

Generating microarrays for protein interaction studies using microfluidic devices

THÈSE N° 5592 (2013)

PRÉSENTÉE LE 11 JANVIER 2013

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE CARACTÉRISATION DU RÉSEAU BIOLOGIQUE
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Sylvie ROCKEL

acceptée sur proposition du jury:

Prof. M. Lutolf, président du jury
Prof. S. Maerkl, directeur de thèse
Dr E. Delamarche, rapporteur
Prof. B. Deplancke, rapporteur
Prof. D. Shore, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013

Develop success from failures.
Discouragement and failure are two of the surest stepping stones to success.
— Dale Carnegie

To my parents. . .

Acknowledgements

Working towards a PhD is a long and challenging journey and I would have not been able to complete this work without the enormous professional and personal support and motivation of so many people, which I would like to thank here.

My first thank goes to Prof. Sebastian Maerkl, who gave me the chance to start working in his lab when it was still in its infancy. Building up the lab equipment and seeing the group grow has been an invaluable experience besides getting first hands-on training one-to-one from Sebastian. Of course, I am also very thankful to all the past and present LBNC members for their support and encouragement, the interesting discussions and wonderful times we had in and outside the lab. You made working in the lab so much more enjoyable: without such a fabulous group spirit it would have been hard for me to pursue this project. In particular, I would like to thank Marcel Geertz for his excellent and patient teaching of fundamental biology, proofreading the paper and thesis, his invaluable advice and his friendship, Nicolas Dénervaud, with whom I had the pleasure to start the lab, who introduced me to EPFL and the French Swiss life and whom I could always ask for help and advice, Luis Miguel Fidalgo for brilliant discussions far beyond research and the fun he brought to the lab, Henrike Niederholtmeier for bringing “female support” in the previously men-driven lab, my bench mate Jose Luis Garcia, Jean-Bernard Nobs for his generous help with programming, Matthew Blackburn, who brews the best beer in the world (says a German!), and Arun Rajkumar for the countless advice and times I could “borrow” lab consumables from him. I also have to thank Helen Chong, who made the administrative questions in the lab so much easier and who always had an open ear.

Special thanks also go to Prof. Bart Deplancke for offering the opportunity to collaborate, and to members of his lab, in particular, Korneel Hens for the excellent collaboration, his patience for answering all my biology questions, Carine (Gubelmann) Delattre and Alina Isakova for the scientific and personal discussions. I very much thank Stefan Kobel not only for his excellent and stimulating scientific (and sometimes not so scientific) discussions but also for his positive attitude, humor and motivation. I am also very grateful to Manuel Bueno from the Biomolecular Screening Facility (BSF) for his assistance with the liquid handling robots and his patience during all the lunch and coffee breaks helping me to improve my French.

Acknowledgements

My greatest thanks also go to Monica Bandera for all her helpful advice and patient listening as a mentor from industry during my last PhD year. I am also very glad to have shared so many good times with member from ACIDE. A big thank you for helping organizing ski weekends and all the fun you brought to it. Special thanks to Tobias Kober, Torsten Mähne, Michela Peisino, Maya Shevlykova, Brammert Ottens, Willem-Jan Zwanenburg not only for their enthusiasm and loyalty but also for their friendship. At this point I would also like to thank the EPFL Toastmaster Club for their great feedback and for helping me to become more confident in public speaking, in particular my mentor Silvana Wasitova, but also Carina Schey, Shashi Bobba, Corinne Spielesoy and again Stefan Kobel. I am forever thankful to Ulrike Lehmann, who introduced me to this club.

Spending more than 4 years in Lausanne I also had the pleasure to meet and get to know many extraordinary personalities, some of which I am happy to call one of my closest friends. Thank you for supporting me in difficult times, not least to mention visiting me countless times in the hospital after my accident in 2010 and at home during the long rehabilitation time. In particular, I would like to thank Sandra Haas for her friendship, her generous help and all the times we went climbing together.

Last but not least, I cannot thank enough my parents and my “adopted” family, who always stood behind me to support and guide me through all my life, no matter how far away I was. Thank you so much! Without you I would not be where I am today.

Lausanne, 10.10.2012

Sylvie Rockel

Abstract

Proteins regulate almost every biological process in living organisms through diverse functional properties. This is achieved by precisely controlled molecular interactions, that act together in a complicated regulatory network. To understand all these complex mechanisms is one of the goals of biology. For this purpose, proteins and their interactions with other molecules have been studied intensively by a variety of techniques *in vitro* as well as *in vivo*. However, information about protein function is still largely missing. While high-throughput screens have been established to identify a large number of protein interaction partners, sensitive techniques are necessary to characterize their molecular and biophysical properties in more detail. Consequently, the quantity and quality of data vary immensely and results are often non-conforming amongst different methodological approaches. Comprehensive characterization of protein interactions thus requires the development of new technologies capable of both, mapping and quantitatively measuring interactions.

This thesis describes the development of alternative approaches to study protein interactions. For this purpose, first two methods to synthesize proteins on a microarray were investigated: Arrays of yeast cells expressing recombinant fusion proteins and expression-ready DNA templates, respectively, were coupled to a microfluidic device with several hundred individually programmable reaction chambers and transformed into protein arrays *in situ*. These protein-on-demand microarrays were assessed thoroughly for protein expression levels, reproducibility and integrity.

Undoubtedly, using genetically engineered cell libraries represented the fastest approach to produce on-chip protein arrays at minimal cost. On the other hand, cell-free protein expression from linear DNA templates showed higher robustness and reproducibility, a preferable criterion for detecting protein interactions under controlled conditions. For this reason, a simple and inexpensive method was developed to generate expression-ready DNA templates directly from cell libraries based on a single-step PCR (polymerase chain reaction). The process was optimized to minimize hands-on time and omit labor as well as resource consuming liquid transfer and purification procedures, a major drawback in many other large-scale protein interaction studies.

Taking cell-free protein expression to a high-throughput level, a DNA template library was con-

Abstract

structed to generate a protein array, containing many of the known *Drosophila melanogaster* transcription factors (TFs). This on-chip TF-array was successfully screened for protein-DNA interactions, technically enabled through an integrated microfluidic platform with a sensitive detection mechanism, which assured highly quantitative data. Detected interactions were characterized in detail using the same platform, which allowed binding affinity and specificity measurements as well as the discovery of new binding motifs. In conclusion, this thesis presents different approaches to generate large-scale protein microarrays, which can be used to identify and characterize protein interactions in high-throughput.

Keywords: protein (micro-)array; protein interactions; TF-DNA interaction; microfluidics; MITOMI; high-throughput; cell-free protein synthesis; cell array; binding affinity; K_d

Zusammenfassung

Proteine regulieren nahezu jeden Ablauf in lebenden Organismen durch verschiedene Funktionen. Dies wird was über unzählige präzise abgestimmte und vernetzte molekulare Interaktionen erreicht. Ein Ziel der Biologie ist daher, diese komplizierten Mechanismen zu verstehen. Daher wurden Proteine und ihr Zusammenspiel mit anderen Molekülen mittels unterschiedlichster Methoden intensiv erforscht, sowohl *in vitro* als auch *in vivo*. Trotzdem sind bis heute nur wenige Funktionen von Proteinen ausreichend aufgeklärt. Zum einen wurden zwar Verfahren entwickelt, um eine große Anzahl von Molekülen zu identifizieren, die an Proteine binden, aber für die Charakterisierung der molekularen und biophysikalischen Eigenschaften dieser Bindung im Detail bedarf es andere, hochempfindliche Techniken. Folglich variieren auch Datenquantität und -qualität enorm und Ergebnisse von methodisch unterschiedlichen Studien stimmen oft nicht überein. Eine umfassende und schlüssige Charakterisierung von Proteininteraktionen erfordert daher die Entwicklung neuer Technologien, die sowohl eine Vielzahl von Wechselwirkung identifizieren als auch quantitativ messen können.

Diese Doktorarbeit befasst sich mit der Entwicklung von alternativen Methoden zur Erforschung von diesen Proteininteraktionen. Dafür wurden zuerst zwei Methoden erforscht, mit denen man Proteine auf einem Microarray synthetisieren kann: Hefezellen, die ein rekombinantes Protein exprimieren, bzw. DNA-Moleküle mit der Information für ein Protein, wurden in einen Mikrofluidik-chip mit mehreren hundert individuell ansteuerbaren Reaktionskammern integriert und anschließend auf dem Chip (*in situ*) zu Proteinen synthetisiert. Diese sogenannten Protein-nach-Bedarf Microarrays wurden eingehend auf ihre Expressionsraten, Reproduzierbarkeit sowie den Proteinzustand getestet.

Zweifelloos ist die Herstellung eines Proteinarrays direkt von genetisch manipulierten Zellen die schnellste und günstigere Variante. Auf der anderen Seite zeigten erste Versuche mit den aus DNA synthetisierten Proteinarrays eine höhere Reproduzierbarkeit und Stabilität, was für die Detektion von Protein Interaktionen unter kontrollierten Bedingungen unerlässlich ist. Aus diesem Grund, wurde weiterhin eine einfache und preiswerte Methode entwickelt, um diese DNA-Moleküle direkt aus Zellen mittels einer einstufigen PCR (Polymerase-Kettenreaktion) zu gewinnen. Die Optimierung der Methode war auf minimale manuelle Arbeit und Kosten ausgerichtet, die oft durch Produkttransfer und Aufreinigungsprozesse in anderen großangelegten

Abstract

Studien dieser Art anfallen.

Um diese Methode auch für einen hohen Durchsatz zu etablieren, wurde ein Proteinarray hergestellt mit einer Vielzahl der identifizierten Transkriptionsfaktoren (TF) von *Drosophila melanogaster*. Mittels eines Mikrofluidik-Chips mit einem empfindlichen Detektionsmechanismus wurden auf dem TF-Array Interaktionen mit DNA-Molekülen detektiert. Die erfassten Interaktionen wurden anschließend mit dem selben Chip detailliert charakterisiert, Stärke sowie Spezifität der Bindung gemessen und spezifische Bindungssequenzen identifiziert.

Im Laufe dieser Doktorarbeit wurden verschiedene Methoden zur Herstellung von Proteinarrays erforscht, die nun dazu dienen können, Proteininteraktionen im großen Maßstab zu identifizieren und quantitativ zu charakterisieren.

Schlüsselwörter: Proteinarray; Proteininteraktionen; TF-DNA Interaktionen; Mikrofluidik; MITOMI; high-throughput screening (HTS); zellfreie Genexpression; Zell-Microarray; Dissoziationskonstante; K_d

Résumé

Les protéines régulent pratiquement tous les processus biologiques des organismes vivants à travers diverses propriétés fonctionnelles. Ceci est rendu possible par le contrôle précis des interactions moléculaires qui forment des réseaux de régulation complexes. Comprendre ces mécanismes est un objectif important en biologie. Pour l'atteindre, les protéines et leurs interactions moléculaires ont été étudiées intensément par différentes techniques *in vitro* et *in vivo*. Cependant, la fonction de ces protéines n'est principalement pas connue à ce jour. Des criblages à haut débit ont permis d'identifier un grand nombre d'interactions. Mais des techniques plus précises sont nécessaires pour caractériser avec plus de détails leurs propriétés moléculaires et biophysiques. En conséquence, la quantité et la qualité des données varient immensément et les résultats ne coïncident pas entre différentes méthodologies. Pour obtenir une meilleure caractérisation des interactions protéinaires, il est donc important de développer de nouvelles technologies, capable de les quantifier et de les cartographier.

Cette thèse décrit le développement d'approches alternatives pour l'étude de ces interactions, basées sur la synthèse de protéines sur micro-puce. Deux méthodes ont été investiguées : des matrices de cellules de levures exprimant des protéines de fusion recombinantes et l'expression de séquence d'ADN *in vitro*. Ces méthodes sont couplées à des dispositifs microfluidiques, comprenant des centaines de chambres de réaction programmables et permettant la génération de matrices de protéines *in situ*. L'expression des protéines, la répétabilité et l'intégrité des micro-puces ont été déterminées de manière approfondie.

L'utilisation de librairie de cellules génétiquement modifiées fut l'application la plus rapide et la moins chère pour produire des micro-puces de protéines. Par contre, la synthèse de protéines à partir de séquences d'ADN démontra une plus grande fiabilité et reproductibilité. Ces critères sont préférables pour obtenir une bonne détection des interactions protéinaires, dans des conditions contrôlées. Pour ces raisons, une méthode simple et bon marché a été développée pour générer des séquences d'ADN prêtes à l'expression, à partir de librairies de cellules et basée sur une PCR (réaction de polymérase en chaîne) à une étape. Ce procédé a été optimisé afin de minimiser l'intervention manuelle de l'expérimentaliste et de limiter la consommation de réactifs et les procédures de purification, qui représentent un désavantage majeur dans les études d'interactions de protéines à large échelle.

Résumé

Une librairie de séquence d'ADN, comprenant les facteurs de transcription (TFs) connus de *Drosophila melanogaster*, a permis de générer des matrices de protéines, en utilisant la synthèse de protéines *in vitro*. Combinée à une plate-forme microfluidique et un système de détection précis, cette micro-puce de TFs a été criblée avec succès, pour identifier les interactions protéine-ADN. Les interactions détectées ont été caractérisées en détail avec la même plate-forme, précisant leur affinité et leur spécificité, et permettant la découverte de nouveau motif de liaison. En conclusion, cette thèse démontre différentes approches pour générer des micro-puces de protéines à large échelle, pour l'identification et la caractérisation systématique des interactions de protéines.

mots clés : micro-puces de protéines ; protein interactions ; interactions protéine-ADN ; microfluidique ; MITOMI ; criblage à haut débit (high-throughput screening, HTS) ; synthèse des protéines *in vitro* ; matrice cellulaire ; constante de dissociation ; K_d

Contents

Acknowledgements	v
Abstract (English/Deutsch)	vii
List of figures	xvi
List of tables	xviii
1 Motivation and Objectives	1
1.1 Motivation	1
1.2 Objectives and Overview	2
2 Introduction	5
2.1 Methods to study protein interactions	6
2.1.1 Two hybrid screens	6
2.1.2 Microarrays	6
2.2 TF-DNA Interactions	9
2.2.1 Experimental methods for the analysis of TF-DNA interactions	9
2.2.2 Gene-centered and TF-centered approach	15
2.2.3 Determining the binding specificity and affinity of TF-DNA interactions .	20
3 Microfluidic devices in biology	25
3.1 Milestones in the development of microfluidics	25
3.2 MITOMI	27
3.2.1 Applications of MITOMI	28
3.2.2 MITOMI versions	29
3.2.3 Modular scale-up of MITOMI	30
4 Yeast cell arrays	33
4.1 Abstract	33
4.2 Introduction	33
4.3 Material and Methods	35
4.3.1 Yeast cell culture	35

Contents

4.3.2	Microfluidic device fabrication an glass slides preparation	36
4.3.3	Spotting of yeast cells and device alignment	36
4.3.4	On-chip experimental procedures for yeast cell arrays	36
4.3.5	Data acquisition and analysis	37
4.3.6	Protein purification (off-chip)	37
4.4	Results and Discussion	39
4.4.1	Spotting efficiency and reproducibility	39
4.4.2	On-chip cell lysis and protein purification	45
4.4.3	Labeling and detection strategies	46
4.4.4	Optimization of induction process	46
4.4.5	Protein immunopurification for flow deposition	50
4.4.6	Co-spotting	52
4.5	Conclusions, potential pitfalls and future directions	52
5	Protein arrays from DNA template	55
5.1	Abstract	55
5.2	Introduction	55
5.3	Material and Methods	57
5.3.1	Linear template generation	57
5.3.2	DNA spotting and device alignment	59
5.3.3	On-chip experiments for linear template arrays	59
5.4	Results and Discussion	60
5.4.1	Template generation	60
5.4.2	On-chip protein expression	63
5.5	Conclusions and Future Directions	65
6	Drosophila transcription factor array	67
6.1	Abstract	67
6.2	Introduction	67
6.3	Material and Methods	69
6.3.1	Plasmid purification	69
6.3.2	Construction of a linear DNA template library	69
6.3.3	Synthesis of Target DNA (Klenow	70
6.3.4	De Bruijn DNA library design and synthesis	70
6.3.5	Microfluidic device fabrication and glass slides preparation	71
6.3.6	DNA spotting and device alignment	71
6.3.7	On-chip experimental procedure	71
6.3.8	Data acquisition and analysis	73
6.4	Results and Discussion	75
6.4.1	Gene-centered approach for the detection of TF-DNA interactions on an on-chip Drosophila TF-array	75

6.4.2	TF-centered approach to characterize TF-DNA interactions	81
6.5	Conclusions	85
6.6	Future Directions	87
7	General discussion and outlook	89
7.1	Limitations and suggestions for the improvement of the current approaches . .	90
7.2	Future applications of microarrays in the field of molecular biology	91
A	MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions	94
A.1	Abstract	94
A.2	Introduction	94
A.3	Materials	96
A.3.1	Mask & Wafer Fabrication	96
A.3.2	MITOMI device fabrication by multilayer soft lithography	96
A.3.3	Epoxy slide preparation	97
A.3.4	DNA synthesis	97
A.3.5	Microarraying / Spotting	99
A.3.6	Microfluidic control elements	99
A.3.7	MITOMI	99
A.3.8	Data acquisition	100
A.4	Methods	100
A.4.1	Mask & Wafer Fabrication	101
A.4.2	MITOMI device fabrication by multilayer soft lithography	103
A.4.3	Glass slide preparation	104
A.4.4	DNA synthesis	105
A.4.5	Microarraying / Spotting	107
A.4.6	On-chip experiment (surface chemistry & MITOMI)	108
A.4.7	Data acquisition & analysis	110
A.5	Notes	111
B	Primer library	112
B.1	Primers for Drosophila TF-array	112
B.2	Primers to generate linear templates from yeast cells	113
C	Protein Purification	114
C.1	Buffers	114
C.2	SDS-PAGE	115
C.2.1	SDS gels	115
C.2.2	SDS PAGE - Buffer System (Laemmli)	116

Contents

D Antibodies	117
D.1 Primary antibodies	117
D.2 Secondary antibodies	118
E Plasmid Purification Protocols	119
E.1 Plasmid DNA Purification using the QIAprep Spin Miniprep Kit	119
E.2 Plasmid DNA Purification using the QIAprep 96 Turbo Miniprep Kit	120
Bibliography	121
Glossary	136
Curriculum Vitae	144

List of Figures

2.1	TF-centered and a gene-centered approach	16
3.1	Schematic of the MITOMI principle	27
3.2	Schematic of a MITOMI chip	28
3.3	Scale-up of MITOMI for large-scale screening applications	31
4.1	Workflow of fabricating a protein array from yeast cells	36
4.2	Spotting tests with yeast cells from the GFP collection with pins of different geometries.	40
4.3	Detection of protein pull-down of 384 yeast GFP-fusion proteins extracted on-chip from yeast cells.	42
4.4	Spotting tests with yeast cells from the GST collection with pins of different geometries.	43
4.5	Spotting test with different blocking reagents	44
4.6	Reproducibility of protein pull-down from frozen arrays	45
4.7	Variability of cell growth between different GST yeast strains	47
4.8	Parameter study to optimize on-chip protein pull-down	48
4.9	On-chip protein pull-down of GST-tagged proteins	49
4.10	Bench-top protein purification	50
5.1	Primer design for one-step colony PCR	61
5.2	Protein expression levels of templates generated by one-step plasmid PCR.	62
5.3	Workflow of cell-free protein synthesis on-chip	63
5.4	On-chip protein expression from DNA templates generated by single-step PCR from lysed yeast cells	64
6.1	Flow chart of selection process of PCR products	70
6.2	PCR performance of generated linear DNA templates	76
6.3	Experimental workflow of the <i>iSLIM</i> approach	77
6.4	Correlation between signal intensities of protein pull-down labeled with a lysine-BODIPY-charged tRNA and a secondary fluorescence GST-antibody	77
6.5	Chip-to-chip reproducibility of protein expression	78

List of Figures

6.6	Distribution of on-chip protein expression	79
6.7	Correlation of protein expression levels and expected protein size	79
6.8	Gene-centered detection TF-DNA interactions on the Drosophila TF-array . .	80
6.9	Concentration-dependent measurements of binding specificity and affinity for selected TFs	82
6.10	Assessment of TF integrity	84
6.11	Determination of DNA binding motifs by de Bruijn analysis	86
A.1	Workflow of a MITOMI experiment	101
A.2	MITOMI chip	104
A.3	Setup of the experiment.	108
A.4	Schematic of the MITOMI principle	110
A.5	Steps to analysis of MITOMI experiments	111

List of Tables

2.1	Commonly used <i>in vitro</i> techniques for analysis of DNA-protein interactions . .	18
2.2	Commonly used <i>in vivo</i> techniques for analysis of DNA-protein interactions . .	19
A.1	Primer sequences used to generate ITT linear templates and target DNA library.	98
B.1	Universal primers for amplification of ORF-GST fragment from pF3A-WG-GST vector	112
B.2	Gene-specific primer sequences of target DNA	112
B.3	Yeast cell MORF-plasmid (BG1805) collection	113
C.1	Buffers for protein purification under native conditions	114
C.2	Buffers for protein purification under denaturing conditions	114
D.1	Primary antibodies for on-chip protein pulldown	117
D.2	Secondary antibodies for on-chip protein detection	118

1 Motivation and Objectives

1.1 Motivation

Molecular interactions occur in biological processes of all living organisms involving numerous molecules at the same time. Many of these interactions take place in the cell nucleus, where they regulate gene expression and have been subject to a variety of investigations to reveal the mechanisms that govern them. As a result we are now able to illustrate complex processes like protein biosynthesis or embryo development.

Key components in these cellular interactions are transcription factors (TFs), a family of proteins that activates or represses genes by binding to specific DNA sequences in response to diverse stimuli. Dysfunction, overrepresentation or alterations in the activity and regulatory specificity of human TFs are responsible for numerous diseases, such as cancer [1] and developmental disorders [2], but are also likely to give rise to phenotypic diversity and evolutionary adaptation [3, 4]. Investigating the mechanisms that govern TF binding and regulation is therefore crucial to understand their role in gene expression. However, despite great interest in research on transcriptional regulation, surprisingly little is known about TFs and their target binding sites, which may be attributed to the complexity associated with studying them: TFs are extremely promiscuous and regulate a variety of genes, including the genes for their own expression. They achieve this alone or together with other proteins in a complex, contributing to an immensely dense network of interactions.

We only begin to uncover the complexity of these interaction networks and systems biology has emerged as an inter-disciplinary field that aims to study the interactions within biological systems from a more holistic perspective. One of its goals is to systematically map all protein interactions in an organism. A considerably more challenging task than cataloguing the identity and abundance of proteins, which was comprehensively done for a vast majority of proteins in model systems such as *Saccharomyces cerevisiae*, *Escherichia coli* or

Drosophila melanogaster. Characterization of protein interactions requires the development of new technologies and the integration of existing techniques, capable of answering complex questions.

Protein and DNA microarrays provide such a multiplexed approach to study molecular interactions in a miniaturized format. Depositing proteins as microarrays and interrogating them with functional assays has been the gold standard to test for binding activity and specificity of proteins in large-scale. This way, the entire yeast proteome was successfully arrayed [5], its protein activity analyzed [6] and binding behavior quantitated [7]. More recently, a similar laborious large-scale effort allowed profiling the human protein-DNA interactome [8].

However, classical protein arrays are extremely labor intensive to produce, have not been integrated with advanced detection mechanisms and the stability of the arrayed proteins is questionable. These difficulties could be largely avoided when using DNA microarrays. Cell-free *E. coli in vitro* transcription/translation (ITT) systems enable the *in situ* synthesis of a protein array from DNA [9]. Generating high-density arrays has been challenging due to diffusional limitations and thus not been applied to map protein interactions. Moreover, traditional arraying techniques missed the detection of interactions with low affinities or very fast off-rates due to loss of specifically bound material during washing.

These limitations have been overcome with the development of a high-throughput microfluidic device based on the principle of mechanically induced trapping of molecular interactions (MITOMI). MITOMI devices allow sensitive measurements of binding energies of transient and low affinity interactions by physical separation and trapping of samples on a glass slide through micromechanical valves. The versatility of this microfluidic approach has been demonstrated in studying a broad range of protein interactions, including those with DNA and RNA, other proteins and small molecules [10, 11, 12, 13].

1.2 Objectives and Overview

Given the potential of MITOMI and the shortcomings of current protein arrays, coupling MITOMI to a new generation of protein arrays would offer a sensitive method to map and characterize protein interaction. Therefore the overall goal of this PhD thesis was the implementation of new experimental approaches to generate a platform to comprehensively study protein interactions in large-scale using a highly integrated microfluidic device. The approach should be fast, inexpensive and reliable for the characterization of protein interactions on a proteomic scale. Particularly, this platform will need to overcome the limitations of current methods applied for interaction studies, which are reviewed in **chapter 2**. It should be able to sensitively map abundant and transient interactions in relatively high throughput, while also

determining binding specificity and affinity.

Chapter 3 of this thesis elaborates on the applications and capacity of existing MITOMI platforms and proposes new scale-up designs to increase the throughput of investigations in a single experiment.

The feasibility of generating a protein array from yeast cells as a first approach to produce an alternative to classical protein arrays is investigated in **chapter 4**. A major focus in this chapter was to develop a protocol for robust array generation and pull-down of on-chip extracted protein using two different types of recombinant yeast fusion libraries.

Chapter 5 describes the generation of a protein array from expression-ready DNA templates obtained from yeast libraries. It includes a new PCR-based method for the fast and low-tech production of expression-ready DNA templates directly from yeast cells.

In **chapter 6**, a high-throughput microfluidic method for integrated systems-level interaction mapping (*iSLIM*) of TF-DNA interactions is presented, which is based on *in situ* synthesis of full length proteins from linear expression templates coding for *Drosophila melanogaster* TFs. A comprehensive approach was developed to demonstrate mapping and characterizing gene regulatory networks with highly accurate, quantitative data. Using a highly integrated microfluidic platform for this approach yielded comparable if not higher detection rates for protein interactions at reduced labor time and cost than other current methods.

Chapter 7 summarizes this work, outlines current limitations and future applications of this microfluidic platform in the field of molecular and cell biology in particular, and gives an outlook for microfluidics in biology related research in general.

2 Introduction

Most biological processes are mediated through the functional properties of protein entities. Key to those properties is often the ability to recognize and bind interaction partners with high selectivity and specificity. Once bound they can accomplish their function, such as catalyzing reactions as enzymes, identifying foreign objects such as bacteria and viruses, mediating signal transduction, activating or repressing DNA transcription or regulating the translation of mRNA into protein. In the nucleus a variety of proteins regulate the crucial process of transcription, including RNA polymerases, histones, chromatin remodeling proteins, general transcription factors, several co-factors, and sequence-specific transcription factors (TFs). Protein interactions are essential for every cellular process and as such have been the focus of many *in vitro* and *in vivo* studies using purified proteins from various sources. Genetic engineering is the state-of-the-art technology to obtain proteins with a tag that facilitates purification. This requires the knowledge of the proteins genetic sequence, which will translate into the protein of interest.

While full-genome sequencing identified a large number of open reading frames (ORFs), information about gene function is still largely missing. Screens for protein function using random complementary DNA (cDNA) libraries and analysis of the biochemical activities of the encoded protein have been traditionally used to shed light on the underlying networks. However, these approaches require tedious protein purifications, a procedure that is difficult to implement on a proteome-wide level. The next section describes common methods for the discovery of protein interactions in high-throughput, including two-hybrid screens (section 2.1.1), classical protein microarrays (section 2.1.2.1) as well as more recent technologies of microarray studies (sections 2.1.2.2 and 2.1.2.3).

2.1 Methods to study protein interactions

2.1.1 Two hybrid screens

Large-scale **yeast two-hybrid (Y2H)** screens are frequently used to study protein function in high-throughput by testing for physical interactions between proteins or protein and DNA, respectively. Uetz *et al.* [14] constructed a library of over 6000 *Saccharomyces cerevisiae* transformants, each expressing a different full-length ORF as a fusion to an activation domain (AD), which they mated with 192 yeast DBD hybrids to screen for protein-protein interactions. They complemented this protein array of activation-domain hybrids with a second approach, in which they generated a similar library of all yeast ORFs, this time fused to the Gal4 binding domain. Each DBD hybrid transformant was then mated to the entire ORF-AD library (5,341 clones) and screened for protein interactions. Despite their unprecedented high-throughput they managed to detect only a fraction of the yeast ORFs in protein-protein interactions, which overlapped only marginally with those detected in another large-scale high-throughput Y2H study by Ito *et al.* [15]. Considering the high false positive/negative rate, dramatic variability and low reproducibility of Y2H systems, the interpretation of the gained results is difficult without further confirmation by other lower throughput assays of higher confidence [16]. A comparative study demonstrated that using a complementary set of multiple Y2H variants generates interaction data of similar quality as five different methods to study protein-protein interactions [17]. Nevertheless, reliably detecting interactions to determine protein function remains challenging and tedious.

