{70} (D); references for the interactions, with subunit specified in cases where interactions with different subunits were established by different references (db indicates that the DIP or Swissprot database reports the interaction citing a source other than Arifuzzaman *et al.* or Butland *et al.*); and functional classification. Genes in italics were not delivered; see below.

Gene	Locus Tag	Partners	Reference	Function
rpoA	b3295	A,B,C,D	Arif(BCD),But(all), db(all)[19][20]	RNAP
rpoB	b3987	A,B,C,D	Arif(ACD),But(all), db(ACD)[19][20]	RNAP
rpoC	b3988	A,B,C,D	Arif(ABD),But(all), db(ABD)[19][20]	RNAP
rpoD	b3067	A,B,C,D	Arif(ABC),But(all), db(all)[19][20][21]	σ factor
ade	b3665	B	Arif	nucleotide metabolism
araC	b0064			transcription factor
arcA	b4401			signal transduction
aspS	b1866	A, C	Arif(C),But(A)	tRNA synthesis
cadA	*b4131*	*B*	*Arif*	*amino acid metabolism*
clpB	b2592	C	Arif	chaperone
crp	b3357	A, D	db[22][23]	transcription factor
cspA	b3556	A,B,C	Arif(BC),But(A)	transcription factor
cspE	b0623	B,C	Arif	transcription factor

Gene	Locus Tag	Partners	Reference	Function
cysB	b1275	A,B	But(B),db(A)[24]	transcription factor
dam	b3387	B,C	Arif	nucleotide metabolism
dhaR	*b1201*			*transcription factor*
dnaX	b0470	C	Arif	DNA polymerase
dppF	b3540	B	Arif	membrane transport
ebgC	b3077	A	Arif	carbohydrate metabolism
elbB	b3209	B	Arif	lipid metabolism
etk	b0981	A	Arif	protein kinase
fabA	b0954	B,C	Arif	lipid metabolism
fecA	b4291	C	Arif	membrane transport
fecI	b4293	A,B,C	Arif(all),db(C)[25]	σ factor
fis	b3261	A, C	But(C),db(A)[26]	transcription factor
fliA	b1922	A,B,C	Arif,db[27]	σ factor
fnr	b1334	A, D	db[28][29]	transcription factor
ftsK	b0890	C	Arif	chromosome segregation
fucR	b2805			transcription factor
fur	b0683	A	Arif	transcription factor
galS	b2151			transcription factor
gcvP	b2903	B	Arif	amino acid metabolism
greA	b3181	A,B,C	Arif(A),But(all)	transcription elongation factor
greB	b3406	A,B,C	But	transcription elongation factor
hns	b1237	B,C	But	transcription factor
hscC	b0650	B, D	Arif(B),db(D)[30]	chaperone
htpG	b0473	C	Arif,But	chaperone
hupA	b4000	A,B,C	But	transcription factor

Gene	Locus Tag	Partners	Reference	Function
hupB	b0440	A	But	transcription factor
hybE	b2992	A	Arif	chaperone
ihf	b1712			transcription factor
ihfB	b0912	C	But	transcription factor
ilvA	b3772	C	Arif	amino acid metabolism
kdgR	b1827			transcription factor
lacI	b0345			transcription factor
lon	b0439	C	Arif	protease
lrp	b0889			transcription factor
malP	b3417	B,C	Arif(B),But(C)	carbohydrate metabolism
malT	b3418			transcription factor
marA	b1531			transcription factor
marB	b1532			transcription factor
marR	*b1530*			*transcription factor*
melR	*b4118*	*A, D*	*db[31][32]*	*transcription factor*
metH	b4019	C	Arif	amino acid metabolism
narL	b1221			transcription factor
norV	b2710	B	Arif	riboflavin metabolism
npr	b3206			membrane transport
nrdR	b0413			transcription factor
nsrR	b4178			transcription factor
nusA	b3169	A,B,C,D	Arif(BC),But(all),db(ABC)	transcription termination factor
nusG	b3982	A,B,C	Arif,But	transcription termination factor
priA	*b3935*	*C*	*Arif*	*helicase*
pta	*b2297*	*A*	*Arif*	*phosphate acetyltransferase*

Gene	Locus Tag	Partners	Reference	Function
ptsN	b3204			carbohydrate metabolism
rapA	b0059	A,B,C	Arif(A),But(all)	helicase
rcnR	b2105			transcription factor
rcsB	b2217	A	But	signal transduction
rhaR	b3906	D	db[33]	transcription factor
rhaS	b3905	D	db[33]	transcription factor
rhlE	b0797	B,C	Arif	RNA helicase
rho	b3783	A,B, D	Arif(D),But(all)	transcription termination factor
rluC	b1086	B,C	Arif,But	rRNA synthesis
rnr	b4179	B,C	Arif	ribonuclease
rplB	b3317	A,B,C,D	Arif(ABC),But(all)	ribosomal
rplO	b3301	A, C,D	Arif(A),But(all)	ribosomal
rpoH	b3461	A,B,C	Arif,But	σ factor
rpoE	b2573	B,C	Arif	σ factor
rpoN	b3202	A,B,C	Arif,But,db[27]	σ factor
rpoS	b2741	A,B,C	But	σ factor
rpoZ	b3649	A,B,C,D	Arif(B),But(all),db(all)[20]	RNAP
rpsA	b0911	A, C,D	Arif(C),But(AD)	ribosomal
rpsD	b3296	A, C,D	Arif(A),But(all)	ribosomal
rpsE	b3303	A,B,C,D	Arif(A),But(all)	ribosomal
rsd	b3995	A,B,C,D	Arif(CD),db(all)[21][27]	anti-σ factor
rutR	b1013			transcription factor
sdhA	b0723	B,C	Arif	carbohydrate metabolism
selD	b1764	C	Arif	amino acid metabolism
soxR	b4063			transcription factor

Gene	Locus Tag	Partners	Reference	Function
soxS	b4062	A	db[22]	transcription factor
speA	b2938	C	Arif	amino acid metabolism
torR	b0995	B	But	signal transduction
usg	b2319	A,B,C	Arif(C),But(all)	amino acid metabolism
uspG	b0607	B,C	Arif	universal stress protein
ydeO	b1499			transcription factor
yiaJ	b3574			transcription factor
zraR	b4004			transcription factor

This last group of 24 consisted mostly of transcription factors; the entire pool was of proteins that had some expectation of being RNAP interactors, either because of previously reported interactions, or because of their functional role. A review of bacterial transcription[34] identifies seven transcription factors—crp, fnr, ihf, fis, arcA, narL, and lrp—as regulating the majority of genes. Of these, crp was recognized as an RNAP interactor by Butland *et al.*, and fnr and fis by the databases; I made sure the remaining four were also in the pool. The review also identified marA as an example of a transcription factor known to operate by a physical interaction with RNAP. AraC, fucR, galS, malT, and rhaS were included as major transcriptional regulators of sugar metabolism[35].

Figure 2.1 presents the information in Table 2.1 in a different format, which allows for a concise impression of the overall connectivity of the network. However it is important to note that because I am ignoring any interactions that don't involve at least one of the four RNAP subunits, the full network involving this pool of proteins is denser with connections.

The quality of the plate of clones that the Venter Institute actually shipped gives an object lesson in the perils of relying on such outside sources. In the first place, only 90 clones were provided (as glycerol-frozen cells in a 96-well plate), and one of the missing ones was rpoB. Sequencing the clones revealed additional errors: several of the clones had a different identity from what was claimed, and in a couple other cases no sequences were found at all. The Institute sent replacements for some of the missing clones, but ultimately six of the 96 genes I had planned to include in the study were unavailable and had to be left out. These are italicized in Table 2.1, and omitted from the network in Figure 2.1.

RpoB would have been a seventh missing gene, but because of its obvious necessity to the study I amplified the gene from the *E. coli* genome myself and used the TOPO TA Cloning Kit (Invitrogen) to insert it into the kit's plasmid. Since I had to do this

for rpoB anyway, I did the same for rpoA, rpoC, and rpoD at the same time.

I made linear expression templates for the screen using the multistep PCR method described in Chapter 1. I added a C-terminal 6-His tag to each of the four preys—rpoA, rpoB, rpoC, and rpoD—and an N-terminal c-Myc epitope tag and a C-terminal T7 epitope tag to each of the 90 baits. I was able to scrape a bit of the cells from the top of each well in the clone plate and use that as template for the first PCRs for most of the baits. The first PCR's primers were targeted to the sequences of

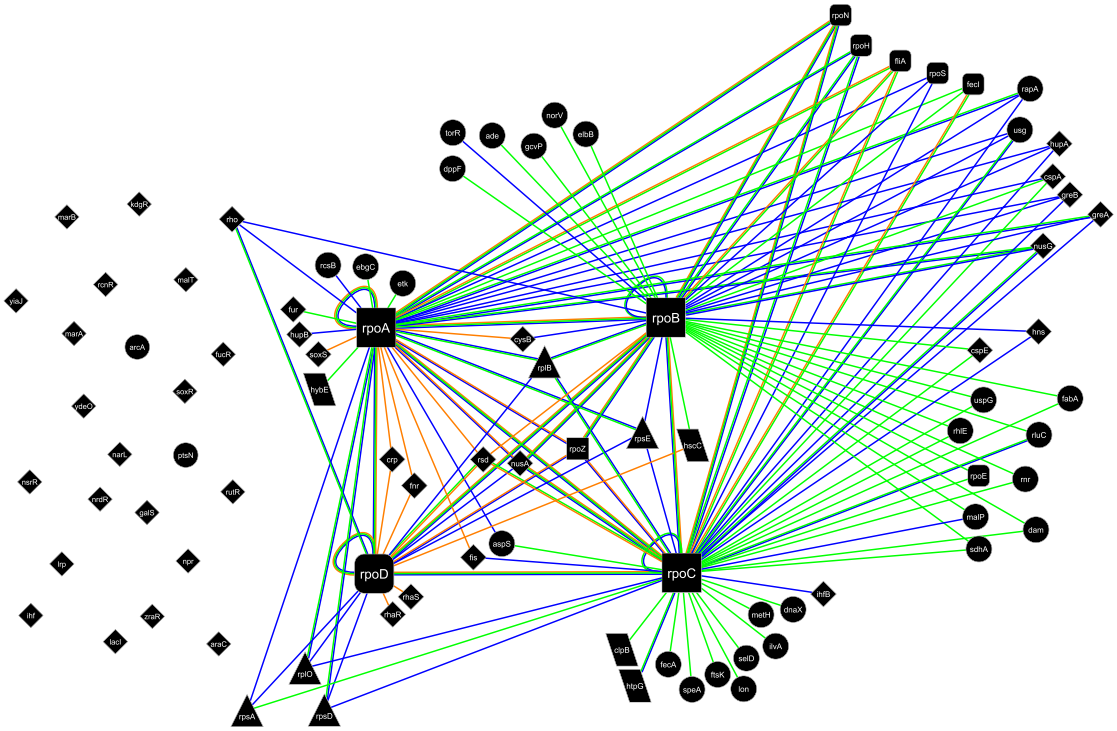

Figure 2.1: Network of previously-reported RNAP interactions. Green edges indicate interactions reported in the high-throughput study by Arifuzzaman et al., blue edges indicate interactions reported by Butland et al., and orange edges indicate interactions deposited in DIP or SwissProt with another reference. Node shapes indicate the functional annotation for each protein. Squares correspond to RNAP subunits, with round-cornered squares being σ factors. Triangles correspond to ribosomal proteins, parallelograms to chaperones, and diamonds to transcription factors. All other proteins are represented by circles.

the Gateway vector flanking the gene insert. I treated rpoA, rpoB, rpoC, and rpoD slightly differently. For the first PCRs of both the bait and prey versions of these four I used the purified plasmid from the TOPO cloning as template, and customized primers targeting the gene insert sequences themselves.

For the interaction screens I used a MITOMI chip with 24 rows of 100 chambers each. The array of linear templates included six replicates of each of the four preys paired with each of the four baits, as well as 36 replicates of each prey with no bait, one copy of each bait with no prey, and six spots containing no templates at all ($6\times4\times90 + 36\times4 + 90 + 6 = 2400$).

I used a biotinylated T7 epitope antibody (Novagen) for bait immobilization and the Roche RTS 500 kit—$25\mu l$ with $0.5\mu l$ of TnT T7 polymerase (Promega) added to boost expression—for ITT. Cy3-conjugated c-Myc epitope antibody (Sigma) was used for labeling expressed bait, and Alexa Fluor 647 Conjugated penta-His antibody (Qiagen) was used for detecting interactions, and the chip was imaged with either an Arrayworx scanner (Applied Precision) or a TECAN LS Reloaded scanner, in both the Cy3 and Cy5 channels.

The amount of measured bait expression in an experiment was rather variable (Figure 2.2). Typically a relatively small number of baits would show a significantly high expression signal, leaving no doubt about successful expression, but most baits gave a signal not significantly greater than that of spots without any bait. That baitless spots would give a nonzero expression signal at all can at least partly be attributed to the nonspecific antibody binding discussed in Chapter 1. To determine if the expression signal for a given bait was significantly higher than that for no bait, I relied on statistical hypothesis tests. I compared the signal from the 25 replicates of each bait to the signal from the 150 spots with no bait. After assigning an expression p value to each bait in this way, I could then crudely group the baits into "definitely expressing" ($p\leq0.01$), "definitely not expressing" ($p>0.05$), and "maybe expressing"

($0.01<p\leq0.05$). Typically 10-15 of the 90 baits would fail the expression hypothesis test in a single trial.

The interaction data had similar characteristics, with a few baits apparently binding the prey strongly, and most of the others giving a less obviously significant signal (Figure 2.3). In this case, the hypothesis tests I did were to determine whether the interaction channel signal from the set of replicates of a given bait-prey pair ($N = 6$) was significantly greater than that from the set of baitless spots with the same prey ($N = 36$).

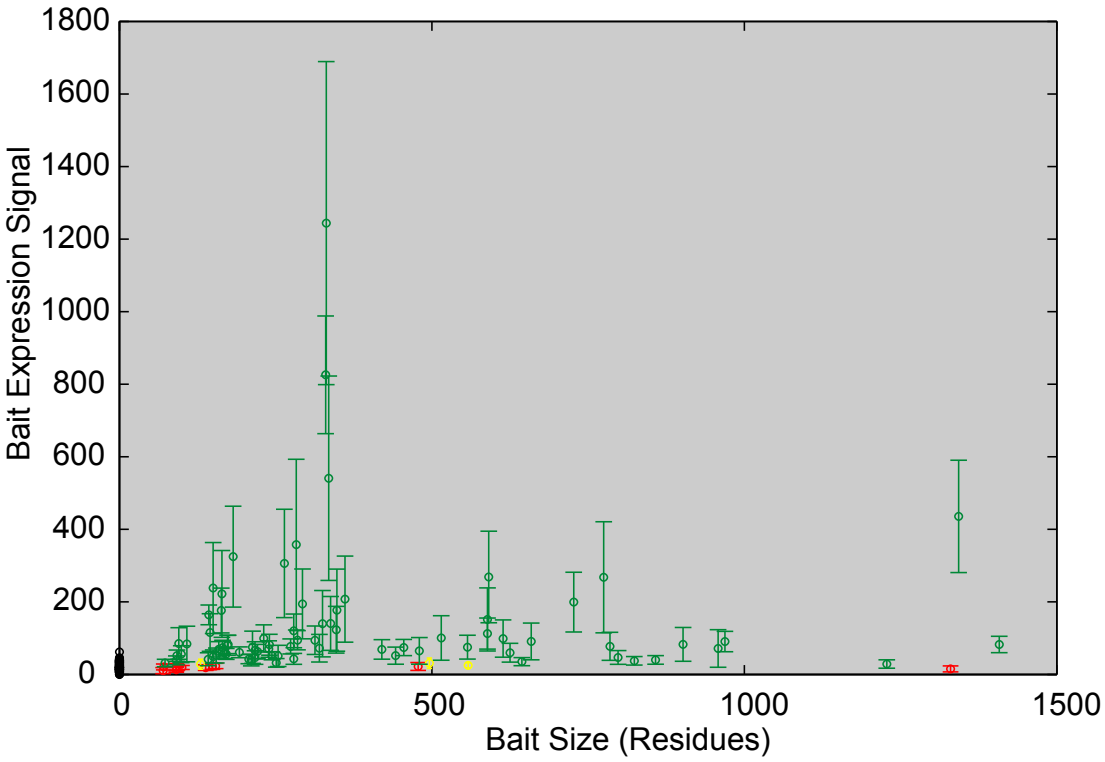

Figure 2.2: Sample expression data from RNAP interaction screen. Mean Cy3 (expression) channel signal ($N = 25$) for each bait plotted against bait size. Error bars are standard deviations. Green indicates passing expression expression hypothesis test with $p\leq0.01$, yellow passing with $0.01<p\leq0.05$, red not passing. The signals from the 150 baitless spots are plotted individually in black.

Although the data are noisy, it was reassuring to note that interaction signal did not appear to be simply proportional to expression signal. Many baits having the lowest ranks of expression signal gave a significant interaction signal. This is some circumstantial evidence that the system was capturing specific physical interactions. On the other hand, it is interesting that the expression signal does not obviously correlate with bait size. One might naively expect that smaller baits would be easier to express, but apparently some other factor limits expression rates.

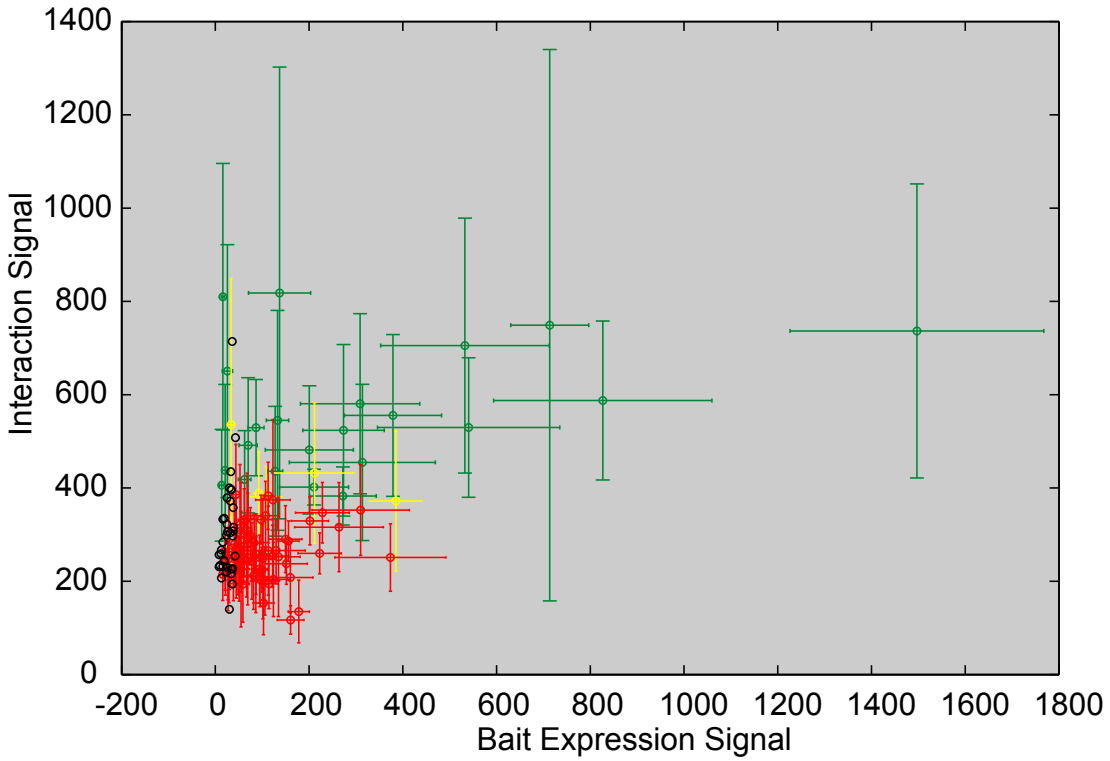

Figure 2.3: Sample rpoC interaction data from same experiment as Figure 2.2. Mean Cy5 (interaction) channel signal ($N = 6$) for each bait paired with rpoC plotted against the corresponding mean expression signal from the full set of 25 replicates of that bait. Error bars are standard deviations. Same color coding as in Figure 2.2, but with respect to rpoC interaction hypothesis test rather than expression hypothesis test. The signals from the 36 baitless spots expressing rpoC prey are plotted individually in black.