The obvious demand for robust platforms to study gene expression and proteins in high-throughput fashion advanced technological developments in other fields and vice versa. **Microarrays**, generated by spotting biomolecules on a solid surface at high spatial density, offer these features and allow the investigation of thousands of targets in parallel.

2.1.2 Microarrays

There are two main types of microarrays: DNA and protein microarrays. *DNA microarrays* were first applied to screen for gene expression profiles [18] before they were used to observe protein interactions, transcriptional regulation or to understand various disease states [19, 20]. However, we cannot easily predict protein expression levels from gene expression, let alone draw conclusion about protein function. To this end, *protein microarrays* found their way into quantitative proteomics through the development of quantitative antibody microarrays [21] and functional protein microarrays [5, 22], respectively. Protein detection is based on immunoassay strategies, but while immunoassays immobilize capture antibodies to the solid support to detect proteins, protein microarrays directly array the proteins.

2.1.2.1 Classical protein microarray

Quantitative protein microarrays, similar to DNA microarrays, measure protein expression levels exploiting the sandwich principle of ELISA: molecules from the sample are captured on the surface with an antibody (or other capture agent) and detected with a second fluorescently labeled antibody which recognizes a different part of the antigen [23]. Quantitative and analytical protein microarrays find applications in biomarker identification (reviewed in [24]), clinical diagnostics [25, 26], or environmental and food safety analysis. Alternatively, in reverse-phase protein microarrays cellular lysate is spotted directly onto the glass slide without the need of a capture antibody and then assayed with a fluorescently labeled antibody [27].

Functional protein microarrays on the other hand probe the activity of a protein against many targets simultaneously. With the development of methods to clone ORFs as fusion proteins into an inducible expression vector it was feasible to express and purify large numbers of proteins, which can then be spotted onto chemically derivatized glass slides using a commercially available contact printing robot. The challenges for that type of array include spotting purified proteins without forfeiting their activity. Functional protein microarrays have been applied to many aspects of discovery-based biology, including protein-protein, protein-lipid, protein-DNA, protein-drug, and protein-peptide interactions as well as examining enzyme activity or substrate specificity.

Among the first who demonstrated the broad applicability of protein microarrays in high-throughput functional assays were the groups of MacBeath and Snyder in 2000. MacBeath *et al.* [5] screened for protein-protein interactions and identified the substrates of protein kinases and protein targets of small molecules by printing proteins with a contact-printing robot on chemically modified glass slides and immersing the sample spots with the respective solutions. Zhu *et al.* took a different approach: a silicon elastomer microwell chip placed onto a microscope slide was used to analyze the biochemical activity of nearly all protein kinases from *Saccharomyces cerevisiae* [22]. Following the printing approach of MacBeath, Snyder and coworkers then printed 5,800 unique yeast proteins and probed these "proteome chips" with labeled liposomes to detect phospholipid-binding proteins [6]. In an advanced study, a phosphorylation map was generated from the results of a proteome chip, which was used to detect the activity and specificity of yeast protein kinases in more than 4000 phosphorylation events [28]. Measuring the binding affinity of protein domains for peptides using protein microarrays uncovered a first quantitative protein interaction network [7, 29]. This mapping approach was extended to the human protein-DNA interactome through the systematic characterization of 4,191 purified human proteins with 460 DNA motifs on a protein array, which identified 17,781 protein-DNA interactions [8].

Although the previous examples demonstrate the potential of protein microarrays for high-

throughput studies, their overall success and feasibility is limited due to challenges in array production and the nature of the detected interactions. Generating protein microarrays is more difficult than DNA microarrays for three reasons:

1. Unlike DNA, which can be amplified by PCR, proteins have to be individually expressed in a suitable cellular host, followed by bench-top purification.
2. The arraying process may alter protein function, conformation, folding, and access to binding sites.
3. Many proteins are unstable or have a very short lifetime when not in their native environments [30].

2.1.2.2 Proteins binding microarray (PBM)

One large-scale, high-throughput approach to acquire information is protein-binding microarrays (**PBMs**), on which double-stranded (ds)DNA oligonucleotides with different sequences are immobilized to glass slides, incubated with TFs or DNA binding domains (DBDs) to saturate the DNA with protein. Subsequently, unbound material is removed in stringent washing steps, bound proteins are detected with fluorescence-labeled antibodies and the binding specificity of the TF is quantified computationally. Developed by Bulyk and coworkers over a decade ago [31], PBMs have greatly increased the throughput for studying protein function by characterizing their binding to DNA, which is addressed below in section 2.2.1.5 relating to TF-DNA interactions.

2.1.2.3 Cell-free protein microarrays

Technical issues associated with conventional protein array production have been partly overcome by *in situ* **protein microarrays** [32, 33, 34]. Cell free expression systems synthesize proteins from expression-ready DNA templates (plasmid or linear DNA) using cell extracts (lysate), which contain all essential components for transcription and translation. Common lysate sources include *Escherichia coli*, wheat germ and rabbit reticulocyte, which are all commercially available as *in vitro* transcription / translation (ITT) systems. Due to their different attributes the system should be chosen considering the protein yield required, protein origin and complexity, downstream processing needs, and cost. Proteins expressed *in situ* typically contain an N- or C-terminal affinity tag for immobilization to a treated glass slide. Constructs of expression-ready DNA can also serve a dual purpose as template for protein synthesis and capture reagent: Chatterjee *et al.* designed an 'on-demand' protein array, where Tus-fusion proteins (Tus...terminus utilization substance) were expressed from plasmid and then bind with high affinity to a 20 bp DNA sequence encoded on the same plasmid [35].

Although, cell-free expression systems facilitate high-throughput studies with proteins, espe-

cially those that are hard to express and purify, high-density arrays still face considerable challenges concerning the integrity of the expressed proteins as well as limitations due to sample diffusion. A drawback of all the previously discussed protein arrays is that many protein interactions tend to show low affinities and/or very fast off-rates and cannot be detected by traditional protein microarray techniques, which involve several washing steps. This problem was recently solved by implementing a novel detection method called mechanically induced trapping of molecular interaction (MITOMI) on a microfluidic device [36], which is explained in more detail in chapter 3.

2.2 TF-DNA Interactions

TFs are proteins that bind to specific DNA sequences via one or more DNA-binding domains (DBDs). They regulate the transcription of genes from DNA to mRNA by either promoting (as an activator) or blocking (as a repressor) the recruitment of RNA polymerase to specific genes. A mechanistic understanding of these transcription processes requires detailed knowledge of binding affinities of all its entities, including their co-operative interactions.

Surprisingly, very little is known of the regulators themselves, let alone their co-operative interactions in a moderate spatial-temporal resolution. Even the presumably trivial question of how many TFs the genome of any given larger organism contains cannot be answered with confidence. Today, the initial estimate of 2,000 to 3,000 sequence-specific DNA-binding human TFs has been corrected to 1,300 - 1,900 TFs, and yet only a third have been annotated in the largest public databases for TFs, Jasper and UniProbe [37]. This discrepancy arises mainly from the limitations that accompany most of the numerous methods, both experimental and computational, which exist to study protein interactions in general including TF-DNA interactions. Existing experimental methods to analyze TF-DNA interactions vary widely in throughput, sensitivity and relative information content per measurement and are summarized for *in vitro* (Table 2.1) and *in vivo* studies (Table 2.2). After a brief description of early methods in the field, more recently developed assays are discussed.

2.2.1 Experimental methods for the analysis of TF-DNA interactions

2.2.1.1 Traditional methods

DNA footprinting was amongst the first methods, which helped to elucidate the binding specificities of proteins to a DNA region of interest. DNA is labeled either with radioactive reagents or fluorophores, incubated with the protein of interest and then subjected to cleavage agents that can degrade DNA, such as DNase I [38] or hydroxyl radicals [39]. Bound DNA regions are protected from cleavage and the resulting DNA fragments generate a typical ladder pattern ("footprint") when separated by gel electrophoresis. The scope of footprinting

applications has been expanded to identify DNA-binding proteins in crude extracts [40] and to determine binding affinities containing multiple binding sites, which are involved in cooperative interactions [41].

Another early method to characterize TF binding sites (TFBS) was electrophoretic mobility shift assay (**EMSA**) [42, 43], which is based on the fact that the mobility of DNA in a gel is slowed when proteins are bound to it and therefore results in a band shift of the product on the gel. Smith and Humphries [44] advanced the conventional protocol with the multiplexed competitor electrophoretic mobility shift assay (MC-EMSA), which uses a series of unlabeled DNA consensus competitors, in combination with a standard electrophoretic mobility shift assay procedure, to identify uncertain DNA-binding proteins. Stenger *et al.* [45] developed a method to identify RNA binding proteins from a dried EMSA gel by mass spectrometry following autoradiography, which increased the sensitivity and accuracy of identified RNA-protein complexes.

Despite its popularity, EMSA is considered a qualitative method and other more quantitative methods have been developed to characterize the binding specificity of TFs to DNA in a less laborious manner.

Most **traditional methods** for DNA-TF binding specificity analysis aim to find target sequences for one TF at a time, often require purified proteins and are generally low-throughput, which made comparison between different experiments difficult.

To overcome these limitations, several methods have been developed with increased throughput and resolution through the incorporation of microarrays and multiplexing; notably *in vitro* selection techniques (section 2.1.3.2), methods based on chromatin immunoprecipitation (ChIP) (section 2.1.3.3), one-hybrid techniques (section 2.1.3.4) and protein microarrays (section 2.1.3.5).

2.2.1.2 *In vitro* selection

One of the first methods to discover and accurately determine TFBS was *in vitro* selection or *in vitro* evolution, a technique developed independently several times in the early 1990s. It is most commonly referred to as **SELEX** (systematic evolution of ligands by exponential enrichment) [46], but also known as CASTing (cyclic amplification and selection of targets) [47] or SAAB (selected and amplified binding sites) [48]. In SELEX purified TFs are allowed to bind to a pool of random DNA oligos, then bound DNA fragments are amplified by PCR, re-incubated with the TF for several cycles until left with the consensus sequence, to which the TF binds with the highest affinity. This selective amplification leads to an overrepresentation of high-affinity binding sites and may exclude other relevant TFBS. One approach to overcome this limitation and to increase throughput, was to couple SELEX to massive parallel

sequencing (SELEX-seq). HT-SELEX requires only one selection round to sufficiently determine consensus sites of enriched DNA fragments and allows a multiplexed analysis approach by indexing samples with different tag-sequences or DNA barcodes [49, 50]. Moreover, using cell extracts Jolma *et al.* circumvented the laborious step of protein purification and limitations related to it. They validated their technology on 19 TFs expressed in mammalian cells from 14 structural classes, showing that it can be used to characterize DNA-binding preferences of proteins that are often excluded from analysis, like proteins that need post-translational modifications or that express poorly [49]. In a similar approach Slattery *et al.* demonstrated the effect of a cofactor on the DNA-binding specificities of all 8 homeobox TFs in *Drosophila*, which all bind to highly related sequences *in vitro*, despite their different functions [51].

Arguably, HT-SELEX represents the method with the highest throughput and the ability to identify much longer high-quality binding profiles (up to 25 bp) than current microarray-based methods at a better cost-efficiency.

2.2.1.3 Methods based on chromatin immunoprecipitation (ChIP)

The biochemical binding specificity of a TF can also be estimated with *in vivo* methods, such as chromatin immunoprecipitation (ChIP) coupled to microarray analysis (**ChIP-chip**) [52, 53, 54, 55] or high-throughput sequencing (**ChIP-seq**) [56, 57]. For ChIP experiments the TF is cross-linked to the DNA *in situ* by formaldehyde fixation and sheared into 100-500 bp fragments. TF-DNA complexes are then precipitated from solution with an antibody specific to the bound TF, before reversing the cross-linking to purify the DNA for either hybridization to DNA microarrays or deep sequencing.

Although ChIP is an extremely important validation tool for *in vitro* methods as it detects full-length TFs in their native environment and locates TFBSs within the genome, it is not a reliable method to measure TF affinities and its accuracy is limited to around 100 bp due to the relatively low resolution of the technique. Another drawback is the requirement of an antibody against the protein of interest, a limitation that was overcome by expressing the protein as a fusion protein with bacterial DNA adenine methyltransferase (**DamID**), which catalyzes methylation of a specific DNA sequence (GATC) upon TF binding [58]. Detection of methylated DNA can be attributed to TF binding nearby since the process of adenosine methylation does not naturally occur in eukaryotes.

DNA immunoprecipitation (**DIP**) [59] is another ChIP-based approach, which borrows elements from *in vitro* selection (SELEX). However, instead of synthetic random DNA oligos it uses purified chromosomal DNA fragments. While genomic DNA is useful for the detection of very long and highly specific TFBSs, which are generally underrepresented in synthetic libraries, DIP is very sensible to analyze shorter TFBSs, which are more evenly covered by

random DNA libraries. Similar to ChIP-Seq the potential to discover new TFBS is limited due to its low resolution.

To increase TFBS resolution of ChIP-seq, a method called **ChIP-exo** was developed to improve ChIP-seq: an exonuclease removes flanking and contaminating DNA from the crosslinked site and bound locations are detected as peak pairs by deep sequencing [60]. Rhee and Pugh applied the method to precisely locate 6,045 transcription pre-initiation complexes (PICs) in *Saccharomyces* [61] as well as to show genome-wide binding of 4 yeast TFs and human transcriptional repressor CTCF (zinc-finger TF) [60]. ChIP-exo can potentially identify novel, low-occupancy binding sites at single nucleotide resolution.

Recently, **Re-ChIP** was described as a sequential immunoprecipitation assay for the identification of multiple proteins on a single DNA sequence, which can determine higher order protein-protein interactions involving chromatin. Re-ChIP is used to analyze multiple, simultaneous, posttranslational modifications to histones to determine the combinatorial pattern of modifications associated with transcriptional status of a gene [62].

2.2.1.4 One-hybrid technologies

One-hybrid systems arose as a variation of the two-hybrid systems. While two-hybrid systems can assess both protein-protein interactions and protein-DNA interactions, one-hybrid systems were developed to specially focus on the latter.

In **yeast one-hybrid (Y1H)** systems a prey protein is fused to the transcriptional activation domain (AD) and direct interaction with the DNA binding domain (DBD) is detected as reporter gene expression [63]. Often a cDNA expression library of DBD sequences is generated and inserted in the promoter region of the reporter gene. Usually, the library is constructed in an *E. coli*-yeast shuttle vector, which produces hybrid proteins consisting of a library protein (prey) and the AD from the yeast GAL4 transcription factor. An increased throughput was achieved in Y2H and Y1H, by combining both systems with the Gateway Technology, where multiple DNA bait sequences were cloned into compatible vectors based on recombination instead of restriction enzymes [64, 65].

The introduction of *E. coli* based methods allowed use of larger libraries (more than 10⁸) due to higher transformation efficiency and faster growth rates. Similar to the yeast based system, **bacteria two-hybrid (B2H)** requires two fusion proteins within the bacterial cell whose interaction stimulates transcription of a reporter gene [66]. The **bacteria one-hybrid version (B1H)** uses a randomized DNA prey library, which are cloned upstream of a positive (HIS3) and negative (URA3) selection marker. The TF (bait) is expressed on a second plasmid as a fusion to subunit of bacterial RNA polymerase. Upon TF binding to the target region (prey), the RNA polymerase is recruited to the promoter and activates the transcription of the downstream reporter genes. Negative selection allows the elimination/removal of

sequences that drive marker expression in the absence of the TF fusion protein. Positively selected clones can be analyzed and sequenced [67]. B1H systems show lower false positive rates than their yeast-based counterparts, allow studies on proteins that would be toxic to yeast and they offer a low-tech alternative to microarray-based technologies.

Several large-scale, gene-centered approaches have used B1H and Y1H systems to characterize TF-DNA interactions in *Drosophila melanogaster* [68], *C. elegans* [69, 70], humans [71], and even plants [72].

One- and two hybrid-systems have significant advantages over other methods that investigate protein-DNA interactions, notably that they do not require specific antibodies and purification of proteins and proteins can bind DNA as monomers and in complexes. However, multiple drawbacks outweigh these advantages and generally, the results are of lower quality and resolution than those acquired with other methods [73].

Coupling Y1H [70, 74] and B1H [75] systems, to high throughput sequencing platforms clearly reduced turn around times and their results indicate that they are useful to uncover previously unknown interactions with putative biological importance. Nonetheless, they also show that other complementary techniques are required for validation and qualitative results.

2.2.1.5 Microarray-related methods

Protein binding microarrays (**PBM**s) have greatly increased the throughput for assessing the *in vitro* binding specificities of TFs to double-stranded DNA microarrays. Taking advantage of existing DNA array technology (here Affymetrix) used for gene expression profiling, Bulyk *et al.* converted a single-stranded into a double-stranded oligonucleotide array [31]. The dsDNA array is incubated with purified epitope-tagged TF DBDs. After stringent washing processes specific binding events are quantified using fluorophore-labeled antibodies. The protein-DNA specificity is determined by assessing the fluorescence intensity of significantly bound DNA spots resulting in binding profiles. In the first-genome-scale PBM study, three yeast TFs were examined for binding to a whole-genome yeast intergenic PCR amplicon microarray, which resulted in the identification of many new binding site motifs [76]. Microarrays constructed with longer DNAs from PCR amplification of genomic regions allows covering a much larger sequence space in fewer spots and representing TF-binding sites with their native genomic flanking sequences.

Based on the mathematical concept of the *De Bruijn* sequence, Bulyk and coworkers [77] designed a maximally compact PBM, which represented all possible DNA sequence variants of a given length k (all k -mers, here $k=10$), covering all permutations of a 10-mer sequence. Given that there are 4^{10} possible 10 bp binding sites, this would give rise to 1,048,576 probes. However, computational segmenting this De Bruijn sequence into 26 distinct, overlapping

10-mer subsequences allowed them to print a ~44,000 single-stranded feature array, which contains more than a million probes of 10 bp. The design also represents sequence variants longer than k as well as gapped k -mers. Therefore, TFs with TFBS longer and shorter than 10 bp can be detected, while TFBS shorter than 10 bp provide an internal control as they are represented several times on the array. In several high-throughput studies, the potential of PBMs was demonstrated by determining the binding specificities for TFs of different structural classes from yeast, worm, mouse and human [77, 78, 79, 80, 81].

Another DNA array approach designed duplex DNA sequences as self-complementary palindromes that form dsDNA hairpins to profile DNA-binding molecules for their relative binding preferences [82]. Coupling this **Cognate Site Identifier (CSI)** microarray platform to DNase I footprinting gave rise to their binding affinities [83].

Recently, a conceptually simple high-throughput sequencing procedure was implemented on a converted Illumina Genome sequencing instrument to directly measure affinity landscape between ~100 million DNA clusters on a PBM and TFs in a microfluidic flow cell [84]. **HiTS-FLIP** (high-throughput sequencing - fluorescent ligand interaction profiling) allows the simultaneous measurement of hetero- and homodimer forms in the same experiment. Given the throughput and depth of this technology (up to ~440 million binding measurement per experiment), HiTS-FLIP is expected to add substantially to the understanding of gene regulation through the contribution of comprehensive DNA affinities.

PBMs and their variations are undoubtedly an important technological development as they offer a rapid, inexpensive, unbiased high-throughput *in vitro* approach to determine binding specificities of TFs without prior knowledge of the conditions in which a TF binds its genomics sites. However, PBMs as well as traditional protein arrays all have an intrinsic problem: They fail to detect interactions with low affinities and/or very fast dissociation rates due to their stringent washing procedure at the end.

2.2.1.6 Surface-plasmon resonance (SPR)

Surface plasmon resonance is an optical technique to study the interaction between an immobilized molecule and an analyte in solution, which can be measured as a change in the reflection angle of light hitting a surface submerged in the solution upon an increase of the surface thickness, such as after analyte binding. For DNA-protein interaction studies, either the DNA-molecule is attached, for instance by means of biotinylation, or the protein is immobilized using tags, such as poly-histidine (His) or glutathione-S-transferase (GST). Since the first SPR immunoassay in 1983 [85] the technique has gained much popularity in biomolecular studies in basic research as well as drug screening because of its advantages over other affinity-based assays: Unlike many other immunoassays SPR is a label-free, real-time assay,

which can measure binding affinities and kinetics simultaneously. With regard to quantitative measurements, this procedure has been the gold standard for protein interaction analysis.

SPR has been adapted as an affinity detection technique in proteomics and genomics studies, such as for biomarker profiling, aptamer and antibody selections and has been used as the gold standard to quantitatively measure biomolecular interactions. However, the experimental throughput is relatively low and the setup not yet compatible with high-throughput microarray technology. Bulyk and coworkers developed a hybrid SPR-PBM to determine the influence of cofactors and complex formation on DNA binding specificity by measuring the absolute binding affinities of a TF versus a TF complex to thousands of individual DNA binding sequences [81].

2.2.1.7 MITOMI

To date, the only quantitative high-throughput method, which can identify TFBS and measure their specificity as well as their absolute affinity, is a microfluidic platform, which detects novel interactions based on the principle of mechanically induced trapping of molecular interactions (MITOMI) [36]. Detailed information on fabrication and recent applications can be found in chapter 3.

2.2.2 Gene-centered and TF-centered approach

The physical TF-DNA interaction can be studied by two conceptually different approaches (Figure 2.1): TF-centered (protein-to-DNA) and gene-centered (DNA-to-protein). While TF-centered methods aim to identify genomic DNA fragments to which a certain TF or set of TFs binds to, gene-centered methods look at one or more regulatory DNA sequences to find TFs that interact with these fragments.

Numerous TF-centered approaches exist for measuring TF binding *in vivo* and *in vitro*. ChIP-based methods have become popular for mapping all genomic binding locations of a transcription factor *in vivo* and gather epigenetic information [87]. Amongst the *in vitro* approaches, protein binding microarrays (PBMs) [80, 78], HT-SELEX [88, 49], HiTS-FLIP [84], and MITOMI [36] have been used in different studies to determine a TF's consensus motif, binding specificities, and affinities.

Gene-centered approaches attempt to derive a comprehensive list of TFs, which can bind to a given DNA element. The dominant method for large-scale gene-centered mapping of

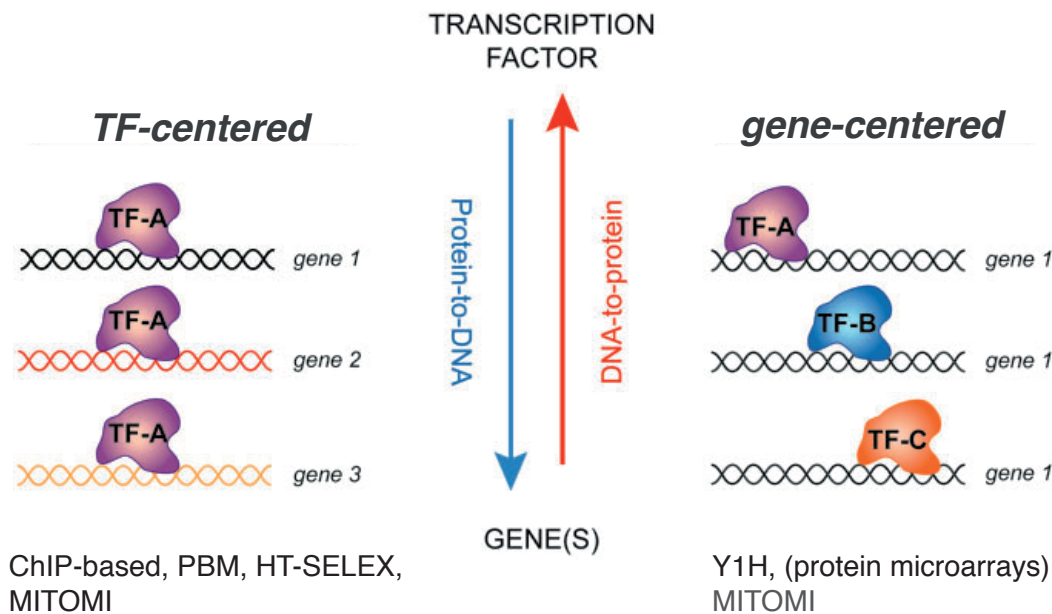


Figure 2.1: TF-DNA interactions can be studied with two approaches: In TF-centered approaches the TF of interest is presented to a wide range of DNA fragments to identify the genetic sequence(s) it binds to. TF-centered *in vitro* approaches such as protein binding microarrays (PBMs), *in vitro* selection techniques (HT-SELEX) and MITOMI are capable of determining the consensus motif, binding specificity and affinity of a TF by quantitative measurements, while *in vivo* techniques are usually limited to map binding locations. Gene-centered mapping of gene regulatory networks (GRN) is predominantly qualitative by identifying the binding preference of various TFs to a given (set of) DNA fragment(s). Currently, only yeast one-hybrid (Y1H) assays are used for this gene-centered approach, but could be equally well feasible on a MITOMI device. Adapted from [86]. ©2009 Oxford University Press

GRNs is based on one-hybrid techniques [89]. Although capable of testing a large number of TFs against a specific DNA element, Y1H approaches have several technical limitations including: (i) a high false positive/negative rate, (ii) intracellular interactions allow no control over the reaction conditions, and (iii) mapping of DNA elements requires ~2 weeks, leading to low turn-around times [74]. Due to some of these reasons Y1H assays have recently been combined with MITOMI for downstream hit validation.

Another approach to gene-centered mapping of GRNs [90] is the use of classical protein arrays, as recently shown by Hu *et al.* who expressed, purified, arrayed, and interrogated 4,191 human proteins [8]. But classical protein arrays remain extremely labor intensive to produce and have not been integrated with advanced detection mechanisms. The development of nucleic acid programmable protein arrays (NAPPA) solved some of the problems associated with generating protein arrays but have not been applied to GRN mapping [9].

Gene-centered and TF-centered approaches are regarded as being highly complementary. However, MITOMI is the only method that could be applied to both approaches on the same platform: TF-DNA interactions are qualitatively mapped by interrogating a protein-array

with specific DNA elements in a gene-centered approach. Protein binding affinity and specificity are then measured using an oligonucleotide array to quantitatively characterize the identified interactions in a TF-centered fashion.

MITOMI is currently the only method which can study gene regulatory networks (GRNs) in two conceptually different ways, from a gene-centered and a TF-centered approach, which will be elaborated in chapter 6.