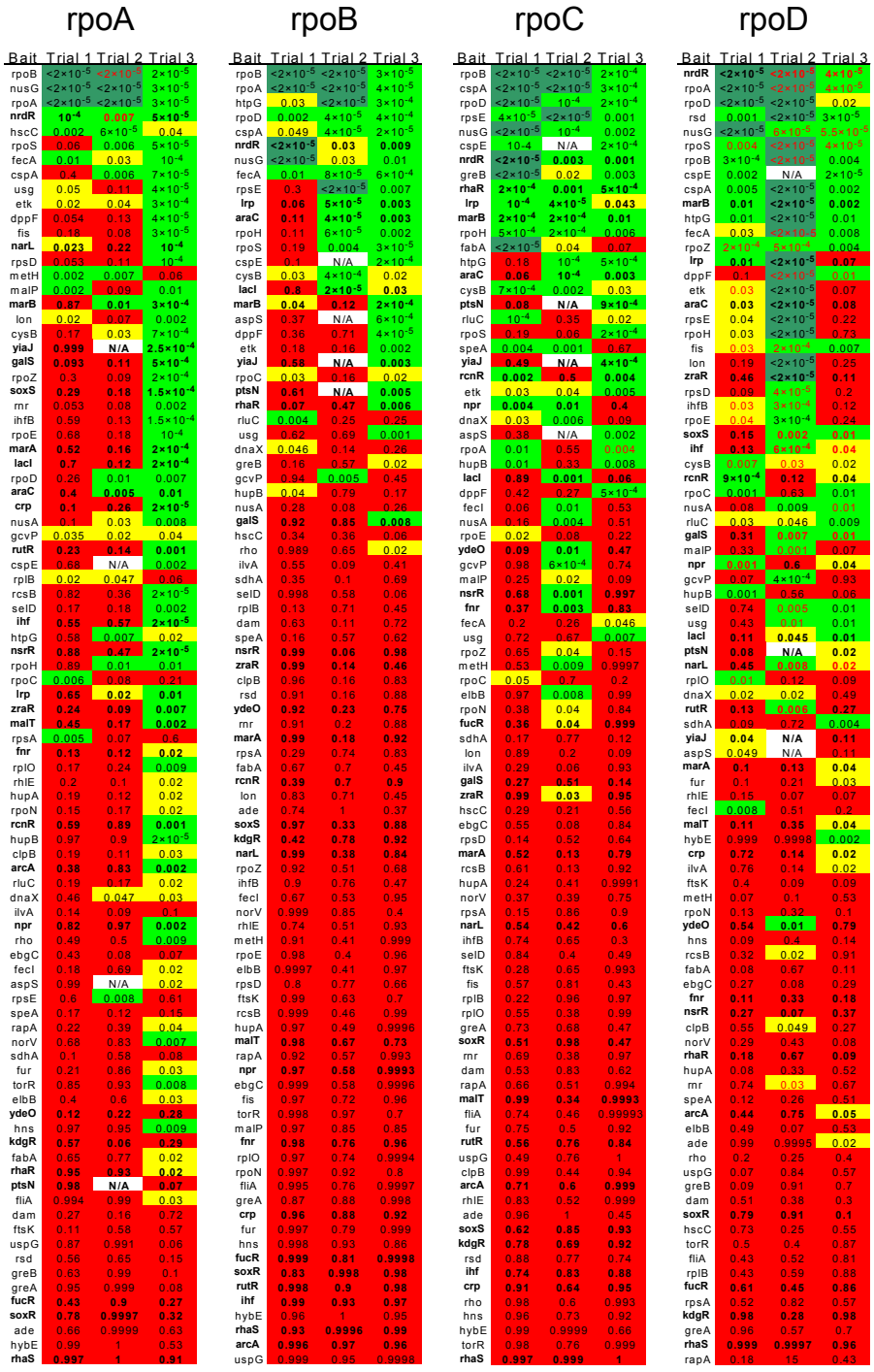

Figure 2.4: RNAP interaction screen results.

Establishing an interaction by comparing to the set of baitless spots expressing the same prey controls for nonspecific binding of both the prey and the interaction labeling antibody to the immobilization antibody, but neglects nonspecific binding of the interaction labeling body to the bait. If I were designing the experiment again, I would include a similarly-sized set of preyless spots for each bait, but analyzing the data after the fact I was limited to one copy of each bait expressed without prey.

Since I could not reasonably do a hypothesis test with an $N = 1$ set, I tried a cruder approach. For each bait-prey combination I calculated a 50000-sample bootstrap estimate of the 95% confidence interval of median interaction signal. If the interaction signal from the single preyless control spot expressing the same bait was greater than the the lower limit of this interval, I rejected the interaction regardless of whether it passed the hypothesis test against the baitless control set.

Even with the use of hypothesis testing to try to separate real signal from background, the relative weakness of the fluorescence leads to somewhat imperfect results. For instance, nine baits—cspA, cspE, ftsK, fur, greA, hupB, ihf, nrdR, and rcnR—failed to pass the expression hypothesis test at the $p \leq 0.05$ level in any of the three experimental trials I used to draw my conclusions. However, cspA, cspE, nrdR, and rcnR consistently passed the hypothesis test for interacting with at least one of the preys, which implies that the expression hypothesis test was consistently giving false negative results. It is easy to imagine an equivalent false negative in the interaction hypothesis testing, with a true interaction being consistently rejected.

It is partly for this reason that I did repeated trials. This brings up the question of how to compare different iterations of the same experiment. The experimental conditions in different trials varied. For instance, I used an Arrayworx scanner with mercury lamp excitation in two of the trials, and a TECAN scanner with laser excitation in the third. The concentration of labeling antibodies also varied, from a 1:50 dilution of the interaction labeling antibody in the first and third experiments to a

1:33$\frac{1}{3}$ dilution in the second. This is not to mention the intangible variables beyond my direct influence. The point is that it would be inappropriate to compare fluorescence data from different trials. This consideration reinforces the necessity of internal controls; I needed to make a decision about the existence of an interaction purely in the context of a single experiment. The p values I calculated for each interaction gave a better basis for comparing different trials.

Figure 2.4 presents the cumulative results of the three trials, grouped by prey. The numerical values are the p values for the interaction of each bait with a given prey in each trial, and the color coding is the same as in Figure 2.3. For each prey, baits are ranked in order of the average logarithm of p values, as an approximate metric of confidence in the interaction. The darker shade of green indicates $p<2\times10^{-5}$, the minimum detectable by the computational hypothesis test I used (see Appendix A); these were assigned a p value of 10^{-6} for ranking purposes. Red font indicates that although the interaction passed the hypothesis test against the set of baitless control spots, it should be rejected because of comparison to the preyless control spot. The baits in bold are the 27 that had no interactions with any of the four preys reported by either Arifuzzaman *et al.* or Butland. *et al.*.

The results support the worry about false negatives from insufficiently strong signals. Despite being essentially a positive control, an interaction between rpoA and rpoC was not consistently detected, whether rpoA or rpoC was the bait. The interaction between rpoC and rpoZ, the ω subunit, though similarly a positive control, barely passed the hypothesis test at even the $p\leq0.05$ level in one trial, and failed in the other two.

Although the most proper way to consider the results is to think of them as falling on a continuum of greater to lesser confidence, the estimation of which is intended by the ranking of p values, I had to interpret the data in a binary way, with a yes-no answer for each potential interaction, to compare to previous studies. I chose to

define an interaction as being detected if it passed the interaction hypothesis test at the $p \leq 0.01$ level in at least one of the trials, though there were other choices I could have reasonably made, such as requiring an interaction to pass the hypothesis test in at least two out of the three trials, or simply choosing a cutoff value for average $log\ p$.

The network defined by this choice (Figure 2.5), while containing roughly the same number of orphans (19) as the previously-established network (22), is not biased against interactions with transcription factors, which immediately lends some support to the idea that MITOMI can more easily detect such interactions. However, it is also worth worrying about the opposite problem, the possibility that these novel

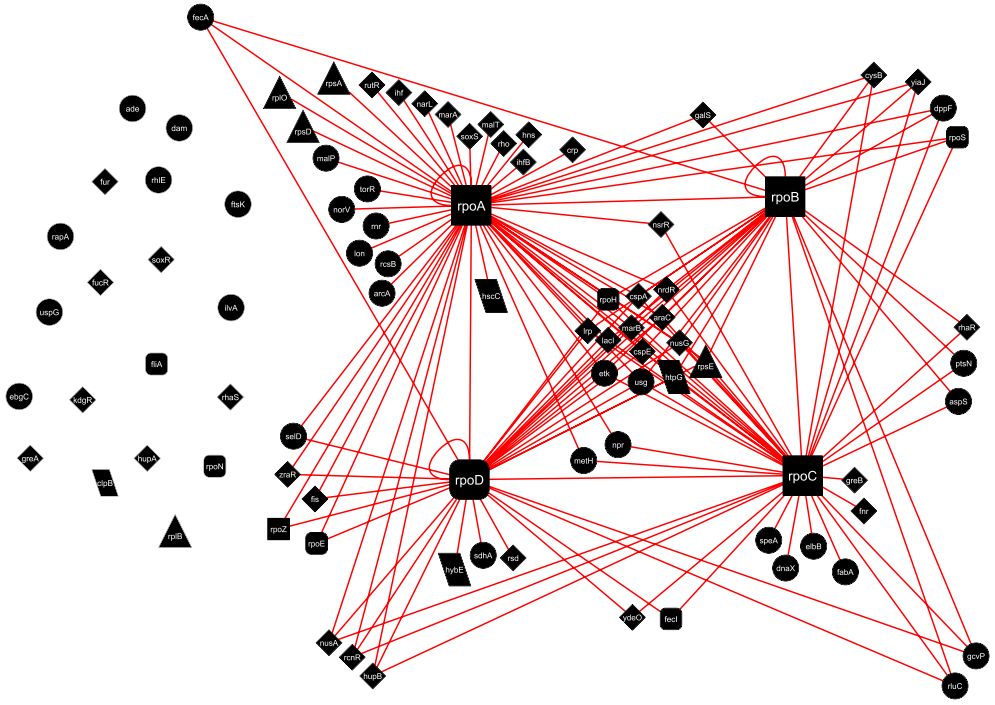

Figure 2.5: Network of RNAP interactions generated by MITOMI, including interactions passing the hypothesis test at $p \leq 0.01$ in at least one trial. Node shapes same as in Figure 2.1

interactions are false positives. In addition to false positives from nonspecific binding, there was another potential source specific to this study but not inherent to the method. The three core enzyme subunits; α, β, and β'; do not exist as monomers in the cell, so genuine physical interactions identified by the binary screen may be "biological false positives" that involve surfaces of the subunits that are inaccessible in the intact enzyme. To address these concerns, it was necessary to validate new interactions with an alternate method, preferably using the assembled enzyme complex.

I chose seven novel interactors for validation: npr, rcnR, nrdR, lrp, narL, rhaR, and zraR. I cloned them into pET-28a(+) vector in One Shot BL21 (DE3) competent cells (Invitrogen), with the addition of C-terminal 6-His tags. I grew the cells at a large scale ($100ml$-$1L$), expressed each bait by inducing with IPTG, and purified them with Ni-NTA resin (Qiagen).

I mixed each of the purified baits at $6\mu M$ with RNA polymerase holoenzyme (Epicentre Biotechnologies S90250) at $1.1\mu M$ and ran the mixture through a Superdex 200 10/300 GL FPLC column with PBS as running buffer. The high molecular weight fractions containing the enzyme complex were sent to the Stanford Mass Spectrometry Facility for analysis. Four of the baits—lrp, narL, rhaR, and zraR—were detected in this way as coeluting with the enzyme, establishing a physical interaction with the complex. Of course, this experiment does not rule out interactions between the complex and npr, rcnR, or nrdR. If the hypothesis that these interactions were not detected by AP-MS because of fast off-rates is correct, the interactions may not have survived for the $\sim 30min.$ time it took for the polymerase to run through the column.

In addition to experimental validation, the MITOMI results can be evaluated by considering prior biological knowledge. For instance, previous work has shown soxS to interact with RNAP specifically through the α subunit[22], exactly as shown by MITOMI. Similarly, rsd's function is to bind to the major σ factor[21], and my

results correctly show it to bind to rpoD but not to the other subunits. On the other hand, rhaS has also been established as interacting with rpoD[33], and the failure of MITOMI to detect this is an example of a false negative.

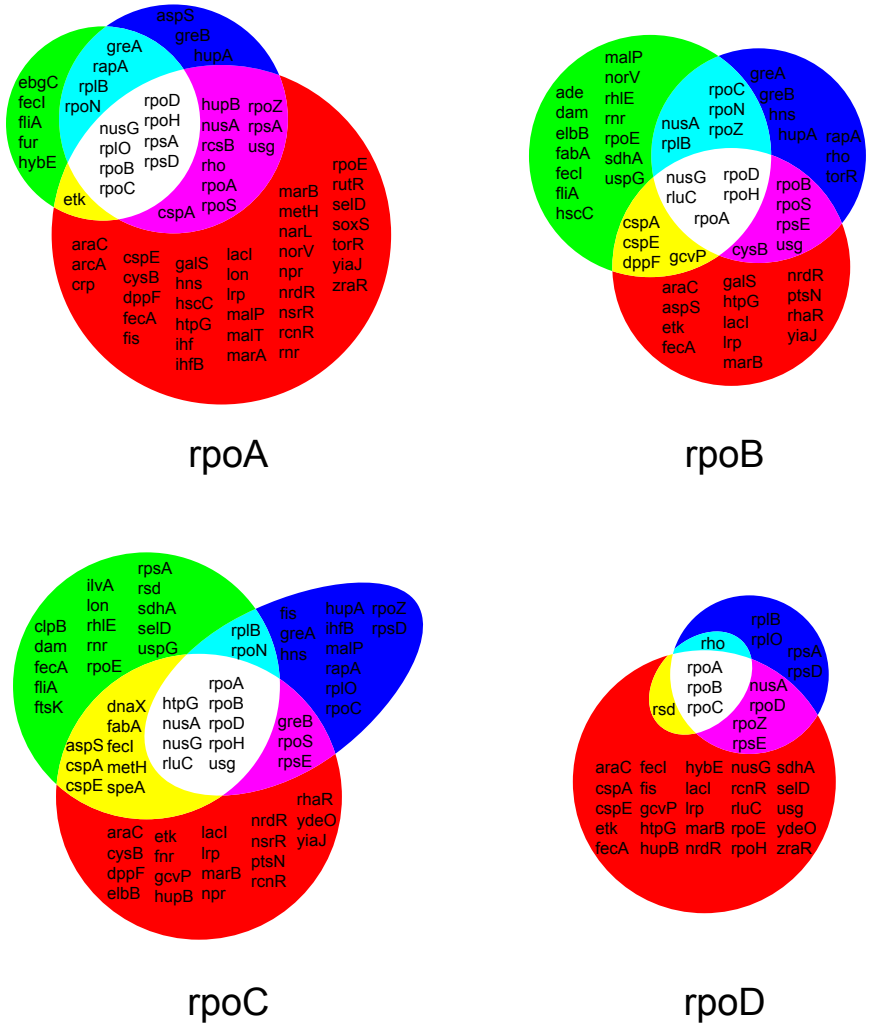

Figure 2.6: Venn diagrams of baits detected as interacting with each of the four preys by Arifuzzaman *et al.* (green), Butland *et al.* (blue), or me (red).

Despite using the same method, the results of the Arifuzzaman *et al.* and Butland *et al.* studies do not agree particularly well with one another, with ∼50 % of the interactions identified by one study undetected by the other (Figure 2.6). The overlap

between MITOMI and either of the AP-MS studies is comparable to that of the two AP-MS studies with one another. This is a clear indication that the AP-MS method suffers from a high false-positive rate, a high false-negative rate, or a combination of the two. However, MITOMI serves as an independent validation of the AP-MS results, and vice-versa, so the 61 interactions that were detected by MITOMI and also one or both of the AP-MS screens can be assigned a particularly high confidence. The FPLC validation of four novel interactors further suggests that a significant fraction of the MITOMI-only results are true positives.

What are the biological implications of the interaction results? With the aforementioned caveat that it may also give a type of false positive result when baits bind to surfaces that are inaccessible in the assembled enzyme, the binary nature of the MITOMI screen allows in principle the identification of which subunit an RNA polymerase binding partner interacts with. For instance, all of the seven major transcription factors were found to bind to the enzyme: arcA, crp, ihf, and narL to the α subunit, fnr to the β' subunit, fis to the α subunit and the major σ factor, and lrp to all four subunits.

In the basic model of bacterial transcription initiation[34], the polymerase binds immediately upstream of the transcription start site, with major contacts at the -35 and -10 sites (i.e. 35 and 10 base pairs upstream of the start site). Transcription can be activated by transcription factors that pre-bind to the DNA promoter region and recruit the polymerase. In the first major type of recruitment, Class I activation, the transcription factor binds well upstream of the -35 element, and the only contact the polymerase can make when binding to the DNA itself is through the flexible C-terminal domains of the two α subunits. Therefore, binding only to rpoA, as arcA, crp, ihf, and narL do, is characteristic of Class I activation. In Class II activation, on the other hand, the transcription factor binds to a target sequence adjacent to the -35 site, where all subunits are accessible. So the binding patterns of fnr, fis, and lrp

are more consistent with this class.

This recruitment model predicts that the transcription factor must bind first to the DNA, before the TF-DNA complex binds to the polymerase. However, lrp, narL, rhaR, and zraR interacted with the enzyme in the absence of DNA in the FPLC experiment, and in the MITOMI experiments, where the only DNA present was linear template sequences designed for T7 RNA polymerase transcription and lacking bacterial promoter sequences, it is highly likely that the interactions were also occuring without DNA mediation.

An alternate mechanism for initiation has been suggested for some transcription factors—including marA and soxS, which were identified by MITOMI as interacting with the α subunit—in which the TF binds to free RNAP first, then the TF-RNAP complex binds to DNA[36][22]. One specific context for this mechanism would be when the number of potential TF binding sites in the genome exceeds the expression level of the transcription factor. It is thought that in this case the transcription factor can more efficiently search the genome by first complexing with RNAP. In this manner, the complex binds only to sites in the genome that contain both the transcription factor binding sequence and sigma factor promoter recognition sequences. The MITOMI results suggest that this mechanism may be more widespread than previously appreciated, and suggests a new line of experimental inquiry to establish the assembly order of TF-DNA-RNAP complexes, and to explore the physiological implications.

This RNA polymerase study establishes that protein-protein interaction mapping with MITOMI can create new insights into biological mechanisms. However, the philosophy I have come to while considering the propensity for false positives and false negatives from MITOMI and AP-MS is that no single method is ultimately adequate on its own. Each will suffer from its own limitations and biases. The greatest value of this method, therefore, is that it potentially represents a third high-throughput

method with biases that are largely independent from those of yeast two-hybrid and AP-MS. A comparative approach of mapping the same interaction networks with different methods will give the truest picture of the interactome.

Chapter 3

A Device to Measure Interaction Kinetics

Applying MITOMI mapping to the RNA polymerase network yielded some potential insights into the the biology of transcription. However, it is possible to do better than a simple binary map. Identifying the quantitative strength of each connection in a network would provide a much more complete understanding of the system[37]. To address this desire, I developed a method to measure molecular interaction kinetics in a MITOMI platform.

Consider a pair of proteins in the context of a MITOMI experiment, with the bait (B) immobilized under the button and the prey (P) in solution ready to interact with the bait and likewise be pulled down. Assuming a 1:1 stoichiometry, the effective concentration of bait-prey dimer, $[BP]$, would increase at a rate proportional to the effective monomeric bait concentration, $[B]$, and to the prey concentration, $[P]$, while the interaction would decay at a rate proportional to $[BP]$ itself:

$$\frac{d[BP]}{dt} = k_{on}[B][P] - k_{off}[BP]. \tag{3.1}$$

The constants of proportionality, known as the on-rate and the off-rate respectively, characterize the kinetics of the interaction. They have the additional feature of determining the affinity of the interaction: at equilibrium the rate of change of dimer concentration vanishes, and the kinetic equation solves to:

$$\frac{k_{off}}{k_{on}} = \frac{[B][P]}{[BP]} \equiv K_d. \tag{3.2}$$

This also makes intuitive sense, since the amount of prey bound at equilibrium can be thought of as determined by a balance between rates of association and dissociation. In addition to giving more detailed information about the interaction, then, kinetics measurements have the added benefit of increasing the efficiency of MITOMI for determination of affinities. Whereas to make an affinity measurement it would be necessary to create a concentration series and measure a binding signal for each concentration, in principle kinetics measurements can measure the affinity of an interaction with a single MITOMI unit cell. In practice, it would of course be necessary to have replicate spots, but it would also be necessary to have replicates at each concentration for an affinity binding curve experiment, so the number of measurements necessary for either method scales proportionally according to the number of desired replicates. Assuming a reasonable number of concentrations probed to make an affinity binding curve to be six, for instance, using kinetics measurements would increase throughput sixfold.

Let F designate the fluorescence signal of bound prey under the button. Assuming that F is proportional to the concentration of immobilized dimer, Equation 3.1 becomes:

$$\frac{dF}{dt} = k_{on}[B][P]\frac{F}{[BP]} - k_{off}F. \tag{3.3}$$

Using the fact that $[B]_0$, the initial concentration of bait under the button, remains

equal to $[B] + [BP]$ throughout the association process, the equation becomes:

$$\frac{dF}{dt} = k_{on}\frac{[B]_0 - [BP]}{[BP]}[P]F - k_{off}F, \tag{3.4}$$

or,

$$\frac{dF}{dt} = k_{on}[B]_0\frac{F}{[BP]}[P] - (k_{on}[P] + k_{off})F = A - (k_{on}[P] + k_{off})F, \tag{3.5}$$

where A must be a constant, since k_{on}, $[B]_0$, $\frac{F}{[BP]}$, and $[P]$ are all constants. This solves to an exponential decay up to a maximum:

$$F = F_0 + C(1 - e^{-(k_{on}[P]+k_{off})t}). \tag{3.6}$$

On the other hand, if instead of beginning the experiment before any prey binds, one begins with a spot of immobilized bait with some amount of prey bound, and no prey in solution ($[P] = 0$), Equation 3.1 becomes

$$\frac{d[BP]}{dt} = -k_{off}[BP]. \tag{3.7}$$

With the same $F \propto [BP]$ assumption as before, this just solves to a simple exponential decay:

$$F = F_\infty + De^{-k_{off}t}. \tag{3.8}$$

With all this in mind, a strategy for a two-part kinetics experiment becomes clear. The first part, the association phase, begins with bait immobilized under the buttons and the introduction of prey at a known concentration. The buttons are lifted for a predetermined amount of time, allowing prey to bind to bait. Then the buttons are closed, and a fluorescence image is taken. Doing this repeatedly generates a series of time points, to which Equation 3.6 can be fit, and the exponential decay constant

$\lambda_{on} = k_{on}[P] + k_{off}$ extracted. Because of the small volume of MITOMI unit cells, it is important to flow fresh prey onto the chip after each association time point, to maintain a constant $[P]$ which might otherwise be perturbed as prey molecules are pulled out of solution.

After the association curve levels off, all the prey is washed out and the dissociation phase begins. The process is exactly the same as with the association phase, except it is fresh buffer that is constantly flowed in to maintain $[P] = 0$. Fitting to Equation 3.8 gives k_{off}, which in combination with the decay constant and prey concentration from the association phase allows k_{on} to be determined.

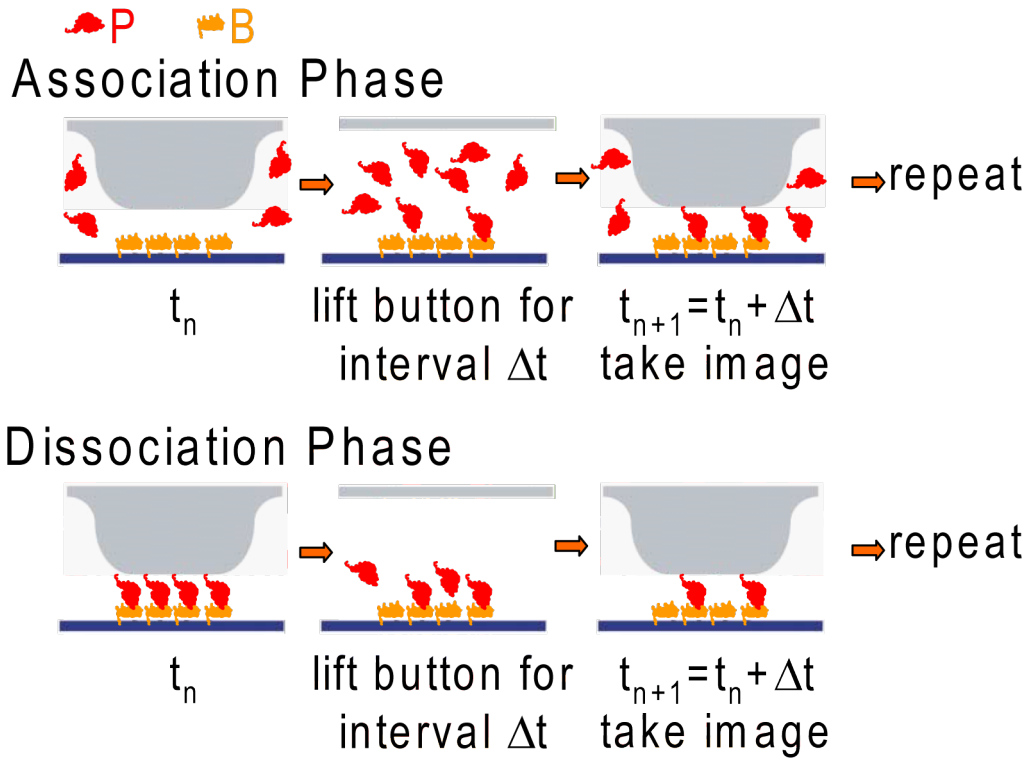

Figure 3.1: The two parts of an interaction kinetics experiment.

Typical on-rates for biomolecular interactions are in the 10^5-$10^6 M^{-1}s^{-1}$ range[38], with off-rates as high as $10s^{-1}$[39]. It is important for the button valves to be able to actuate fast enough to capture these kinetics. The volume flow rate, Q, for control

channels can be estimated with:

$$Q = \frac{wh^3}{12\mu L}\Delta P, \tag{3.9}$$

with w, h, and L being the channel width, height, and length, μ being the fluid viscosity, and ΔP the driving pressure[40].

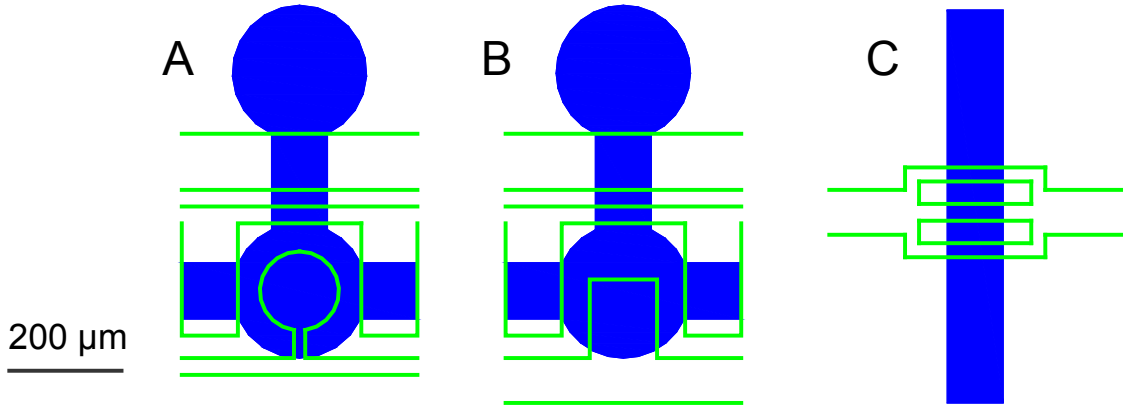

Figure 3.2: First attempt at quick buttons. Flow layer in blue, control layer in green. **A.** Unit cell of standard MITOMI chip. **B.** Unit cell of Quick Buttons chip. **C.** Channel-crossing structure for Quick Buttons chip. The middle branch is $30\mu m$ wide, and the other two branches $25\mu m$ each.

The obvious first thing to change from the standard MITOMI design was to remove the constriction connecting each individual button to the common button control line (Figure 3.2A and 3.2B). In fact, it was preferable for the control line not to narrow at all between the button control port (i.e. the point at which the tubing coupling the button control line to the off-chip pressure manifold enters the chip) and the buttons themselves. This presented a problem. The button control line was $80\mu m$ wide, and it would need to cross flow channels to get to the buttons from the control port, but at that width it would create an unintended valve. According to standard design rules a control line crossing a flow channel should be at most $30\mu m$ in width to avoid an unwanted valve, though the standard MITOMI chips get away as much as $40\mu m$. So

I designed a channel-crossing structure that split the control line into three branches wherever it crossed a flow line, with no single branch being more than $30\mu m$, but with the total $80\mu m$ width maintained (Figure 3.2C).

The redesigned buttons increased w in Equation 3.9; to increase ΔP I concurrently tried another strategy, connecting a vacuum pump to the valve manifolds controlling the chip. Since the control lines actuating MITOMI valves are run at $\sim$15-20psi gauge pressure, switching between pressure and vacuum instead of pressure and atmosphere can nearly double ΔP.

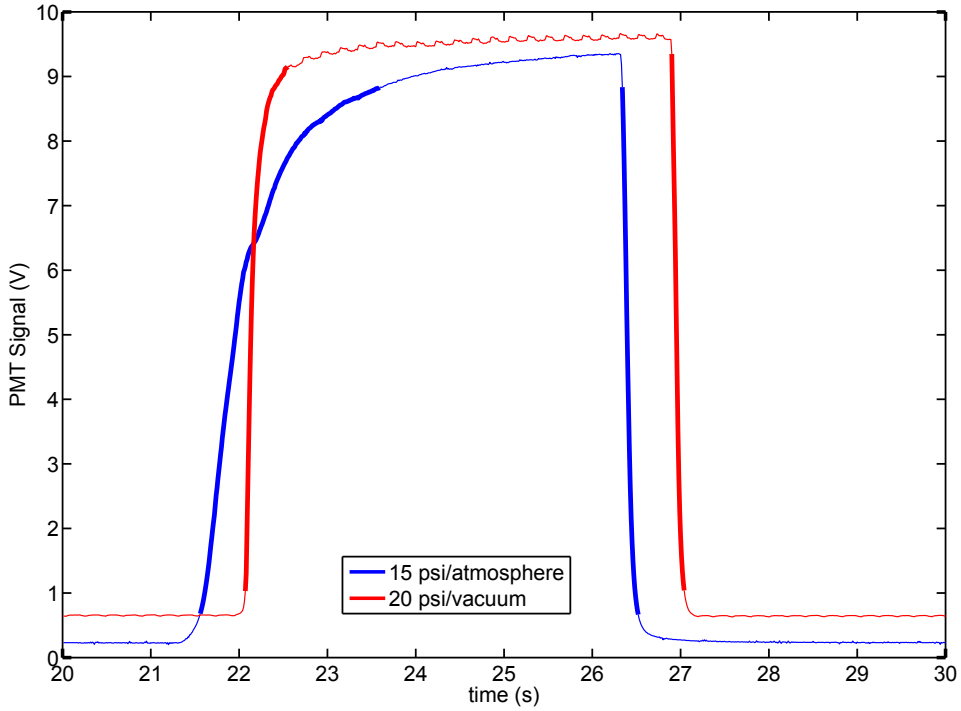

Figure 3.3: Example of a single cycle ($10s$ period) for a single button, actuated either by switching between $15psi$ and atmosphere or between $20psi$ and vacuum. Opening and closing portions of the curves in bold. $15psi$/atmosphere voltage signal multiplied fivefold from raw data for ease of comparison.

I developed a strategy to characterize button actuation speeds fluorescently. I filled the flow channels of the chip with $100\mu M$ fluorescein, and excited with a mercury

lamp filtered through a Nikon B-2E/C filter cube ($480\pm15nm$ bandpass excitation, $535\pm20nm$ bandpass emission). If I precisely positioned the excitation spot under the button through a $40\times$ objective, the state of the button was converted to fluorescent readout. With button open, fluorescein filled the space under the valve, emitting light, but closing the button squeezed the fluorescein out, making it go dark. I wrote a Labview program to repeatedly open and close the button for a chosen number of cycles—typically three—at a chosen period—typically 4-10s—and measured the fluorescence with a photomultiplier tube whose output was read with Labview Signal Express. An ideal button would give a perfect square-wave output under these conditions.

To define opening and closing times, I first defined a "high baseline" for each plateau and a "low baseline" for each trough of the PMT signal by finding the mean voltage in a $1s$ window, positioned (by eye) shortly before the next opening in the case of a trough or before the next closing in the case of a plateau. I then considered the channel to be opening or closing when the voltage was between 5% and 95% of the way from one baseline to the next. In the case of the button in Figure 3.3, the baseline windows were from 20 to $21s$, from 25 to $26s$, and from 29 to $30s$ for the atmosphere condition, and from 20.5 to $21.5s$, from 25.5 to $26.5s$, and from 29 to $30s$ for the vacuum condition. Without vacuum, opening took $2.02s$ and closing took $0.17s$ while with vacuum, opening took 0.47 s and closing took 0.13 s.

The improvement in button performance from the use of vacuum was clear. However, even with vacuum the performance of the first version of Quick Buttons was somewhat lacking. To prevent errors in data interpretation, the interval the buttons are open for each time point in a kinetics experiment (Δt in Figure 3.1) should be large compared to the total time spent actually effecting the opening and closing time: $\Delta t \gg (t_{open} + t_{close})$, or, to put a numerical value on it, Δt should be at the bare minimum $10\times(t_{open} + t_{close})$. If $(t_{open} + t_{close}) = 0.6s$, this limits Δt to a minimum of

$6s$. To properly capture the exponential decay of a dissociation curve (Equation 3.8), ideally one wants Δt to be less than the dissociation half-life, $\tau_{1/2}(=\frac{\ln 2}{k_o ff})$, so that each octave is captured by several time points.

This button, then, would only be good for off-rates of up to $\sim 0.1 s^{-1}$, which would fail to capture the off-rates of the fastest-dissociating interactions. The bigger problem, though, was that the channel crossing structures (Figure 3.2C) tended to collapse and block the flow channels, which became readily apparent when the flow channels were filled with fluorescein for the button characterization experiment. So I decided to start from scratch and completely rethink the chip design.

I realized that the best way to solve the problem of button control lines crossing flow channels was to avoid it entirely. The second generation of Quick Buttons chip was designed so that the button control entry ports were topologically inside all flow channels, eliminating the need for channel crossings. Each button control port could control the buttons for a pair of rows of MITOMI unit cells if they were designed in a back-to-back orientation, which necessitates multiple button control ports for each chip. The need to widen the distance between the two rows of the pair to make room

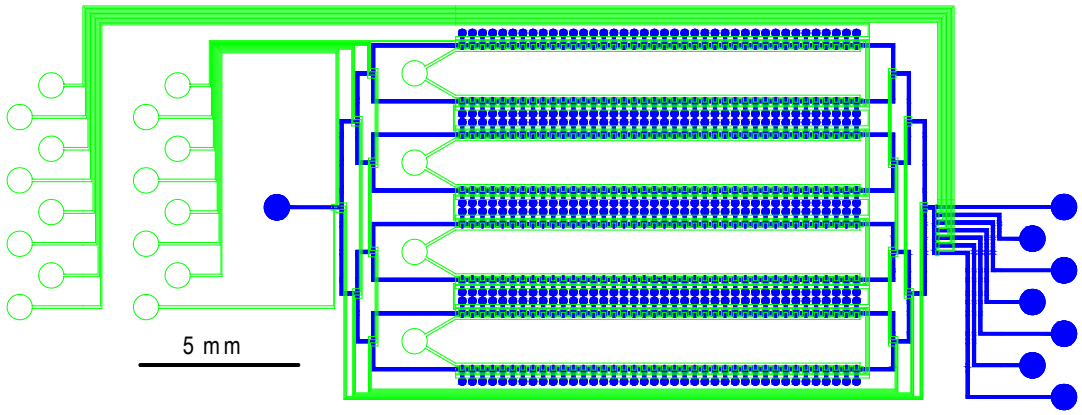

Figure 3.4: Quick Buttons 2 chip. Flow layer in blue, control layer in green. This version has eight rows of forty chambers each.

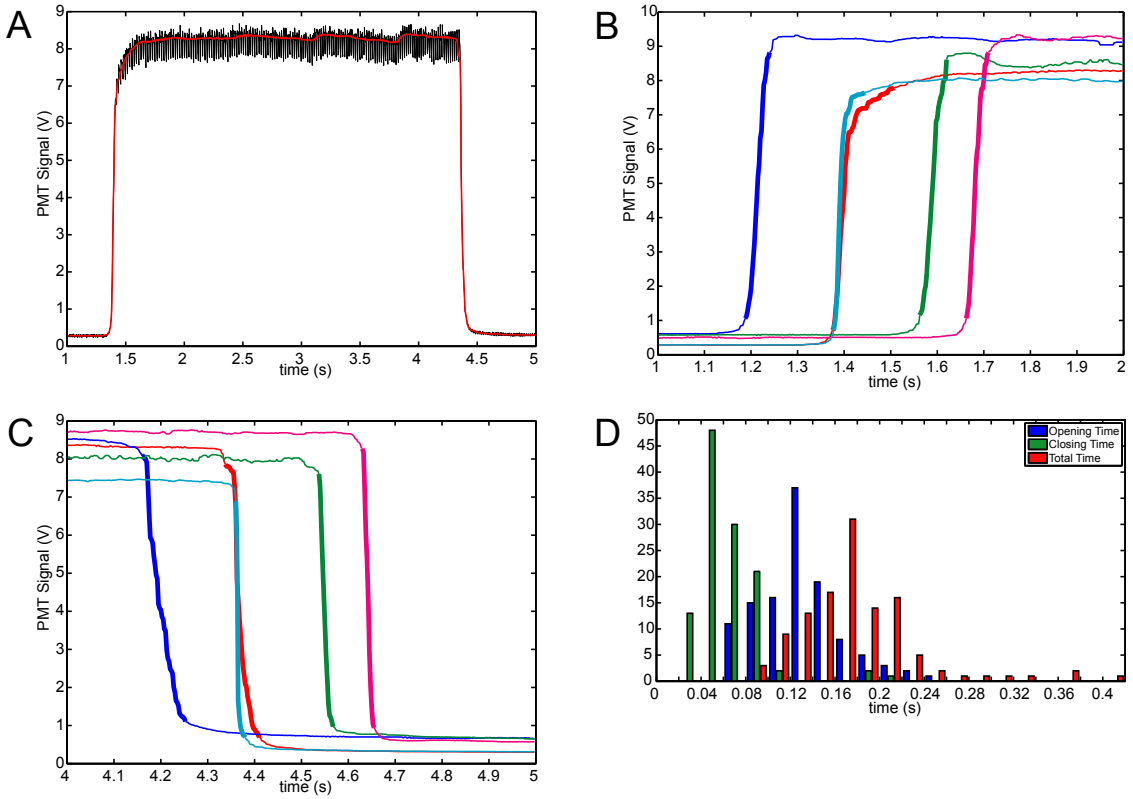

Figure 3.5: **A.** Sample data before (black) and after (red) smoothing. **B.** Sample opening and **C.** closing of chamber 1 (red), 11 (blue), 20 (green), 30 (magenta), and 40 (cyan) in a single chip row. Opening and closing portions of curves bolded as in Figure 3.3. Signal from chambers 11, 20, and 30 multiplied twofold from raw data for ease of comparison. **D.** Histogram of button speed characterizations. $N = 117$: 3 trials of opening and closing for each of 39 buttons.

for the control port limited chip density, but this was an acceptable tradeoff. In the first place, multiple independently controlled buttons are a useful feature for a kinetics experiment in which nothing is known about the rates. In light of the need to have an appropriate Δt compared to $\tau_{1/2}$, it is helpful to be able to cover a wider range of Δts. On the other hand, the reduced density is not such a problem, since the highest density MITOMI chips are optimized for large-scale screens of potential interactions. A chip for kinetics would be used once interactions of interest are already identified, when a somewhat reduced throughput is more acceptable.

Applying the button characterization methodology to the new chip showed improved performance for the buttons. Since button performance might vary with distance from the button control port, I considered buttons in five different positions along each chip row, in Chambers 1, 11, 20, 30, and 40 (by convention chambers are numbered in the direction of flow, or from right to left in Figure 3.4, so chamber 40 is closest to the button control port). Each of those five buttons in each of the eight rows was actuated for three cycles of opening and closing, with the PMT recording at $400Hz$. The raw data were very noisy, so they were smoothed by replacing each data point with the median voltage in a $0.1s$ window (41 data points, $400Hz$) centered at that data point (Figure 3.5A). Then the opening and closing times were defined similarly to before, except that I used $0.25s$ baseline windows, and chose a pair of windows flanking each rise or fall in voltage. Without even using vacuum, opening time averaged $0.11s$ and closing time averaged $0.05s$, with an average of $0.16s$ spent opening and closing in a single cycle, which pushes the measurable off-rate up to $\sim 0.4s^{-1}$ (Figure 3.5D).

I used this chip to study the kinetics of epitope tags interacting with their antibodies. Antibody interactions with 6-His (HHHHHH), T7 (MASMTGGQQMG), and c-Myc (EQKLISEEDL) epitope tags are used for bait immobilization and bait and prey labeling in MITOMI protein-protein interaction screening experiments (see Chapter 1). So the antibodies—biotinylated penta-His antibody (Qiagen 34440), biotinylated T7 epitope antibody (Novagen 69968), and biotinylated c-Myc epitope antibdy (Sigma B7554)—were conveniently available to act as the bait in the context of a kinetic experiment, and it was clearly useful to characterize these interactions that were an integral part of the MITOMI method. Furthermore, any epitope tag could be easily incorporated into eGFP by the two-step PCR method for generating a linear expression template, which provided for a straightforward way to fluorescently measure the interaction.

Since for the purposes of interpreting on-rate data it was also obviously important to know the concentration of the prey being introduced from off-chip, every time I expressed tagged eGFP I created a calibration curve by measuring the fluoresence of a dilution series of commercial eGFP of a known concentration (Biovision) using a spectrofluorophotometer at $480nm$ excitation, $509nm$ emission. Fitting a linear regression to fluorescence as a function of commercial eGFP concentration allowed me to calculate a conversion factor for determining concentration of epitope-tagged eGFP expressed with the ITT kit. Typically these $28\mu l$ reactions gave yields on the order of $1\mu M$.

The procedure I ultimately developed was to set up the surface chemistry in the standard way, depositing the appropriate antibody under the button to play the role of the bait. Then, with all buttons closed introduce unpurified ITT reaction of a known eGFP concentration—one of the major virtues of MITOMI is the ability to work with unpurified protein. Open three of the four button valves in succession, each for a different predetermined interval, then take an image and repeat. Introduce fresh eGFP during the time each image is taken. After finishing the association measurements, open all four button valves and allow the system to come to equilibrium, then lower them, wash out the remaining eGFP with buffer, and begin dissociation measurements. Similarly to the association measurements, leave the fourth valve closed the entire time, and introduce fresh buffer while taking each image. Then fit Equation 3.6 to the assocation data and Equation 3.8 to the dissociation data for each spot.

Leaving one button closed during the experiments was an important control, establishing that the interaction is effectively trapped, with no significant change in signal from protein leaking in or out, and that photobleaching of eGFP from repeated excitation is negligible (Figure 3.6). The data also fit well to the theoretical equations, and that the association curves have similar functional forms for each of the three

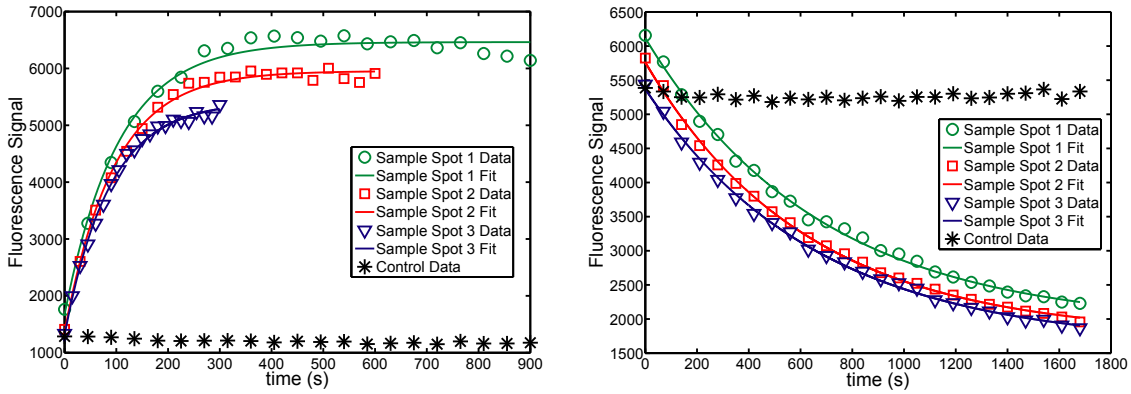

Figure 3.6: Sample data from T7-epitope/antibody interaction kinetics experiment. Left panel data fit to Equation 3.6, and right panel data fit to Equation 3.8. Control data arbitrarily plotted on same timescale as Sample 1 to demonstrate lack of change in signal.

chosen time resolutions supports the association and dissociation being dependendent on the time the button is raised and increases confidence that the observed changes in signal are not artifacts of either repeated mechanical or optical stimulation.