Chapter 2. Introduction

Table 2.1: Commonly used *in vitro* techniques for analysis of DNA-protein interactions

Assay (synonym)	Method of interaction detection	Advantages	Limitations	Through-put (max. sites)	Resolution	Ref.
EMSA (Electro-phoretic mobility shift assay)	Band shift in gel upon binding of protein	<ul style="list-style-type: none"> - technically simple assay - semi-quantitative analysis - multiplexed competition of large sets of consensus sequences allows for identification of TFs 	<ul style="list-style-type: none"> - low throughput - no immediate information on involved proteins or TFBS - weak or unstable reaction are often not detected 	~10	Few binding sites	[43] [91] [44]
DNA footprinting	Localization of binding site on gel after cleaving of DNA around binding site	<ul style="list-style-type: none"> - simple technical setup - binding affinity measurements by quantitative footprinting - detection of methylation sites with <i>in vivo</i> footprinting 	<ul style="list-style-type: none"> - low throughput - incomplete binding often results in unclear footprint - labor-intensive, time-consuming 	~10	Single-base	[92] [93] [40] [41]
HT-SELEX (High throughput systematic evolution of ligands by exponential enrichment)	Selection of target sequences (often coupled to next generation sequencing)	<ul style="list-style-type: none"> - very high throughput - simultaneous analysis of multiple TFs - TFBS up to 35 bp (longer sites) 	<ul style="list-style-type: none"> - sensitivity: selects for strongest binding (high affinity TFBS) - only relative K_d - PCR and sequencing can introduce bias to some sequences and artefacts - usually requires <i>in vitro</i> synthesis or purification of recombinant proteins 	$> 2 \times 10^5$	Nucleotide resolution	[46] [50] [49] [51]
SPR (Surface plasmon resonance)	Optical measurement of reflection angle change upon binding	<ul style="list-style-type: none"> - absolute affinity measurements (absolute K_d, k_{on}, k_{off}) - real-time recording of association and dissociation constants - label-free 	<ul style="list-style-type: none"> - low throughput - special immobilization required - requires special equipment 	~100	Few binding sites only	[94] [81]
Microarrays						
PBM (Protein-binding microarray)	Intensity readout of proteins bound to ds DNA microarray	<ul style="list-style-type: none"> - very high throughput - determination of relative binding affinities (relative K_d) - microarray equipment available at relatively low cost 	<ul style="list-style-type: none"> - no determination of absolute affinities - stringent washing procedure (removes weakly bound proteins usually purification of proteins) 	~1 million	Nucleotide resolution	[76] [80] [81] [77]
CSI (Cognate Site Identifier)	Intensity readout of proteins bound to ds DNA hairpin microarray	<ul style="list-style-type: none"> - simultaneous measurements of affinity and specificity landscapes - microarray equipment available at relatively low cost 	<ul style="list-style-type: none"> - stringent washing procedure - usually requires <i>in vitro</i> synthesis or purification of recombinant proteins 	~1 million	Nucleotide resolution	[82] [83] [95]
HiTS-FLIP (High-throughput sequencing – fluorescent ligand interaction profiling)	Intensity readout of proteins bound to double-stranded DNA microarray	<ul style="list-style-type: none"> - very high throughput - microarray equipment available at relatively low cost - direct measurement of affinity landscapes - simultaneous measurement of hetero- and homodimer forms - possibility to multiplex 	<ul style="list-style-type: none"> - stringent washing procedure - usually requires <i>in vitro</i> synthesis or purification of recombinant proteins 	~100 million	Nucleotide resolution	[84]
MITOMI (Mechanically induced trapping of molecular interactions)	Mechanical trapping on microfluidic device	<ul style="list-style-type: none"> - detection of low and high-affinity binding sites - determination of relative and absolute binding affinities (abs. K_d, k_{on}, k_{off}) - PWM/motif discovery - gene-centered and TF-centered approach 	<ul style="list-style-type: none"> - requires integration with microfluidic setup (special equipment) - currently only medium throughput 	~4000	Nucleotide-resolution	[36] [10] [13] [96] [11] [97]

2.2. TF-DNA Interactions

Table 2.2: Commonly used *in vivo* techniques for analysis of DNA-protein interactions

Assay (synonym)	Method of interaction detection	Advantages	Limitations	Through-put (max. sites)	Resolution	Ref.
Y1H & B1H (Yeast one-hybrid & bacterial one-hybrid system)	Cloning into & expression of TF in host	<ul style="list-style-type: none"> - high throughput - detection of binding sites - semi-quantitative analysis process - can be automated - no protein purification required 	<ul style="list-style-type: none"> - high false positive/negative rate - intracellular interactions allow no control over reaction conditions - long turn-around times (mating and incubation times) 	All genomic sites	-	[67] [68] [65] [70]
ChIP-based (Chromatin immunoprecipitation)						
ChIP-chip (ChIP coupled to microarray)	Pulldown of cross-linked DNA-protein complexes on a microarray	<ul style="list-style-type: none"> - allows detection of post-translational modifications - high throughput - no protein purification required 	<ul style="list-style-type: none"> - dependent on antibody availability and specificity - low resolution of binding sites - could detect non-specifically bound DNA due to abundance of available DNA sequences 	All genomic sites	100–500 bp	[52] [53] [54] [55]
ChIP-seq (ChIP coupled to high-throughput sequencing)	Pulldown of cross-linked DNA-protein complexes followed by sequencing	<ul style="list-style-type: none"> - allows detection of post-translational modifications - high throughput - no protein purification required 	<ul style="list-style-type: none"> - dependent on antibody availability and specificity - low resolution of binding sites - could detect non-specifically bound DNA due to abundance of available DNA sequences 	All genomic sites	100–500 bp	[56] [57]
DIP-chip (DNA immune-precipitation)	Microarray analysis of enriched DNA fragments	<ul style="list-style-type: none"> - uses purified chromosomal DNA instead of synthetic oligos 	<ul style="list-style-type: none"> - resolution limited by sheared fragment size - DNA purification steps - bias towards high-affinity sites 	All genomic sites	100–500 bp	[59]
DamID (DNA adenine methyltransferase identification)	TF mediated DNA methylation	<ul style="list-style-type: none"> - no antibody required 	<ul style="list-style-type: none"> - resolution limited to distance between 2 methylated binding sites 	All genomic sites	100–500 bp	[58]
Reverse ChIP/ PICh (proteomics of isolated chromatin)	mass spec identification of TF-DNA composition at genomic loci	<ul style="list-style-type: none"> - identification of proteins bound to regulatory sequence based solely upon the identity of that DNA sequence 	<ul style="list-style-type: none"> - low throughput 	1 genomic site	-	[98]
Digital genomic footprinting / DNase-chip	Identification of DNase I hypersensitive sites using DNA tiling arrays	<ul style="list-style-type: none"> - unbiased, genome-wide mapping of protein binding 	<ul style="list-style-type: none"> - resolution limited by sheared fragment size & DNase I hypersensitive sites (~250 bp) - exclusion of repetitive DNA - requires large cell numbers 	All genomic sites	Nucleotide resolution possible	[99] [100] [101]

2.2.3 Determining the binding specificity and affinity of TF-DNA interactions

Many of the new high-throughput methods discussed above rapidly deliver comprehensive information on DNA-protein binding with detailed views of regulatory networks in cells. Identifying the specific DNA sequences and proteins that bind to them elucidates only a fraction of the underlying genomic regulation. Far more insightful are the dynamics, which govern this spatial-temporal regulation of GRNs.

The basis for this understanding lies in the process of specific DNA recognition by a certain amino acid motif in DNA binding proteins, predominantly TFs. The majority of TFs make contacts to the bases exposed in the major groove of the negatively charged DNA via a positively charged alpha-helical segment [102]. Some TFs are dimeric with two alpha-helices: the short helix binds to DNA, while the longer helix mediates dimerization by folding and packing against another helix, which in turn regulates the activity of the TF. However, the specificity of a TF to a DNA sequence arises from its base-specific contact. The strength with which a TF binds to the preferred binding sequence can be quantitated as relative affinity and specificity.

2.2.3.1 Binding affinity

Binding of a TF to a particular DNA sequence (S_i) can be described as a two-state process, which is regulated by an on-rate (k_{on}) for the complex formation and an off-rate (k_{off}) for its dissociation.



The binding affinity of the TF is then defined as the dissociation constant K_d , the concentration of free TF [TF] for which half of the free DNA in solution [S_i] is bound in a TF-DNA complex [TF][S_i].

$$K_{d,i} = \frac{k_{off,i}}{k_{on,i}} = \frac{[TF][S_i]}{[TF \cdot S_i]}; \Delta G^\circ = RT \ln(K_{d,i}) \quad (2.2)$$

Equation (2.2) represents the concentration of all entities at equilibrium, where ΔG° is the Gibbs standard free energy, R the gas constant and T the temperature in Kelvin.

The binding affinity of individual sequences is usually experimentally determined by generating saturation binding curves from either TF or DNA concentration-dependent binding measurements. The dissociation equilibrium constant is calculated by performing a nonlinear

regression fit using a one-site binding model to the data plotted as the ratio of surface bound target DNA (in RFU) to surface bound protein concentration (in RFU) as a function of total target DNA concentration (RFU). The relative K_d (in RFU^{-1}) is then transformed into absolute K_d (in M^{-1}) using a calibration curve, generated by measuring known concentrations of the fluorescence-labeled DNA. K_d values for several hundred TF-DNA interactions can be easily obtained from binding curves measured with a MITOMI device, as is shown in section 6.4.2.1.

In an environment of excess TF, the probability of the sequence S_i being bound to a TF is:

$$P(\text{bound}) = \frac{[TF \cdot S_i]}{[TF \cdot S_i] + [S_i]} = \frac{1}{1 + \frac{1}{K_{d,i}[TF]}} \quad (2.3)$$

However, determining the binding affinity of one TF to a single finely dosed DNA sequence does not give a realistic picture of the circumstances *in vivo*, where each TF is submerged in a pool of highly concentrated DNA sequences at any given time. For a TF to properly regulate gene expression, it is therefore crucial to distinguish its specific regulatory site from the pool of competing non-regulatory sites [73]. Determining the relative binding affinities for all potential binding sites is referred to as specificity of a TF, which can be defined by the sum of all K_d values:

$$\text{Spec} \equiv \sum_{S_i} \frac{K_d(S_i)}{\sum K_d(S_i)} \ln \left(\frac{K_d(S_i)}{\langle K_d(S_i) \rangle} \right) \quad (2.4)$$

2.2.3.2 Representation of binding motifs

With the advances in high-throughput technology described in previous sections (2.2.1.2, 2.2.1.4, 2.2.1.5 and 2.2.1.7) it became much easier to obtain the probability of binding and binding affinities of all possible TFBSs. Nonetheless, having a model to describe and visualize how a TF binds to DNA, facilitates their communication among researchers. Arguably, the simplest approach to display the sequence to which a TF binds with highest affinity is called **consensus sequence**. More quantitative information contains a position weight matrix (**PWM**), where a score is assigned to each of the four possible nucleotide bases at every position in the binding site. The sum of all position-specific scores for each base results in the final PWM score, which is graphically represented as a **sequence logo** [103]. The early basic model of additive PWMs assumes that each base contributes independently to binding. While more complex models with higher-order contributions can undoubtedly determine more accurately the binding site specificity, additive models provide very good approximations in the discovery and prediction of TFBSs in genomic DNA [104]. A PWM is constructed from each TF interacting with the PBM by applying a *Seed-and-Wobble* algorithm, which

first identifies the top-scoring 8-mer (with the greatest PBM enrichment or E-score) as a seed and then systematically tests the relative preference of each nucleotide variant at each position, both within and outside the seed [105]. Using PBMs and a modified algorithm searching for additional secondary binding motifs to assess the binding specificity, Badis *et al.* [80] revealed a complex binding landscape, which suggested that some TFs have a more complicated DNA recognition mode and that most binding profiles can be much better represented by multiple motifs than by a single motif. These assumptions were challenged by a recent method, which derives PWMs from PBMs, termed BEEML-PBM (binding energy estimation by maximum likelihood for PBMs) [106]. The conclusions still remain open but the lively scientific discussion [107] on the topic of DNA binding specificity demonstrates that is crucial to employ optimized analysis methods to maximize the information obtained from PBMs.

2.2.3.3 Complementarity of *in vitro* and *in vivo* studies

The enormous amount of data produced with current high-throughput technologies *in vitro* as well as *in vivo*, has considerably improved our understanding of TF-DNA binding specificity on a systems level and allows us to predict gene regulation. Though *in vitro* methods have fostered the discovery of protein interactions, identification of consensus sequences of TFBSs, measurement of binding energy landscapes and quantification of biophysical properties, they are difficult to translate directly into *in vivo* function. Many *in vitro* approaches cannot account for secondary effects *in vivo* like masking of TFBSs by competing TFs or nucleosomes, the widespread nonfunctional binding, or the presence of other proteins and post-translational modifications, which mediate interactions [108]. Here, *in vivo*-based methods can help to evaluate newly identified TFBSs in the relevant biological context, albeit, at the expense of resolution and identifying the binding causalities. ChIP-based methods are able to indicate the location to which a specific TF binds in different conditions *in vivo*, but the resolution is generally not precise enough to determine the exact TFBS and the specificity of a TF. Moreover, motif discovery algorithms often fail to reveal the accurate TF binding motif from ChIP-data because of the confounding factors that occur *in vivo*. The binding preference and mode of a TF depends very much on the context, like cell type, developmental stage, environmental conditions, and other cooperating or competing binding partners. Inversely, the occupancy of a binding site sheds light on the accessibility of the site and the binding behavior of TFs. While some TFs predominantly bind DNA indirectly through protein partners, others interact directly as well as indirectly. The complexity of parameters that determine each individual binding event in the cell limits the throughput of *in vivo* measurements.

Currently neither *in vivo* nor *in vitro* methods can characterize TF-DNA interactions sufficiently. Despite the abundance of experimental methods, there is still a lack of reliable and coherent data sets on TF-DNA interaction. This bottleneck is compounded by the expensive and laborious analysis of TF-DNA experiments as well as a high rate of false predictions. Thus, in order to gain a quantitative understanding of GRNs, a serial combination or hybrid approach of *in vitro* and *in vivo* measurements combined with computational modeling appears to be the best strategy.

In conclusion, there has been progress in elucidating the different aspects of protein activities on a global scale with various studies by massive parallel screening of the proteome of different organisms, which in concert with biocomputational and modeling efforts contribute towards our current understanding of cellular networks. Nevertheless, the sheer complexity and precision of regulation require ever more sophisticated techniques and ways to analyze them. And certainly, advances in proteomics and other fields are intimately linked with the development of new and improvement of existing technologies, which will help to answer some questions but probably raise even more.

3 Microfluidic devices in biology

Progress in molecular biology in the last two decades, particularly the high-throughput measurements in genomics and proteomics, moved our understanding of biology to a systems-level and gave rise to the new field of Systems Biology around the year 2000. These advances in scale would not have been possible without the development of innovative technologies from interdisciplinary research at the cutting edge of science and engineering.

One of these emerging technologies is microfluidics, which started to grow from the fields of analytical chemistry and microtechnology in the early 1990s with the development of microfluidic devices for capillary electrophoresis [109, 110, 111]. The term microfluidics refers to the science and technology of systems that process or manipulate very small amounts of fluids, exploiting special characteristics of fluids in flow channels at the micro- or even nanoscale. Scaling down existing bench-top assays to a postage-sized footprint allows researchers in biology, chemistry and other fields to perform multiple complex reactions simultaneously with smaller volumes and therefore at lower costs.

3.1 Milestones in the development of microfluidics

Two developments in microfabrication techniques can be seen as milestones in the evolution of integrated microfluidic systems: The first one was the introduction of soft lithography to fabricate replicate structures from a master template using moldable elastomeric polymers, such as polydimethylsiloxane (PDMS). The master templates are typically manufactured using photolithography and etching steps, common techniques in the semi-conductor industry. The second milestone was a simple method for the fabrication of pneumatically activated components by multilayer soft lithography (MSL), developed in the Quake lab [112], which generates devices with two or more stacked, micro-patterned layers. Key component for fluid-handling functionality in these devices is a monolithic micromechanical valve, generated by two actuated crossing channels, which are separated by a thin PDMS membrane between

the two distinct layers and may be deflected up or down, depending on which channel is pressurized.

Incorporating valves allowed for more complexity and the design modularity of PDMS devices fabricated by MSL gave rise to microfluidic functional elements such as fluid input trees, pumps [112], sieve valves [113], multiplexers [114], and freestanding membranes [36]. The intention to integrate and automate all lab processes in chip-format, led to the development of more complex next-generation microfluidic devices, which coined the term microfluidic large-scale integration (MLSI). MLSI devices can integrate thousands of micromechanical valves on a very small footprint, which allows for hundreds of assays to be run in parallel.

The application of microfluidic devices shows obvious advantages that can be attributed to a smaller scale: (i) smaller volumes reduce sample/reagent consumption; (ii) shorter distances decrease mass and heat transport time and thereby reduce reaction times; (iii) denser packaging of microfluidic devices allow higher degrees of parallelization. As a consequence, microfluidics provides novel processes and detection mechanisms that cannot be achieved on larger length-scales.

Owing to its obvious benefits, the field of microfluidics has an ever-increasing impact on all fields of biology. Today, microfluidic devices have been applied to almost all classical biological methods, which can be seen on the growing list of applications:

- Protein crystallography [115]
- PCR [116, 117, 118] and RT-PCR [119]
- Single cell analysis [113, 120]
- ChIP [121, 122], ELISA [123]
- Two-dimensional electrophoresis [124]
- gene transfection [125] and cloning [126]
- separation and purification of DNA [127, 126], proteins [128], and other small particles [129]
- SPR [130, 131]
- *in vitro* transcription/translation (ITT) for protein synthesis from DNA [10, 13]
- high-throughput screening of protein interactions [36, 13], drugs [11], enzymes [132], antibodies [133] and cells [134, 135]
- cell trapping [136, 137], sorting [138, 139], and 3D-growth characterization [140]
- point-of-care diagnostic and clinical assays [141, 142, 140, 143]

Microfluidics has already been incorporated as an important tool in life science research, especially systems biology, and is likely to become a standard practice in the near future as highly integrated microfluidic screening platforms promise to overcome limitations in scaling and high-throughput problems in the research world of “-omics”.

3.2 MITOMI

During the past decade ever more complex microfluidic devices have evolved and their application advanced progress in almost all areas of life science. The development of a high-throughput microfluidic platform with a novel mechanism to detection molecular interactions in 2007 by Maerkl *et al.* introduced a completely new way studying molecular binding events [36]. Microfluidic devices based on mechanically induced trapping of molecular interactions (MITOMI) allow for direct measurement of binding affinities and specificities in relative high-throughput. MITOMI can therefore also capture low affinity and transient interactions at equilibrium, which are usually missed by most other methods that detect binding due to stringent washing processes. Principally, in a MITOMI experiment, a protein is bound to an antibody immobilized to a restricted region on the surface of a glass slide. Interactions with the molecule(s) of interest, such as proteins, DNA or drugs, are trapped by collapsing a deflectable membrane to trap interactions in equilibrium before washing off non-bound material (see Figure 3.1).

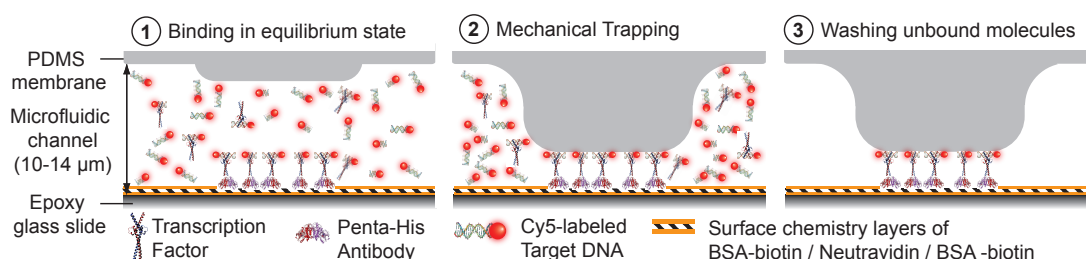


Figure 3.1: Schematic of the MITOMI principle. The gray structure at the top of each panel represents the deflectable button membrane that can be brought into contact with the glass surface. (1) His-tagged TFs are localized to immobilized penta-His antibody at the epoxy-coated glass slide. Specific binding between Cy5-labeled target DNA and TFs are at steady state when (2) the button membrane is actuated and brought into contact with the surface. Any molecules in solution are expelled while surface-bound material is mechanically trapped. (3) Unbound material that was not physically protected is washed away, and the remaining molecules are quantified.

The MITOMI device harbors several hundred to thousands of dumbbell-shaped unit cells (Figure 3.2 A), each individually programmed with samples and controlled by three valves (Figure 3.2 B): the neck or chamber valve separates the sample chamber from the detection area, the sandwich valves compartmentalize adjacent unit cells, and the button membrane is a circular, freestanding membrane hovering over the detection area, which can be collapsed to protect the area underneath (Figure 3.2 E).

In order to program each unit cell with a different sample, microarrays are spotted with small pins within a microarrayer from a multiwell plate. Solutions and samples are introduced via inputs, which are arranged in a tree-like fashion (Figure 3.2 C), each branch controlled by a separate valve. Additional control of fluid control and flexibility in sample choice can be achieved by adding a multiplexer to the inlet tree (Figure 3.2 D).

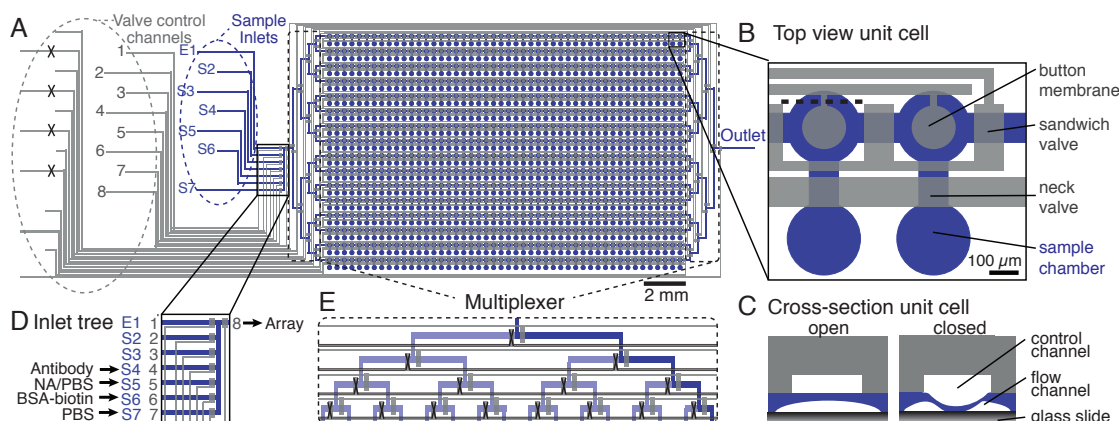


Figure 3.2: Schematic of a MITOMI chip (A) Drawing of a MITOMI device with 768 unit cells in the flow layer (purple) which are controlled (grey) by 2,388 valves. Resistance equalizers toward solution inlets and outlet ensure equal flow velocities and derivatization in each row of the channels. (B) The magnified view shows two individual unit cells, each controlled by three separate micromechanical valves: Each dumbbell-shaped unit cell can host a different sample, contained by the neck valve. Sandwich valves compartmentalize individual unit cells to prevent diffusional cross-contamination of samples during incubation periods. The freestanding button membrane can be collapsed to protect the circular area underneath. (C) A cross-section of a unit cell is shown to illustrate the detection mechanism based on mechanically induced trapping of molecular interactions (MITOMI). The deflectable membrane is pushed down by actuating the water-filled control channel to trap bound molecules. (D) Different sample solutions are loaded via the sample inlet. Loading of the device with samples via the inlets (S2-S7) is controlled by opening and closing the corresponding control valves (1-8). (E) A multiplexer allows individual addressing of each flow channel with additional valves acting on consecutive stages of the flow inlet tree, which separate in two at each conjunction. The clockwise-rotated zoom of the multiplexer illustrates flow through the first channel (dark purple), but not through all the other 15 channels (light purple) on the array after closing a combination of control valves (indicated with a cross).

3.2.1 Applications of MITOMI

The potential of MITOMI has been exploited in a number of studies. In their seminal publication, Maerkl and Quake [36] used 17 MITOMI chips to map the binding energy landscapes of four eukaryotic TFs from the basic helix-loop-helix (bHLH) family by measuring their binding affinities to 464 different DNA variants, which allowed them to test binding specificity hypothesis and predict *in vivo* function of these TFs. In a second publication, the Quake group programmed the device with target RNA sequences instead of DNA and validated the platform for measuring binding constants of interactions between membrane-bound protein and RNA. They expressed the trans-membrane protein *in vitro* using a cell lysate supplemented with microsomal membrane to resemble natural folding conditions. Having determined the binding specificity and affinity of the trans-membrane protein, they screened a compound library of potential target drugs for their effect on the protein and identified inhibitors of protein interactions [11]. Gerber *et al.* [10] expanded the range of possible interaction studies by characterizing all pair-wise interactions between 43 bacterial proteins, applying an on-chip *in vitro* transcription-translation procedure to synthesize proteins *in situ* from a spotted library of expression-ready DNA templates. The scope and sensitivity of MITOMI studies were highlighted in a study on the evolvability of a TF family, where Maerkl

and Quake [12] measured the detailed DNA-binding behavior of all 95 single-point mutations in five DNA-binding sites of a human TF within this TF family. They were able to predict 92% of the naturally occurring diversity at these positions, but also pointed out that not all functional mutations could be found *in vivo*, which led to their conclusion that mutations are not solely selected based on their functionality. In an unpublished study, Maerkl demonstrated the potential of the MITOMI device to generate live yeast cell arrays and yeast protein arrays, respectively.

All the above studies demonstrate the variability and flexibility of MITOMI-based studies. The principal can be applied to screen virtually any molecular interactions as long as the molecules of interest can be solubilized. MITOMI has been recognized in the literature as a reliable and robust method to determine the binding affinity and specificity of protein to DNA [73] and is currently used to validate and complement other detection methods [74, 144, 96, 145].

3.2.2 MITOMI versions

Since the inception of multilayer soft lithography by the Quake lab in 2000 [112] ever more complex fluidic devices were fabricated by MLSI with more and more micromechanical valves on a postage stamp footprint. The fast increase of valve density on microfluidic devices has even outpaced Moore's law, which in electronics describes the exponential growth of transistor counts on integrated circuits with transistor counts doubling approximately every two years. While conventional methods to produce integrated circuits in the semi-conductor industry reaches its physical limits, soft lithography allows a sustainable, more rapid growth with valve density doubling every 4 months (source: <http://www.fluidigm.com>).

The increasing demand for screens of large sample libraries in biology and chemistry asks for more functionally complex microfluidic devices, integrating thousands of valves while minimizing the number of external ports and incorporating strategies for parallelization, metering and multistep processing.

The modularity of many high-throughput devices such as MITOMI allows for easy adjustments of the total capacity of the array to host more unit cells by adding more or less rows / columns to the design or varying the size and therefore the volume within each unit cell. Prototype MITOMI designs hosted 16 flow rows with 40 chambers each, allowing the device to be programmed with 640 different samples. A scale-up to 2,400 unit cells, controlled by 7,233 valves, was used for many of the published applications [36, 10, 11]. Obviously, a larger version allows for a higher throughput of larger libraries but also for investigating replicates on the same chip to control for variations in sample preparation and external conditions. However, depending on the application and process feasibility in mind, the geometry has to respect certain design rules, such as pitch size (between unit cells), channel routing as

well as channel and valve geometry. Increasing the number of unit cells on a similarly small footprint may impede functionality or ease of handling.

In 2010 Fordyce *et al.* published a paper featuring a second-generation MITOMI device (MITOMI 2.0), which contained 4,160 unit cells and ~12,555 valves. Several design changes were implemented compared to the first scale-up to ensure function at this scale [13], such as the homogenous flow across the chip and efficient bonding of the chip to the glass slide. To address these aspects, they changed the geometry of the flow channels and increased the pitch between individual unit cells, respectively. In addition, an alternative channel routing and input tree provided more space for the printed array on the slide.

Very recently, Araci and Quake [146] launched the era of microfluidic very large scale integration (mVLSI) devices by integrating up to a million valves as small as $8 \times 6 \mu\text{m}^2$ on a footprint of 1 cm^2 using a MSL design with three layers instead of two. Reducing the valve dimensions by more than a magnitude allowed them to increase valve density by more than two orders of magnitude compared to the current state of the art microfluidic MLSI devices. Arguably, mVLSI devices would be favorable for single-cell assays of prokaryotic cells, as the channel dimensions get closer to the size of these cells.

However, programming several thousand to over a million of unit cells/chambers with different sample solutions might be another question to be addressed. Currently, programming high-throughput microfluidic devices for large-scale screens is typically accomplished by depositing samples on a glass slide either via contact printing using pins, inkjet printing [147, 148] or droplet formation technologies (non-contact printing). Despite their experimental variability spotted microarrays have been the most popular due to wide availability, high flexibility and relatively low cost [149]. Using small pen-like devices or pins the sample solution is picked up from a reservoir such as a microwell plate and transferred to a glass slide, where the deposited liquid remains as a spot. Spot diameters can vary and depend on the diameter of pin, its geometry (pointed or blunt end, solid pin tip or pin with cavity), the viscous properties of the sample solution as well as the surface chemistry of the glass slide. However, the resolution of pin-aided contact printing is limited and precision at the submicron to nano-scale might be better achieved with technologies like the direct-write dip-pen-nanolithography (DPN) [150], which allows feature sizes of less than 15 nm by molecular self-assembly.