The off-rates, on-rates, and affinities I reported [41] for each epitope tag interaction came from averages across multiple experiments. For each button I extracted k_{off} from the dissociation fit, then extracted the association decay constant λ_{on} and calculated $k_{on} = \frac{\lambda_{on} - k_{off}}{[eGFP]}$ as well as $K_d = \frac{k_{off}}{k_{on}}$. The reported values were the averages and standard deviations of these kinetic constants and affinities calculated spot-by-spot , as opposed to something like averaging λ_{on} across all spots and then calculating k_{on} and K_d (Table 3.1).

With the assistance of the Stanford Protein and Nucleic Acid Facility, I tried to validate the measured His and T7 kinetic constants using surface plasmon resonance spectroscopy (Biacore 3000). The SPR surface was first loaded with the appropriate antibody dialyzed into HBS-EP buffer ($0.01M$ HEPES, pH 7.4, $0.15M$ NaCl, $3mM$ EDTA). The dialysis was necessary because using undialyzed antibody in a previous attempt at the experiment caused a bulk shift in the resonance signal that would have

	$k_{off}\pm SD(s^{-1})$	$k_{on}\pm SD(M^{-1}s^{-1})$	$K_d\pm SD(nM)$	n
His MITOMI	$3.3\times10^{-4}\pm2.8\times10^{-5}$	$4.4\times10^{4}\pm6.5\times10^{3}$	7.5±0.73	240
His Biacore	3.0×10^{-4}	9.8×10^{4}	3.1	
T7 MITOMI	$1.9\times10^{-3}\pm6.6\times10^{-4}$	$4.4\times10^{4}\pm1.0\times10^{4}$	45±14	492
T7 Biacore	8.1×10^{-3}	1.1×10^{5}	72	
T7 MITOMI (HBS-EP buffer)	$5.7\times10^{-3}\pm8.5\times10^{-4}$	$4.3\times10^{4}\pm8.6\times10^{3}$	130±33	120
c-Myc MITOMI	$2.7\times10^{-3}\pm2.5\times10^{-4}$	$1.4\times10^{4}\pm2.9\times10^{3}$	195±49	80

Table 3.1: Summary of epitope-tag/antibody interaction kinetics measurements.

swamped out the desired interaction signals. Then a concentration series of unpurified eGFP ITT reaction was flowed in, association and dissocation was monitored using HBS-EP as running buffer, and kinetic constants were determined by a global fit across the concentrations (Table 3.1).

The Biacore kinetic constants agreed reasonably well (within a factor of two) with the MITOMI kinetic constants. However, I wondered if I could do better, and considered that the differences might be partly explained by the difference in buffers between the two experiments, which is known to be a factor that can potentially alter interaction kinetics[42]. So I did additional kinetics measurements with the T7-epitope antibody, using HBS-EP as buffer instead of the usual PBS. The results were not significantly different with respect to the on-rate, but the off-rate was brought closer to agreement, though not completely. This suggests that the differences are not entirely accounted for by the buffer, but that buffer choice does indeed have a measurable effect.

The discussed method of measuring k_{off} and λ_{on} in the same experiment and using them to calculate k_{on} is the most straightforward. However, there is another way of considering the data: $\lambda_{on} = k_{on}[P] + k_{off}$ is a linear equation with one tunable parameter. If λ_{on} is determined for different experiments with different $[P]$, the slope and intercept of the regression line should be k_{on} and k_{off} respectively.

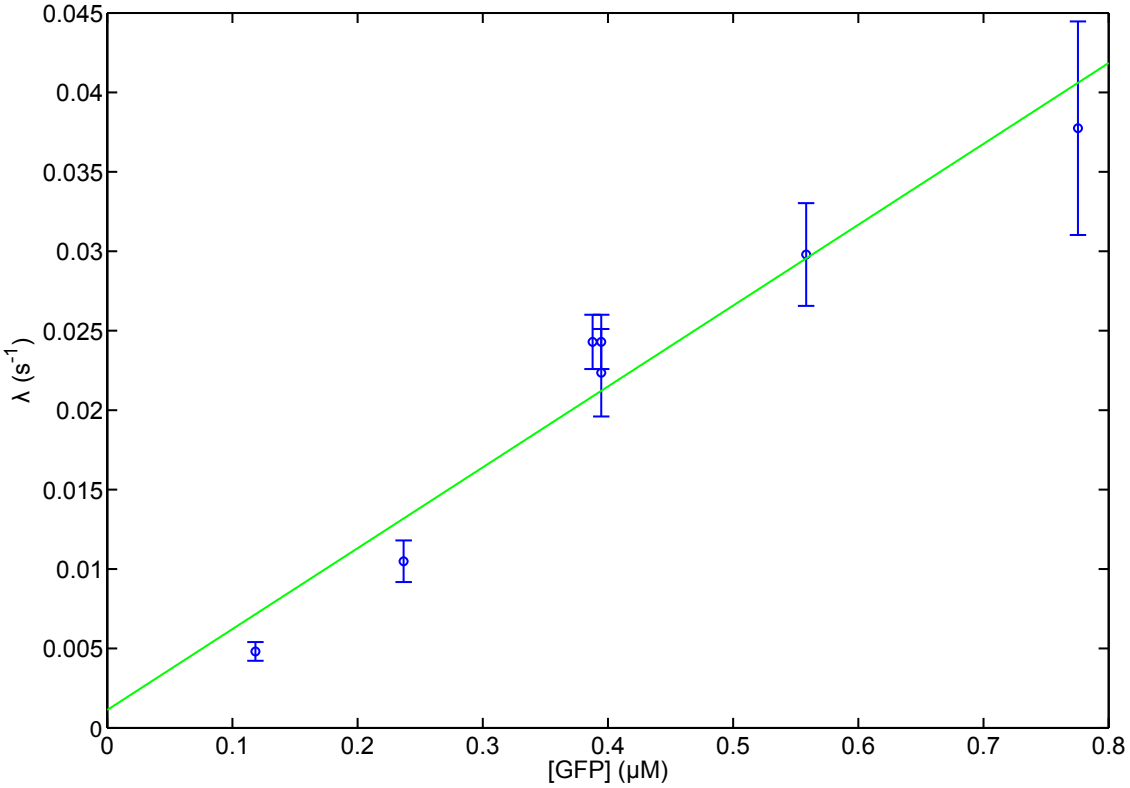

Figure 3.7: T7-epitope/antibody λ_{on} as a function of [P] for seven different experiments. Error bars are standard deviations from averaging over all buttons on the chip.

I had done enough measurements of the T7-epitope/antibody kinetics with enough eGFP concentrations that I was able to apply this analysis (Figure 3.7). The linear regression gave a slope of $5.1\times10^4\pm5.7\times10^3 M^{-1}s^{-1}$ and an intercept of $1.1\times10^{-3}\pm 2.6\times10^{-3}s^{-1}$, making a K_d of $22\pm51nM$. With such large error bars, the k_{off} and K_d calculated by this method are useless, but it was reassuring to see that the slope agreed reasonably well with the previously determined k_{on}.

The problem with analyzing the data this way can be understood by considering an alternative expression for λ_{on}: $k_{on}([P]+K_d)$. For $[P]\gg K_d$ (which is definitely true for all but the lowest two eGFP concentrations considered, and still arguably true for

those two), the intercept is negligibly small. I would have had to probe concentrations much closer to the origin to get useful data on k_{off}. In any case, for any experiment in which λ_{on} can be measured, it is just as easy to measure k_{off} directly, making k_{on} and K_d calculable in a single experiment. So this linear regression analysis would be of limited practical use anyway, though an interesting curiosity.

I attempted to apply the method to measuring kinetics of interactions involving *E. coli* RNA polymerase I had identified (Chapter 2). To make kinetic measurements I needed prey with a covalently attached fluorescent label, and I was able to obtain some RNAP with GFP fused to the C-terminus of the β' subunit[43]. While the epitope tag kinetics experiments did not make use of the DNA chambers or on-chip expression, the RNAP experiments were intended to start taking advantage of MITOMI's arraying and multiplexing capabilities for high throughput.

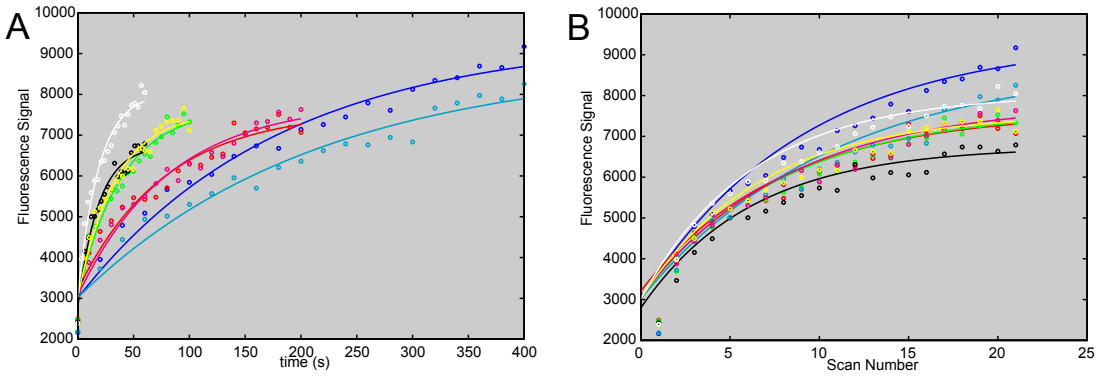

Figure 3.8: **A.** Association curves of RNAP holoenzyme interacting with greB, probed at $20s$ (blue and cyan), $10s$ (red and magenta), $5s$ (green and yellow), or $3s$ (black and white) resolution and fit to Equation 3.6. **B.** The same data replotted as a function of number of images taken.

However, the results were extremely problematic. I was able to get what appeared to be good association and dissocation curves for the interactions between polymerase and several of the baits, but on further investigation the curves turned out to be functions of number of scans rather than of time. The example of RNAP's interaction

with the transcription elongation factor greB—expressed on-chip and pulled down with His-tag antibody—illustrates the problem (Figure 3.8), but it is present for every RNAP interaction that I measured kinetics for. Compare the curves with differing time resolutions in Figure 3.6 to that in Figure 3.8A: the former clearly have similar exponential decay constants, and the curves in Figure 3.8B, where the decay constant is now a function of number of images, match this behavior a lot better.

The dissociation curves show the same symptoms. If it had been the dissociation curves alone, I would have hypothesized that the multimeric nature of RNAP was to blame, and that button actuation was mechanically disrupting the complex, speeding dissocation. However, in the association phase, the curves still seem to be approaching an equilibrium with some RNA polymerase bound. The best explanation I can come up with is that the kinetics are much faster than the buttons can capture, even at the highest time resolution. So $3s$ resoltion doesn't look significantly different from $20s$ resolution compared to the speed of the interaction. The only reason the interaction did not reach equilibrium in the first interval every time is that RNAP was pulled out of solution so quickly as it bound to the bait that the solution was locally depleted, and more binding could occur only after fresh prey was introduced for the next cycle of button actuation. This explanation isn't entirely satisfying, but the idea that the problem depends on fast interaction kinetics is supported by similar results with transcription factor interaction kinetics.

It is well-known that the off-rates for transcription factors interacting with their DNA sequences tend to be large[44]. Nevertheless, transcription factor kinetics were still an attractive target for the system, partly because it would complement other work in the lab on transcription factor affinities as well as my own studies of RNAP protein-protein interactions, but also because it was so much easier to use fluorescently-labeled DNA as prey compared to fluorescently-labeled protein. Whereas the latter case requires engineering a fusion with a fluorescent protein and

determining that the fusion protein folds properly and does not have its interactions disrupted, it is easy to purchase a dye-conjugated oligonucleotide that can be used in a linear extension or amplification to produce full-length labeled prey.

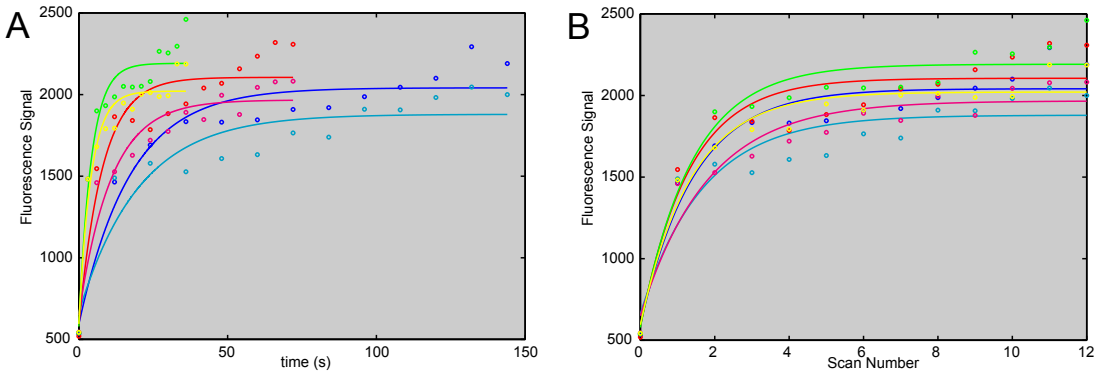

Figure 3.9: **A.** Association curves of DNA containing Erf7 consensus sequence interacting with Erf7, probed at $12s$ (blue and cyan), $6s$ (red and magenta), or $3s$ (green and yellow) resolution and fit to Equation 3.6. **B.** The same data replotted as a function of number of images taken.

One of the interactions I considered was the *Arabidopsis* transcription factor Erf7 binding to its consensus binding sequence: GCCGCC. I expressed 6-His-tagged Erf7 off-chip and introduced it to a chip prepared with His antibody to act as bait. Prey was a Cy5-labeled 52mer that I had prepared by Klenow extension from a template using a Cy5-labeled primer. Unfortunately, the association kinetics showed the same problems as those of RNA polymerase and its binding partners (Figure 3.9). Kinetic measurements of the yeast transcription factor Pho4p binding to its consensus sequence suffered from the same problems.

In this case, the approach to equilibrium is also a lot more rapid, mostly taking place in the first interval, so that the data do not fit an exponential nearly as well. This supports my proposed explanation of fast kinetics combined with local prey depletion: the $\sim 30kDa$ DNA 52mer would be much more mobile by diffusion than the $\sim 450kDa$ RNA polymerase holoenzyme, and able to more completely approach

equilibrium before depletion in the first measurement interval.

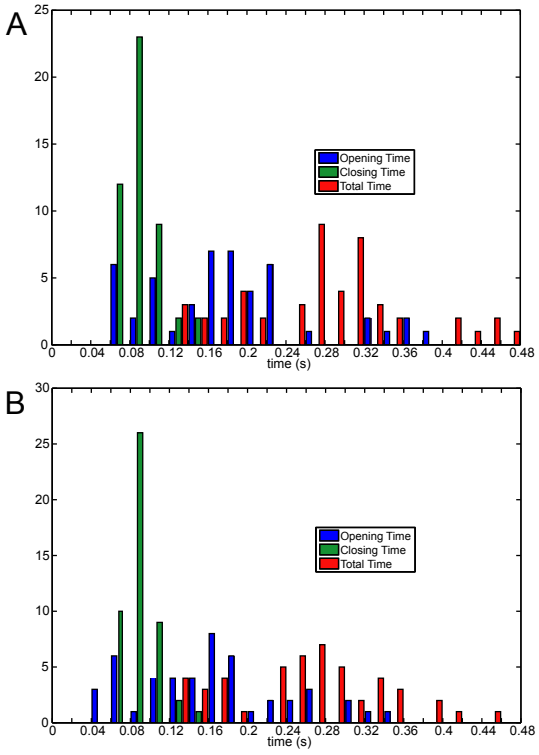

Figure 3.10: Histograms of button speed characterizations from actuating with air. **A.** No vacuum **B.** Vacuum. $N = 48$: three trials of opening and closing for each of 16 buttons.

With this in mind, I made another attempt to speed up the buttons. Using the same chip design I tried actuating the button valves with air rather than water, which would increase Q in Equation 3.9 by decreasing μ. I went through the same button speed characterization procedure as before, using a $20\times$ objective, actuating each button tested for three cycles with a $2s$ period (switching between either $20psi$ gauge pressure and atmosphere, or $20psi$ and vacuum) and recording at $1kHz$. I also maintained an $8psi$ back-pressure in the flow line (by closing off the outlet but continuing to apply pressure to the fluorescein inlet), which should speed up button opening. The data was smoothed with a $0.1s$ window as before, and $0.2s$ windows

flanking each rise and fall in voltage were used to establish baselines for opening and closing. Opening, closing, and total actuation times averaged $0.19s$, $0.09s$, and $0.28s$ respectively for the atmosphere condition and $0.17s$, $0.09s$, and $0.26s$ for the vacuum condition, which was actually a worse performance than before.

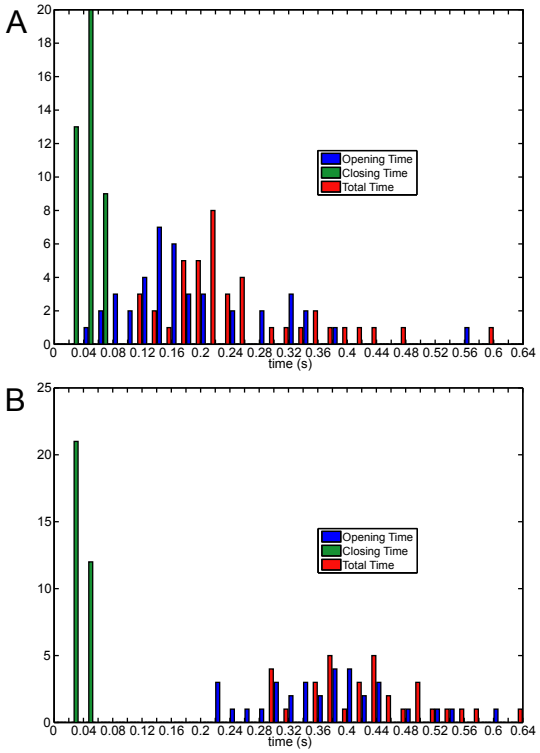

Figure 3.11: Histograms of final button speed characterizations. **A.** No vacuum **B.** Vacuum. $N = 42$ for **A** and $N = 33$ for **B**: 3 trials of opening and closing for each of 14 and 11 buttons, respectively.

After this I made one final attempt. Because of the cubic dependence on h in Equation 3.9, raising button control channel height should be more effective than any of the other approaches. So I designed a modified chip with the four control channels for the buttons $\sim 100\mu m$ high instead of the usual $\sim 20\mu m$. Height for the rest of the control channels was unchanged. On the suggestion that the Tygon tubing used to couple the pressure manifold to the chip was flexible enough to create excess

capacitance that would slow down the button response, I switched to using stiffer PTFE tubing for the button valves. I continued to use air in the button control lines, and actuated with $18psi$ control pressure, maintaining $10psi$ back-pressure in the flow line, though instead of maintaining flow pressure in the fluorescein line I let the pressure build up before button actuation, then closed the sandwich valves to "seal in" the pressure. The results this time were no better than before: $0.20s$, $0.05s$, and $0.24s$ average opening, closing, and total times without vacuum, and $0.38s$, $0.04s$, and $0.42s$ with. It seems that it is not possible yet to probe the fastest—and most biologically relevant—interaction kinetics using microfluidics.

Chapter 4

A Device to Generate Titration Curves

One strength of the approach of making non-equilibrium measurements of an interaction is that determining the on- and off-rates also automatically also yields the interaction affinity, since $K_d = \frac{k_{off}}{k_{on}}$. However, there are cases in which an interaction is not amenable to kinetics measurements, but may be suited to equilibrium measurements. A general method for efficiently generating affinity binding curves in high throughput would be useful for these cases.

My original goal for the RNA polymerase project had been to follow the initial binary mapping with measuring the kinetics of the discovered interactions, using the device I'd developed for the purpose. However, the enzyme turned out to be unsuitable for kinetics measurements, possibly due to its multimeric nature, or to the magnitude of the interactions' off-rates (Chapter 3). Nevertheless, I still had some of the RNA polymerase that was engineered with a GFP label[43], so I began planning a slightly less ambitious set of measurements: generating equilibrium binding curves for RNA polymerase interacting with selected binding partners I had identified, using MITOMI in the same way as it had been in its original application[11].

However, I had received a limited amount of enzyme: 200 μl core enzyme ($\alpha_2\beta\beta'$) and 100 μl holoenzyme ($\alpha_2\beta\beta'\sigma^{70}$), each at $\sim$0.4 $\frac{mg}{ml}$, or about 985 nM core and 840 nM holoenzyme. This presented a problem of resource management that affinity measurements with far more readily synthesized labeled oligonucleotides do not have. Generation of a binding curve would require interrogating multiple concentrations, reasonably at least six but preferably more. I could have done this in a single MITOMI device, by using its addressing capabilities to fill subsets of the MITOMI matrix with different concentrations in series. Typical flow rates onto a chip are $\sim 1.2\frac{\mu l}{min.}$, and when introducing reagents, I would preferably flow at least $3min.$ to make sure the reagent fills the entire row. Assuming a dilution series of 1, 0.8, 0.6, 0.4, 0.2 and 0 times the initial enzyme concentration, such an experiment would require $(1 + 0.8 + 0.6 + 0.4 + 0.2) \times 1.2\frac{\mu l}{min.} \times 3min. = 10.8\mu l$, a significant fraction of the enzyme I had to work with. Taking into account the need for several replicates of each affinity measurement experiment, the possibility that the experiments would not work during the first several attempts, and the fact that I had already used up some of the initial quantity of enzyme during the kinetics experiments, I barely had enough RNA polymerase to do anything with.

However, PDMS microfluidic chips are capable of precisely metering and mixing nl-scale fluid volumes[45]. My material limit problems could be alleviated by generating a dilution series of prey on-chip and loading each row of the MITOMI matrix with a different concentration, using only a small ($\sim$2-3μl) volume initially loaded from off-chip. This strategy would rely on combining three major microfluidic design elements: a peristaltic pump, a rotary mixer, and the MITOMI matrix itself.