3.2.3 Modular scale-up of MITOMI

In view of future experiments another scale-up of MITOMI is desirable to accommodate entire fusion protein libraries for full-proteome scale interaction studies on a single device. Building on previous efforts, a larger device was designed in 2009 with 7,200 unit cells covering an array size of 54 mm x 18 mm (Figure 3.3 A). The scale-up therefore contained an order of magnitude more unit cells than the most commonly used MITOMI chip (768 unit cells) in the

Maerkl lab at only a 4.5 times larger footprint, which still allowed the use of arrays printed on a regular 3 x 1 inch glass slide (76 x 26 mm). However, increased unit cell density came at a price: difficulties associated with such small delicate features rendered chip fabrication impractical (Figure 3.3 B). Moreover, high feature density called for smaller pitches (distance between unit cells), which reduced the surface area available for bonding the chip to a glass slide and hence may result in poor adhesion. Increasing the number of unit cells may also result in decelerated response times in the control layer with longer fluid channels to be filled and actuated. Longer channels will result in higher resistance at the same flow rate, compromised mass transfer and longer surface modification times. The implementation of additional fluid control components such as separated modules with individual fluid input trees and multiplexer elements can avoid overly prolonged operation times for on-chip experiments and allow more control. However, increasing the feature density to ~ 740 unit cells/cm² and $\sim 2,200$ valves/cm² requires some practice in aligning the features of the device and might therefore be less suitable as a tool for common biology labs.

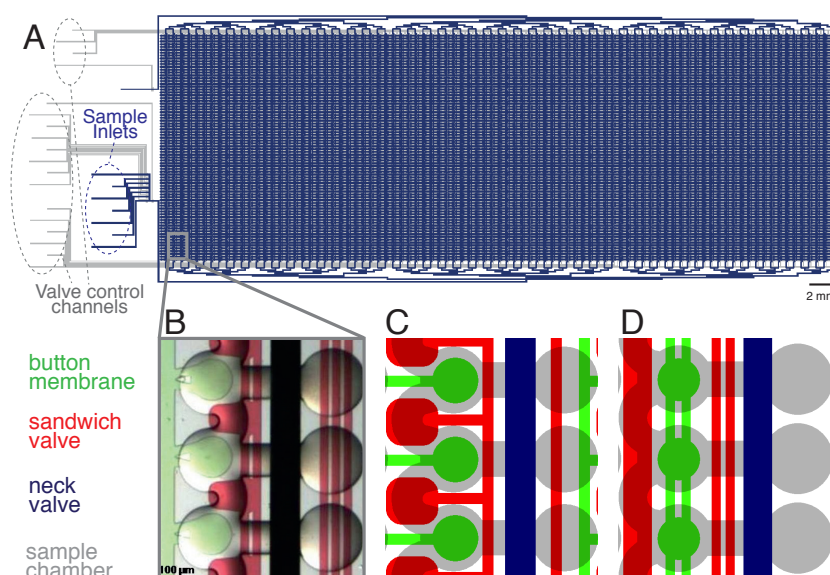


Figure 3.3: (A) Drawing of a device with 7,200 unit cells. (B) Picture of three unit cells from the same device, showing the challenges when aligning small control elements with detailed structures. Simpler feature geometries (C), (D) will overcome the shortcoming of the previous large-scale design but still require a lot of practice to manually align a microfluidic device of this capacity.

Revised versions of the scale-up favored simple feature geometry (Figure 3.3 C, D), a modular design, and an inlet tree multiplexer, which allowed individual addressing of each of the 4 sub-arrays with 1,800 unit cells as well as each row within it. The 7,200 unit cells are controlled by $\sim 21,720$ valves on a footprint of 9.72 cm² (54 mm x 18 mm array size), which was to our knowledge the highest valve density (2,235 valves/cm²) on a functional microfluidic chip at that time. Challenges with such a high feature density include programming the device with samples and control of valves. The latter one can be addressed by implementing

computationally controlled valves [151, 97].

However, reducing the size of the chamber volume also limits the initial seeding density of samples, which might be critical for (live) cell arrays, where deposition of multiple layers using pins with a small diameter is not an option due to time constraints. In addition, decreasing the chamber size by 20% requires a precise technology for depositing sufficient biological material. While well-established methods are available for generating microarrays of homogenous solutions such as bacteria, DNA, proteins and other small molecules, spotting larger cells may require some adjustments to the spotting procedure, which will be discussed in section chapter 4. Additionally, higher flow channels and unit cells as well as a modified sample chamber shape can help to accommodate larger volumes.

4 Yeast cell arrays

4.1 Abstract

Protein arrays have been the gold standard to study the binding activity and specificity of proteins with other molecules in large-scale functional assays. Depositing purified proteins on glass slides was used to array the entire yeast proteome [5], analyze their activity in interaction studies [6] and obtain quantitative networks [7]. However, large-scale mapping of interactions using these classical protein arrays has two main drawbacks: the extremely laborious protein purification process and a bias towards strong binding proteins due to a lack of advanced detection mechanisms. This chapter describes an alternative approach to generate a protein array on a microfluidic chip from an array of whole yeast cells, each expressing a different protein, which are purified and detected *in situ*.

4.2 Introduction

The budding yeast *Saccharomyces cerevisiae* has been used by human kind since ancient times, first in the processes of winemaking, brewing and baking; and with the beginning of the biotechnology era as a cell factory for the production of fuel, chemicals, pharmaceuticals and more. Today, *Saccharomyces cerevisiae* is one of the most intensively studied eukaryotic model organisms in molecular and cell biology. Many genome-wide technologies were first established in yeast before they were applied to other organisms. A number of regulatory pathways are conserved between this yeast and humans, and thus yeast has served as a model organism for detailed molecular studies. Direct parallels between the two organisms allowed conclusive studies in lipid metabolism [152], prion development and neurodegenerative diseases such as Alzheimer's [153, 154], mitochondrial metabolism [155] or aging [156]. Within the emerging field of systems biology, *Saccharomyces cerevisiae* has been successfully used for mapping complex regulatory networks and resolving the dynamics of signal transduction pathways.

The *Saccharomyces* Genome Database (SGD) lists 6,717 genes, of which 5,818 are known or putative protein-coding ORFs and ~75% thereof characterized. Several yeast fusion libraries with different affinity tags and knockout strains have been constructed to facilitate protein analyses on a proteome-wide scale for a growing research community. The ORF can be either genomically integrated in its intrinsic chromosomal location or encoded separately on a plasmid. In an enormous effort the labs of O'Shea and Weissman generated two genomic collections by systematically inserting a tag and a selectable marker gene (His) at the C-terminus of each ORF. They are tagged with green fluorescence protein (GFP) for protein localization [157] and tandem affinity purification (TAP) [158], respectively. Both ORF collections comprise ~75% of the entire yeast proteome and are now commercially available. Detected levels of fusion proteins in genomic libraries correspond to the physiological abundance of that protein in the cell as protein expression is under the control of a native promoter.

In order to yield higher and constant protein quantities, a number of plasmid-based fusion libraries were designed using high-copy number, inducible vectors with an integrated selection marker that enables robust protein over-expression. Two commercially available plasmid-based collections were considered for the generation of protein arrays from cells: the glutathione-S-transferase (GST)-tagged ORF-fusion protein collection [6] and a moveable ORF collection (MORF), where each ORF is fused to a versatile tandem fusion tag that allows a range of detection and affinity purification strategies [159]. Both collections contain more than 80% of all *S. cerevisiae* ORFs and protein expression is under control of a GAL-promoter, inducible by adding galactose to the yeast cells. Most of the described yeast ORF collections were interrogated in extremely labor-intensive large-scale screens using laborious techniques such as purifications for the production of protein microarrays [6, 159, 160] and Western blot analysis [158, 159] or more technologically advanced approaches, such as imaging [157, 160] and FACS-based sorting [161, 160]. Clearly, these large-scale screens had two main shortcomings: array production and limitation of detection methods.

To overcome these limitations, we generated protein arrays directly from deposited yeast cells. Coupling this cell array to a microfluidic device allowed cell lysis and protein extraction, pull-down and sensitive detection based on the MITOMI principle. The *in situ* generated protein arrays can be used to perform protein interaction studies at high-throughput on a very small footprint with minimal reagent consumption.

4.3 Material and Methods

4.3.1 Yeast cell culture

Cells from the yeast **GFP clone stock collection** (Invitrogen, Switzerland), containing 200 μ l yeast peptone dextrose (YPD) medium (Sigma-Aldrich, Switzerland) supplemented with 15% glycerol were inoculated into 96-well plates with 150 μ l fresh YPD medium using a 96-floating pin tool replicator (V&P Scientific). Plates were covered with Breathe-Easy sealing membranes (Sigma-Aldrich, Switzerland) and cultured overnight at 30 °C in an orbital incubator shaker (Gyromax 737R, Amerex Instruments, Inc.) to an OD₆₀₀ of around 0.8-1.2, before centrifuging them at 2400 rpm for 3 minutes to obtain a compact cell pellet for spotting.

The full collection of **GST-tagged yeast**, created by the Andrews lab is now distributed by Open Biosystems (now Thermo Scientific, Switzerland) in 76 microwell plates (96-well), containing CM broth (chopped meat) with glucose minus uracil (SD-ura/0.5% glucose) with 15% glycerol. Plates were replicated by inoculating cells into 150 μ l fresh selective growth medium (SD-ura/0.5% glucose) with a 96-pin replicator tool, covered with Breathe-Easy sealing membranes and grown for 2 days in an orbital incubator shaker at 30 °C and approximately 250 rpm. Plates were then either supplemented with 65 μ l of a 50/50 glycerol/growth medium mixture (total glycerol 15%) for maintenance at -80 °C or re-inoculated into 100 μ l SC-ura/2% raffinose to grow cells in a glucose-free medium before induction. Confluent overnight cultures were re-inoculated into fresh 100 μ l SC-ura/2% raffinose and grown at 30 °C (250 rpm) to an OD₆₀₀ between 0.6 and max. 1.0. Cells were then induced by adding 50 μ l 3x YP + 6% galactose (2% total galactose) induction medium (50 μ l SC-ura/2% raffinose was added to a negative control plate), grown for another 3-5 hours at 30 °C while shaking and centrifuged at 3,000 rpm for 10 min at 4 °C before washing the pellets with 100 μ l cold water. Resuspended cells were spun again at 3,000 rpm for 10 min at 4 °C and resuspended in 100 μ l YPD for spotting. Alternatively, blotted pellets of induced cells could be either directly frozen at -80 °C or resuspended in YPD medium containing 25% glycerol before storing them at -80 °C.

Cells from the **MORF yeast collection** (distributed by Open Biosystems, now Thermo Fisher, Switzerland) were also maintained in 96-well flat bottom plates filled with SD-ura/0.5% glucose containing 15% glycerol and grown overnight at 30 °C in growth medium (SD-ura/0.5% glucose) before re-inoculation into SC-ura/2% raffinose. Overnight cultures were diluted 1:25 into 100 μ l fresh SC-ura/2%raffinose medium (start OD₆₀₀ = 0.3) and grown to OD₆₀₀ 0.8-1.2 for induction with 50 μ l induction medium (3x YP + 6% galactose, total 2% galactose), while 50 μ l SC-ura/2% raffinose was added to another plate, serving as negative control (non-induced). Cells were induced for 6 hours at 30 °C while shaking, then harvested and washed like the GST cells for spotting or storage of induced cell plates.

4.3.2 Microfluidic device fabrication and glass slides preparation

Microfluidic chip design and fabrication as well as coating of glass slides with epoxysilane were performed as described in Appendix A [162].

4.3.3 Spotting of yeast cells and device alignment

The fabrication of a yeast cell arrays for *in situ* protein pull-down is illustrated in Figure 4.1.

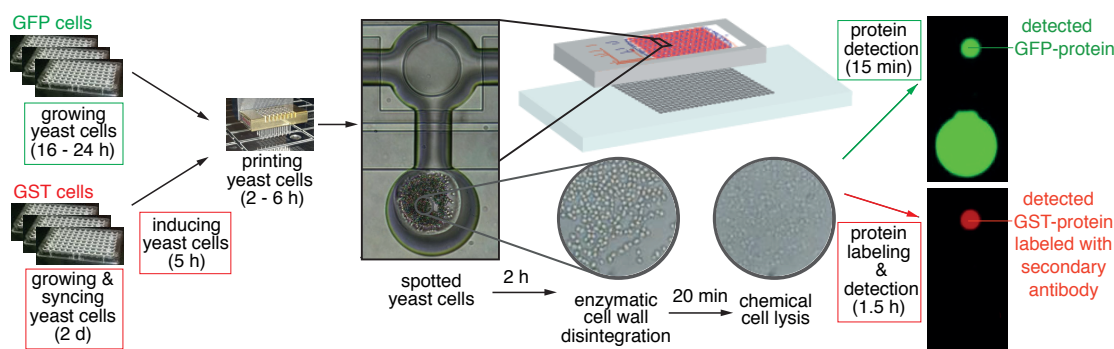


Figure 4.1: Workflow of fabricating a protein array from yeast cells: Grown yeast cells are spotted on glass slides, aligned and bonded to a MITOMI microfluidic device and then lysed *in situ* for protein extraction. Pulled down protein is detected using an antibody immobilized. Cells from the GFP collection can be printed directly from the growth plate, while the GST yeast cells have to be re-inoculated to synchronize cell growth before inducing protein expression with a galactose-containing medium. A secondary antibody is used to label GST protein pull-down for detection.

Yeast cells were pelleted, resuspended in low-sugar medium or PBS, and then spotted onto epoxy-coated microscope slides with 373 μm column and 746 μm row spacing using a QArray2 microarrayer (Genetix) and metal pins of different geometries (ArrayIt). Spotted yeast arrays were manually aligned to microfluidic PDMS devices containing 768 unit cells using a Nikon SMz1500 stereoscope and bonded for at least 2 hours at 40 °C.

4.3.4 On-chip experimental procedures for yeast cell arrays

The surface area was modified by depositing layers of BSA-biotin and NeutrAvidin/PBS as described in Appendix A [162]. The glass slide surface modification was finished by immobilizing biotinylated antibody in PBS (see Appendix D) to the area beneath the button membrane, followed by washing with PBS.

In situ cell lysis was initiated by first introducing an enzymatic solution of 10 mg/ml Zymolyase (amsbio, Switzerland), 1% BME (2-Mercaptoethanol), in 0.1 M Tris-HCl, which breaks down the cell walls. Following a 30 min incubation period at room temperature, a detergent solution containing Y-PER (Pierce, now Thermo Scientific, Switzerland), 1 μM PMSF (phenylmethyl-

sulfonyl fluoride), 1 protease inhibitor cocktail tablet (Roche, Switzerland) was loaded into the chamber and allowed to lyse the cell spheroblasts within 2 hours with the chambers closed. A variation of this 2-step (cell wall degradation and lysis) lysis protocol was to combine the two solutions and load them together into the cell chamber by dead-end filling, i.e. with the outlet closed. The button membrane was kept closed during the loading of the lysis solutions to prevent a) degradation of the antibody underneath the button membrane and b) cross-contamination by cells that were flushed from the chamber. After complete cell lysis at $\sim 4^{\circ}\text{C}$, extracted molecules were allowed to diffuse from the chamber to the immobilized antibodies, while the sandwich valves were kept closed. After 30 min incubation at room temperature the glass surface was blocked with a solution of PBS containing 1% milk and 1% BSA, opening the button membrane at the end, and followed by a PBS wash. The button membrane was collapsed for imaging of the device on a modified ArrayWoRx (Applied Precision) microarray scanner, when using the GFP collection. In experiments with the GST and MORF library, protein pull-down was detected flowing a Penta-His Alexa Fluor 647 conjugate (see Appendix D) on chip. After a final PBS washing step, devices were scanned with the button membrane closed. If not otherwise noted, chemicals and reagents were purchased from Sigma-Aldrich, Switzerland.

4.3.5 Data acquisition and analysis

Slides were scanned on an ArrayWorx scanner after the final PBS washing step with 1.0 s with the following wavelengths: 488 nm for GFP-tagged yeast proteins and 685 nm for antibody-labeled GST-tagged proteins. A reference image was taken with 0.1 s at the reflective channel and at 488 nm to detect cell spots in the chamber before cell lysis. For detection of surface bound protein a microarray grid was aligned to each button membrane area. For correlation of spotted cell density local backgrounds were subtracted for all channels by moving the grid just next to the chamber, outside the channels. Unit cells with bad features or signal intensities below the background signal were excluded from further analysis, which was done with a code written in Mathematica (Wolfram Research). Graphs were plotted with Prism 5 (GraphPad Software).

4.3.6 Protein purification (off-chip)

For protein purification, cells were grown as described above for on-chip experiments in either 96-deep well plates (1 ml/well) or in 10-ml microvials. Alternatively, individual strains were streaked from the yeast glycerol stock onto agar plates containing SD-ura/0.5% glucose, incubated at 30°C until visible colonies (2 mm) had formed (~ 24 h) and then parafilm sealed kept at 4°C until they were inoculated into glucose free medium (SC-ura/2% raffinose). Inoculated cultures were grown overnight at 30°C under agitation. Confluent overnight cultures were re-inoculated into fresh SC-ura/2% raffinose, grown at 30°C (250 rpm) to an

OD₆₀₀ between 0.6 and max. and then induced by adding 3x YP + 6% galactose (2% total galactose) induction medium (SC-ura/2% raffinose was added to a negative control plate). Induction was quenched after 3-5 h (GST-tagged strains) and 6 h (MORF), respectively, by washing the cell pellets (centrifugation: 4,000 rpm, 10 min, 4 °C) with ice-cold water. Pellets of induced and non-induced cells were kept at -80 °C until lysis.

Proteins from GST and MORF yeast strains were purified under native and denaturing conditions using Ni-NTA spin columns (Qiagen) for the purification of His-tagged proteins, with adaptations to yeast cells. Buffers and reagents for protein purification can be found in Appendix C.

4.3.6.1 Cell lysis and protein purification

Thawed cell pellets were resuspended in 1 ml lysis buffer, 0.5 ml transferred into a 2 ml Eppendorf tube and spun at 3,000 rpm for 4 min at 4 °C. The pellet was resuspend in 0.5 ml lysis buffer and 0.5 ml Y-PER and sonicated with added zirconium beads (diameter 0.5 mm, Sigma) 5 times for 40 seconds with 2 minutes intervals on ice in between sonication rounds. Sonicated lysates were centrifuged at 4,000 rpm for 45 min at 4 °C and 0.6 ml of clear lysate (L) was loaded onto a Ni-NTA spin column (Qiagen), which had been pre-equilibrated with 0.6 ml lysis buffer. The flow through (FT) after centrifugation (2000 rpm, 2 min, room temperature (RT)) was collected for SDS-PAGE analysis. The Ni-NTA column was washed twice with 0.6 ml wash buffer and centrifuged (2000 rpm, 2 min, RT). Both wash fractions (W1, W2) were saved for SDS-PAGE analysis. The bound protein was eluted twice from the Ni-NTA column by centrifugation (2000 rpm, 2 min, RT) with 0.2 ml elution buffer and the eluate collected after each step (E1, E2). All samples collected during the purification process (L, FT, W1, W2, E1, E2) were visualized by loading them on 13.5% acrylamide gels and running a SDS-PAGE and stained with Coomassie Brilliant Blue 250.

4.3.6.2 Western blotting

SDS-PAGE gels were washed with deionized water and transferred to a PVDF membrane using the standard settings (7 min at 20-25 V) of the iBlot Dry Blotting system (Invitrogen). The membrane was washed twice with deionized water for 5 minutes, blocked with a blocking buffer (TBST with 5% non-fat dry milk powder) overnight at 4 °C, incubated with a 1:2,000 dilution of primary biotinylated antibody (see Appendix D) for 2 h on a shaker at room temperature and washed four times with TBST for 10 minutes. The membrane was then incubated in a 1:50,000 dilution of streptavidin-horseradish peroxidase conjugate in TBST with 1% non-fat dry milk powder while shaking. The reaction was stopped with cold water when bands were sufficiently visible. The membrane was washed thrice with cold water for 2 minutes and scanned.

4.3.6.3 InVision His-Tag In-Gel Stain

The InVision His-tag In-gel Stain (Invitrogen, Switzerland) contains a fluorescent dye (Ex 560 nm / Em 590 nm) conjugated to Ni^{2+} : nitrilotriacetic acid (NTA) complex, which binds specifically to the oligohistidine domain of His-tagged fusion proteins. All incubations were performed at room temperature and on an orbital shaker set to low agitation, following a modified version of the suppliers' manual. After electrophoresis, the gel was fixed for 1-2 hours in a fixing solution (50% methanol, 40% ultrapure water, 10% acetic acid), washed twice for 10 minutes each with ultrapure water to remove the fixative, stained with the ready-to-use solution of InVision His-tag In-gel Stain for at least 1 hour, washed twice for 10 minutes each with PBS and then visualized at a fluorescence scanner (Typhoon).

4.4 Results and Discussion

In a proof-of-principle experiment we programmed a MITOMI device with 192 randomly chosen strains from the yeast GFP library and successfully detected more than 80% of 192 GFP fusion proteins with a signal intensity of at least 20 relative fluorescence units (RFU), which represented the signal threshold for non-specific detection (2 s.d. above mean of signal for cell-free chambers). However, we noted a considerable variation of the amounts of cells printed on the glass slide across the chip and between different batches of spotting. The inhomogeneous depositing of cell spots eventually contributed to an even larger variation of detected protein signals between on-chip experiments from different spotting batches. There are many interlinked factors that play into the final spotting outcome, mainly (1) the spotting solution or suspension, (2) the settings for the spotting procedure and (3) the parameters of the metal pin, which transfers the suspended cells from the source plate to the glass slide.

4.4.1 Spotting efficiency and reproducibility

Obviously, spotting yeast cells differs from spotting a molecular solution. Many factors, such as cell density, pellet quality, viscosity and sugar content of the suspension medium have an impact on the spotting result. Obtaining sufficient protein pull-down for interaction studies requires a critical number of cells per spot. To pick up enough cells from the source, GFP strains were initially spotted with a slit pin, which exploits the mechanisms of surface tension and adhesion to transfer the sample from the source plate to the glass slide. Yeast cell suspension are loaded into the pin channel (slit) by capillary forces and form a thin layer at the end of the flat tip, which then contacts the glass slide surface to expel the suspension (Figure 4.2 B).

Gently releasing small volumes of the loaded cell suspension with this ink stamping mechanism allows printing multiple spots from one sample-loading step. However, multiple stamping deposited varying number of cells in consecutive spots and usually only the first stamp pro-

duced a spot of sufficient cell density. Washing the pin after each deposition step increases the arraying time considerably. This is particularly an issue for large-scale spotting (the whole yeast library consists of 48 plates) because cells in the plate as well as on the glass slide start to dry out. On the other hand, re-inking into the cell suspension tended to clog the pin channel, which led to inconsistent spotting patterns and the pin was subsequently difficult to clean. Finally, using a solid pin with a larger blunt tip diameter produced spots of consistent size and sufficient cell density, as shown in Figure 4.2 A.

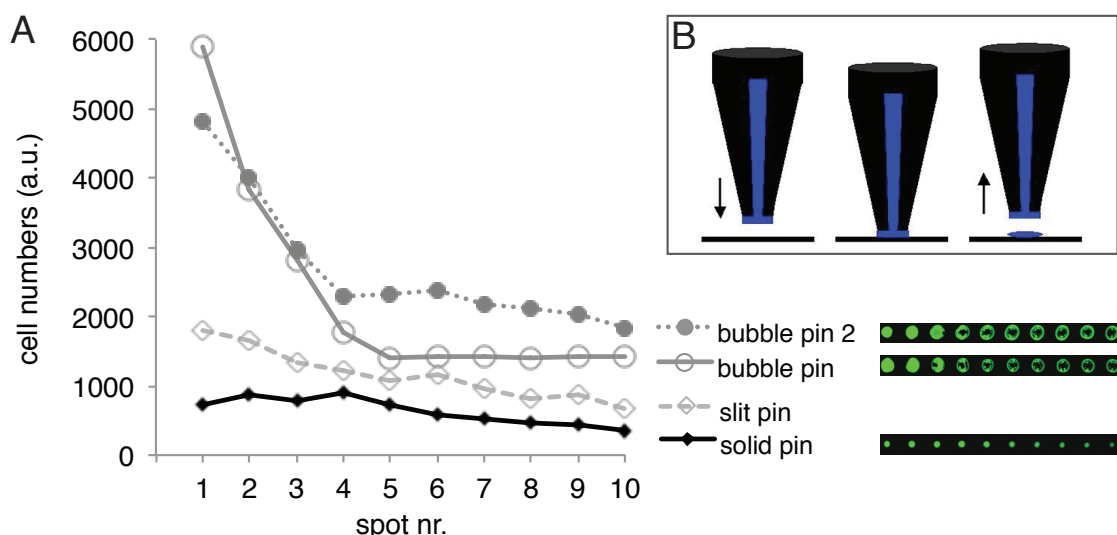


Figure 4.2: Spotting tests with yeast cells from the GFP collection with pins of different geometries. (A) Spot reproducibility: Each pin was loaded once with yeast cell suspension (same strain) and spotted on a glass slide ten times without re-loading. Fluorescence intensity of cell spots was detected with a scanner and is plotted as a function of spot number. (B) Principle of depositing solutions with a metal pin.

To account for the differences in signal intensity due to varying cell-seeding numbers, we sought to normalize the detected protein signal to the number of cells that gave rise to it. However, while the fluorescence signal of the protein detected by an antibody, which is immobilized on the glass surface, can be easily correlated to the protein concentration, correlating the signal of deposited cells or counting cell numbers on the chip would assume a monolayer of cells, yet the printing resulted in several stacked cell layers. Thus inferring the total number of cells per chamber from the detected signal is not a reasonably sensible correlation and can only serve as a rough indicator. A more sensitive normalization might be to measure the signal of the GFP-fusion protein in solution after cell lysis. However, this approach only works for yeast strains with an intrinsic label.

Besides the obvious advantage of an intrinsic fluorescence signal, the GFP yeast collection has a major drawback: In response to a native promoter GFP yeast strains express protein at physiological levels, which naturally vary. These signal variations render analysis of detected protein from GFP fusion strains difficult and impractical to study protein interactions. Thus quantitative analysis of protein pull-down requires normalization of the spotted cell numbers

to the detected signal intensity after protein pull-down and correlation to measured *in vitro* concentrations derived from Western blots of the TAP-tagged yeast fusion library [158]. The correlation of a protein array, generated from 384 different GFP strains (4 plates), to the protein concentrations derived from the TAP-library was only moderate (Figure 4.3 C). A correlation of our protein signals to the GFP abundance measured by fluorescence activated cell sorting (FACS) using the same GFP fusion strains [161] showed a considerably more linear trend (Figure 4.3 D). The observed plateau for proteins of higher abundance could have two reasons: First, the antibodies used to pull-down proteins were saturated with excessive proteins in solution and second, the protein intensity was outside of the dynamic range of the fluorescence scanner. Experimental effects, such as a low number of deposited cells or incomplete cell lysis may contribute to the detection of weak GFP signals (outliers in the lower left field) despite its physiological higher abundance. In any case, correlation would benefit from a more sensitive normalization of on-chip protein intensity to the number of cells giving rise to it.

Nevertheless, Figure 4.3 B demonstrates that the method is very well reproducible (average goodness of the linear regression fit $r^2=0.94$) if yeast cells were spotted under optimized conditions. We detected most of the selected strains at levels that are considered reasonable to study protein interactions (Figure 4.3 A).

Admittedly, low or no expression of certain proteins under the given conditions makes on-chip experiments with deposited GFP fusion strains extremely difficult. Promising alternatives are plasmid-based collections where the ORF is fused to a versatile tandem-tag (MORF library) or GST- and His-tag (GST library). Over-expression of these libraries can be induced in the presence of galactose, giving rise to higher protein yields. First test experiments showed pull-down of proteins from a subset of both libraries, albeit slightly better results were obtained from the GST collection, which was chosen for most of the following experiments. A critical point was the relatively low robustness of the method, which could be traced back to the induction procedure and spotting efficiency. On-chip protein pull-down results from yeast GST cell arrays, generated under the same conditions, showed poor reproducibility. In addition, some yeast strains expressed much more protein. Thus, before moving on to quantitative interaction studies, the following parameters had to be optimized for reliable protein binding data: (i) spotting efficiency, (ii) on-chip cell lysis, and (iii) labeling and detection strategies. Last but not least, a more robust protocol for off-chip induction of protein over-expression needed to be elaborated.