A peristaltic pump can be effected simply by placing three or more valves in sequence and actuating them in the right pattern[8]. The amount of liquid pumped per cycle is determined by the dead volume under the middle of the three valves[40]. I used one particular pattern, the 120° pattern. With "1" representing a valve's closed

state and "0" representing its open state and the valves ordered from left to right in the direction of intended fluid flow, the 120° pattern is: 100, 110, 010, 011, 001, 101. Each valve is actuated for half the period of each cycle, but with the beginning of the actuation periods staggered 120° out of phase with one another.

Mixing of fluids is a challenge for microfluidics. Because of the low Reynolds number regime operative in the chips' channels, laminar flow is dominant. Streams of two different fluids will flow in parallel without any mixing except by diffusion. One strategy to deal with this is the rotary mixer[46], or ring mixer, which uses a flow channel in the shape of a ring (Figure 4.1). If fluid is driven around this ring, such as by peristaltic pumping, a given fluid volume element will complete a circumnavigation at a rate varying continuously depending on its position along the channel's cross-section, due to the difference between the inner and outer circumferences. Fluid closer to the ring's center will have a slightly faster angular velocity.

Now if different sections of the ring are filled with two different fluids, A and B, such that the interfaces between the two are initially perpendicular to the channel, the angular velocity differential as the fluids are driven around the ring will cause

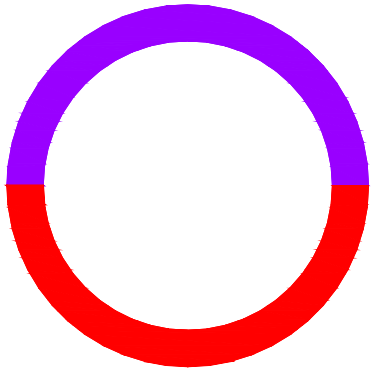

Figure 4.1: Schematic of ring used for mixing, with $400\mu m$ inner diameter and $500\mu m$ outer diamater. Half of the ring is filled with fluid A (violet) and half with fluid B (red), in preparation for mixing.

the interfaces to elongate in the direction of the flow. The separate fluid phases will essentially fold into each other, and the distance between any given volume element of each fluid and the nearest interface will go to zero. As this limit is approached, diffusion will take over, completing the mixing in a finite time.

My basic approach for one-shot generation of a dilution series on chip was clear: use a peristaltic pump to meter a selected volume of undiluted prey into a ring mixer, where it can be diluted by mixing with buffer. Then use valve addressing to select one row of the MITOMI matrix, meter the diluted prey into the row, and repeat with a different dilution and a different row.

Since I was returning to equilibrium measurements, I no longer had a need for rapid button actuation, and returned to the style of buttons used in standard MITOMI chips (Figure 4.2). I designed the MITOMI matrix with 8 rows of 48 chambers each. Since the diluted prey that would fill the ring after mixing would have to be sufficient to fill an entire row of the chip, the ring would have to be much larger than in previous rotary mixer designs. I would be filling each chip row with the neck valves closed, in which context the footprint area of each MITOMI unit cell can be estimated by

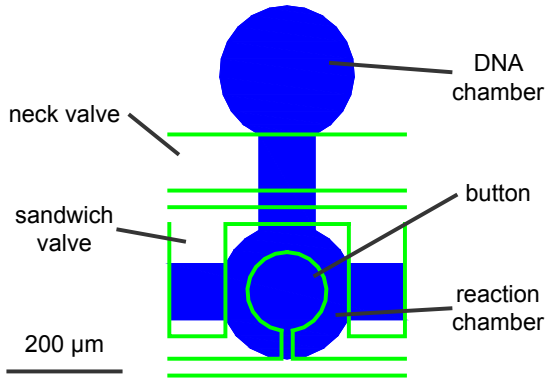

Figure 4.2: Unit cell of MITOMI matrix used in dilution series chip. Flow layer in blue, control layer in green.

approximating the reaction chamber as a $120\mu m$ radius circle and the connection to the next chamber (underneath the sandwich valve) as a square with a side of $100\mu m$.

The area I would need to cover is therefore approximately

$$48\times(100\mu m\times 100\mu m + \pi\times(120\mu m)^2)\approx 2.65\times 10^6\mu m^2. \tag{4.1}$$

The ring I designed had an inner radius of $8940\mu m$ and an outer radius of $9060\mu m$. The footprint area of the ring can then be estimated by its channel width multiplied by the channel-center circumference:

$$2\pi\times 9000\mu m\times 120\mu m\approx 6.79\times 10^6\mu m^2, \tag{4.2}$$

more than enough to comfortably cover a chip row.

This ring size actually approaches a size limit in chip design. The standard size of the glass epoxy microarray slides that MITOMI chips are bonded to is $1in.\times 3in.$ ($\sim 25.4mm\times 76.2mm$). The $\sim 18mm$-diameter ring needs an additional $\sim 1mm$ tolerance distance from the edge of the chip on either side, necessitating a $\sim 20mm$-wide chip. Add another couple mm on each side between the chip edges and the slide edges, necessary for the chip to fit into the holder of the scanner during imaging, and this ring is already at about the maximum size that will fit onto a chip that is bonded to a standard slide. Designing a matrix with a much larger number of chambers per row—more than about 120 or so—would require a larger ring and a larger slide to bond to.

On the other hand, the large ring size offers an advantage in the efficient use of "chip real estate." The interior of the ring is large enough to accomodate all of the flow inlets and the valves controlling those inlets, in addition to the valves constituting the chip's two pumps.

The overall chip design is shown in Figure 4.3. The seven inlets join together into

a common inlet channel that enters the ring from its interior. As with all MITOMI chips, each inlet is controlled by its own valve and, as is optional but common for such chips, there is additonally a common inlet valve controlling entry of fluid into

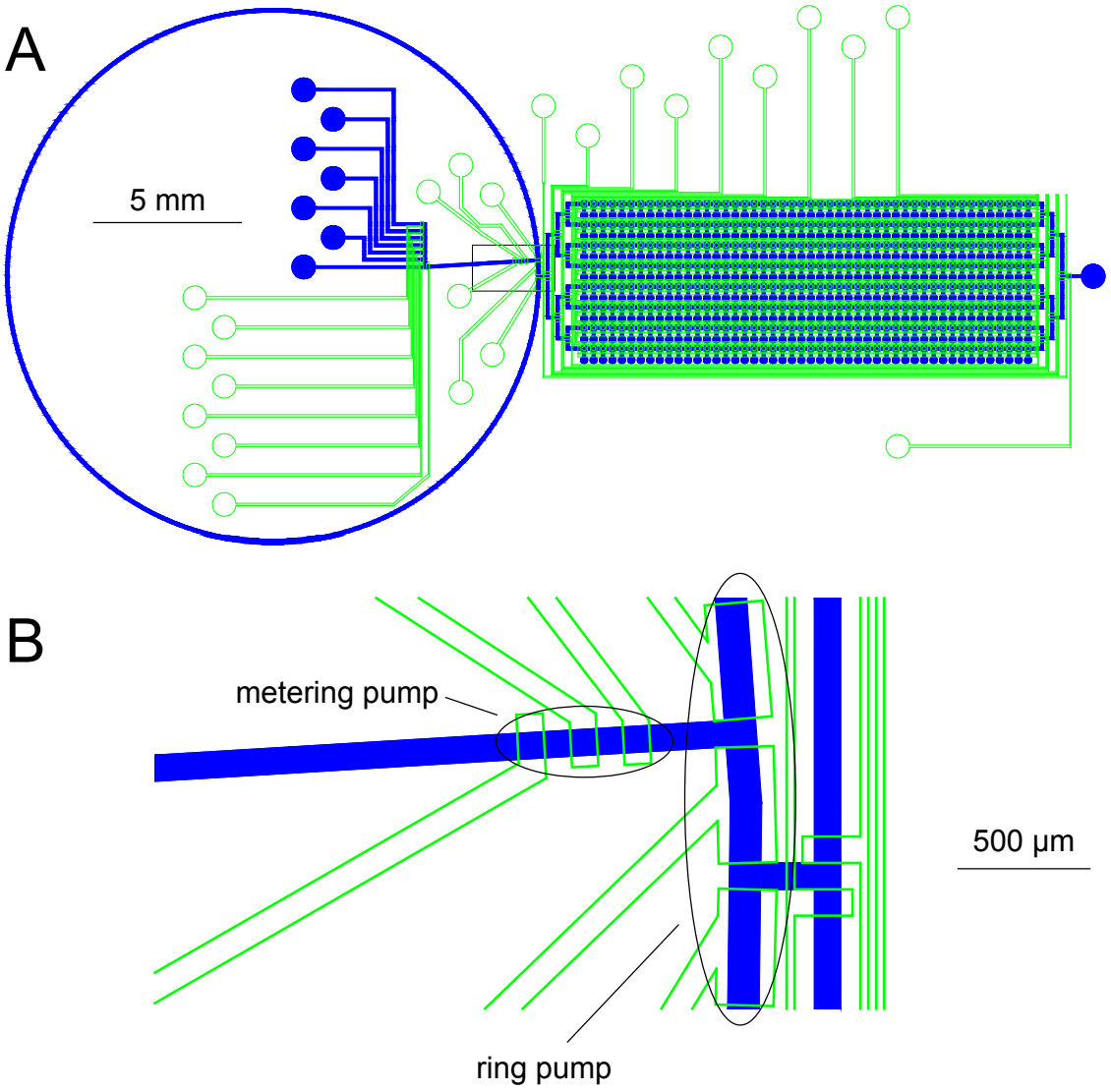

Figure 4.3: **A.** Overall layout of dilution series chip. Flow layer in blue, control layer in green. MITOMI rows are numbered increasing from bottom to top, and unit cells in each row are numbered increasing from left to right. **B.** Blowup of boxed region in **A** showing metering and ring pumps.

the common inlet region from any of the individual inlets. The metering pump used to inject defined amounts of fluid into the ring is placed immediately before the entry point of the common inlet into the ring.

The metering pump valves cross the flow channel with a $100\mu m$ width (i.e. in the direction parallel to flow), the same size as the valves used for inlet control and addressing. On the other hand, the ring pump valves have a $440\mu m$ width, much larger than is typical, to increase the unit volume pumped by each ring pump cycle and thereby increase the speed at which fluid is driven around the ring to effect efficient mixing. The common inlet's entry point into the ring is angularly displaced from the ring's exit point into the MITOMI matrix region of the chip, with the middle of the ring pumps in between.

The chip can operate in any of three modes with respect to the ring. In "ring bypass mode," the middle ring pump valve is open while the other two ring pump valves are closed, and fluid can enter the MITOMI region directly from the common inlet. In "ring entry mode," the closure state of each ring pump valve is reversed, and fluid must travel counterclockwise around the entire ring to get from the common inlet to the MITOMI region. Finally, in "mixing mode" the ring is isolated by the adjacent metering pump valves and address valves.

There are two additional considerations important to the use of these peristaltic pumps. The first is due to the flexibility of the PDMS. This constitutes a pressure capacitance that would reduce pump performance, since a large part of the local pressure increase that is intended to drive fluid movement would be absorbed by swelling the flow channel. Although the pump can function at zero flow pressure, then, it is aided by the technique of applying external pressure through both the inlet and outlet. I would typically run the inlets of all fluids to be applied during mixing (i.e. the buffer and the solution to be diluted) as well as the outlet off the same macroscopic pressure manifold at $\sim$7-10psi, and allow pressure to build up,

pre-swelling the PDMS in the flow channels before beginning metering and mixing. With the flow channels already pressurized, the capacitance problem is alleviated. The control channels driving the pump valves themselves were actuated at $\sim 19 psi$. It was important to make sure that the lines controlling the pump valves were filled with water and devoid of bubbles, since air's compressibility would create capacitance of its own and limit valve performance.

The second consideration relates to choosing the optimal speed to run the pumps at. It would obviously be desirable to have as many pump cycles per second as possible, given the number of pump cycles necessary for metering and mixing multiple dilutions. On the other hand, there is a limit to the speed of valve response, and driving the pumps at a rate approaching that limit would lead to incomplete closing of the pump valves, which would decrease the amount of fluid driven per cycle and ultimately defeat the purpose of increasing pumping speed. With some trial and error, I settled on a standard pumping rate of $40 Hz$, which, it is important to note, means $\frac{1}{40} s$ for each of the six elements of the 120° pattern. The actual rate in terms of pump cycles is $\frac{40}{6} Hz$. Because the performance of the metering pump might vary according to factors that would potentially affect the flow pressure experienced in the channels, for consistency's sake I would always set the address valves such that only Row 8 was open during the metering of fluid into the ring prior to mixing, and I would keep buttons actuated during the entire mixing procedure.

I first tested to see if the metering and mixing capabilities worked as I'd planned, using an earlier version of the chip shown in Figure 4.3. This chip had a $9940 \mu m$ inner radius and a $10060 \mu m$ outer radius for the ring, and was designed before I realized it was just slightly too large a chip to consistently fit onto a slide without problems inserting the slide into the scanner. I calibrated the pump with food coloring and found that complete loading of the ring in ring entry mode, from the common inlet entry point to the exit point into the MITOMI region, corresponded to $\sim 650 - 700$

metering pump cycles. It took an additional ∼300 cycles to push the fluid at the exit point to the end of Row 8.

I then generated a dilution series of Cy3-conjugated c-Myc antibody, starting with a concentration tenfold diluted from the antibody's initial concentration in the tube. I would meter a given number of cycles of antibody into the ring and chase it with 100 cycles of PBS to clear the common inlet. Then I would isolate the ring by closing all the metering pump valves as well as the outermost address valves and mix with 2000 ring pump cycles clockwise followed by 2000 cycles counterclockwise. I would address the appropriate row, and load the diluted antibody with 500 metering cycles in ring entry mode. Then I would switch to addressing Row 8, clear the ring and Row 8 with 1000 metering cycles of PBS, and repeat for the next dilution. I loaded 5 of the 8 MITOMI rows with dilutions this way, and after the 5th, instead of the usual PBS clear through Row 8, I switched to metering undiluted antibody into a 6th row, for 1000 cycles in ring entry mode. Finally I cleared the ring and Row 8 one last time, with 1500 metering cycles of PBS.

Then I imaged the chip in the Cy3 channel of the TECAN LS Reloaded scanner. In the Genepix analysis I measured the fluorescence in the neck of each chamber, immediately adjacent to the neck valve (which had remained closed during the entire process since the DNA chambers were not used). I found that despite the clearing of Row 8, a signficant amount of fluorescence remained in that row. I suspected this was due to the antibody binding nonspecifically to the glass surface, which had not been protected with the usual surface chemistry steps. However the other six rows showed a consistent fluorescence signal across the entire row, demonstrating the completeness of the mixing, and the average fluorescence in each row linearly depended on the number of pump cycles used to load the ring before mixing, demonstrating precise metering (Figure 4.4).

Next I attempted to actually measure an affinity of a biomolecular interaction.

For my first test case I looked at the interaction between the yeast transcription factor Pho4p and an oligonucleotide containing its optimal binding sequence, CACGTG, the affinity of which had previously been characterized by MITOMI[11]. I prepared a chip

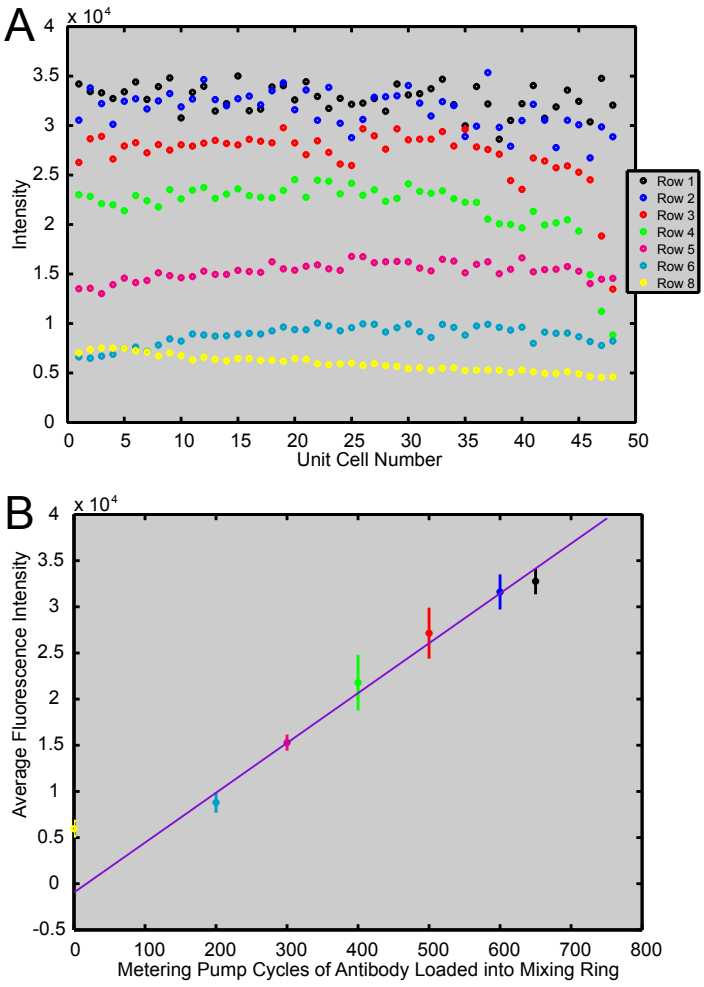

Figure 4.4: **A.** Cy3 intensity measured in each chamber of each row in initial test of chip. **B.** Average intensity in each row, ± standard deviation, as a function of amount of antibody metered into ring before mixing the dilution for that row. Row 1, which received undiluted antibody without mixing, was assigned a value of 650 cycles, as the estimate of cycles corresponding to entirely filling the ring. Linear regression (omitting Row 8 outlier) gives $r^2 = 0.988$.

with the usual surface chemistry steps, depositing biotinylated penta-His antibody under the buttons. Instead of an on-chip ITT reaction, I introduced Pho4p that had been expressed off-chip as bait, and proceeded with generating a concentration series as before, starting with an initial concentration of $40nM$ Cy5-labeled oligonucleotide. The one difference was that in the previous experiment Row 8, the row I used as a path to outlet during metering and ring clearing, was intended to ultimately have a final prey concentration of 0. For this and future experiments, I intended for Row 1 to have no prey, and avoided metering anything into it, to prevent any more nonspecific binding problems. Row 8 was now assigned undiluted prey, so after loading the other rows with their dilutions I simply metered undiluted prey into Row 8 without clearing. I then equilibrated for $\sim 1hr.$ at room temperature before imaging Cy5 fluorescence.

To calculate the affinity, I needed to fit to the equilibrium binding equation

$$F = \frac{F_{max}[P]}{[P] + K_d} + F_0, \tag{4.3}$$

with F_{max}, K_d, and F_0 the undetermined parameters. During the Genepix analysis, in addition to measuring the non-background-subtracted signal under each button (the F in Equation 4.3), I again measured the fluorescence in the neck region of each chamber, F_{neck}, as a proxy for the local concentration of oligonucleotide prey (the $[P]$).

To get an intelligible K_d value in units of concentration, I needed to convert each F_{neck} to a concentration value. Knowing the undiluted prey concentration, $40nM$, I measured the fluorescence of points along the branching path from the ring's exit point to the beginning of Row 8, since this path contained undiluted prey from the final loading of Row 8. The average signal at these points, $F_{ceiling}$, corresponded to $40nM$. To determine the baseline fluorescence signal corresponding to no prey, F_{floor}, I averaged the signal from selected spots in the DNA chamber of Row 1. Then, for

each MITOMI unit cell,

$$[P] = 40nM \times \frac{F_{neck} - F_{floor}}{F_{ceiling} - F_{floor}}. \tag{4.4}$$

The data points followed roughly the expected functional form, but a significant amount of inherent error is apparent, which can be seen for instance from the fact that some of the calculated local prey concentrations are greater than the theoretical maximum of $40nM$ (Figure 4.5). In addition, the data points from Row 8 behaved oddly and did not fit the pattern of the others; I excluded all Row 8 data points from

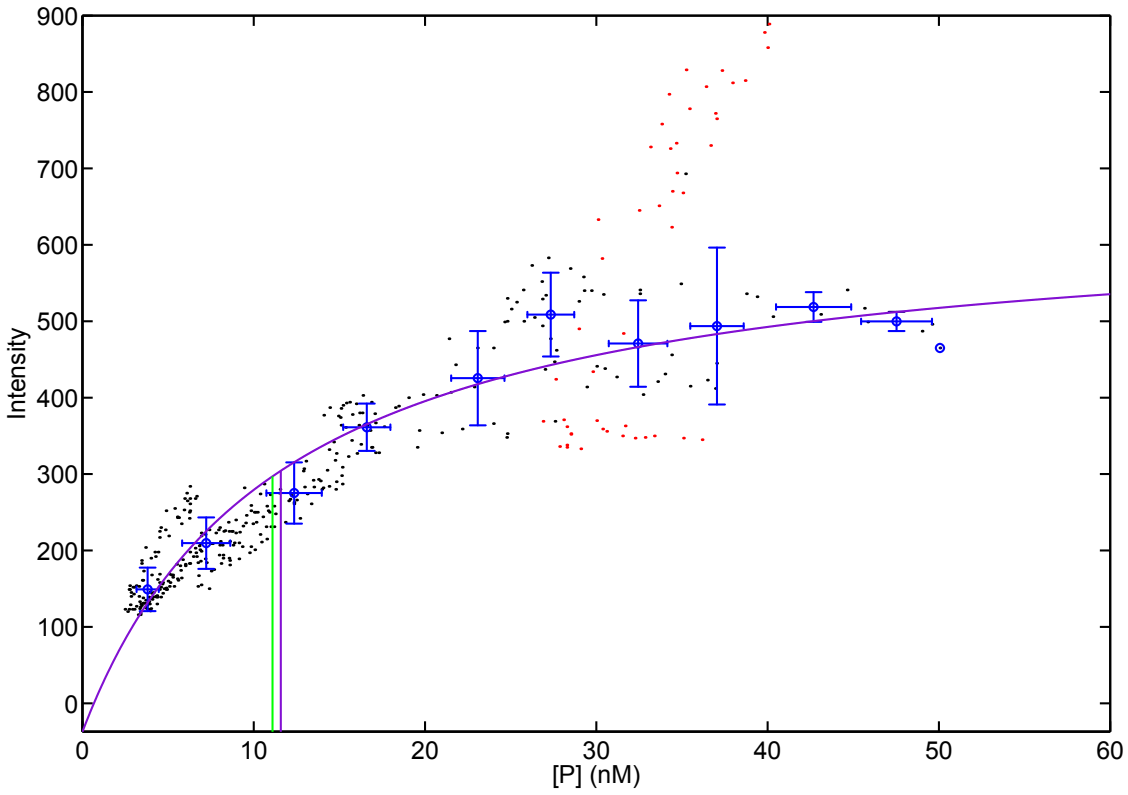

Figure 4.5: Binding curve from Pho4p affinity experiment. Data points from Row 8 in red. Bin averages and standard deviation error bars in blue. Fit to bin averages and fit K_d in violet. Canonical K_d in green.

further analysis. Then to try to reduce the impact of noise, I divided my concentration range into $5nM$ bins, and found the average $[P]$ and F for each bin. I fitted these bin averages to Equation 4.3, and determined an experimental K_d of $11.6nM$, in excellent agreement with the published value of $11.1nM$.

Some experience working with this style of chip made me realize that its consistent successful operation depended on solving a problem of calibration. The number of metering pump cycles required to fill the ring, and to fill MITOMI rows, varied from chip-to-chip, even if I took care to keep constant the pump rate, flow driving pressure, and control pressure variables. This is possibly due to random variation in the heights of the channels, but repeated experiments indicated that I had simply gotten lucky that the chip I had used to make the Pho4p interaction measurements had similar flow characteristics to the chip I had used to test the ability to generate a concentration series, and did not require recalibration.

The technique I had used of visualizing flow with food coloring would not work well for calibrating a chip for an affinity binding experiment. These experiments relied on fluorescence readout, and food coloring contains small-molecule chromophores that would be susceptible to being absorbed into the PDMS where they could interfere with the imaging. Instead I decided to rely on fluorescent dyes with well-defined spectra. When designing an experiment, I could select a calibration dye with an emission spectrum orthogonal to those of the bait and prey labels.

The next affinity I attempted to measure was that of the T7 epitope (MASMTG-GQQMG) to its antibody (Novagen 69968), using the same C-terminal T7-tagged eGFP I'd used for kinetics experiments (Chapter 3). The calibration dye I selected was 5-carboxy-X-rhodamine (ROX), whose emission spectrum does not overlap substantially with that of GFP (Figure 4.6). I determined that I could visualize the chip's flow channels under a 20× objective with a Chroma 41021 filter set ($565\pm15nm$ bandpass excitation, $620\pm30nm$ bandpass emission) when they were filled with $20\mu M$ ROX

in PBS, but that this concentration gave negligible signal in the GFP channel of the TECAN LS Reloaded scanner. One problem I did identify was that once a section of channel had had ROX pumped through it, the fluorescence signal was resistant to clearing by pumping buffer through the channel. This indicated that the small-molecule dye was absorbing into the PDMS. In future experiments I planned to use calibration dye in the form of a DNA oligonucleotide conjugate instead of free dye, since even a relatively short DNA sequence would be large and hydrophilic enough

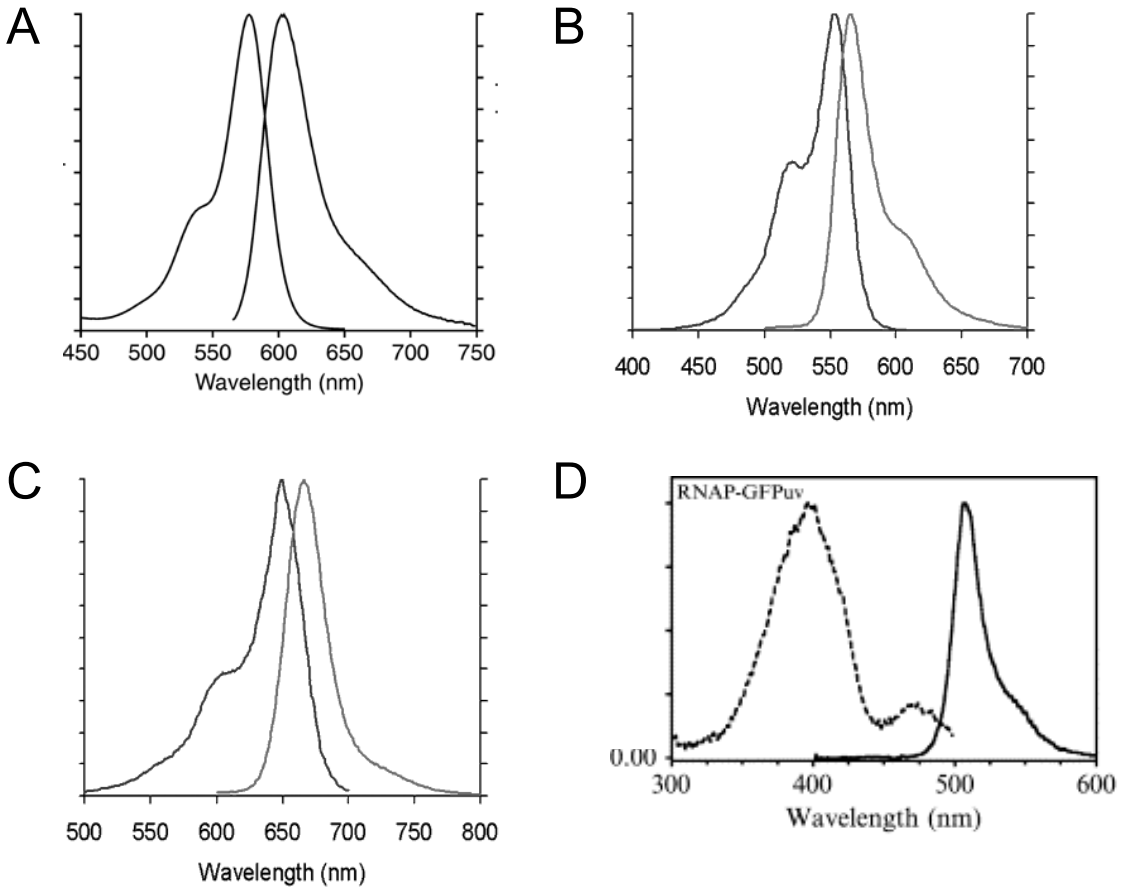

Figure 4.6: Excitation and emission spectra for **A.** ROX, **B.** Cy3, **C.** Cy5, and **D.** the GFP used to label RNA polymerase, which has its excitation spectrum shifted into the UV compared to standard GFP, with its emission spectrum relatively unchanged. **A-C** spectra from Molecular Probes, **D** reproduced from Cabrera and Jin[43].

to avoid absorption.

Nevertheless, because of the lack of ROX interference with the GFP channel, I could go ahead with the experiment using free ROX for calibration. By this time I was using the mixing ring design with $8940\mu m$ inner radius and $9060\mu m$ outer radius. I was also making sure to run the chip in ring entry mode for all surface chemistry steps through the second biotinylated BSA step, to ensure that the floor of the ring was chemically blocked in preparation for the protein that would be pumped through it. For the T7-epitope antibody deposition, I switched to ring bypass mode.

I had problems with the pump performance this time, and I had to experiment with

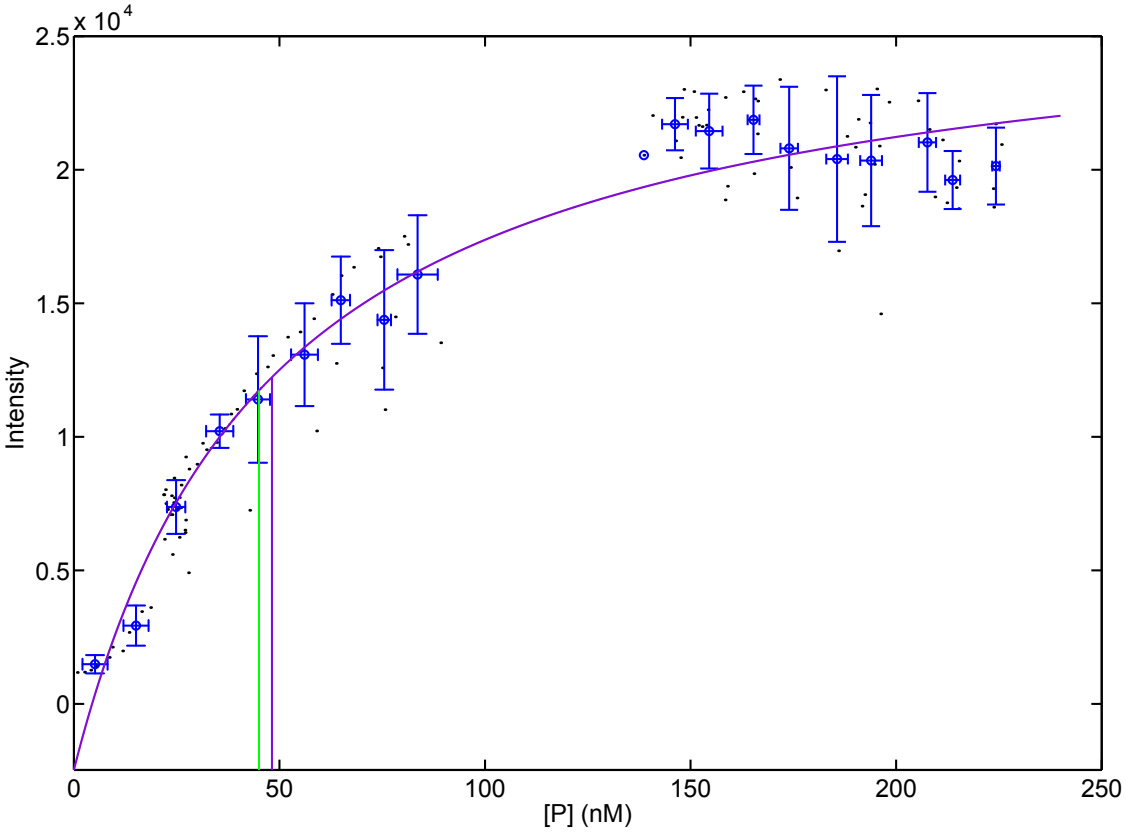

Figure 4.7: Binding curve from T7-epitope/antibody affinity experiment. For this experiment I used $10nM$ bins instead of $5nM$. Bin averages and standard deviation error bars in blue. Fit to bin averages and fit K_d in violet. Canonical K_d in green.

lowering the pumping rate to $30Hz$ for part of the concentration series generation. In addition, I realized too late that during the antibody deposition step an air bubble had been blocking flow to one of the rows, rendering the row useless. These issues prevented me from taking advantage of the full MITOMI matrix, and I only ended up making use of MITOMI unit cells in four of the rows. One thing I noticed that would be correctible in the long run is that even if all address valves were closed during mixing, the force of the large ring pump valves would push prey through, with some making it as far as the first few unit cells in certain rows. In future generations of this design, I widened the outermost address valves to $200\mu m$ in order to prevent this leakage.

In any case, I was still able to obtain enough data points to fit a binding curve for the interaction (Figure 4.7). The K_d calculated from this experiment was $48.2nM$, which matched well the $45nM$ value I had calculated from the measured k_{off} and k_{on}[41]. Rather than trying to do any better by redoing the experiment to get more data points, it was more worthwhile to move on from proofs of principle and attempt to measure the unknown affinities of the RNA polymerase interactions I had discovered.

An affinity experiment with baits expressed on-chip presented additional challenges compared to an experiment in which the bait was introduced from outside. One hidden assumption of my use of Equation 4.3 up to this point is that the spots under the buttons are uniformly loaded with bait. This is more likely to be true if the bait is flowed over the spots from off-chip, since the spots will probably saturate. Nevertheless, some variation in bait loading likely contributed to experimental noise in the proof of principle experiments. My experience measuring on-chip expression levels demonstrated that there could be substantially high variance in spot loading, even just comparing spots with the same bait (see Figure 2.2). Data points on a true binding curve have as their ordinates the fractional occupancy, the percentage

of bait molecules on the surface with a prey molecule bound. Although I would fit the data to Equation 4.3 as before, I would have to be careful to interpret F as the prey binding signal normalized to the amount of bait.

I selected seven RNA polymerase binding partners for affinity characterization: araC, kdgR, lrp, marB, nusG, rpoA, and rpoS. I designed templates for expressing N-termminal c-Myc-, C-terminal 6-His-tagged versions of the baits, and arrayed them in a repeating pattern that included an empty spot as a negative control, repeating six times per MITOMI row, for a total of 48 replicates of each gene.

I proceeded with the surface chemistry steps as in the T7-epitope affinity experiment, and after depositing penta-His antibody under the buttons I introduced the ITT mix in ring bypass mode, flooded the DNA chambers, and expressed for $\sim 260 min.$ Since I had to use Cy3-conjugated antibody to measure bait expression levels, I needed a redder calibration dye that would not interfere, so I used a Cy5-conjugated DNA 15mer, and a Chroma 41019 filter set ($615 \pm 22.5 nm$ bandpass excitation, $679 \pm 30 nm$ bandpass emission) for visualization. The prey loaded onto the chip consisted of $67 \mu M$ calibration oligo and $700 nM$ RNA polymerase holoenzyme.

I determined that filling the ring corresponded to $\sim$840 metering pump cycles of prey, and, with N designating the MITOMI row number, I generated the dilutions for N =2-7 by metering $(N-1) \times 120$ cycles of prey into the ring and mixing for a total of 4000 mixing cylces, alternating mixing direction for every 1000 cycles. I left Row 1 unloaded, and completed the process by loading undiluted prey into Row 8.

After opening the buttons and neck valve and closing the sandwich valves, I equilibrated for $\sim 64 min.$ then closed buttons and neck and opened sandwich valves again and imaged the chip. In ring bypass mode, I washed the excess prey out with PBS and imaged again, then loaded the MITOMI matrix with undiluted Cy3-conjugated c-Myc antibody and equilibrated again, for $\sim 76 min.$ Finally I imaged the chip twice more, pre- and post-wash.

I used undiluted antibody to measure the bait expression instead of the 1:33$\frac{1}{3}$ or 1:50 dilutions I had used in the RNAP interaction screens. This was important because unlike in those experiments, I needed quantitative bait labeling. The antibody concentration in the tube is listed as 0.5-2$\frac{mg}{ml}$, or as little as 3$\frac{1}{3}\mu M$ using the typical IgG molecular weight of $150kDa$. From the kinetics experiments (Chapter 3) I knew the affinity of the antibody for its epitope tag to be $\sim 195\mu M$. Any significant dilution would potentially take the antibody concentration out of the $[Ab] \gg K_d$ regime and into a range where the antibody binding could no longer be assumed to be proportional to the amount of bait.

I used the pre-wash image after the first equilibration (i.e. before the introduction of the labeling antibody) to establish local prey concentration for each unit cell, with the same procedure as in the previous affinity experiments. The "raw" signal of prey binding under each button, R, was taken from the post-wash image after the first equilibration. I used the post-wash image after the second equilibration to quantify the amount of bait under each button, with the Cy3 signal under each button, C. As with the conversion of neck fluorescence to local prey concentration, there was a nonzero signal corresponding to no bait, which I estimated by averaging the C's for baitless spots, to get C_{floor}. Then the normalized prey binding signal, $\frac{R}{C-C_{floor}}$, played the role of F when fitting to Equation 4.3.

One potential problem with this procedure is that it implicitly assumes that the amount of bait measured after the introduction of the antibody and subsequent equilibration is the same as was present when the prey binding signals were measured. Since the free bait in the reaction chamber of each unit cell was washed out as part of the process of introducing the antibody, during the second equilibration a new equilibrium between free and bound bait was also presumably established. However, there was still excess free bait trapped in the DNA chamber at the beginning of this equilibration. The assumption is that diffusion reestablished a uniform concentration

across the reaction and DNA chambers, and that this concentration remained high enough that the spot under the button was saturated and there was no significant difference in the amount of bound bait molecules.

This is obviously an imperfect assumption, but the alternative would be to introduce the labeling antibody as part of the prey mixture. In this case, the antibody would be subject to the same dilution series as the prey, with varying concentrations at different positions in the MITOMI matrix. In this case I would not trust that the antibody concentration would remain high enough to effect quantitative labeling.

Unfortunately, even at best, the data for each bait behaved like that for nusG (Figure 4.8). At the higher concentrations there is a rise in normalized signal, but

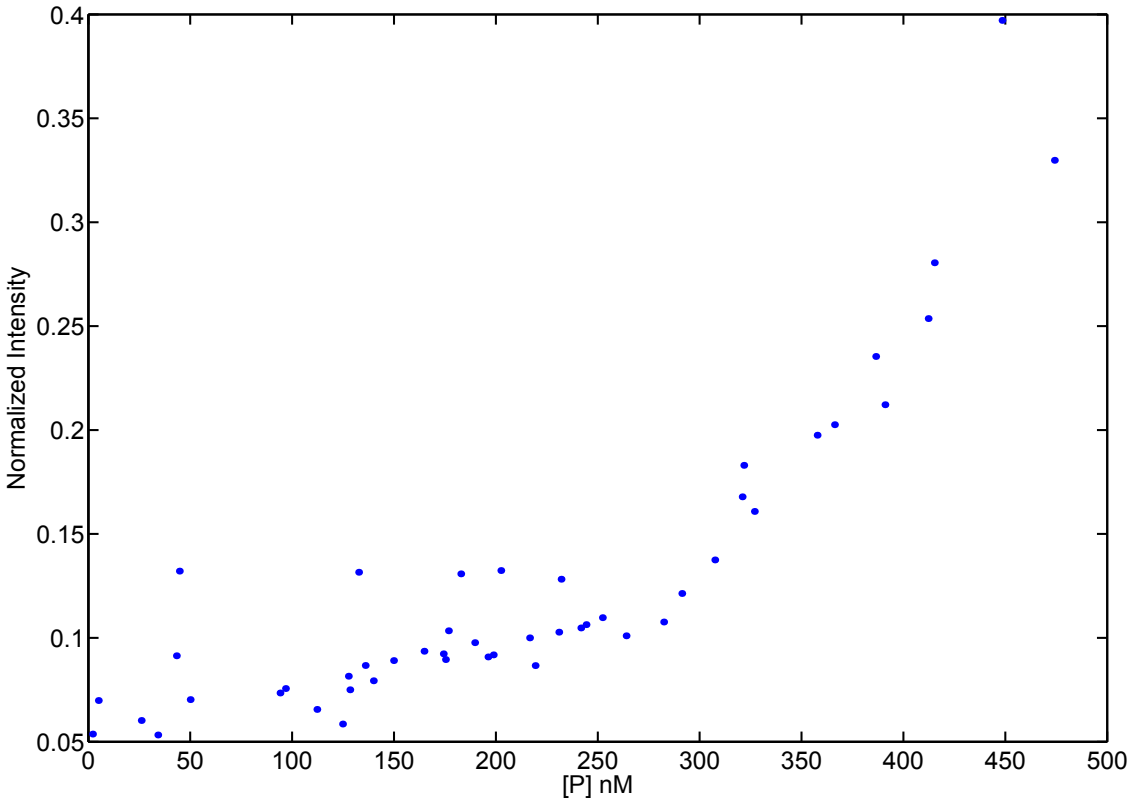

Figure 4.8: Attempted binding curve of nusG interacting with RNA polymerase.

the "curve" never levels off as it should for concentrations substantially higher than the K_d. This is not entirely surprising, since the affinities of biologically relevant interactions would be expected to be in the range of a few μM. However, it is not possible to reasonably fit a binding curve given only data points from concentrations less than the K_d so measuring these affinities are tantalizingly out of reach without a higher concentration of RNA polymerase.

Despite my inability to apply this technique to the interactions I was most interested in, it has potential as a more general method. Situations in which adding a fluorescent tag while maintaining native folding and function presents a substantial protein engineering challenge are not uncommon, particularly in instances such as RNA polymerase where the protein is a multimeric complex. In such cases, the labeled product would be precious, and an approach that would conserve it and put it to the most efficient use would be advantageous.

These proof of principle experiments demonstrate the factors that need to be taken into account to use on-chip mixing for minimization of reagent consumption in MITOMI affinity measurements. First, prepare the chip with the standard MITOMI surface chemistry steps, and express an array of baits on-chip. Next create a dilution series of fluorescently labeled prey, starting with a sufficiently high initial concentration to generate binding curves for the interactions of interest. A minimum of $10\mu M$ should be sufficient for most biologically relevant interactions. For precise calibration, include with the prey a fluorescent dye—conjugated to a large, polar carrier molecule such as DNA to avoid adsorption to the glass and PDMS, and with a spectrum that does not interfere with that of the bait and prey labels—to monitor metering and mixing in real time. After equilibration of the interactions, imaging, and if necessary, bait label introduction followed by another round of equilibration and imaging, the binding curves can be calculated by fitting to Equation 4.3, with local prey concentration determined using the discussed calibration procedure and Equation 4.4, and

the fractional prey binding, F, determined by normalizing prey binding signal with bait labeling signal.

There are some further improvements that could be made to the procedure. The address valves could be further widened to better prevent prey leakage from the ring during mixing, and further engineering may be able to achieve higher pumping rates. An alternate bait labeling method would be useful if it could avoid the problems associated with the extra step to introduce bait labeling antibody. My work on this problem, in combination with these additional considerations, lays the groundwork for a successful application of this strategy.

Chapter 5

Conclusion and Future Directions

The results of my work on protein-protein interactions are disappointing. My overarching goal was to comprehensively screen a biologically interesting network of potential interactions, identify the actual interactions, and then quantify the strengths of those interactions by measuring their kinetics or affinities. This would be an ideal way to demonstrate the utility of MITOMI, since there are no comparable systems that can detect interactions in high throughput and quantitatively characterize them using the same platform.