Even though inducible yeast ORF libraries overexpress the protein of interest, the seeding cell number on the array influences largely the success of protein detection after pull-down of the overexpressed and extracted protein. Just as with the GFP strains, the initial arraying of induced strains of the GST and MORF collection showed heterogeneous spotting patterns with a large strain-to-strain variability across experiments. Crucial parameters for efficient

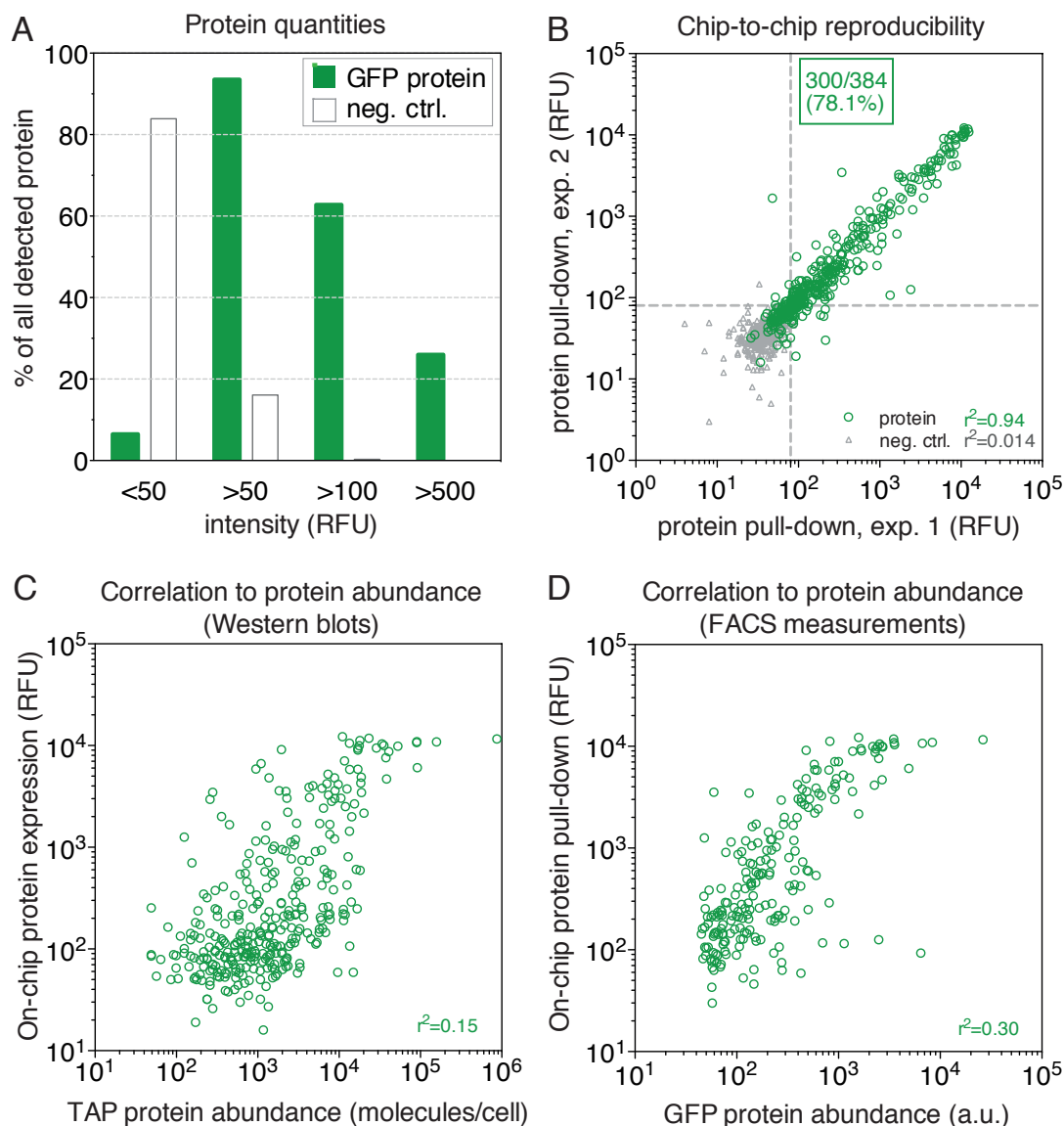


Figure 4.3: Detection of protein pull-down of 384 yeast GFP-fusion proteins extracted on-chip from yeast cells. (A) Pull-down efficiency: The percentage of clones giving rise to a minimal RFU (relative fluorescence units) value of 50, 100, and 500 is plotted. Signals detected in unit cells without spotted cells served as negative controls. (B) Reproducibility of pull-down between two on-chip experiments. Dotted gray lines indicate the background threshold (2 s.d. above the mean of the negative control). (C,D) RFU signals from on-chip protein expression were compared to values of molecular abundance of the same protein, which was determined by (C) Western blotting TAP-tagged fusion strains [158] and (D) FACS measurements of the GFP-tagged fusion strains [161], respectively.

cell deposition, namely pin geometry, resuspension medium and the washing procedure of the pin are addressed in the following section.

Spotting tests with different pin geometries revealed, that a solid pin, delivering one spot of

cells at a time, produced the most homogeneous yeast cell array with a rather consistent cell spot size and density (Figure 4.4 A, left). However, cell spots were very small, contained very few cells and the throughput remained low even with altered spotting routines (multiple pins and multiple prints on the same cell spot). Cells in the source wells started to dry up during the long spotting routine, which proofed to be uneconomical for spotting larger batches of arrays covering the entire library. More efficient cell transfer was observed for pins with a cavity (Figure 4.4 A, center).

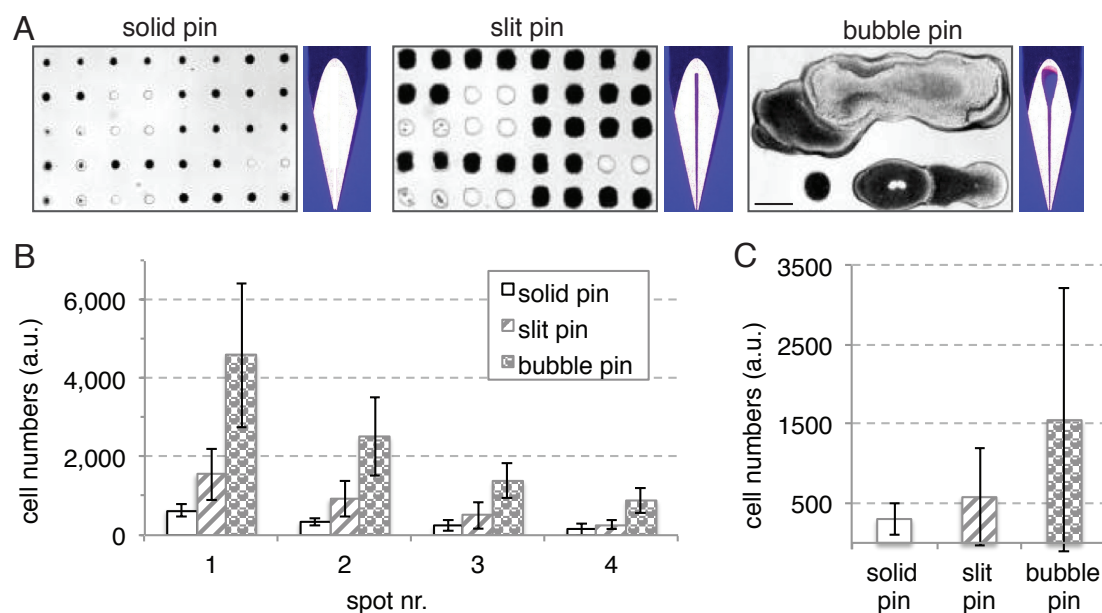


Figure 4.4: Spotting tests with yeast cells from the GST collection with pins of different geometries. (A) Print pattern after spotting different strains with pins of different geometry. Spot size differences after multi-stamping of one GST without re-inking (B) and after printing 96 different strains on an array with a single stamp (C). (Scale bar for all spotting images: 500 μ m)

Although having a tip diameter (print area) comparable to their solid pin or slit-pin counterparts, they are capable of taking up more than double the volume in the bubble-shaped cavity (0.6 μ l vs. 0.25 μ l in the slit) and delivering almost 30% more volume per spot (between 0.7 and 1.4 nl). However, with increasing uptake and delivery capacity of the pin, spots became more irregular-shaped due to sudden releases of large cell agglomerations, leading to merging of cell spots (Figure 4.4 A, right).

Consequently, three issues were addressed in spotting tests: (i) a more stringent washing routine of the pin to overcome congestion of the channel, (ii) different spotting routines/procedures and (iii) various buffers to resuspend the cells in after induction in order to avoid agglomeration of cells. Ideally, pins with a bubble-shaped channel should stamp several spots of cells before reloading the pin at the source well. Practically, consecutive spots contained fluctuating cell numbers (Figure 4.4 B) and were poorly reproducible for the entire array (Figure 4.4 C). Reloading after each stamp increased the probability of congestion in

Chapter 4. Yeast cell arrays

the pin cavity. Hence, while being less efficient, pins were washed after each spot. Multiple rounds of wash-dry with water for several seconds improved the washing effect but proofed again too time-consuming and uneconomical. Rigorous washing procedures with a mild detergent (0.05% Tween-20 in PBS or water) eventually helped to clean adhesive cells from the pin cavities.

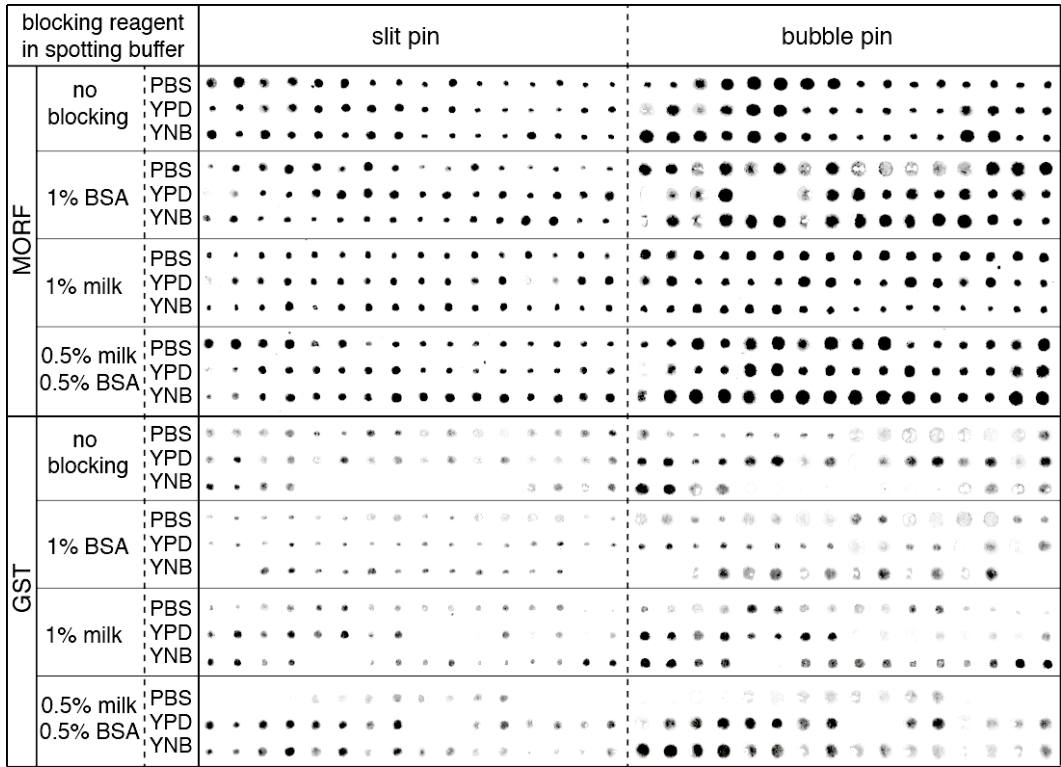


Figure 4.5: Spotting test with different blocking reagents added to the cell media before printing an array of 48 induced GST and MORF strains using a slit pin and a bubble pin, respectively.

To reduce the agglomeration of cells, induced cells were washed several times with cold water or PBS and finally resuspended in different blocking solutions. The results after can be seen in Figure 4.5: Neither buffer showed sufficient improvement when spotting several strains, although induced yeast cells seemed to agglomerate and thus block the pin less.

Analysis of protein pull-down reproducibility of serially spotted slides (with cells from the same plate) after freezing pre-assembled chips would be an advantage for screening varying conditions through the entire proteome to avoid batch-to-batch variations of spotted cells and moreover to reduce total handling time in bulk fabrication. However, we observed only moderate reproducibility for protein pull-down experiments on cell arrays, which had been spotted from the same plates, aligned to the microfluidic device and kept at -80 °C in order to maintain the yeast cell protein integrity (Figure 4.6 A) when compared to frozen GFP strains (Figure 4.6 B).

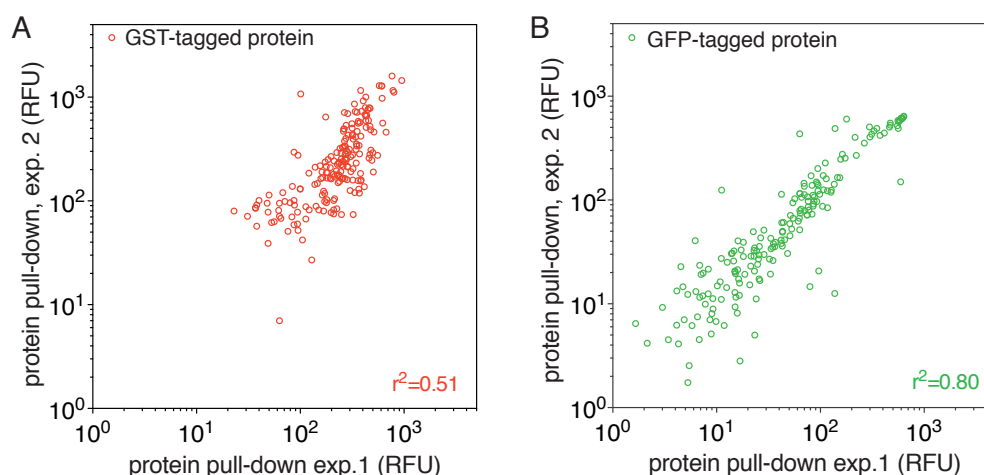


Figure 4.6: Reproducibility of protein pull-down for arrays kept at -80 °C for 6 weeks with 96 yeast strains from the (A) GFP and (B) GST collection.

Another possible scenario to improve on-chip performance would be to grow spotted cells on the device, which has several advantages: 1. The total amount of expressed protein to be purified increases with growing cell numbers. 2. The viability and morphology of cells can be monitored. Assuming that only healthy cells produce functional proteins in sufficient quantity, detected protein signal intensity could be correlated to number and growth rate of the cells. 3. *In situ* induction of plasmid-encoded collections reduces loss of cells that is typically associated with washing and pelleting in a 96-well plate. However, observing cell density and behavior is critical to induce expression during the right stage of cell growth as well as a fast and complete media exchange after over-expression.

4.4.2 On-chip cell lysis and protein purification

On-chip yeast cell lysis involved two steps: Cell wall disintegration by an enzyme (Zymolyase) followed by lysis of the remaining spheroblasts using detergent-based solution (Y-PER). The two-step cell lysis could be potentially combined. However, we noted that cells in chambers closer to the fluid inlet tree degraded faster and more efficiently than those located in chambers towards the outlet, which contained almost intact cell walls even after long time incubation with the lysis mix. The observed gradient effect was almost completely avoided by re-loading the enzyme cocktail after 10 minutes. However, in the case of a one-step lysis procedure, re-loading of the lysis cocktail may wash the already lysed cells (closer to the inlet tree), out of the chamber and could potentially contaminate adjacent chambers. In addition, if cell lysis is carried out at room temperature, the integrity and functionality of some proteins may be disturbed. Consequently, the whole chip was transferred to ice water for cell lysis and protein pull-down. However, cell wall disintegration with Zymolyase works best at 37 °C. Keeping the two lysis steps separated, cell walls were broken down efficiently at room

temperature, before complete lysis and protein extraction was allowed to occur at 4 °C.

4.4.3 Labeling and detection strategies

With exception of the GFP collection, detection of on-chip purified protein is only possible by introducing a second antibody conjugated to a fluorescence marker. However, implementing two or more antibodies to immobilize and detect interacting proteins may affect the sensitivity of detection as a result of crowding and steric hindrance of the molecules.

To detect interactions several approaches could be considered. First, using chemically or intrinsically labeled TFs or proteins, which are bench-top purified. This provides a reliable labeling method and avoids incubation with another antibody on-chip. A drawback is that labeling proteins in solution often requires several purification steps. Moreover, potential binding sites might be blocked by the chemical dye. Thus, to ensure accessibility of binding sites, the protein interaction could be captured using an antibody binding to the affinity tag of the non-immobilized protein. This second approach may even circumvent constraints of detecting both proteins with a fluorescence-labeled antibody. However, sufficient and reliable protein pull-down is an essential pre-condition, which had not been established at that stage in the project. Therefore, applying purified proteins from the GFP fusion library would be the method of choice for proof-of-principle interaction detection in the future.

4.4.4 Optimization of induction process

One of the advantages of the chosen plasmid-based fusion libraries is the possibility to induce protein over-expression independently from intrinsic gene regulation, as each ORF is under the control of a *GAL1* promoter, which initiates transcription of the gene of interest only in the presence of the sugar source galactose. To induce robust protein expression, the yeast cells have to be depleted of any glucose in the medium because glucose acts as a repressor of ORF-fusion protein expression.

Most culturing and downstream conditions of previous large-scale protein induction and purification studies were setup in a 96-well box (2 ml/well) with a minimum volume of 1 ml [6, 159]. However, to facilitate the generation of yeast cell arrays, our yeast cell cultures needed to be compatible with the spotting routine, which required regular 96-well microplates with a maximum volume of 300 µl/well. Consequently, we inoculated cells using a slit 96-pin tool replicator (with slot/channel), which delivers on average 1 µl of liquid in a liquid-to-liquid transfer. However, similar to the slit spotting pins, pelleted yeast cells can congest the channel and this makes the delivery volume of a yeast suspension much more variable. Inoculating yeast cells into small volumes of media thus lead to different cell densities within one 96-well plate, which makes it difficult to induce in the recommended OD₆₀₀ range (Figure 4.7 A). Repeated re-inoculation of confluent yeast cell cultures into fresh medium using a solid

pin 96-replicator tool (0.4 μl /transfer) was optimal to obtain synchronized growth, with an induction window of 1h, the time in which most GST strains were in the recommended range of OD_{600} 0.6 and 1.0 (Figure 4.7 B).

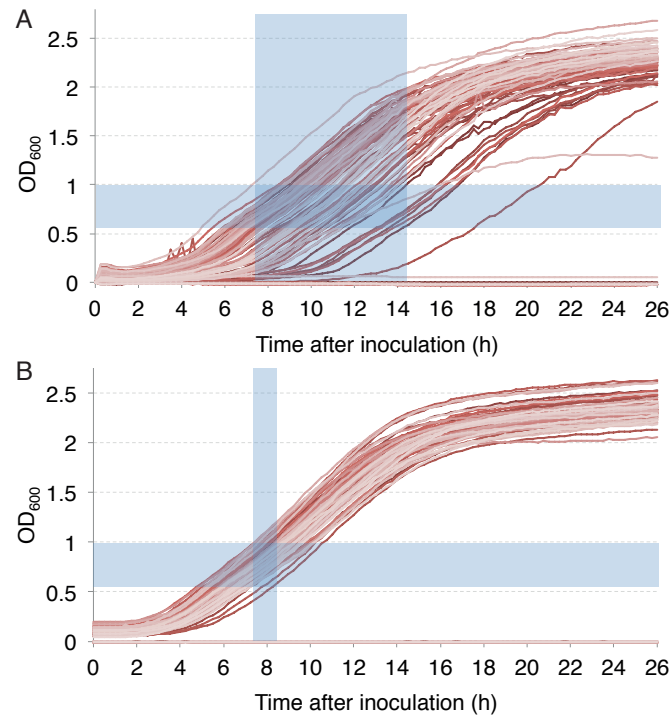


Figure 4.7: Variability of cell growth between different GST yeast strains in the same 96-well plate (A) after inoculation from glycerol stock into growth medium (B) and with synchronized growth after double-inoculation. Indicated in the blue boxes are the intervals, in which yeast strains reached the recommended cell density (OD_{600} 0.6-1.0) for induction.

Knowing the optimal induction time for protein expression is crucial, as over-expressed protein can be toxic to the cells. In order to determine the most robust induction variables, two GST fusion strains were selected, whose optical density 7.5 hours after double-inoculation was at the upper (SYN8, OD_{600} 0.98) and lower end (ACS1, OD_{600} 0.62), respectively. The two strains were grown to four different cell densities ranging from OD_{600} 0.6 to 1.2, before induction and then allowed to express the protein of interest for 3, 4 and 5 hours, respectively. Figure 4.8 shows that detected protein expression levels were higher for cell cultures, in which expression was induced at lower optical densities (OD_{600} = 0.6 and 0.8) and allowed to last for 4 or 5 hours, whereas those cultures induced at higher cell densities (OD_{600} = 1.0 or 1.2) cease protein expression over prolonged induction periods. This indicates, that cells are not as efficient in protein expression, once they have reached a critical growth phase. Interestingly, very high cell seeding numbers (10-times concentrated) are not reflected in the intensity of the detected protein signal, particularly, when detected pull-down is normalized to cell numbers. This effect is probably due to incomplete cell lysis, which had been observed before for very dense cell spots.

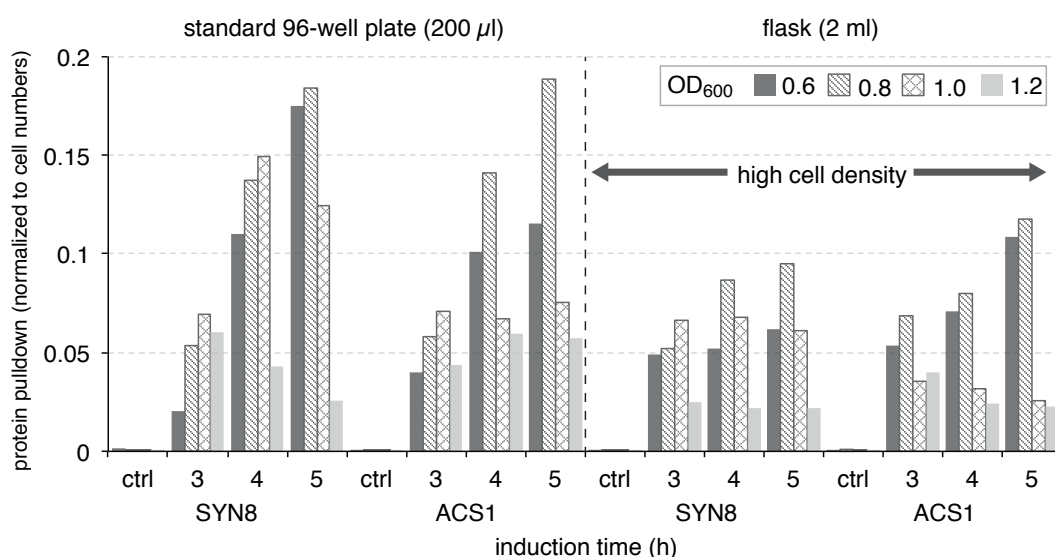


Figure 4.8: Parameter study to optimize on-chip protein pull-down. Yeast cells from two different GST fusion strains (SYN8: 55.1 kDa; ACS1: 105.1 kDa) were grown to different cell densities (OD_{600}) before expression was induced. Induction was quenched after 3, 4 and 5 hours, respectively. A MITOMI chip was then programmed with normally resuspended and 10-times concentrated cell spots (high cell density) and proteins extracted *in situ*. Protein pull-down was normalized to the number of cells spotted and is plotted for all parameters studied.

To summarize, we demonstrated that it is possible to generate a protein array on-chip from spotted induced GST-fusion yeast cells in a large-scale format that is compatible with micro-arraying conditions. Several changes in array production resulted in improved on-chip reproducibility and increased protein detection to levels, which are considered sufficient for future protein interaction studies (Figure 4.9).

The most significant improvements were observed with the following parameters:

1. Yeast cell cultures from a glycerol stock should be transferred at least twice into 150 μ l/well fresh growth medium (glucose-free) in a regular 96-well plate (300 μ l/well) with a solid 96-pin replicator tool and grown to stationary phase ($OD_{600} \sim 4.0$)
2. Cell density should be between OD_{600} 0.6-0.8 and not exceed 1.0 for induction with induction medium (containing galactose). Cell cultures in a 96-well plate may need to be monitored with a plate reader to ensure synchronized cell growth.
3. Longer induction times (up to 5 hours) yield higher protein expression levels.
4. Cells are to be washed three times with PBS or cold water after induction to remove any remaining induction medium.
5. Cell pellets should be resuspended in PBS or sugar-free medium for spotting and pelleted by short low-impact centrifugation (for max. 1 min at max. 1000 rpm).
6. A solid spotting pin (no cavity) results in a cell array with the most reproducible cell number/spot. Using pins with a larger tip diameter (pin tip > 125 μ m) will increase cell numbers transferred to the array. Note: Initial seeding cell numbers do not signifi-

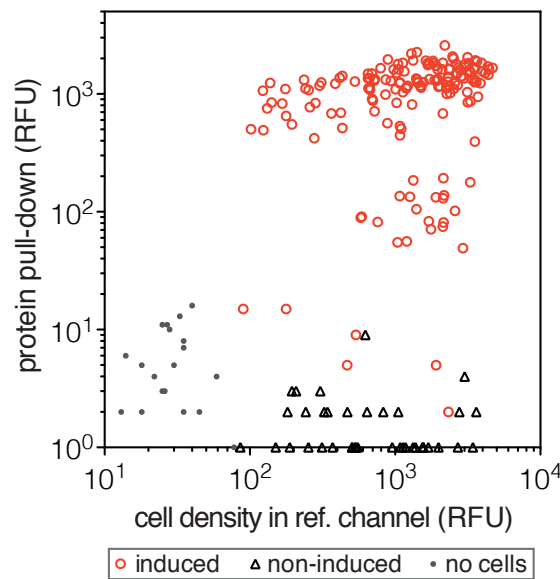


Figure 4.9: On-chip detection of protein expression from 96 yeast GST fusion strains after optimization of array generation. Protein pull-down was detected with a secondary antibody and plotted over the intensity of cell density after spotting (ref. channel).

cantly affect protein pull-down on-chip. Instead, very high cell numbers per chamber on the array can impede cell lysis and protein extraction. Nevertheless, a sensible normalization of detected protein pull-down to the cell number given rise to it should be considered for quantitative analysis.

7. Two-step cell lysis is more efficient than combining the enzymatic break down of cell walls, which works better at room temperature, and lysis of the yeast spheroblasts with a detergent. *In situ* protein extraction should be performed at $\sim 4^{\circ}\text{C}$ (in iced water) to ensure protein integrity.
8. Washing the chip with blocking solution (BSA/milk) after protein pull-down for 45 min helps to reduce non-specific binding of the fluorophore-labeled detection antibody.

Arguably, synchronized growth of all strains in a 96-well plate in order to induce them at the right time seemed to be one of the biggest challenges to generate GST-tagged yeast cell arrays. Precise monitoring of cell growth on-chip is very difficult. Therefore, the option of on-chip growth and induction was not pursued.

Instead, successfully induced plates could be frozen to circumvent the issues of repeatedly monitoring and synchronizing cell growth after inoculation from frozen plates into fresh medium. For that, thoroughly washed pellets were either shock-frozen directly or resuspended in glucose-free medium containing 15% glycerol and then stored at -80°C . Thawed plates were pelleted, washed and resuspended in sugar-free medium or PBS for spotting. However, spotting of thawed plates with resuspended cells produced a very low cell density compared to spotting from freshly processed plates. Alternatively, non-viscous cryoprotectant such as DMSO could be added to the yeast cell plate directly before spotting if protein yields

per device drop significantly after storage at -80 °C.

Reliably expressing and detecting proteins *in situ* from a large collection of yeast strains is crucial to perform protein interaction studies on-chip, which can be accomplished in two different ways using our MITOMI device: First, flowing off-chip purified proteins or cell lysate over the protein array, generated from lysed yeast cells, and second, co-spotting different yeast cell strains. Both approaches will be discussed in more detail in the following two sections.