However, my choice of system to focus on was perhaps not the best. As a test subject, RNA polymerase has the advantage of high biological importance, and its multimeric state allows for a "few prey against many bait" experimental design scale of exactly the size I wanted. However, this fact of a multisubunit complex is a double-edged sword. The α, β, and β' subunits are not normally present *in vivo* except in association with one another, so treating them as distinct preys in the MITOMI experiments creates the possibility of "biological false positives" in an interaction screen, comprising interactions that involve subunit surfaces that would be buried in the complex. This is in addition to any false positives intrinsic to MITOMI—for example the "sticky bait" issue; see Figure 1.6—worsening a worrying potential

problem. For this reason, it may have been preferable to focus on σ factor interactions, since σ factors are present in the cell in free solution at significant concentrations[47]. This screen would have included as prey all *E. coli* σ factors—fecI, fliA, rpoD, rpoE, rpoH, rpoN, and rpos—and as a "seven against many" instead of a "four against many" screen, could have maintained a similar scale.

As mentioned in Chapter 1, the underdevelopment of fluorescent labeling techniques has been a major problem. This issue is particularly important for kinetics measurements, which require covalently labeled prey to be produced in large quantities off-chip. Properly, I would have had the help of a protein chemist dedicated to engineering folded, functional preys with preferably small covalent labels. However, I was the only person working on this project, and dealing with a problem of this magnitude—which would be similar in scale to a thesis in its own right—would have been beyond the scope of what I was able to do.

The unaddressed need for labeled prey was the single major cause for my failure to quantitatively characterize the RNA polymerase interactions. An attempted collaboration to produce labeled preys failed due to a lack of interest and competence on the collaborator's part, and the best I was able to do was obtain the previously produced labeled core enzyme and holoenzyme discussed at the beginning of Chapter 4. However, in addition to being of too small a quantity for me to effectively troubleshoot the problems with the measurements, this enzyme was several years old, and may have lost its activity, compounding the problems in measuring binding. Any future attempts to apply the kinetic and affinity measurement techniques I have developed to a protein-protein interaction network must begin by addressing this labeling problem.

Finally, MITOMI has turned out to be a much messier system than hoped. It is not clear that its false-positive and false-negative rates are any better than that of the yeast two-hybrid and mass spectrometry methods that constitute the standards in the field. Meticulous care in designing a MITOMI experiment, in identifying the

sources of bias, and in choosing the best analysis techniques is crucial to obtain useful data. Related to this is the understanding that MITOMI results are best interpreted in the context of independent sources of information, especially orthogonal interaction detection methods and the proteins' functional annotations.

Despite these downsides, MITOMI offers a great deal of flexibility. As discussed in Chapter 1, bait and prey can be independently either expressed on-chip or introduced from off-chip, as is most appropriate. In general, bait and prey are coexpressed on-chip in the context of binary interaction mapping experiments, while bait is expressed on-chip and prey introduced from off-chip in the context of quantitative measurements, but other ways of handling bait and prey introduction may be preferable depending on the needs of the particular experiment. Additionally, the benefits of the arraying of expression templates cannot be overstated. The templates can be simply created in a large scale from a library using universal primers, which also allows for easy programming of epitope tags appropriate for the designed experiment. These tags can be straightforwardly changed for variations on the same experiment. For instance, to help weed out false positives in a mapping experiment, it is useful to redo the screen with the roles of bait and prey switched, which is readily implemented by a switch in tags[48].

There are additional flexibilities. MITOMI, in contrast to yeast two-hybrid and AP-MS, can probe the interactions of membrane-bound proteins by introducing liposomes to the chip to support them[49]. This ability to control the elements present in the proteins' environment on-chip is relevant to cytosolic proteins as well. As seen in Table 3.1, an interaction's strength may be affected by a change in buffer. A class of potential projects would involve studying how a protein's binding properties are altered by systematically decreasing and increasing the buffer's ionic strength, by changing the identity of the ions, or by adding cofactors. Another option I find particularly interesting, entirely unexplored so far, is the possibility of incorporating *in*

vitro methods of post-translational modification. This could be useful, for instance, to map the interaction kinetics of a MAP kinase network, or to study the interactions of glycoproteins. There is also potential in focusing, to the exclusion of the mechanical trapping element, on the MITOMI chips' capability of arraying independent reactions. For instance, an array of expressed proteins or small molecules can be studied for its effects on cells (see Appendix B).

Despite its problems, then, MITOMI continues to have promise as a useful method in proteomics. To best develop this promise, I have strived always to take into consideration MITOMI's place in the context of other methods for studying interactions. In the case of simple binary mapping experiments, this has led to my belief that each method gives a different "picture" of an interaction network, and that results from different methods need to be compared to give the best idea of the "true picture." In the case of kinetic and affinity measurements, on the other hand, I was consciously developing a capability that doesn't exist at all in any comparably high-throughput systems. With this guiding principle, it is my hope that my work lays the groundwork for MITOMI to reach its fullest potential.

Appendix A

Numerical Method for Hypothesis Testing

I developed a Matlab program to perform hypothesis tests computationally, the development and justification of which I will briefly describe in this appendix. I first wrote the program for evaluating data from valve speed characterization experiments (see Chapter 3) when I was trying to increase valve speed to extend the limits of kinetics measurements. Later, when I needed to do hypothesis tests to determine the presence or absence of an interaction based on my RNA polymerase interaction screening experiments (see Chapter 2) I streamlined and automated the program to do the 360 hypothesis tests (one for each bait-prey combination) needed for each experiment.

In all cases I started with the null hypothesis that the two sets of data (measuring either valve speed or fluorescence signal) to be compared come from the same distribution, with the alternative hypothesis that one of the data sets came from a distribution with a greater average than that of the other. The obvious, most conventional option for statistical testing was the Student's t test, but in both experimental cases I was worried about the effect of outliers. Also, I was uncomfortable with

assuming that the data was represented by a t distribution.

The next thing to consider was a nonparametric test, such as the ranksum test, also known as the Wilcoxon ranksum test, the Mann-Whitney test, or by some combination of the terms, such as Mann-Whitney-Wilcoxon test; the terminology is inconsistent. Analogously to the t test, the ranksum test calculates a test statistic from the data, and converts the test statistic to a p value. Starting with two samples to be compared, sample x $(x_1, x_2, ..., x_{N_x})$ and sample y $(y_1, y_2, ..., y_{N_y})$, pool the $N_x + N_y$ values, order the values from least to greatest, and assign a rank from 1 to $N_x + N_y$ to each data value. Now each sample has a set of ranks ($r_{x1}, r_{x2}, ..., r_{xN_x}$ and $r_{y1}, r_{y2}, ..., r_{yN_y}$) associated with it. The test statistic used is simply the sum of the ranks for sample x,

$$T_x = \sum_{i=1}^{N_x} r_{xi} \tag{A.1}$$

. The corresponding sum of the ranks of the other sample, T_y, is an equally valid test statistic. Unlike the t distribution, the distribution used to convert T_x to a p value has no closed form and is defined recursively[50].

In any case, I was able to use the Matlab function ranksum, which simply takes the two samples as arguments and outputs the p value, without the need for an intermediate calculation of T_x or T_y. The only difficulty was that Matlab would only perform a two-tailed ranksum test. Since I was interested in determining if one sample was significantly greater than the other, not merely different, I had to halve the p value given by Matlab, under the assumption that the distribution was symmetric.

However, I was still not entirely satisfied with this solution. Although the ranksum test relaxes many of the assumptions needed for a t test and is more robust to outliers, I did not feel confident that I could rule out hidden assumptions creeping in. If possible, it was preferable to get the p value even more straightforwardly.

Inspired by a website of statistics tutorials by Professor David C. Howell[51], I

realized the appropriate way to look at the problem. If the two samples come from the same distribution (that is, if the experimental difference has no effect on the values measured), then the fact that a given data point is in one sample and not the other is due entirely to chance. Therefore, repeated random reassigment of each data point to the two samples should generate a probability distribution to determine how likely the actual sample assignments were to occur by chance. This completely gets around any parametrizations or intermediate steps to calculate a p value directly.

This is very general; to actually implement such a method I needed to choose a specific metric of comparison. The difference in the two samples' medians ($m = median(y) - median(x)$) seemed the most natural, though there may be situations where something else such as the difference in means would be preferable. Assuming x is the control data set and y the experimental data set, for each trial randomly assign N_x of the $N_x + N_y$ data points to a "simulated control" data set, with the remaining N_y data points constituting a corresponding simulated experimental set. Then calculate the simulated difference of medians and repeat for as many trials as desired to generate a distribution of these simulated differences, M. Finally calculate a p value by determining where the actual observed m falls in this distribution. I was able to straightforwardly implement this in Matlab, in a program I called shuffletest:

```
function p=shuffletest(x,y);
%The way the arguments are ordered, the question being asked is if y
%comes from a distribution with a greater median than x
%A p value less than the significance level answers in the affirmative
rand('state',sum(100*clock)); %seed pseudorandom number generator
n=50000; %number of trials chosen; this defines a minimum calculable p value of 1/n.
Nx=length(x);
Ny=length(y);
m=median(y)-median(x);
z=[x;y]; %the most straightforward way to implement the random assignment
%into simulated x and y data sets is to concatenate x and y into a single set z and for
```

```
%each trial randomly permute the elements of z. The first Nx elements of the permuted
 %z are the simulated x for that trial.
 Nz=length(z);
 s=zeros(Nz,n);
 for i=1:n
    s(1:Nz,i)=z(randperm(Nz)); %each row of s is a simulated z
 end
 xsample(1:Nx,1:n)=s(1:Nx,1:n);
 ysample(1:Ny,1:n)=s(Nx+1:Nz,1:n);
 M=median(ysample)-median(xsample); %M is the distribution of simulated m's
 hist(M,50); %this line and the next three are optional and plot a histogram of the
 %distribution with location of the experimental m marked
 hold on;
 plot(m,0,'.r');
 hold off;
 p=(n+sum(sign(M-m))+sum(not(M-m)))/(2*n);
```

As the example in Figure A.1 illustrates, in contrast to bootstrap methods, the distributions generated by shuffletest are not normal, or even necessarily symmetrical or unimodal. These characteristics do not present any problems since the p value is calculated without needing prior knowledge about the form of the distribution, in effect by numerical integration.

An obvious question is how much difference the choice of statistical test makes. Figure A.2 shows that the three tests I considered give the same answer the majority of the time, as should be expected. The interesting cases are the marginal ones which pass one or two, but not all, of the tests. I considered in greater detail the eleven such corner cases for the rpoD interactions.

My evaluation of each statistical method came down to making judgment calls on whether a given distribution of data values "should" pass a hypothesis test. In the examples shown in Figure A.3, the four interactors that passed only the t test look very clearly to me like they should have been rejected, while the two interactors that failed only the t test, narL and nrdR, clearly should have been accepted. I felt justified in my initial hestitancy about using the t test, since it gives false positives

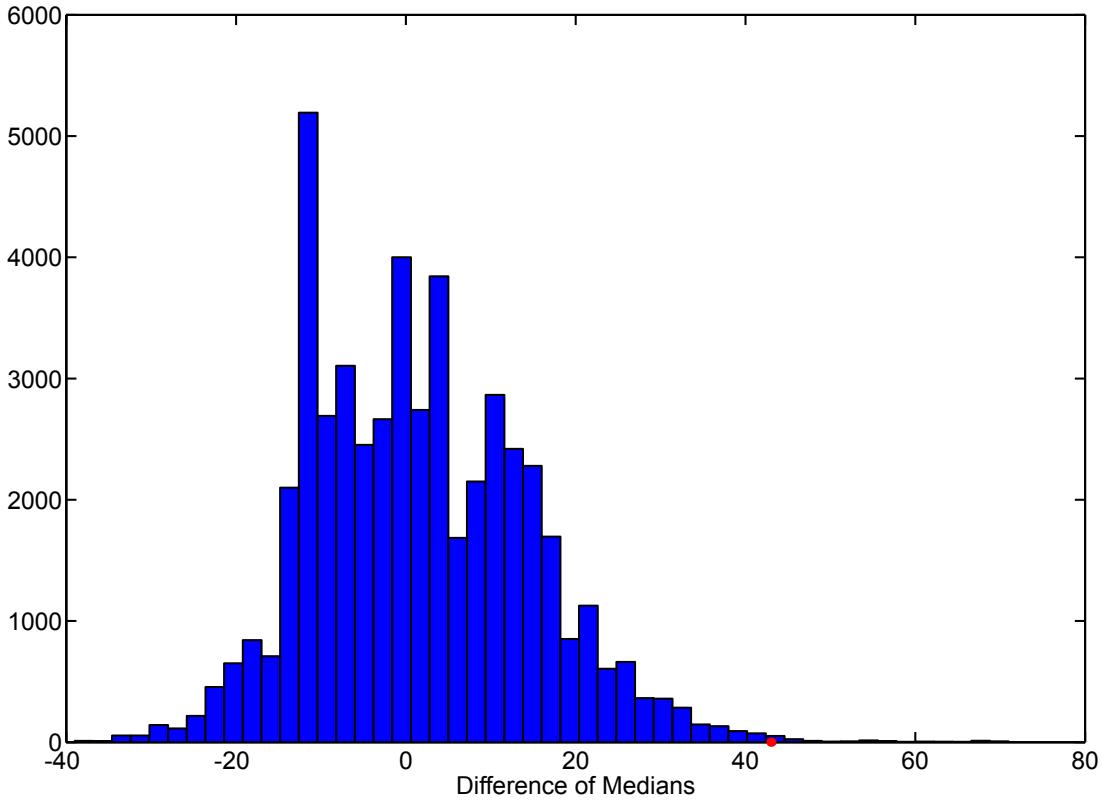

Figure A.1: Example of histogram generated by shuffletest with MITOMI data (rpoD-soxS interaction). The red dot indicates the observed difference in medians.

and false negatives that the other two tests do not.

The choice between the ranksum test and shuffletest was harder and in the context of these examples is equivalent to making a choice between rejecting the interactions with crp, ebgC and rhlE (if I decided on the shuffletest) or rejecting the intereactions with rnr and soxS (if I decided on the ranksum test). Arguably, all five should be accepted, and I considered using both methods before deciding it was best to make a single choice for simplicity's sake. Ultimately I preferred to accept as passing data sets like those of soxS, with a large apparent increase in both variance and average compared to control, instead of data sets like rhlE, with little apparent change in variance and a smaller increase in the average. I used only the shuffletest for any

hypothesis testing from that time forward.

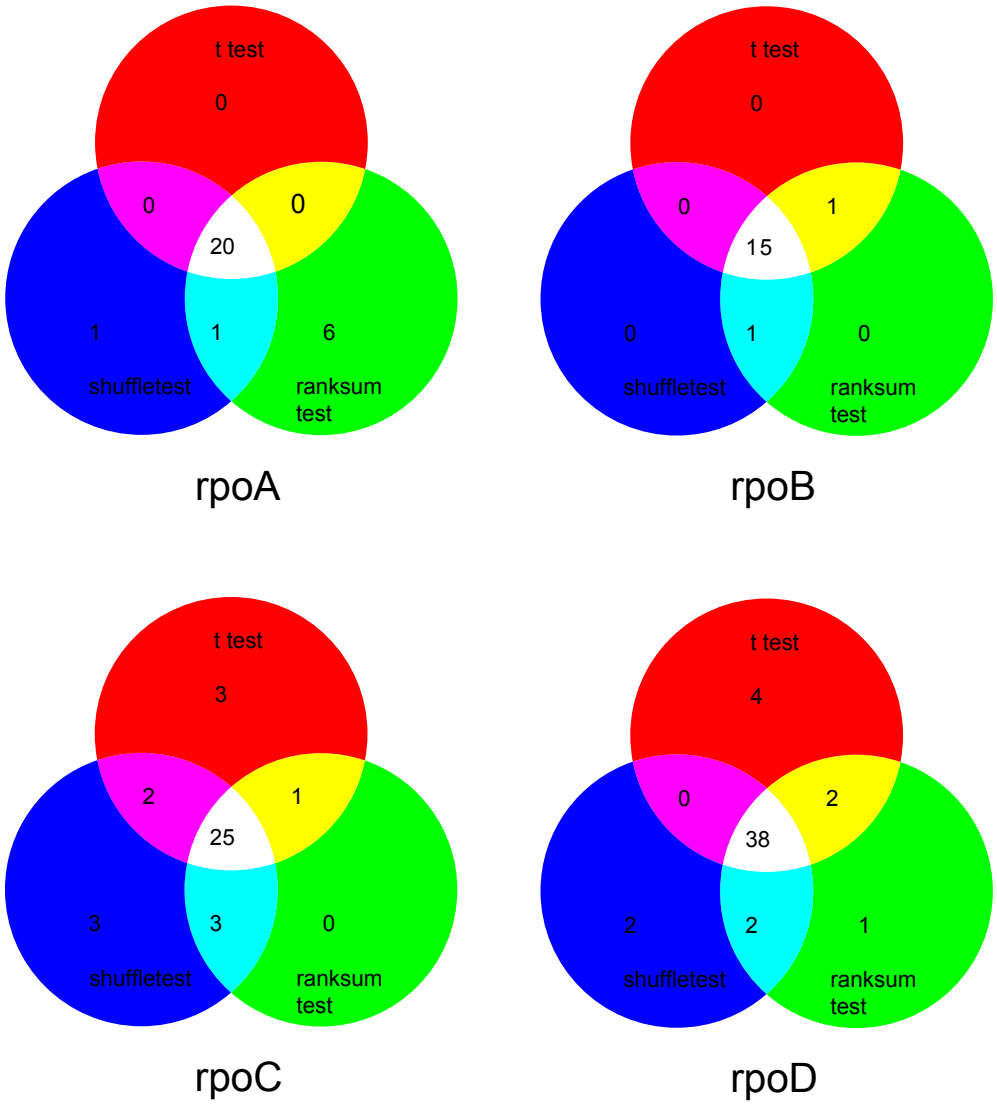

Figure A.2: Number of interactions established at $p \leq 0.05$ level in a single MITOMI experiment, with statistical analysis done by either t test, ranksum test, or shuffletest.

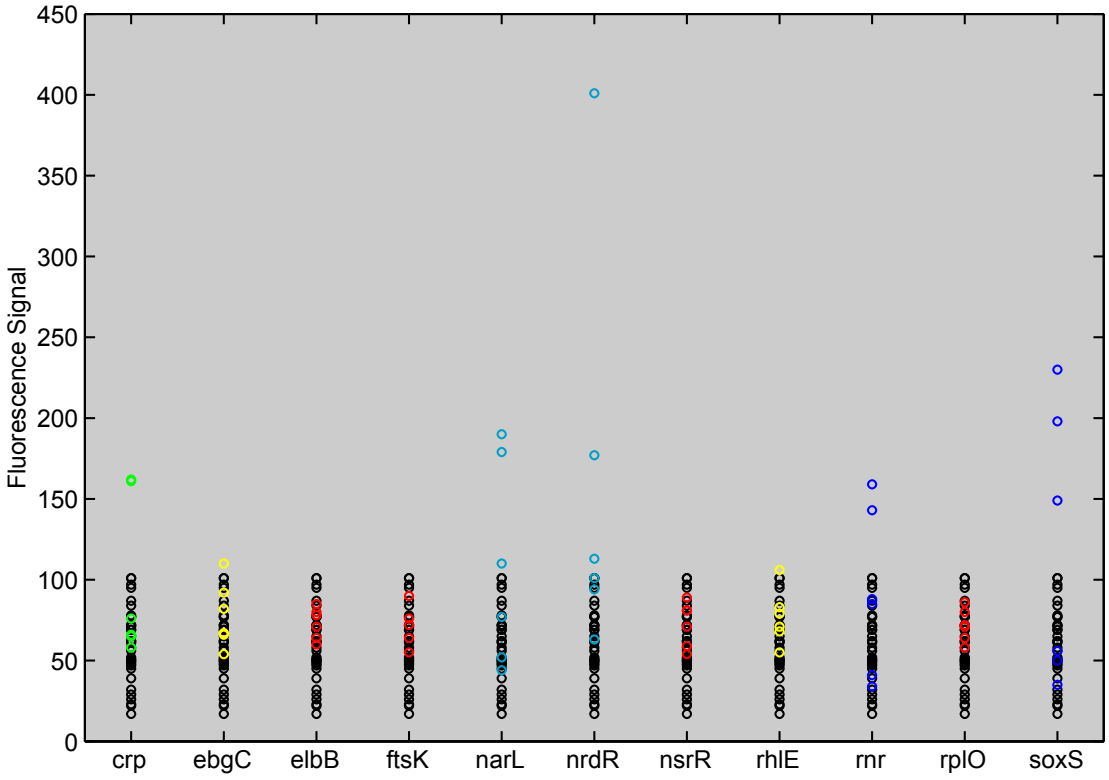

Figure A.3: Raw fluorescence data of eleven selected rpoD interactions. Black circles are from the baitless rpoD spots used as a negative control, and colored circles are from the spots with rpoD and the appropriate bait. Red indicates that the interaction passed only the t test at the $p \leq 0.05$ level, green only the ranksum test, and blue only the shuffletest. Yellow indicates the interaction passed the t test and ranksum test but not the shuffletest, while cyan indicates the interaction passed the ranksum test and shuffletest but not the t test.

Appendix B

Growing Cells on a MITOMI Chip

Aside from its applications in the context of an *in vitro* approach to proteomics, the ability to independently express an array of genes at high density that the MITOMI chip represents has wider potential uses. One side project I tried involved the expression of an array of genes to test for antibiotic activity.

The motivation was the observation that certain genes appeared to be resistant to cloning into *E. coli* in the context of large-scale sequencing projects. The hypothesis was that these genes had antibiotic properties that inhibited the growth of the clones[52]. By using a MITOMI chip, I could express a large set of these genes *in vitro* and determine if they had any antibiotic effect when applied to *E. coli* cells externally.

For these experiments I used a standard MITOMI chip of 16 rows with 40 unit cells each. I determined that it was relatively straightforward to get *E. coli* to grow in the chip. I had a strain of BL21 cells expressing GFP from a plasmid under ampicillin control, making it possible to monitor cell growth fluorescently on-chip (Figure B.1).