4.4.5 Protein immunopurification for flow deposition

Generating a protein array to study protein-protein interactions generally requires large-scale bench-top protein purifications. For this purpose, engineered recombinant proteins with a specific affinity tag are expressed and purified using affinity chromatography methods. Affinity tags such as glutathione-S-transferase (GST) and hexahistidine (6xHis) bind with high affinity to immobilized glutathione agarose and nickel or cobalt ions, respectively, and are then displaced with excess competing compounds during a process called elution. His-tagged proteins from selected GST and MORF library strains were purified using commercially available affinity columns based on nickel-nitrilotriacetic acid (Ni-NTA) metal affinity chromatography matrices. Native conditions with mild lysis conditions and enzymatic inhibitors were preferred to reduce binding of contaminants and prevent protein degradation, respectively. The results of protein purification and detection of the common 6xHis-tag and a library-specific affinity tag with antibodies are shown in Figure 4.10.

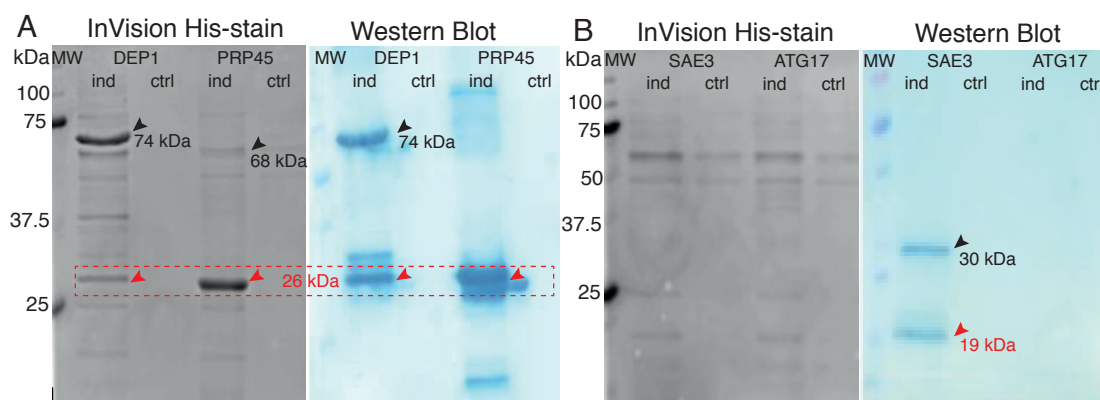


Figure 4.10: Bench-top protein purification from two strains (A) of the GST (expected protein size: 74 kDa for DEP1 and 68 kDa for PRP45) and (B) the MORF (expected protein size: 30 kDa for SAE3 and 68 kDa for ATG17) collection. Protein expression of induced cells (ind) was detected with His-antibody (left, InVision His-staining), (A) GST-antibody (right, Western blot) and (B) Protein A-antibody, respectively. Black annotations indicate detected bands at the expected protein size, red the size of the respective fusion tag. Non-induced cell served as negative control (ctrl).

Contaminants visible as additional bands on the electrophoresis gels and Western blots remained in the purified product even after adding reducing reagents (beta-mercaptoethanol) and mild detergents (Tween 20 and Triton-x). A co-purified protein of ~26 kDa was repeatedly detected for GST-tagged proteins (Figure 4.10 A, red dotted box), indicating that expression terminated prematurely after the N-terminal fusion-tag or proteins were degraded. Contaminants of ~60 kDa were pronounced for MORF-fusion proteins detected by His-straining (Figure 4.10 B, left). Western blotting with a specific antibody against another affinity tag detected the right antigen, but also indicated that the eluate contained a mixture of full size proteins as well as fragments of the affinity tags. Nickel columns seldom return a pure protein and non-specific bands in His-tag stained gels can be an indicator of highly basic proteins and divalent metal binding proteins, which cross-react with the stain producing non-specific bands. Eukaryotic cells are known to contain proteins with neighboring histidines as well as endogenous proteins with metal-binding sites that also bind to the Ni-NTA matrix to a considerable extent, which is then detected in the eluate as high background. In addition, yeast cells may acidify the culture medium and both the cells and the medium may contain certain compounds that influence binding of 6xHis tags to Ni-NTA matrices. Therefore, eukaryotic proteins may require a tandem affinity purification using two affinity tags to yield higher purity [163]. To that end, glutathione sepharose columns are available to purify GST-tagged strains and IgG sepharose for protein A fusion proteins in MORF strains, respectively, but this possibility was not further investigated.

While immunopurification of tagged fusion proteins requires an antibody or other capturing reagent, crude lysate could be used, which contain a mixture of proteins, including the protein of interest. This option is particularly attractive to obtain GFP fusion protein cause most current GFP purification methods yield low concentrations. However, complex formation, non-specific binding or interactions with other proteins in the lysate may hinder detection of the protein of interest, potentially of low-expressing proteins.

Synthesizing proteins can be seen as an alternative approach to obtain the protein of interest. However chemical synthesis of proteins is expensive and laborious. Expressing proteins from a DNA expression template using a cell-free *in vitro* transcription/translation (ITT) system can overcome many of the current limitations inherent to conventional protein production and is described in more detail in chapter 5. Generally, flowing samples over the array, limits the combinatorial possibilities and throughput of the interaction study on the current MITOMI device as no more than one sample can be flown at a time. Alternatives are the integration of a multiplexer, as shown in Figure 3.2 E, which allows applying m different samples consecutively in different rows.

4.4.6 Co-spotting

Admittedly, the investigation of protein-protein interactions using purified proteins on a protein array platform remains suboptimal for the analysis of dynamic and short-lived protein interactions, such as those between enzymes and their substrates. Additionally, optimizing the entire procedure of protein purification from induced cell strains is a tedious and time-consuming process, which could be easily circumvented by co-spotting yeast strains with different affinity tags. Co-spotting allows the device to be programmed in a variety of combinations by depositing multiple solutions on the same microarray spot. Furthermore, there is no need to stabilize purified proteins as *in situ* extracted proteins from different source cells can interact directly on-chip.

Cell arrays of induced 96 GST strains and a single clone from the GFP collection showed high cell density. We detected pull-down for GST-tagged proteins, visualized by a secondary antibody, but no signal for GFP proteins for any of the tested strains. Arguably, some of the presented GFP strains were expressed in low numbers and therefore may have not reached a critical level to be detected in interactions. However, using GST and GFP-antibodies in different parts of the chip for protein pull-down detected sufficient GFP signal for two tested GFP strains. To further assess the integrity of this approach, well-studied protein-protein interactions will have to be selected in future experiments.

An alternative to generating protein arrays from whole yeast cells is *in situ* synthesis of proteins from DNA templates by cell-free expression. For this path the existing plasmid-based yeast cell libraries may be used to obtain the DNA templates, which is described in chapter 5.

4.5 Conclusions, potential pitfalls and future directions

In this project, an alternative approach to generate a protein array was investigated, that can potentially overcome limitations of conventional technologies and which involves processing of the generated microarray on a microfluidic device with controllable valves, which allow surface modification and mechanical trapping based on the MITOMI principle [36].

Whole yeast cells were deposited and proteins purified *in situ* by cell lysis. We tested two different types of commercially available yeast ORF-fusion libraries for the generation of a protein array that would allow interaction measurements: The GFP-tagged collection with the ORF integrated in its intrinsic chromosomal location and two plasmid-based libraries (GST, MORF). We repeatedly detected protein pulldown for more than 78% of the studied GFP strains with an expression level that significantly higher than negative controls. However, quantitative studies would require normalization of the intrinsically varying GFP signals to cellular abundance of the expressed protein. Unfortunately, we observed only moderate

4.5. Conclusions, potential pitfalls and future directions

correlation, particularly for proteins with a low GFP signal, even after normalization to the number of cells giving rise to the protein expressed on chip. Weak GFP signal from proteins that are normally expressed at high or moderate levels in the cell could result from initial low cell seeding numbers on chip, incomplete cell lysis or the effect that the GFP-tag may have on protein synthesis, transport and function.

As a consequence of these limitations when working with a fusion library with protein expression at a physiological level, we focused the project on the generation of a protein array from plasmid-based yeast cell collections, where all strains express proteins at high levels after induction with a galactose medium. More than 95% of the induced GST strains expressed proteins at levels significantly higher than non-induced control cells, regardless of the number of cells giving rise to this expression. However, on-chip protein expression was not very robust, which could be due to variable cell seeding numbers, incomplete cell lysis and unsynchronized growth of different strains at the time of induction. The conditions for most efficient protein expression were optimized. Nevertheless, large-scale production of yeast cell arrays with the GST collection for protein interaction studies remains difficult. In addition to problems related to array fabrication, the yeast GST library contains numerous non-transformed strains as well as duplicates throughout the 96-well plate collection and several strains repeatedly failed to grow or grew only minimally. The low reproducibility of the library on the level of replication probably also contributed to the overall low reproducibility of the yeast cell arrays. In view of future protein interaction studies, the entire GST collection or a sub-collection of targeted strains should be re-arrayed, as done for all known yeast TFs, to facilitate array production.

For on-chip protein-protein interactions two approaches were envisioned: interaction with purified proteins as well as co-spotting of different yeast strains. Although considered as the gold standard, protein purification has several disadvantages: despite the availability of more convenient purification spin columns and possibilities to automate the procedure, protocol optimization is time-consuming and the process of purification remains laborious as well as challenging for certain proteins. Moreover, purified proteins are instable and their integrity/functionality may be compromised. Given the low performance rate as well as questionable quality (contaminants and truncated proteins) of purified proteins from the MORF and GST collection proteins, we considered co-spotting of yeast cells expressing different fusion proteins as a more promising path. Different strategies were discussed and should be tested with selected strains of well-studied protein interaction sub-networks.

In conclusion, problems for the generation of whole yeast cell array were encountered on the level of cell culture, array fabrication and on-chip purification, which could be mostly solved through step-wise parameter optimization. Challenges included homogeneity of the cell array and reproducibility of on-chip protein expression, especially in large-scale studies

involving several hundred different yeast fusion strains. In particular, the use of plasmid-based fusion libraries requires strenuous monitoring of cell growth and synchronization to establish robust conditions prior to induction of protein expression.

In addition to array fabrication, other factors can influence the detection of protein interaction with the described method: Obviously, after on-chip protein extraction an abundance of other cellular proteins, genomic and plasmid DNA diffuses through the unit cell, which may bind to the protein of interest. These binding events can mask the antigen-binding site for protein pull-down or even the potential protein-binding site of the target protein, which is probably present at lower concentrations. In the future, both experimental approaches to generate a protein array are attractive alternatives to study protein-protein or protein-DNA interactions. The re-arrayed yeast GST-fusion collection containing all known TFs can be used to study protein interactions on a yeast TF-array, before moving on to a larger scale covering the yeast proteome.

5 Protein arrays from DNA template

5.1 Abstract

Producing proteins from pre-arrayed DNA or RNA directly on-chip using cell free expression systems has been a promising approach to overcome issues of conventional protein arrays such as protein on-chip stability, protein purification and efficient expression. However, generation of expression-ready DNA templates often involves multiple cloning, amplification and purification steps. Here we developed a simple one-step PCR-based approach to generate linear DNA templates directly from cell cultures to synthesize proteins on a microfluidic device, suitable for large-scale protein interaction studies.

5.2 Introduction

Cell-free protein synthesis or *in vitro* transcription/translation (ITT) has been established as a technique to rapidly and efficiently express virtually any protein outside the cellular environment from an existing DNA template. Because *in vitro* expressed proteins do not require cell culture and protein purification, it is considerably faster than *in vivo* expression of recombinant proteins in cellular hosts, such as *E. coli* or *S. cerevisiae*. Thus, potentially all types of protein can be generated and analyzed by cell-free protein expression, including those which are toxic to a host cell, normally produced in low abundance, difficult to transport across membranes or prone to proteolytic degradation. Cell-free systems can be used to express a single gene or a DNA library and to investigate affects of single point mutations by rational DNA template design. ITT also offers many advantages over chemical synthesis of proteins as it is not limited by protein size and can be easily scaled to proteomic throughput. Applications range from structural proteomics such as basic understanding of the molecular principles of mRNA translation into functional proteins [164], high-throughput functional proteomics in protein interaction studies [36, 10, 13] and the experimental determination of regulatory and evolutionary plasticity of proteins [12, 165, 96], or drug target discovery [11],

diagnostics and therapeutics (reviewed in [166]).

Cell-free protein expression is based on two main ingredients: the genetic template (DNA or mRNA), which encodes the target protein, and the transcriptional and translational molecular components, which are usually provided by cell extracts from various cell sources. Systems based on cell lysates of *E. coli*, wheat germ and rabbit reticulocytes are most common, but cell-lysate-free systems with individually purified and reconstituted components become more popular for highly controlled reaction conditions and easy purification of untagged proteins [167]. Choosing the appropriate system depends on the origin and nature of the protein as well as required yields and downstream applications. Although generally lower in yields than *E. coli* based systems, eukaryotic systems provide better platforms to study post-translational modifications. Transcription and translation in ITT systems are by definition coupled and provide a cost-, time- and resource-efficient tool for high-throughput protein synthesis, which is suitable to on-chip screening applications. Cell-free systems can be easily supplemented with further components, including labeling for downstream protein detection, reagents to improve protein folding and post-translational processing, or modified tRNA to incorporate non-natural or chemically modified amino acids in the protein.

Expressing proteins with cell-free protein synthesis systems requires expression-ready DNA templates of the open reading frame (ORF), either in the form of plasmids or linear DNA fragments generated by PCR, which contain a promoter (typically T7, SP6 or T3) and a translation initiation signal such as the Shine-Dalgarno (prokaryotic) or Kozak (eukaryotic) sequence. A transcription and translation termination region can be incorporated to increase expression levels. Cloning the ORF of interest into plasmids involves PCR steps, digestion by restriction enzymes, ligation and transformation into a suitable host, ideally followed by sequencing to ensure that the insert is correct. Obviously, this laborious procedure makes plasmid-based templates unfeasible for large-scale applications. Although plasmids provide a more robust template, generating linear expression templates is generally the method of choice.

5.3 Material and Methods

5.3.1 Linear template generation

Linear expression templates were generated by two PCR methods: from purified plasmids of the yeast MORF fusion collection or directly from liquid cell culture of the same collection, hosted in *S. cerevisiae* and *E. coli*, respectively (all distributed by Open Biosystems, now Thermo Fisher, Switzerland).

5.3.1.1 Plasmid purification from MORF-fusion strains

This procedure was adapted from the suppliers protocol for isolation of plasmid DNA from *Saccharomyces cerevisiae* strains using the QIAprep Spin Miniprep Kit. Selected strains were inoculated from the frozen glycerol stock into 5 ml growth medium, grown overnight (16-24 h) at 30 °C with low agitation, harvested by centrifugation (4,000 rpm, 10 min, 4 °C), washed with 250 µl PBS and pelleted again (4,000 rpm, 10 min, 4 °C). Cells were lysed in either of the following two procedures: In the first one, following the instruction manual, the pellet was resuspended in 250 µl buffer P1 containing 0.1 mg/ml RNase A and vortexed for 5 min with 100 µl sterile glass beads (diameter 0.5 mm, Sigma). In the alternative procedure, the pellet was resuspended in 50 µl Lyse-and-Go PCR Reagent (Pierce) and incubated for 5 min at 95 °C. Suspensions of both methods were transferred to another tube containing 250 µl lysis buffer P2 and inverted gently 4-6 times to mix. After 5 min incubation at RT 350 µl neutralization buffer N3 was added to the tube and inverted immediately but gently 4-6 times. The lysate was centrifuged (14,000 rpm, 10 min, RT) and the supernatant transferred to a QIAprep Spin Column inside a 2 ml collection tube, centrifuged (14,000 rpm, 60 sec, RT) and the flow-through discarded. The spin column was washed with 0.75 ml of Buffer PE, centrifuged (14,000 rpm, 60 sec, RT) and the flow-through discarded. Residual wash buffer was removed by additional centrifugation for 1 min. The spin columns were transferred to a fresh tube before plasmids were eluted by adding 25 µl of Buffer EB (10 mM Tris-Cl, pH 8.5) to the center of each spin column, letting it stand for 1 min and a final centrifugation for 1 min. The quality of the purified plasmids was then assessed by running 2 µl sample on a 1% agarose gel and measuring their concentration and purity with a NanoDrop 1000 spectrophotometer (Wileg AG, Switzerland). Purified plasmids were stored at -20 °C for further processing.

5.3.1.2 Linear template synthesis by two-step PCR

Initially, linear expression templates were generated by a slightly modified version of the two-step PCR method described by Maerkl *et al.* [36], in which the first step amplifies the target sequence and the second step adds the required 5'UTR and 3'UTR components for

efficient ITT as well a fluorescence label through a third set of primers. The first step PCR reaction contained 0.45 μM of each gene specific primer in 25 μl ready-to-use High Fidelity PCR Master mix (Roche). Template was added by pipetting 20 ng/ μl purified plasmid or by transferring cell template from a glycerol stock ($\text{OD}_{600}=4.0$) with a 96-well metal solid pin replicator tool (V & P Scientific). The reaction mix was cycled for 2 min (plasmid template) or 10 min (cell template) at 94°C, followed by 10 cycles of 10 sec at 94°C, 70 sec at 50°C (annealing) and 2 min at 72°C (elongation), then another 20 cycles of 15 sec at 94°C, 30 sec at 50°C (annealing) and elongation for 2 min + 5 sec for each successive cycle at 72°C, followed by a final extension at 72°C for 7 min. The second PCR contained 25 nM of each extension primer in 23.5 μl ready-to-use High Fidelity PCR Master mix, to which 1 μl of the unpurified product from the first PCR step was added as template. The reaction was cycled for 2 min at 94°C followed by 10 cycles of 10 sec at 94°C, 70 sec at 50°C and 2 min at 72°C followed by a final extension of 72°C for 7 min. After extension 0.5 μl of final primer mix (0.5 μM final concentration for 5'finalCy5 and 3'final in dH_2O) was added to each reaction tube. In a 96-well plate setup, products of the extension PCR were transferred to another PCR plate containing 0.5 μM of each final primer in a 25 μl or 50 μl ready-to-use High Fidelity PCR Master mix (Roche) with a 96-well metal solid pin replicator tool (V & P Scientific). Cycling was then continued immediately for 2 min at 94°C, followed by 10 cycles of 10 sec at 94°C, 70 sec at 50°C (annealing) and 2 min at 72°C (elongation), then another 20 cycles of 15 sec at 94°C, 30 sec at 50°C (annealing) and elongation for 2 min + 5 sec for each successive cycle at 72°C, followed by a final extension at 72°C for 7 min. Unpurified products of each reaction step were assessed by running 0.5 μl sample on a 1% agarose gel. The same settings for the extension and final PCR were used to generate expression templates directly without the first PCR. In this 1.5-step PCR 0.5 μl plasmid (20 ng/ μl final concentration) or 1 μl yeast cells served as template for the extension PCR.

5.3.1.3 Linear template synthesis by single-tube colony PCR

Linear expression templates were generated by single-tube PCR in a 96-well plate format by transferring cell template from a glycerol stock ($\text{OD}_{600}=4.0$) with a 96-well metal solid pin replicator tool (V & P Scientific) to 50 μl ready-to-use High Fidelity PCR Master mix containing 0.45 μM of each extension primer (as used for the 2-step PCR). The mix was heated at 94°C for 10 min, followed by 10 cycles of 10 sec at 94°C, 70 sec at 50°C (annealing) and 2 min at 72°C (elongation), then another 20 cycles of 15 sec at 94°C, 30 sec at 50°C (annealing) and elongation for 2 min + 5 sec for each successive cycle at 72°C, followed by a final extension at 72°C for 7 min. Unpurified products were assessed by running 0.5 μl sample on an 1% agarose gel.

5.3.2 DNA spotting and device alignment

Solutions of unpurified expression-ready DNA templates were spotted onto epoxy-coated microscope slides with 373 μm column and 746 μm row spacing using a QArray2 microarrayer (Genetix) and 946MP2 slit-pins (ArrayIt). Sample solutions contained 2(w/v)% BSA/dH₂O to prevent covalent linkage of DNA to the epoxy surface and to visualize DNA spots for alignment to the microfluidic device. Spotted DNA arrays were manually aligned to microfluidic PDMS devices containing 768 unit cells using a Nikon SMz1500 stereoscope and bonded overnight at 40°C.

5.3.3 On-chip experiments for linear template arrays

The surface area was modified by depositing layers of BSA-biotin and NeutrAvidin/PBS as described in Appendix A [162]. The surface derivatization was finished by immobilizing biotinylated Penta His-antibody or HA-antibody in PBS (see Appendix D) to the area beneath the button membrane, followed by washing with PBS.

Expression of linear templates was first tested by setting up a test tube containing 12 μl of either TNT T7 Coupled Wheat Germ Extract System (Promega) or TnT SP6 High-Yield Wheat Germ Protein Expression System (Promega) with 0.5 μl tRNA-Lysine-Bodipy-Fl (Promega) to which 0.5 μl unpurified expression template from the 1.5-step PCR were added. The ITT mix was incubated for 90 minutes in a PCR cycler (BIO-RAD) at 30°C (T7) and 25°C (SP6), respectively, before flowing them on the prepared chip surface.

For on-chip protein expression, an *in vitro* transcription-translation (ITT) mix containing 12 μl TnT T7 Coupled Wheat Germ Protein Expression System, spiked with 0.5 μl tRNA-Lysine-Bodipy-Fl was loaded into the chambers of the device to solubilize the linear DNA. The outlet valve and the button membrane were kept closed during the loading of the ITT mix to protect the antibody underneath the button membrane and to prevent cross-contamination by DNA samples that were flushed from the chamber. After complete resuspension of the DNA in the chamber, the neck valve was closed again and the channels washed with ITT mix for 5 min to remove potentially flushed out DNA. The reaction then incubated for 2 h at room temperature with the sandwich valves closed and the neck valve and button membrane open to allow the expressed proteins to diffuse to the antibody. A final washing step with 1% BSA in PBS for 45 min with the button membrane and neck valve closed removed excess ITT mix from the channels for both experimental approaches (on-chip and off-chip ITT), after which the device was imaged on a modified ArrayWoRx (Applied Precision) microarray scanner with 1.0 s at 488 nm.

5.4 Results and Discussion

5.4.1 Template generation

We applied a PCR-based approach to generate linear DNA templates from the yeast MORF-fusion library, described in chapter 4. Compared to the GST-fusion library, which was the library of choice for the yeast cell arrays, the MORF collection offers several advantages: ORFs were tagged at their C-terminal with a versatile triple affinity tag unlike the N-terminal GST-6x-His-tag in the GST collection, which could result in the detection of truncated proteins (detection of tag without ORF). Moreover, the quality of the commercially available GST yeast library was lower than the MORF library, with respect to cell library and data base maintenance.

Wheat germ ITT (Promega) was chosen over *E. coli*-based systems for this study on yeast proteins for two reasons. Multi-domain proteins, often found in eukaryotes, tend to fold into their correct conformations much better in eukaryotic rather than prokaryotic translation systems [168]. In addition, PCR-generated DNA templates are generally transcribed and translated less efficiently in *E. coli* systems because mRNA and DNA-degradation enzymes from the cells decrease the stability of the templates and this reduces yields. Moreover, wheat-germ-based expression was optimized for higher yields using a PCR-based approach developed by Sawasaki *et al.* [169]. Although PCR is the fastest way to generate linear DNA templates, most approaches involve a time-consuming multi-step PCR. On the other hand, they are extremely flexible and modular; and linear templates can be generated from genomic DNA, cDNA clones and other plasmid-based ORFs. Nevertheless, optimization of the PCR and usually purification of the PCR products are required to minimize the presence of contaminants, such as primer dimers, in the final template. To easily generate linear expression-ready DNA templates from genomic DNA, Maerkl *et al.* developed a two-step PCR method, in which the first step amplifies the target sequence and the second step adds elements to the 5'UTR and 3'UTR (untranslated region) for efficient *in vitro* transcription/translation [36]. With the short gene-specific primers in the first PCR they added the translational initiation sequence (start codon and a Kozak sequence in the case of eukaryotes) and a double-stop codon to the template as well as overhang sequences, which allowed downstream priming in the second PCR using a set of extension primers. The 5' extension primer contained the required promoter sequence and a beta-globin sequence to increase expression. A Poly-A tail and a terminator were added to the 3' extension primer to enhance mRNA stability and transcription. After 10 cycles of extension PCR, a set of short final primers was added, which not only amplified and labeled the already functional linear templates but should also reduce the chance of primer dimer formation.

5.4.1.1 Template generation by single-step PCR using purified plasmids

One of the goals of this project was to find faster, less expensive and more direct ways to generate large scale protein arrays to enable protein interaction studies on a proteome-wide level. However, the 2-step PCR still required numerous handling steps, like transferring PCR products and adding the final primer set, which were impractical for the generation of templates from a library consisting of almost 6,000 ORF-fusion strains. Even with the aid of facilitating tools, such as a 96-replicator tool and liquid dispensing systems, the entire process would become rather laborious, reagent consuming and prone to manual errors in a large-scale 96-well-microplate format.

Therefore, the set out to design a new streamlined PCR to overcome these limitations, if possible as a one-tube, one-step reaction. Critical points addressed in the original 2-step PCR were still considered, such as a highly modular PCR design, which uses commercially available primer sets with a maximum length of 135 bp. Additionally, linear templates should be compatible with on-chip processing without the need of spin column based purification steps. Two methods were developed and compared to the 2-step PCR results. In a first approach, the 5'UTR and 3'UTR elements were added to the target gene/region of interest during the first round of PCR using long extension primers (between 88 and 123 bp), as shown in Figure 5.1.

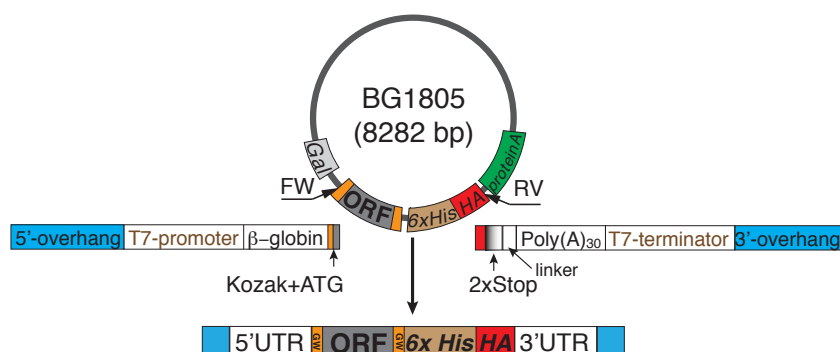


Figure 5.1: Primer design to generate a linear DNA template by colony PCR from a plasmid-based cell collection. Primers were designed, which contained all the required elements for efficient cell-free protein expression using a commercial *in vitro* transcription-translation system: a T7 promoter and a T7 terminator were added to be compatible with commercial *in vitro* transcription systems using bacterial RNA polymerase to initiate gene transcription. In addition, a beta-globin sequence was added to increase the translational rate as well as a Poly-A tail to stabilize the mRNA after transcription. The translational initiators, Kozak sequence and Start codon were already present in the plasmid, but a double stop codon was added after the HA-tag. The blue-colored 5'- and 3'-overhangs can be used for further amplification of the resulting linear template in another PCR step to add a fluorescence tag to the sequence and reduce the chance of primer dimer formation in the PCR mix, if needed.

The resulting linear DNA product was then amplified and labeled with a Cy3 fluorescence tag in a final PCR by adding a short primer set. The fluorescence label allows visualizing the linear DNA template on a microarray after spotting onto a glass slide. In a proof-of-principle experiment using purified plasmids from the MORF yeast library all templates were amplified

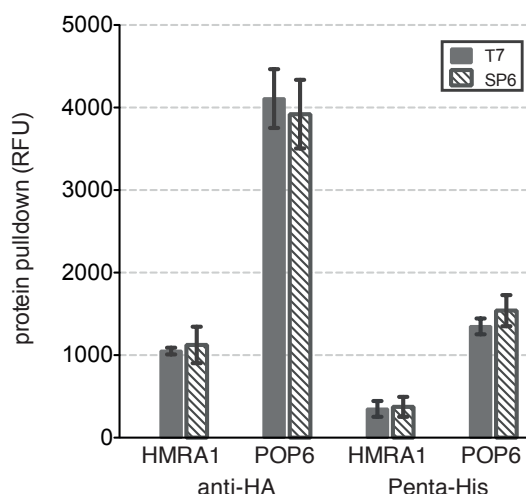


Figure 5.2: Protein expression levels of templates generated by one-step plasmid PCR. Two proteins of similar size (protein size: 34 kDa for HMRA1 and 37 kDa for POP6) were synthesized off-chip using the SP6 and the T7 ITT expression system, respectively, from linear DNA templates generated by a double-step plasmid PCR. The protein-ITT-mix was then flown on-chip and detected by either anti-HA antibody or Penta-His antibody.

at the correct size in sufficient quantity for downstream processing (dilution for spotting). On-chip detection of two proteins (HMRA1 and POP6) synthesized showed slightly similar signals for proteins expressed off-chip with the SP6 and T7 system, respectively (Figure 5.2).

Although, skipping the gene-specific PCR amplification saved one transfer step, it still required adding 0.5 μ l of final primer to each well in a 96-well plate or copying into another plate. The final PCR serves three functions: (i) the amplification of the generally low number of extension PCR product due to the length of the primer set, (ii) reducing the chance of primer dimer formation (iii), and the possibility to incorporate a fluorescence tag with a shorter primer pair, which is much cheaper than adding the tag to the long extension primers. However, in principal the PCR products are already functional after the extension step and thus the final PCR is not strictly required. Running the linear template on a gel showed that most samples were already sufficiently amplified after the extension PCR step without significant primer dimer contamination. In a second approach, the PCR conditions were modified to enable a one-step, one-tube PCR by using only extension primers. In addition, a single-step PCR reduces the total number of PCR cycles and therefore the risk of accumulating point mutations due to the error rate of the DNA polymerase.

5.4.1.2 Template generation by colony PCR

To further optimize the PCR conditions, templates ought to be generated directly from liquid cell culture, which circumvents the need of plasmid purification. Commercially available spin column systems generally facilitate plasmid purification from liquid cell culture or colonies. Although the yield and purity of plasmids purified from yeast cells was only moderate, it proved

sufficient for PCR amplification. Arguably, if very low plasmid concentrations allowed the generation of linear expression DNA templates, it would be worthwhile to test the amplification after lysing yeast cell cultures. In order to keep handling steps to a minimum, boiling yeast cells for 10 min before PCR was the preferred cell lysis method over adding detergents as a heating step can be easily implemented in the PCR cyclor without additional transfer of lysed yeast cells. DNA band signals on the gel after single-step PCR from lysed yeast cells were weaker than after single-step PCR from purified plasmids, probably due to lower plasmid concentrations after cell lysis. PCR failed repeatedly for 20-33% of the lysed yeast strains, even though plasmid PCR had worked for some of them. This high failure rate during colony PCR could indicate too low cell numbers as starting material or incomplete cell lysis. The first hypothesis was ruled out as it proved that highest DNA yields were obtained in PCR reactions using the lowest cell numbers, which had been transferred with a 96-pin replicator tool. Cell lysis with detergent or enzymatic reagents, such as Y-PER, Lyse-and-Go PCR Reagent or Zymolyase even seemed to impair the quality of the PCR. The MORF plasmids were also transformed in *E. coli* cells, but the library format/setup differed from the yeast collection, which made plate-to-plate comparison difficult. Single-step PCR from lysed *E. coli* cells instead of yeast showed an overall higher success rate (always more than 85%). However, when comparing PCR results between the same transformants, *E. coli* was only slightly better than yeast. Figure 5.4 A shows that DNA fragments were amplified equally well for a DNA length ranging from 369 to 1400 bp.

5.4.2 On-chip protein expression

Linear DNA templates can be easily transferred to a microfluidic device and converted into a protein array by introducing a cell-free expression system, without the need of further purification (Figure 5.3).

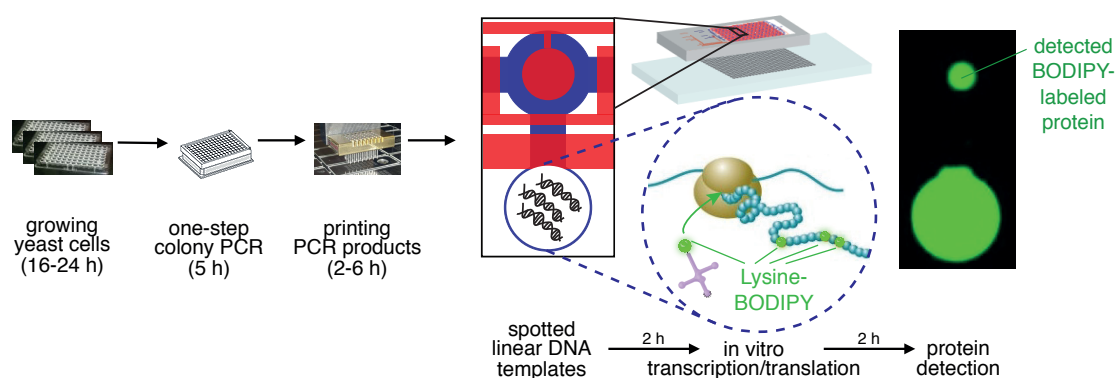


Figure 5.3: Workflow of cell-free protein synthesis on-chip: Products from the one-step colony PCR were printed on a glass slide and aligned to a MITOMI device. An *in vitro* transcription-translation (ITT) mix was then loaded into the chamber to drive protein expression. The mix also contained a tRNA carrying a BODIPY-labeled lysine amino acid, that is incorporated into the final protein, allowing its detection once it is bound to the detection antibody.

Chapter 5. Protein arrays from DNA template

As a proof-of-principle on-chip experiment, 16 MORF-fusion proteins were expressed from spotted expression templates, which had been generated by single-step colony PCR of yeast and *E. coli* cells using a wheat-germ *in vitro* transcription/translation (ITT) system. In a scale-up experiment the MITOMI device was programmed with 64 linear templates, which were obtained by a single-step colony PCR in a 96-microwell plate. The device was then loaded with T7 wheat germ ITT mix and incubated for 2 hours to allow protein synthesis and pull-down to an immobilized HA-antibody. Protein expression signals were detected for almost 70% of the 51 samples, for which a correct PCR product was observed on the electrophoresis gel, with protein sizes ranging from 153 to 496 amino acids (6.3-44.1 kDa) as shown in Figure 5.4 B. The relatively high signals of spots programmed with incorrect PCR products (no or very weak signal, double band, too small fragments) could be caused by non-specific binding or by samples with a dubious DNA template. Moreover, wheat germ extracts contain endogenously biotinylated proteins, which may bind non-specifically.

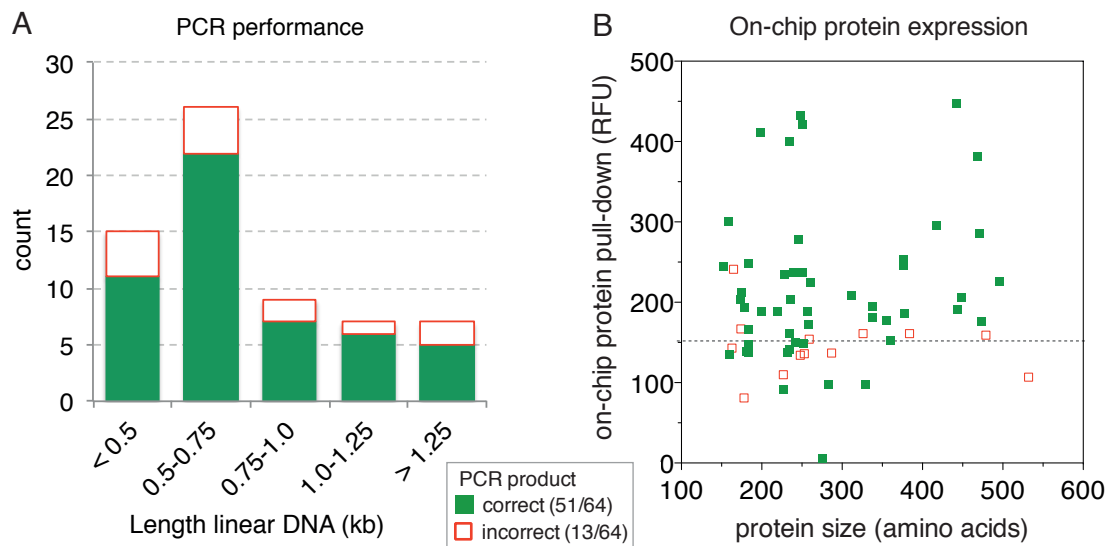


Figure 5.4: On-chip protein expression from DNA templates generated by single-step PCR from lysed yeast cells: Expression levels as a function of protein size for 64 MORF-fusion strains (o = correct size PCR product, x = no or incorrect size of DNA product detected on gel). The signal of almost 70% of expressed proteins was higher than the average signal of the spots without a PCR product (— negative control cutoff).

5.5 Conclusions and Future Directions

In this project, the MITOMI device was charged with linear DNA templates encoding an ORF, from which proteins were synthesized by a cell-free *in vitro* transcription/translation system. Linear expression templates were generated from a plasmid-based cell library. The PCR was designed as a single-step, one-tube reaction that allows amplification and tagging of ORFs as well as incorporating all the 3'UTR and 5'UTR elements necessary for transcription and translation through appropriate primer design. By optimizing the PCR conditions the method was used to generate linear DNA expression templates from cell cultures of yeast and *E. coli* cells. The circumvention of any purification steps renders this approach very useful for large-scale applications. Spotted linear templates were successfully expressed *in situ* using the wheat germ ITT system and proteins detected on-chip. However, using only a single set of primers increases the risk of primer dimer formation, which contain all the necessary 3'UTR and 5'UTR elements and therefore could compete for ITT resources during cell-free protein expression. In the presented study, primer dimer formation was not observed at detectable levels. Undoubtedly, cell-free ITT systems offer a fast, convenient and simple approach to synthesize proteins from DNA. However, wheat germ expression systems contain cellular debris, which include necessary molecular components for protein synthesis. These particles tend to block the microfluidic channels when the ITT mixture is loaded into the chambers with spotted expression templates. Clogging might be a bigger problem when a scale-up version of MITOMI with more unit cells is used, as proposed in 3.2.3 for high-throughput, large-scale studies because the channel dimensions are even smaller. A possible solution is a chip design with higher flow channels, which would allow more efficient purging of fluids in the device.

6 **Drosophila transcription factor array**

6.1 Abstract

Mapping gene regulatory networks (GRNs) is a significant challenge in systems biology, yet only a few methods are currently capable of systems-level identification of transcription factors (TFs) that bind a specific regulatory element. We developed a microfluidic method for integrated systems-level interaction mapping (*iSLIM*) of TF-DNA interactions, generating, and interrogating an array of 423 full-length *Drosophila* TFs. With *iSLIM* it is now possible to rapidly and quantitatively map GRNs of higher eukaryotes.

6.2 Introduction

The fruit fly, *Drosophila melanogaster*, has been instrumental as a model organism in biological research, particularly in genetics and developmental biology, and today it is one of the most widely used and genetically best-known organisms among eukaryotes. Its genome, released in 2000 [170], contains more than twenty times less base pairs (139.5 Mbp) than the human genome (3,200 Mbp) on only four chromosomes, coding for more than 15,000 genes. Many basic processes are shared between *Drosophila* and human, which has made it a great genetic model system to study human diseases, including the neurodegenerative disorders Parkinson's, Huntington's and Alzheimer's disease, as well as developmental defects and other malfunctions [171]. Today, 77% of all known human disease genes were found to have a fly homologue, which can be searched in an online database (<http://homophila.sdsc.edu>) [172]. *Drosophila* is also being used to study mechanisms underlying aging and oxidative stress, diabetes, immunity, and cancer, as well as drug addiction. Moreover, studying fly homologs of various human genes provides new insight into fundamental aspects of protein function. The importance of these studies for human health was recognized with the Nobel Prize in 1995.

Despite the considerable progress in elucidating the function of each gene in the fly genome, we still know little about the regulatory sequences and corresponding gene regulatory networks (GRNs), which control gene expression. This lack of knowledge results from the complexity of GRNs and the difficulties associated with studying them. GRNs consist primarily of transcription factors (TFs), which can bind to many DNA sequences as well as to other TFs or proteins. However, while only 20% of all fly TFs have currently been associated with target genes, most of them were found to bind to multiple cis-regulatory modules on the DNA (source: REDfly database). The sheer number of combinatorial possibilities seems infinite, given the number of TFs encoded by the genome and the potential TF binding sites (TFBS) on target genes. Moreover, TF regulation is often dependent on higher-order interactions: many TFs require dimer formation to bind to their respective target DNA sequence. To complicate matters even more, expression levels are often mediated by the recruitment of transcriptional co-factors (co-regulators) and concentration-dependent binding affinities and dynamics multiply the quantitative impact on regulation of gene expression. Understanding the full regulatory repertoire will be a considerable and tedious task. It would first require identifying all binary TF-TF interactions and their regulatory interaction partners as well as determining their transcriptional properties. This implies a need to measure affinities of individual interactions at varying conditions and integration of all regulatory interactions with their dynamic properties into a multidimensional GRN. Mapping all interactions that constitute a GRN to elucidate central regulatory mechanisms is one of the challenges of systems biology. Experimental and computational methods can help to generate complex regulatory models, but they require large-scale quantitative measurements of TF-DNA interactions. However, the growing information on interaction networks remains predominantly qualitative and binary in nature.

MITOMI is a versatile platform capable of mapping and characterizing 768 independent molecular interactions through physical confinement of individual interactions and a trapping mechanism that eliminates stringent washing steps. This avoids cross-contamination and allows the detection of low-affinity and transient binding, typically found in TF-DNA interactions. MITOMI was shown to be able to map interactions between 43 prokaryotic proteins and generate a small-scale protein-interaction network [10]. Quantitative measurements allowed a TF-centered mapping of binding energy landscapes of 4 eukaryotic TFs to short oligos [36] and determination of binding specificities and affinities of 28 yeast TFs to various DNA-binding domains (DBDs), presented as an oligonucleotide array, which covered the entire genomic DNA sequence space [13].

However, mapping and characterizing all TF-DNA interactions of an organism using the same platform would provide a rapid and inexpensive approach to generate a comprehensive and quantitative regulatory network. The nearly complete repertoire of 755 sequence-specific *Drosophila melanogaster* TFs is ideal to (i) assess the capacity of MITOMI to synthesize hundreds of functional eukaryotic TFs from expression-ready DNA *in situ*, (ii) investigate

its ability to detect specific interactions on a systems level and (iii) measure their binding specificity and affinity in order to map and characterize TF interactions in a GRN.

6.3 Material and Methods

6.3.1 Plasmid purification

Plasmids from overnight *E. coli* cultures were isolated for microvials (5 ml) and 96-well plates following exactly the suppliers' protocols QIAprep Spin Miniprep Kit and Plasmid DNA Purification using the QIAprep 96 Turbo Miniprep Kit, respectively (both Qiagen, see Appendix F).

6.3.2 Construction of a linear DNA template library

A collection of open reading frames (ORFs) coding for all known *Drosophila melanogaster* TFs was previously cloned into Gateway compatible destination vectors [74] with a C-terminal GST-tag.

A universal set of primers (IDT) was designed framing the Gateway cassette on the pF3A-WG-GST plasmid with a 5' primer (see Appendix B) located 10 bp upstream of the SP6 promoter and a Cy3-labeled 3'-primer (see Appendix B) located 24 bp downstream of the T7 terminator. We performed colony PCR in standard 96-well PCR plates for 575 samples (for which plasmids had been successfully transformed into *E. coli*). About 0.4 µl cell template from a glycerol stock were transferred into 50 µl of a ready-to-use High Fidelity PCR Master reaction mix (Roche) with a 96-well metal solid pin replicator tool (V & P Scientific). The reaction mix was cycled for 2 min at 94 °C, followed by 10 cycles of 10 sec at 94 °C, 70 sec at 54 °C (annealing) and 2 min at 72 °C (elongation), then another 20 cycles of 15 sec at 94 °C, 30 sec at 54 °C (annealing) and elongation for 2 min + 5 sec for each successive cycle at 72 °C, followed by a final extension at 72 °C for 7 min. Following the High Fidelity PCR Master manual, two PCR cycling protocols with different elongation times and temperatures (2 min at 72 °C and 4 min at 68 °C, respectively) were used for each clone to cover most of the expected PCR fragment sizes ranging from 1.5 kb to 7.2 kb.

Correct size of the PCR products was checked on E-Gel 96 1% agarose gels (Invitrogen). A MatLab-based image analysis program automatically detected and analyzed PCR product bands. The readout listed the number of detected DNA fragments (number of peaks) per gel lane, their size and intensity. PCR products for the final TF library were selected based on expected transcription size (+/- 33% from expected size), band intensity ($Int_i > 300$ a.u.), and purity of the PCR product ($P = \frac{Int_i}{\sum Int_i} > 0.8$), as shown in Figure 6.1. Selected PCR products were diluted 1:3 with 2(w/v)% BSA/dH₂O without previous purification and transferred into two 384-well plates using a liquid handling robot (Tecan).

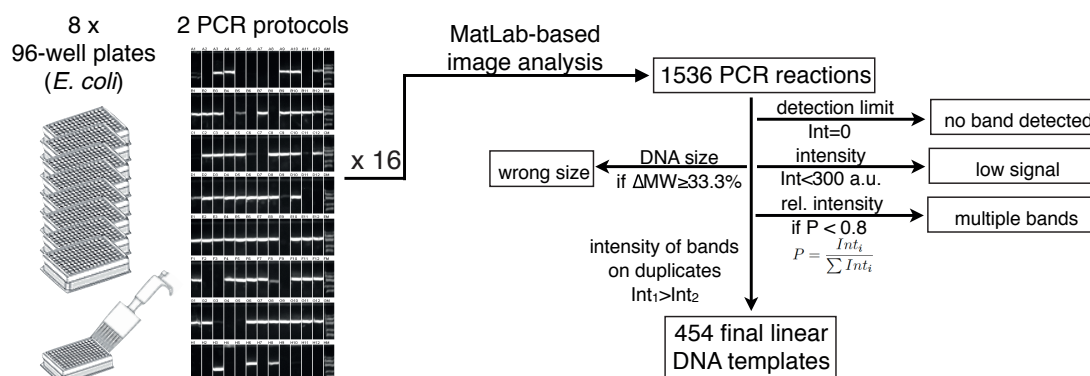


Figure 6.1: PCR read-outs, which did not match our selection criteria for size ($\pm 33\%$ from expected size), band intensity and purity of the PCR product, were excluded from the candidate pool. For samples with successful PCR products in both methods we selected the sample with the higher intensity value on the electrophoresis gel. We double-checked for false negatives within the pool of size-excluded candidates for samples without a positive read-out from both methods.

6.3.3 Synthesis of Target DNA (Klenow)

Single-stranded Cy5-labeled DNA oligo sequence templates for 12 TFs from the even-skipped GRN (bcd, D, gt, kni, Kr, Mad, Med, pan, prd, tin, ttk, twi) [173] were ordered from IDT (see Appendix B). To synthesize double-stranded DNA target sequences, we prepared a 20 μ l primer mix containing 2 μ l of 150 μ M target DNA sequence, 0.4 μ l 5' Comp Cy5 primer (see Appendix B), 2 μ l 10x NEBuffer 2 (NEB) and 15.6 μ l PCR-grade water. The primer mix was denatured for 3 minutes at 94 $^{\circ}$ C and then slowly cooled down (0.1 $^{\circ}$ C/sec) to 37 $^{\circ}$ C before adding 10 μ l of a Klenow-fragment-nucleotide mix containing 1 μ l Klenow Fragment (3' \rightarrow 5' exo-, NEB), 3 μ l dNTP (10 mM PCR nucleotide mix, Roche), 1 μ l 10x NEBuffer 2 (NEB), and 5 μ l PCR-grade water. The final 30 μ l mix was then incubated for 60 minutes at 37 $^{\circ}$ C and 20 minutes at 72 $^{\circ}$ C before slowly cooling it to 30 $^{\circ}$ C at a rate of 0.1 $^{\circ}$ C/sec.

6.3.4 De Bruijn DNA library design and synthesis

De Bruijn sequences have essentially been designed and generated as previously described [174]. All possible 8mer DNA sequences (a total of 65536) were computationally assembled into a DNA library of 1523 individual 68-mer oligonucleotides and one 72-mer oligonucleotide. Oligonucleotides with five or more consecutive guanines were replaced by its reverse complement to avoid problems during primer extension. Each oligonucleotide contained multiple, overlapping 8-bp sequences, a 5'-CGC-3' clamp at the 5' end, and a 5'-CTCCGGCGGTATGAC-3' sequence at the 3' end for hybridization to fluorophore labeled extension primer. Double stranded target DNAs were essentially generated as outlined above. However, instead of using only one Cy5-labeled extension primer, we used an additional Cy-3-labeled primer. DNA library was split into equal size of 762 oligonucleotides and extended

with either Cy3- or Cy5-labeled primer. Following completion of extension a pair of Cy3- and Cy5-labeled double-stranded DNAs were pooled and diluted in 1% BSA for spotting.

6.3.5 Microfluidic device fabrication and glass slides preparation

Microfluidic chip design and fabrication, as well as coating of glass slides with epoxysilane, were performed as previously described (see Appendix A) [162]. All experiments were performed with a MITOMI device containing 768 unit cells and a microfluidic multiplexer capable of controlling each flow channel row individually (see Figure 3.2 E).

6.3.6 DNA spotting and device alignment

DNA sequences were spotted onto epoxy-coated microscope slides with 373 μm column and 746 μm row spacing using a QArray2 microarrayer (Genetix) and 946MP2 slit-pins (ArrayIt). Samples with expression-ready DNA templates sample solutions contained 2(w/v)% BSA/dH₂O to prevent covalent linkage of DNA to the epoxy surface and to visualize DNA spots for alignment to the microfluidic device. Target DNA dilutions and the De Bruijn library contained only 1% BSA. Spotted DNA arrays were manually aligned to microfluidic PDMS devices using a Nikon SMz1500 stereoscope and bonded overnight at 40 °C.

6.3.7 On-chip experimental procedure

MITOMI experiments were performed with flow and control channel pressures set to 3 and 17.5-22.5 psi, respectively. All control lines were filled with water. The surface area was modified by depositing layers of BSA-biotin and NeutrAvidin/PBS as previously described [162]. The surface derivatization was completed by immobilizing 1 $\mu\text{g}/\text{ml}$ biotinylated GST antibody (ab71283, abcam, see Appendix D) in 1% BSA/PBS to the area beneath the button membrane and blocking the surface using a blocking solution of 0.5% non-fat dried milk powder, 1% BSA in PBS and 0.1 mg/ml sheared salmon sperm DNA (Sigma).

All TFs were produced using an *in vitro* transcription-translation (ITT) mix containing 5 μl TnT SP6 High-Yield Wheat Germ Protein Expression System (Promega), 0.25 μl tRNA^{Lys}-Bodipy-Fl (Promega) and 3 μl of nuclease-free water (Promega). The channels were first equilibrated with the ITT-mix before solubilizing the DNA spots in the chamber by opening the neck valve without flow. Neck valves were closed again and the channels flushed with the remaining ITT mix to remove DNA molecules that diffused out of the chamber. All ITT-loading steps were performed with the button membrane activated (closed) to protect the antibody. On-chip protein synthesis was incubated for 3 h at room temperature with the sandwich valves closed, which allowed the expressed GST-tagged TFs to diffuse to the antibody. The device was imaged on a modified ArrayWoRx (Applied Precision) microarray scanner with the button membrane open and closed before and after a final wash step, respectively.

6.3.7.1 Detection of TF-DNA interactions

For interaction studies, we pooled all 12 synthesized target DNA and diluted them 1:10 in PCR-grade water. The ITT mix was prepared containing 5 μ l TnT SP6 High-Yield Wheat Germ Protein Expression System (Promega), 0.25 μ l tRNALys-Bodipy-FI (Promega), 0.8 μ l of pre-diluted pooled target DNA mix, 0.32 μ l sheared salmon sperm DNA (Sigma) and 1.63 μ l of nuclease-free water (Promega). Loading of the ITT-DNA mix, incubation and image acquisition were the same as in the above procedure.

6.3.7.2 Correlation of signal detection from BODIPY lysine

For the experiment in which we correlated the amount of BODIPY-labeled lysines that were incorporated into the proteins to the amount of detected proteins, we let a second GST-antibody bind to the BODIPY-labeled surface immobilized TFs. A monoclonal fluorescently labeled GST-antibody (Hilyte Fluor 647, ab64370, see Appendix D) was diluted 1:200 in 1% BSA/PBS to a final concentration of 5 μ g/ml and loaded on chip for 5 min with the button membrane closed. After closing the sandwich valve again and re-opening the button membrane the antibody solution was allowed to diffuse and bind to the GST-tagged proteins for 30 min, after which the device was washed once more for 5 min with the membrane buttons closed and the sandwich valves open and scanned a third time.

6.3.7.3 Concentration-dependent binding measurements (Klenow)

For the concentration-dependent binding measurements, we printed 6 serial dilutions in 1% BSA for each target DNA on epoxy-coated glass slides, aligned and bonded them to the MITOMI device. Surface chemistry was as described for the TF detection studies. Transcription factor proteins were produced off-chip by adding 0.5 μ l linear DNA template (unpurified PCR product) to 5 μ l TnT SP6 High-Yield Wheat Germ Protein Expression System (Promega), 0.25 μ l tRNALys-Bodipy-FI (Promega) and 1.75 μ l of nuclease-free water (Promega) and incubating the ITT-TF mix for 2 h at 25°C.

DNA binding was measured in two different modes: For the monomeric measurements, the supernatant of the ITT-TF mix was perfused over the chip, allowing the TFs to bind to the GST-antibody beneath the button valve. Excess protein was washed off with 1%BSA/PBS and the button membrane collapsed before solubilizing the DNA spots in the chamber by opening the neck valve without flow. DNA was allowed to diffuse to the TF underneath the button area, where we captured interactions by collapsing the button membrane after 60 minutes. For the measurements in dimer mode, the area beneath the button membrane was protected by collapsing the membrane before equilibrating the channels with ITT-mix. DNA spots were solubilized by passive diffusion of the ITT-mix into the chambers. Neck valves were then closed again and the channels flushed with remaining ITT mix to remove

DNA molecules that diffused out of the chamber. During the 1-hour incubation period, TFs were allowed to interact with the resuspended target DNA in the chambers before binding to the immobilized antibody underneath the button membrane. The device was imaged on a modified ArrayWoRx (Applied Precision) microarray scanner with the button membrane open and closed before and after a final wash step, respectively.

6.3.7.4 8-mer library (De Bruijn) on-chip experimental procedure

We programmed the MITOMI device with a library of 1524 oligonucleotides. Surface chemistry and modification as well as off-chip TF production were prepared as described above. The supernatant of the ITT-TF mix was loaded into the chambers with button membranes and outlet valve closed, which allowed the DNA to interact with the TFs. The neck valves were then closed again and the channels flushed with ITT mix, before closing the sandwich valves and allowing the solution to diffuse to the area beneath the button. After 1 hour of incubation, the button membrane trapped surface-bound TF-DNA-complexes and a final PBS wash removed unbound TFs and DNA from the channels.

6.3.8 Data acquisition and analysis

For each experiment, slides were scanned on an ArrayWorx scanner with 0.5 s and 1.0 s exposure times before and after the final wash step at the following wavelengths: 488 nm for BODIPY-labeled proteins, 685 nm for Cy5-labeled target DNA (or anti-GST antibody Hilyte Fluor 647, see Appendix D) and 595 nm for the Cy3-labeled in the De Bruijn experiments. For each experiment two images were analyzed with GenePix6.0 (Molecular Devices). The first image (before washing) determined the available DNA concentration in solution, the second (after washing) the amount of surface bound DNA and protein (see Figure A.5 A). The parallel recording in the DAPI channel allowed a quick detection of the area below the button membrane, which had been labeled with NeutrAvidin Alexa Fluor 350 conjugate. For detection of surface bound protein and DNA a microarray grid was aligned to the stained circular NA spots and used as a template for the detection of protein pull-down and DNA binding events. Local backgrounds were subtracted for all channels by moving the grid just next to the chamber, outside the channels. We excluded unit cells with bad features or insufficient signals from further analysis, which was done with a code written in Mathematica (Wolfram Research).