I would grow the cells to log phase in LB media, introduce them to the chip, and allow them to grow on-chip by placing the chip's epoxy slide on a hot plate set to $37°C$. I imaged the chip in the GFP channel every $30min$. During the $\sim 2min$. scan

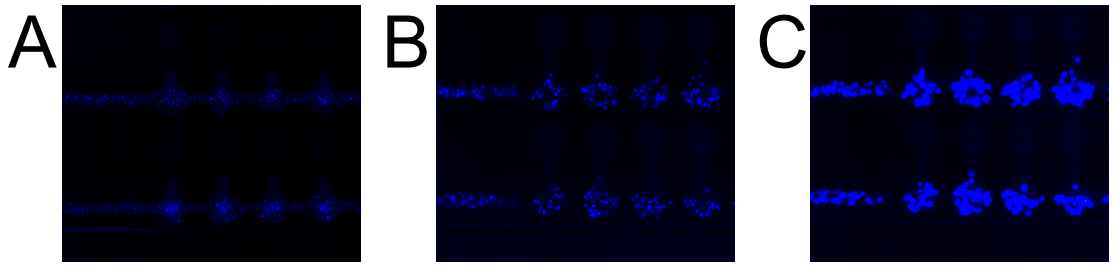

Figure B.1: Images of cells growing in MITOMI chip at **A.** $t = 0$, **B.** $t = 1hr.$, and **C.** $t = 3.5hr.$

time, the chip would have to be removed from the hot plate, but this periodic drop in temperature did not show any evidence of substantially inhibiting cell growth.

I characterized the growth by defining analysis spots in Genepix that covered the

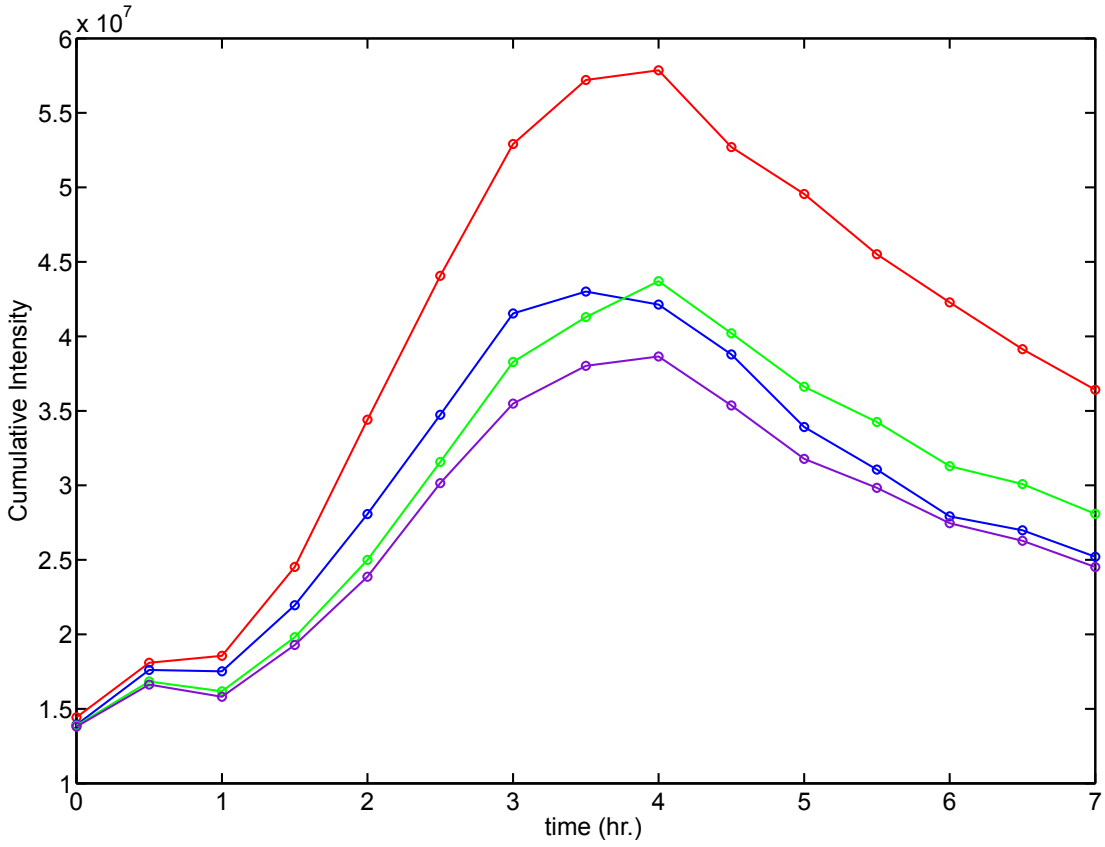

Figure B.2: On-chip cell growth curves.

entire reaction chamber of each MITOMI unit cell. The cumulative signal intensity in each analysis spot was used as a measure of amount of cells. Plotting signal intensity as a function of time showed that the on-chip cell growth was self-limiting. It appeared that after $\sim$3-4hr. the LB media would become depleted, the cells would begin dying, and the fluorescence signal would begin decaying (Figure B.2). However, the exponential growth in the first $\sim$3hr. would suffice as a baseline against which antibiotic activity could be compared.

The genes to be tested were in the form of a clone library filling a 96-well plate, from which I made expression templates with the usual PCR method (see Chapter 1).

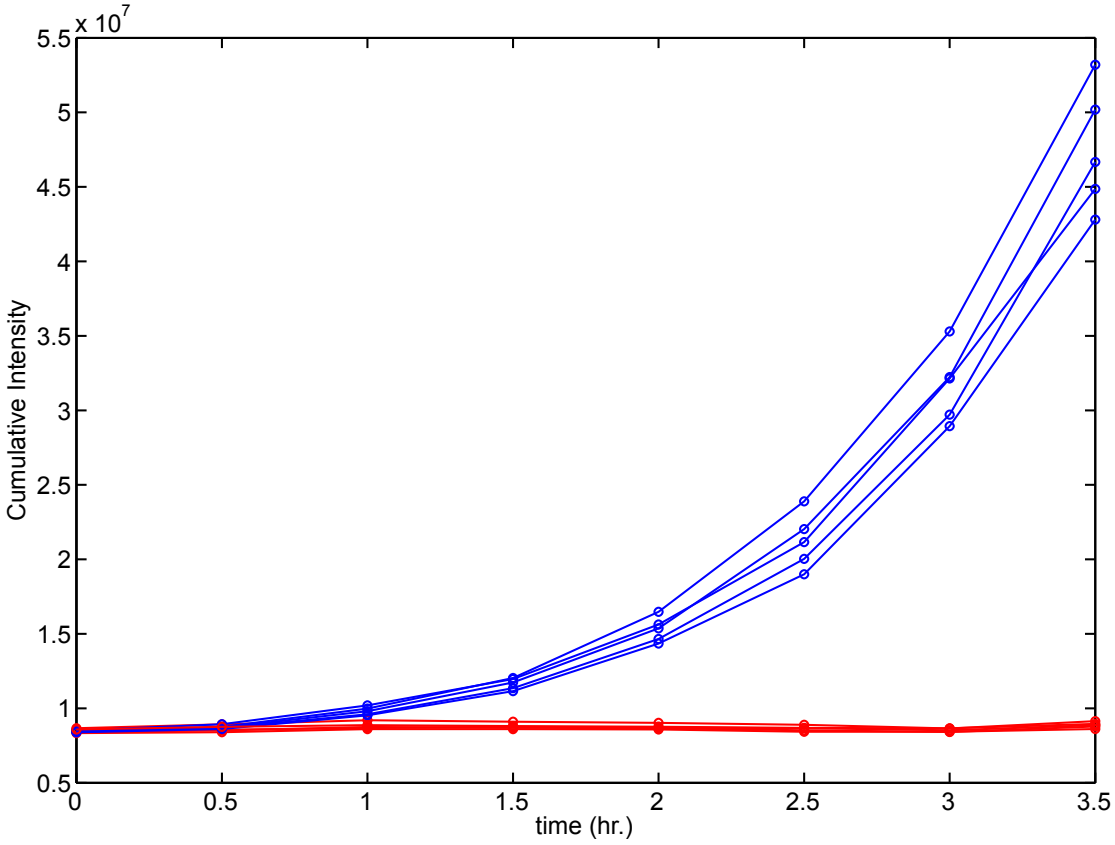

Figure B.3: Antibiotic positive control experiment. Data from unit cells with colicin in red; data from unit cells with blank ITT mix in blue.

Also following the usual interaction screening procedure, I would array the templates onto an epoxy slide and bond a MITOMI chip over the array. I would flow BSA through the matrix to block the floor of the flow channels, and since further surface chemistry steps were unnecessary, I would proceed with a standard on-chip ITT, and then close the neck valves, trapping expressed protein in the DNA chambers. I would then introduce the cells and proceed with cell growth incubation with the sandwich valves closed, isolating the chambers, and the neck valve open, allowing the protein in each DNA chamber to diffuse into the reaction chamber with the cells.

None of the genes from the library showed any significant effect on cell growth. As a positive control, I obtained some template for the antibiotic protein colicin, and expressed it in an off-chip ITT reaction. If I introduced the ITT reaction to DNA chambers on a MITOMI chip, filled the corresponding MITOMI reaction chambers with cells, and allowed the cells to grow with neck valves open and sandwich valves closed, the colicin inhibited growth (Figure B.3). If I substituted a "blank ITT reaction," a reaction that had gone through the expression incubation but lacked a DNA template, the cell growth was unperturbed.

I stopped trying to test the potential antibiotic genes in a microfluidic chip, and instead started expressing the library in a 384 well plate, adding cells, and incubating them in an automated plate reader at $37°C$ with optical density measured every $10min.$. Under these conditions, the test genes continued to have no robust effects on cell growth. Although this particular experiment was not a success, the general technique of introducing cells to a MITOMI chip, where they can grow in the presence of expressed genes to test their effects on growth, may be of wider interest and use.

Bibliography

[1] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[2] Stanley Fields and Ok kyu Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.

[3] G. Traver Hart, Arun K. Ramani, and Edward M. Marcotte. How complete are current yeast and human protein-interaction networks? *Genome Biology*, 7(11):120.1–120.9, 2006.

[4] Barry Causier. Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrometry Reviews*, 23(5):350–367, 2004.

[5] Andreas Bauer and Bernhard Kuster. Affinity purification-mass spectrometry powerful tools for the characterization of protein complexes. *The FEBS Journal*, 270(4):570–578, 2003.

[6] Haiyuan Yu, Pascal Braun, Muhammed A. Yıldırım, Irma Lemmens, Kavitha Venkatesan, Julie Sahalie, Tomoko Hirozane-Kishikawa, Fana Gebreab, Na Li, Nicolas Simonis, Tong Hao, Jean-François Rual, Amélie Dricot, Alexei Vazquez, Ryan R. Murray, Christophe Simon, Leah Tardivo, Stanley Tam, Nenad Svrzikapa, Changyu Fan, Anne-Sophie de Smet, Adriana Motyl, Michael E. Hudson, Juyong Park, Xiaofeng Xin, Michael E. Cusick, Troy Moore, Charlie

Boone, Michael Snyder, Frederick P. Roth, Albert-László Barabási, Jan Tavernier, David E. Hill, and Marc Vidal. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, 2008.

[7] Mark O. Collins and Jyoti S. Choudhary. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Current Opinion in Biotechnology*, 19(4):324–330, 2008.

[8] Marc A. Unger, Hou-Pu Chou, Todd Thorsen, Axel Scherer, and Stephen R. Quake. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science*, 288(5463):113–116, 2000.

[9] Carl L. Hansen, Emmanuel Skordalakes, James M. Berger, and Stephen R. Quake. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16531–16536, 2002.

[10] Frederick K. Balagaddé, Lingchong You, Carl L. Hansen, Frances H. Arnold, and Stephen R. Quake. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science*, 309(5731):137–140, 2005.

[11] Sebastian J. Maerkl and Stephen R. Quake. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, 315(5809):233–237, 2007.

[12] Marta Fernández-Suárez, Hemanta Baruah, Laura Martínez-Hernández, Kathleen T. Xie, Jeremy M. Baskin, Carolyn R. Bertozzi, and Alice Y. Ting. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotechnology*, 25(12):1483–1487, 2007.

[13] B. Albert Griffin, Stephen R. Adams, Jay Jones, and Roger Y. Tsien. Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Methods in Enzymology*, 327(1):565–578, 2000.

[14] Mohammad Arifuzzaman, Maki Maeda, Aya Itoh, Kensaku Nishikata, Chiharu Takita, Rintaro Saito, Takeshi Ara, Kenji Nakahigashi, Hsuan-Cheng Huang, Aki Hirai, Kohei Tsuzuki, Seira Nakamura, Mohammad Altaf-Ul-Amin, Taku Oshima, Tomoya Baba, Natsuko Yamamoto, Tomoyo Kawamura, Tomoko Ioka-Nakamichi, Masanari Kitagawa, Masaru Tomita, Shigehiko Kanaya, Chieko Wada, and Hirotada Mori. Large-scale identification of protein-protein interaction of Escherichia coli K-12. *Genome Research*, 16(5):686–691, 2006.

[15] Sergei Borukhov and Evgeny Nudler. RNA polymerase holoenzyme: structure, function and biological implications. *Current Opinion in Microbiology*, 6(2):93–100, 2003.

[16] Leonid Minakhin, Sechal Bhagat, Adrian Brunning, Elizabeth A. Campbell, Seth A. Darst, Richard H. Ebright, and Konstantin Severinov. Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 98(3):892–897, 2001.

[17] Catherine E. Vrentas, Tamas Gaal, Wilma Ross, Richard H. Ebright, and Richard L. Gourse. Response of RNA polymerase to ppGpp: requirement for the ω subunit and relief of this requirement by DksA. *Genes & Development*, 19(19):2378–2387, 2005.

[18] Gareth Butland, José Manuel Peregrín-Alvarez, Joyce Li, Wehong Yang, Xiaochun Yang, Veronica Canadien, Andrei Starostine, Dawn Richards, Bryan

Beattie, Nevan Krogan, Michael Davey, John Parkinson, Jack Greenblatt, and Andrew Emili. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, 433(7025):531–537, 2005.

[19] Resham D. Kulkarni and Anne O. Summers. MerR Cross-Links to the α, β, and σ^{70} Subunits of RNA Polymerase in the Preinitiation Complex at the *merTP-CAD* Promoter. *Biochemistry*, 38(11):3362–3368, 1999.

[20] Leopold L. Ilag, Lars F. Westblade, Caroline Deshayes, Annie Kolb, Stephen J.W. Busby, and Carol V. Robinson. Mass Spectrometry of *Escherichia coli* RNA Polymerase: Interactions of the Core Enzyme with σ^{70} and Rsd Protein. *Structure*, 12(2):269–275, 2004.

[21] Lars F. Westblade, Leopold L. Ilag, Andrew K. Powell, Annie Kolb, Carol V. Robinson, and Stephen J. W. Busby. Studies of the *Escherichia coli* Rsd–σ^{70} Complex. *Journal of Molecular Biology*, 335(3):685–692, 2004.

[22] Ishita M. Shah and Richard E. Wolf Jr. Novel proteinprotein interaction between *Escherichia coli* SoxS and the DNA Binding Determinant of the RNA Polymerase α Subunit: SoxS Functions as a Co-sigma Factor and Redeploys RNA Polymerase from UPelement-containing Promoters to SoxS-dependent Promoters during Oxidative Stress. *Journal of Molecular Biology*, 343(3):513–532, 2004.

[23] Virgil A. Rhodius and Stephen J. W. Busby. Interactions between Activating Region 3 of the *Escherichia coli* Cyclic AMP Receptor Protein and Region 4 of the RNA Polymerase σ^{70} Subunit: Application of Suppression Genetics. *Journal of Molecular Biology*, 299(2):311–324, 2000.

[24] Anna Lochowska, Roksana Iwanicka-Nowicka, Jolanta Zaim, Malgorzata Witkowska-Zimny, Krystyna Bolewska, and Monika M. Hryniewicz. Identification of activating region (AR) of *Escherichia coli* LysR-type transcription factor CysB and CysB contact site on RNA polymerase alpha subunit at the cysP promoter. *Molecular Microbiology*, 43(3):791–806, 2004.

[25] Susanne Mahren and Volkmar Braun. The FecI Extracytoplasmic-Function Sigma Factor of *Escherichia coli* Interacts with the β' Subunit of RNA Polymerase. *Journal of Bacteriology*, 185(6):1796–1802, 2003.

[26] Sarah M. McLeod, Sarah E. Aiyar, Richard L. Gourse, and Reid C. Johnson. The C-terminal Domains of the RNA Polymerase α Subunits: Contact Site with Fis and Localization during Co-activation with CRP at the *Escherichia coli* proP P2 Promoter. *Journal of Molecular Biology*, 316(3):517–529, 2002.

[27] Miki Jishage and Akira Ishihama. A stationary phase protein in *Escherichia coli* with binding activity to the major σ subunit of RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 95(9):4953–4958, 1998.

[28] K. Derek Weber, Owen D. Vincent, and Patricia J. Kiley. Additional Determinants within *Escherichia coli* FNR Activating Region 1 and RNA Polymerase α Subunit Required for Transcription Activation. *Journal of Bacteriology*, 187(5):1724–1731, 2005.

[29] Michael A. Lonetto, Virgil Rhodius, Karin Lamberg, Patricia Kiley, Stephen Busby, and Carol Gross. Identification of a Contact Site for Different Transcription Activators in Region 4 of the *Escherichia coli* RNA Polymerase σ^{70} Subunit. *Journal of Molecular Biology*, 284(5):1353–1365, 1998.

[30] Mohammad Arifuzzaman, Taku Oshima, Shinsuke Nakade, and Hirotada Mori. Characterization of HscC (Hsc62), homologue of Hsp70 in *Escherichia coli*: over-expression of HscC modulates the activity of house keeping sigma factor σ^{70}. *Genes to Cells*, 7(6):553–566, 2002.

[31] David C. Grainger, Tamara A. Belyaeva, David J. Lee, Eva I. Hyde, and Stephen J. W. Busby. Transcription activation at the *Escherichia coli* melAB promoter: interactions of MelR with the C-terminal domain of the RNA polymerase α subunit. *Molecular Microbiology*, 51(5):1311–1320, 2004.

[32] David C. Grainger, Christine L. Webster, Tamara A. Belyaeva, Eva I. Hyde, and Stephen J. W. Busby. Transcription activation at the *Escherichia coli* melAB promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase σ subunit. *Molecular Microbiology*, 51(5):1297–1309, 2004.

[33] Jason R. Wickstrum and Susan M. Egan. Amino Acid Contacts between Sigma 70 Domain 4 and the Transcription Activators RhaS and RhaR. *Journal of Bacteriology*, 186(18):6277–6285, 2004.

[34] Douglas F. Browning and Stephen J.W. Busby. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1):57–65, 2004.

[35] Shai Kaplan, Anat Bren, Alon Zaslaver, Erez Dekel, and Uri Alon. Diverse Two-Dimensional Input Functions Control Bacterial Sugar Genes. *Molecular Cell*, 29(6):786–792, 2008.

[36] Robert G. Martin, William K. Gillette, Nicholas I. Martin, and Judah L. Rosner. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. *Molecular Microbiology*, 43(2):355–370, 2002.

[37] Richard B. Jones, Andrew Gordus, Jordan A. Krall, and Gavin MacBeath. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439(7073):168–174, 2006.

[38] Gideon Schreiber. Kinetic studies of protein-protein interactions. *Current Opinion in Structural Biology*, 12(1):41–47, 2002.

[39] Sebastian J. Maerkl. *Microfluidic Large Scale Integration and its Application to Systems Biology*. PhD thesis, California Institute of Technology, 2007.

[40] Carl L. Hansen. *Microfluidic Technologies for Structural Biology*. PhD thesis, California Institute of Technology, 2004.

[41] Steven R. Bates and Stephen R. Quake. Highly parallel measurements of interaction kinetic constants with a microfabricated optomechanical device. *Applied Physics Letters*, 95(7):073705–1–073705–3, 2009.

[42] Karl Andersson, Daphne Areskoug, and Emilia Hardenborg. Exploring buffer space for molecular interactions. *Journal of Molecular Recognition*, 12(5):310–315, 1999.

[43] Julio E. Cabrera and Ding Jung Jin. Construction, Purification, and Characterization of *Escherichia coli* RNA Polymerases Tagged with Different Fluorescent Proteins. *Methods in Enzymology*, 370:3–10, 2003.

[44] Daryl S. Spinner, Shaohua Liu, Shao-Wen Wang, and Jakob Schmidt. Interaction of the Myogenic Determination Factor Myogenin with E12 and a DNA Target: Mechanism and Kinetics. *Journal of Molecular Biology*, 317(3):431–435, 2002.

[45] Carl L. Hansen, Morten O. A. Sommer, and Stephen R. Quake. Systematic investigation of protein phase behavior with a microfluidic formulator. *Proceedings of the National Academy of Sciences of the United States of America*, 101(40):14431–14436, 2004.

[46] Hou-Pu Chou, Marc A. Unger, and Stephen R. Quake. A microfabricated rotary pump. *Biomedical Microdevices*, 3(4):323–330, 2001.

[47] Akira Ishihama. Functional Modulation of *Escherichia coli* RNA Polymerase. *Annual Review of Microbiology*, 54(1):499–518, 2000.

[48] Doron Gerber, Sebastian J. Maerkl, and Stephen R. Quake. An in vitro microfluidic approach to generating protein-interaction networks. *Nature Methods*, 6(1):71–74, 2009.

[49] Shirit Einav, Doron Gerber, Paul D. Bryson, Ella H. Sklan, Menashe Elazar, Sebastian J. Maerkl, Jeffrey S. Glenn, and Stephen R. Quake. Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nature Biotechnology*, 26(9):1019–1027, 2008.

[50] Sheldon M. Ross. *Introduction to Probability and Statistics for Scientists and Engineers*. John Wiley & Sons, 1987.

[51] David C. Howell. Randomization test on two independent samples. `http://www.uvm.edu/~dhowell/StatPages/Resampling/Random2Sample/twoindependentsamples.html`, 2001.

[52] Rotem Sorek, Yiwen Zhu, Christopher J. Creevey, M. Pilar Francino, Peer Bork, and Edward M. Rubin. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318(5855):1449–1452, 2007.