6.3.8.1 K_d determination for concentration-dependent binding

The concentration of available target DNA for binding in each unit cell was determined by measuring the Cy-5 fluorescence in the chamber after incubation. For this purpose we generated a calibration curve, for which we filled the device with serial dilutions of Cy5-labeled

target DNA in 1% BSA at 8 concentrations (1.25 μ M, 0.625 μ M, 0.156 μ M, 0.078 μ M, 0.039 μ M, 19.5 nM, 9.8 nM and 4.9 nM) and scanned the device at 635 nm with 0.5 s and 1.0 s exposure time, respectively.

Measured Cy5 chamber intensities were converted to concentrations using the calibration curve described above (see Figure A.5 B). For each protein, measured fluorescence ratios r ($r = \text{DNA-Cy5 intensity} / \text{protein-BODIPY-FL intensity}$) were fitted as a function of total available DNA concentration ($[D]$ n RFU) to a single-site binding model to determine the equilibrium binding constant (K_d) using Prism 5 (GraphPad Software):

$$r = \frac{B_{\max} \cdot [D]}{[D] + K_d} \quad (6.1)$$

For each protein, we performed a global nonlinear regression fit for all target sequences over all concentrations with the same B_{\max} , because this maximum intensity ratio, at which all binding sites are occupied, should be the same for all sequences.

6.3.8.2 De Bruijn Analysis and Motif Discovery

The de Bruijn analysis was performed essentially as described by Fordyce *et al.* with the exception that MEME was used for the initial motif discovery instead of fREDUCE [13]. We first normalized the background adjusted Cy3 and Cy5 DNA pull-down signals by dividing each data point by the mean of each dataset. The combined datasets (Cy3 and Cy5 channels) were rank ordered and all oligos above the mean plus two standard deviations were used in an initial MEME search. The MEME search was conducted using the standard MEME tool (v. 4.8.1) using a maximum motif width of 14 bps and the remaining standard settings (zero or one motif per sequence, min motif width 6, max number of motifs to find: 3). If MEME returned a statistically significant motif it was used as a seed motif for MatrixREDUCE [175]. We then ran MatrixREDUCE (REDUCE Suite v2.0) on the dataset using standard settings and the seed motif derived from MEME. In some instances REDUCE could be used directly to provide a seed motif by running the algorithm first without providing a seed motif. The motifs returned by REDUCE and MEME were plotted using LogoGenerator (included with REDUCE Suite v2.0).

6.4 Results and Discussion

6.4.1 Gene-centered approach for the detection of TF-DNA interactions on an on-chip *Drosophila* TF-array

The efficiency/versatility of cell-free protein synthesis from expression-ready DNA templates has been demonstrated in a broad range of biomolecular interaction studies using the MITOMI platform, including protein-protein [10], protein-DNA [36, 13], protein-RNA and protein-small molecule [11].

Expression-ready DNA templates for *in vitro* protein synthesis can be provided as plasmid or linear cDNA containing of the open reading frame (ORF), an appropriate promoter (typically T7, SP6 or T3), a translation initiation sequence (Kozak for eukaryotes) and optionally a transcription and translation termination sequence. For that a collection of 722 Gateway-compatible Entry vectors containing the open reading frame (ORF) for 96% of all 755 predicted *Drosophila* TFs had been generated by incorporating existing cDNA and *de novo* cloning [74]. The 647 sequence-verified clones were then transferred into a pF3A-WG-GST vector by Gateway cloning, containing an SP6 as well as a T7 promoter, a T7 terminator, the Kozak sequence and a GST-affinity tag. More than 75% of TF-ORFs (575 of 755) had been successfully transformed in *E. coli*, which allows replication and plasmid purification to generate an expression-ready DNA template.

6.4.1.1 TF-array from plasmid DNA

Plasmids containing the TF-ORF purified from 96-well plate bacterial cultures using commercial purification systems (QIAprep 96 Turbo Miniprep Kit) yielded low concentrations, even when pooled from cell cultures grown to saturation in up to three plates containing 200 μ l/well (20-40 ng/ μ l) or 96-deep-well plates (max. 1 ml/well). We failed to observe protein expression on-chip using the TNT SP6 High-Yield Wheat Germ Protein Expression System (Promega) for *in situ* synthesis from spotted templates as well as for ITT reactions set up in larger volumes in a test tube and captured on-chip. Indeed, the SP6 ITT instruction manual recommends at least 10 times higher template concentrations for efficient protein synthesis. Using the same experimental setup we detected concentration-dependent signals from 10 mouse TFs and two human receptor proteins, which were expressed on-chip from plasmids. However, the ORF sequences had been cloned into a different vector, which yielded much higher plasmid concentrations after purification (~ 200 ng/ μ l).

It should be noted that purification of plasmids from clones containing the pF3A-WG-GST vector would require large bacterial culture volumes to yield sufficient DNA concentrations.

6.4.1.2 Construction of a linear DNA template library of TFs

Considering that large-scale plasmid purification under the given conditions was uneconomic and too labor intensive, we decided to apply an alternative approach by amplifying the ORFs by colony PCR using a universal Cy3-tagged primer pair for the pF3A-WG-GST vector (Figure 6.3 A). We used two PCR cycling protocols for each clone to efficiently amplify the large range of ORF sizes (1.5-7.2 kb). All PCR reactions were validated by gel electrophoresis and automatically assessed for quality using a MatLab-based image analysis program, which returned the number of detected DNA fragments for each PCR product, its size and intensity on each 96-well PCR gel. PCR performance was assessed based on expected transcription size ($\pm 33\%$ of expected size), band intensity and purity of the PCR product. We successfully amplified 454 of 575 ORFs (79%), spanning the entire size range of up to 7.2 kb (Figure 6.2 A) and distributed uniformly among all major TF families (Figure 6.2 B).

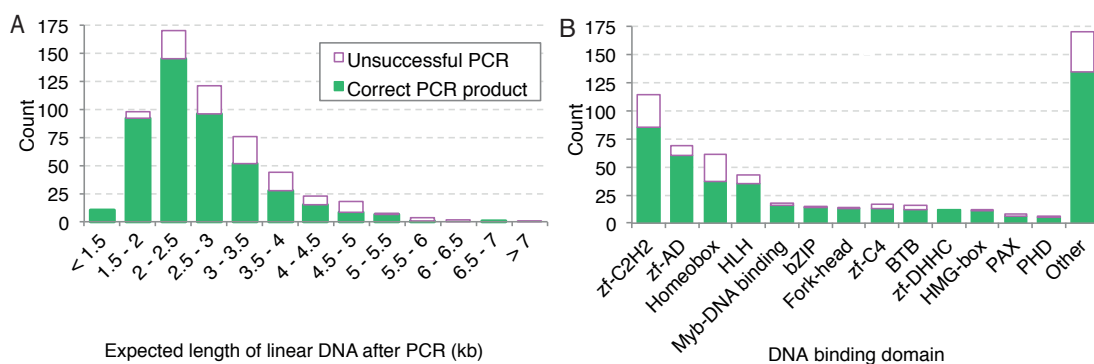


Figure 6.2: PCR performance of generated linear DNA templates after quality assessment. Histograms of PCR success rate of (A) expected DNA fragment size, and (B) DNA-binding protein domains within the Drosophila TFs.

The linear template library generated by PCR is sufficient to spot dozens of arrays from PCR products in BSA. The PCR reactions can be used directly for array generation without requiring additional purification steps, which further streamlines the approach. In addition, spotted template arrays can be easily visualized because of the Cy3-fluorophore that was added to the PCR product via the reverse primer.

6.4.1.3 Generation of an on-chip Drosophila TF-array

We programmed the MITOMI device consisting of 768 unit cells with the linear template library (Figure 6.3 B) and expressed proteins *in situ* using the SP6 ITT kit.

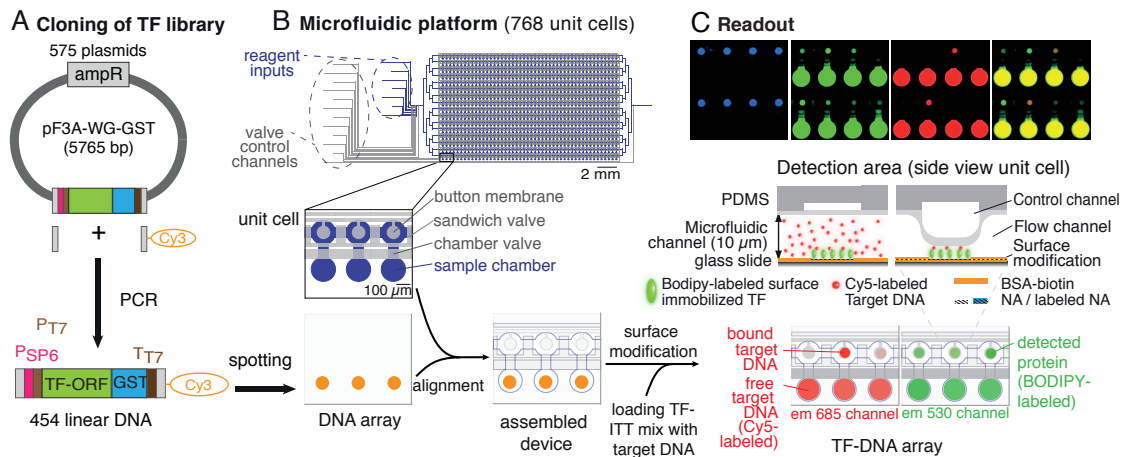


Figure 6.3: Experimental workflow of the *iSLIM* approach: (A) Linear DNA templates were generated from plasmids by PCR. (B) PCR reactions were directly spotted onto a glass slide and aligned to a microfluidic device containing 768 unit cells. After surface modification and on-chip protein synthesis, TF-DNA interactions are detected by scanning the device on a fluorescence DNA microarray scanner. (C) A typical readout shows the TF-DNA pull-down area, stained with AlexaFluor350-labeled NeutrAvidin (blue), TF expression (green), bound target DNA (red) and the TF-DNA interaction of the merged channels.

Protein expression levels were characterized by relating BODIPY intensities under the button valve area to the number TFs carrying a lysine-BODIPY-charged tRNA, which were bound to the surface-immobilized antibody. This correlation was confirmed when additionally detecting protein pull-down with a secondary Hilyte Fluor 647 labeled GST-antibody (Figure 6.4). Detected protein signals (in relative fluorescence units, RFU) can thus be correlated to expression levels after normalization to their lysine counts (Figure 6.7).

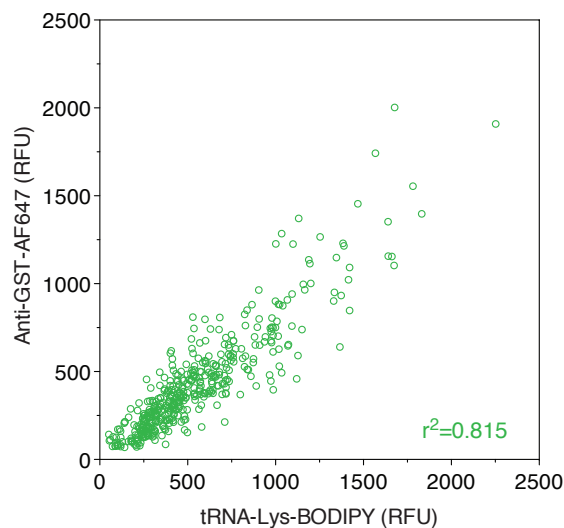


Figure 6.4: Correlation between signal intensities of protein pull-down labeled with a lysine-BODIPY-charged tRNA and a secondary GST-antibody (Hilyte Fluor 647) on the same TF array.

The sensitive detection of different BODIPY signals allows a quantitative analysis of the amount of protein present for each TF, unlike in Y1H methods where TF expression levels are not assessed.

Detection of TF expression was considered positive for BODIPY signals that showed 2 s.d. above average noise level of negative controls (only BSA spotted). We detected protein expression that was significantly higher than the negative controls (only BSA) on average for more than 90% of the DNA templates with very good reproducibility ($r^2=0.85$) as shown in Figure 6.5.

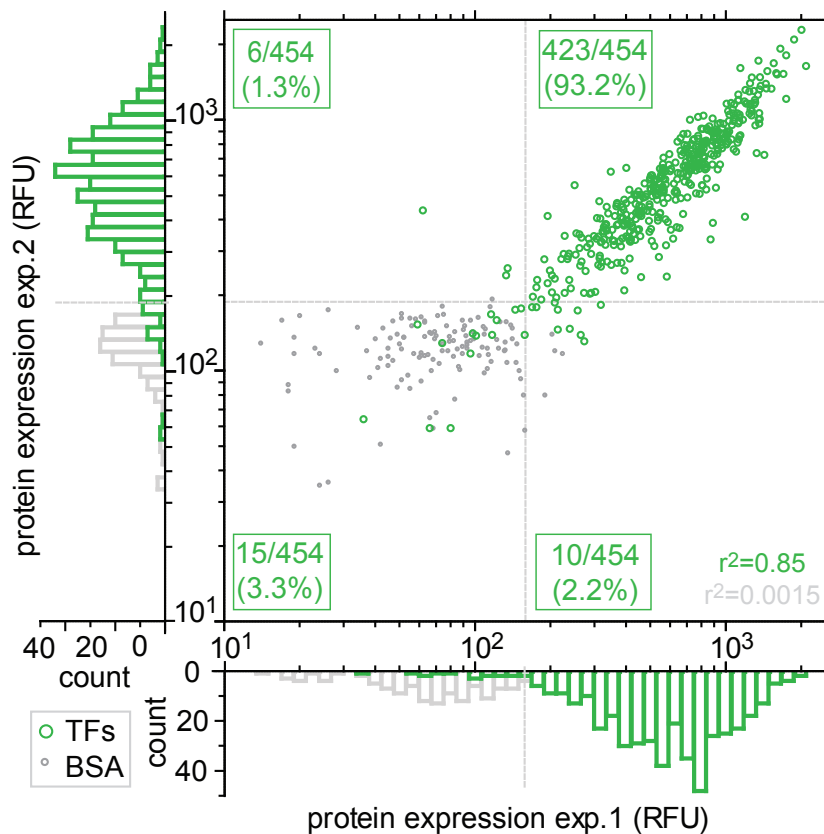


Figure 6.5: Correlation of protein expression levels (TF) and negative controls (BSA) from two different chips. Dotted gray lines indicate the background threshold (2 s.d. above the mean of the negative control).

Overall, we successfully detected 423 full-length drosophila TFs on-chip (Figure 6.5), which constituted 56% of all 755 predicted fly TFs. Protein expression was uniformly detected among all major TF families (Figure 6.6 B). We expressed 93.2% of the 454 linear DNA template library with protein sizes ranging from 317 to 2078 amino acids (37-231 kDa) (Figure 6.6 A). By comparison, in a previous study 43 *S. pneumonia* proteins were detected ranging in size from 35-757 amino acids (4-83 kDa) on a microfluidic device [10]. The protein array generated in this study thus increases throughput by an order of magnitude and consists of large eukaryotic proteins.

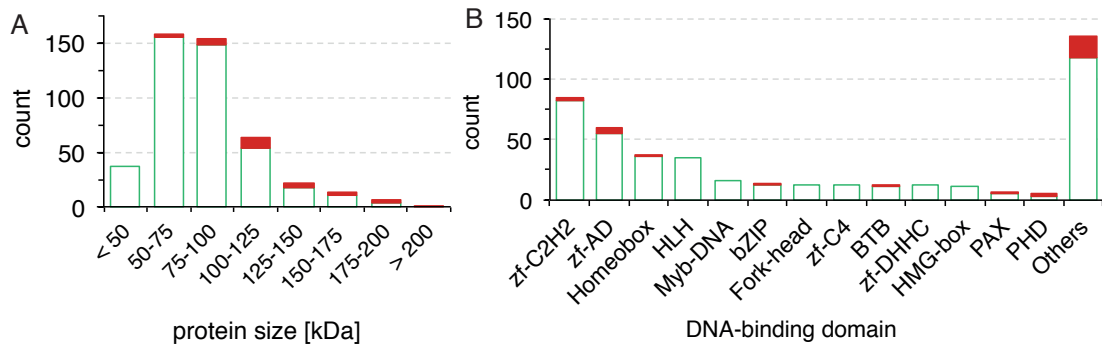


Figure 6.6: Distribution of on-chip protein expression: Histograms of (A) protein size (B) and DNA-binding domains of the expressed *Drosophila* TFs.

Contrary to the study on 43 prokaryotic proteins, we repeatedly observe a correlation between protein size and expression levels (Figure 6.7), which could have several reasons: Larger proteins take longer to synthesize and their size may prevent more proteins to bind to the surface-immobilized antibody. In addition, the accessibility of the GST-tag may be compromised for binding to the antibody.

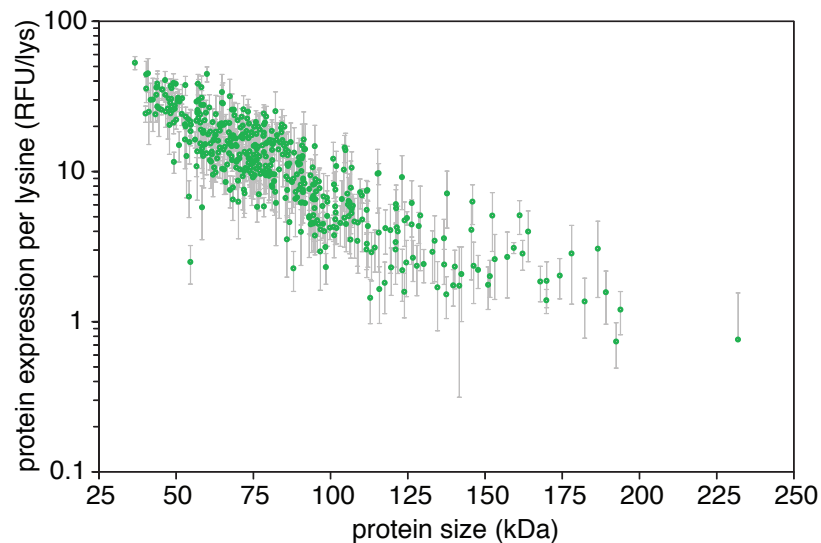


Figure 6.7: Correlation of protein expression levels and expected protein size: protein expression levels were normalized to the number of lysine residues per protein and plotted as a function of protein size for all 454 TF (error bars: standard deviation; $n=6$).

6.4.1.4 Detection of TF-DNA interactions on chip

After obtaining quantitative protein expression levels, the TF-array was assessed for the ability to detect TF-DNA interactions. Recent interrogations of TF binding specificities by ChIP-seq and *in vitro* methods suggest that TFs exhibit considerable cross-reactivity and that promiscuous binding can be of functional significance [176]. To investigate the extent of

Chapter 6. Drosophila transcription factor array

TF promiscuity on a systems-level, we used the MITOMI platform to measure the binding of 12 DNA consensus motifs for TFs involved in the well-studied even-skipped GRN (bcd, D, gt, kni, Kr, Mad, Med, pan, prd, tin, ttk, twi) [173] against all 423 TFs present on our array. The 12 double-stranded DNA motifs were synthesized by isothermal primer extension of single-stranded Cy5-labeled oligonucleotide sequences. We loaded all 12 DNA sequences as a pool together with the ITT mix into the chambers containing linear DNA templates to synthesize TFs. Adding the target DNA motifs as a pool to the ITT resembles closer the situation inside the cell, where TFs have to discriminate between an abundance of genomic DNA. The detected Cy5 intensity of TF-bound DNA correlates to the number of DNA molecules bound to surface immobilized TF, which is reflected by the BODIPY intensity in the same area. Interrogating all 12 DNA motifs as a pool allowed us to test a total of 5,076 interactions in a single experiment (423 TFs x 12 DNA motifs). For 11 of these 12 motifs, the targeted TF was present on our array. We observed DNA binding by 4 TFs (Kr, gt, prd, D), while 7 TFs failed to bind DNA above background (bcd, kni, ttk, tin, Mad, twi, pan) as shown in Figure 6.8.

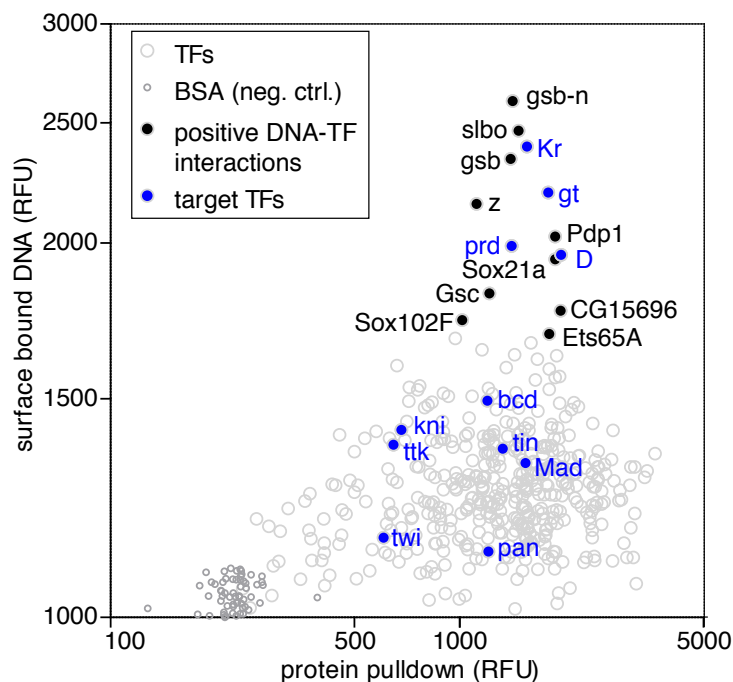


Figure 6.8: Gene-centered detection TF-DNA interactions on the Drosophila TF-array: scatter plot showing the DNA binding as a function of protein expression of all 423 TFs on the array to a pool of 12 DNA consensus motifs for TFs involved in the even-skipped gene regulatory network (GRN). Interactions are represented as grey circles. Highlighted are targeted TFs expressed on the array (blue) and additional TFs that interacted with the consensus sequences (black).

Amongst these 7 TFs, twist (twi) is known to require dimerization with another basic helix-loop-helix (bHLH) domain for efficient DNA binding [177], and Mad was shown to bind DNA only in a truncated form [178]. Given that Mad and twi were not expected to be functional

we arrive at 4 out of 9 TFs (44%) that functionally bound to DNA in the context of this particular set of proteins. This relatively low detection rate still compares well to a recent large-scale Y1H study, based on the same ORF library [74], which identified 26% of the control interactions, as well as other Y1H screens on *C. elegans* (25-33%) [86, 69].

6.4.2 TF-centered approach to characterize TF-DNA interactions

6.4.2.1 Affinity and specificity measurements

Interestingly, in addition to previously known interactions, we also identified 11 TFs (gsb-n, gsb, Pdp1, slbo, z, Sox21a, Sox102F, Gsc, Ets65A, bun, CG15696) that reproducibly bound to one or more of the 12 consensus motifs on the TF-array (Figure 6.8). Three of these TFs (Pdp1, Sox21a, bun) had no prior PWM associated with them. As we initially interrogated the 12 consensus sequences as a pool, we next set out to determine the binding specificities and affinities of each of the 11 TFs individually. For these quantitative measurements we took a TF-centered approach by programming the MITOMI device with known concentrations of the 12 target DNA motifs and detecting the binding preferences to the selected TF. We obtained saturating binding curves for each TF-DNA sequence combination by fitting the measured DNA/protein ratio signals over available consensus DNA (in solution) to a single-site binding model (Figure 6.9 A, B, left panel). The ratio of DNA fluorescence intensities to BODIPY intensities was used as a dimensionless factor of how strongly DNA was bound by protein as the number of protein and DNA molecules beneath each button can differ due to fluid flow variations in the channels. Binding affinities to each consensus motif were then plotted as K_d value (Figure 6.9 A, B, right panel).

Overall we obtained K_d values for 228 TF-DNA interactions from binding curves consisting of 6,192 data points. The detailed affinity analysis showed that all 11 TFs were able to bind DNA, indicating that the false positive rate of on-chip TF-DNA interaction detection with MITOMI is low. Out of these 11 TFs, 8 TFs bound specifically and with high affinity to one of the 12 consensus motifs: i) gsb, and gsb-n bound the prd motif, ii) slbo, and Pdp1 bound the gt motif, iii) Gsc bound the bcd motif, iv) Ets65A bound the Mad motif, and v) Sox21a and Sox102F bound the D motif. Z bound to both the bcd and prd motif with high affinity. CG15696 binding was weak but showed a slight preference for the gt motif, while bun was entirely non-specific.

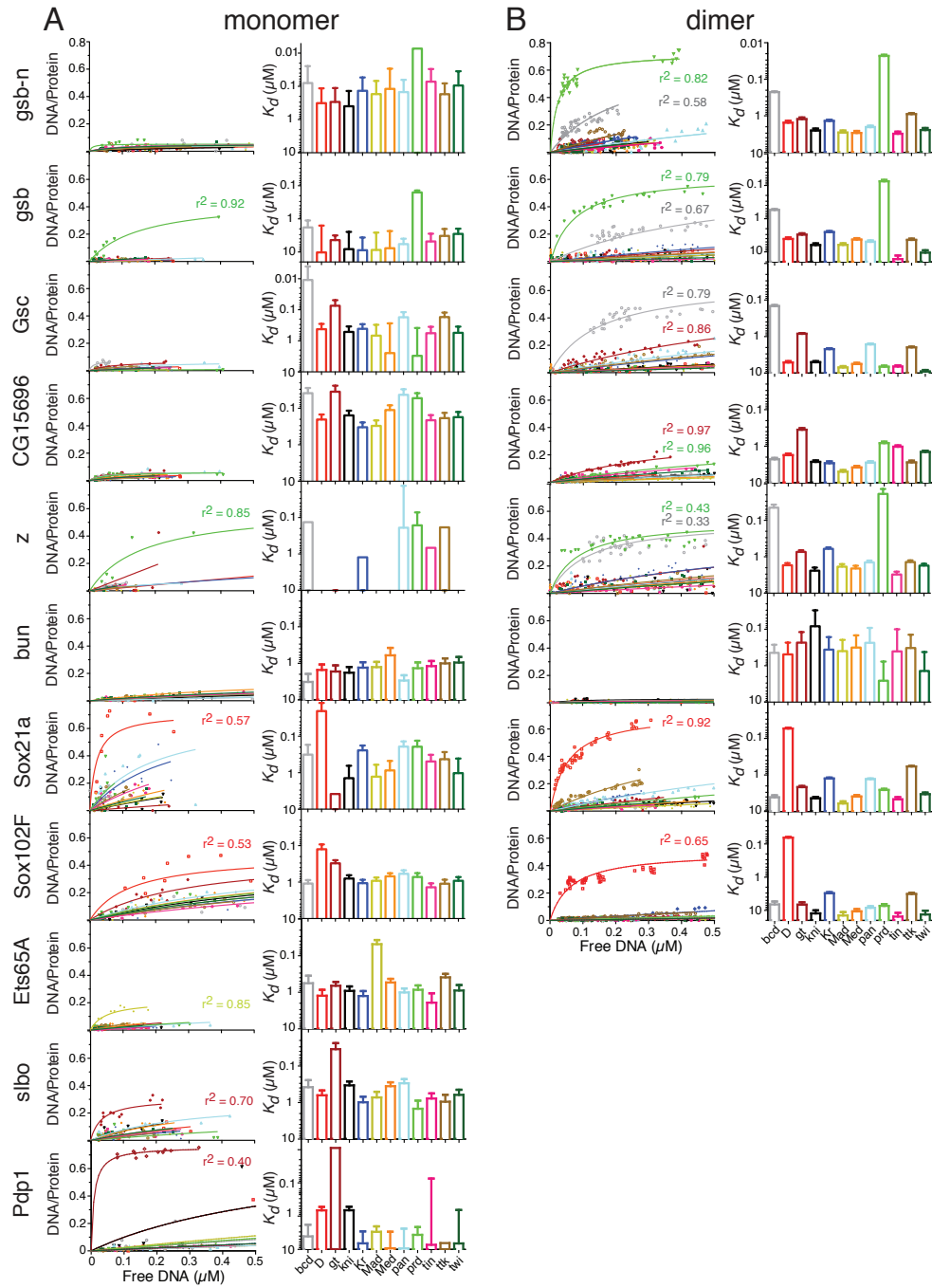


Figure 6.9: Binding specificities and affinities of TFs for which binding was observed were measured for all 12 DNA consensus motifs in two different modes: (A) For monomer-DNA binding, target DNA could bind only to surface-immobilized TFs. (B) In the second mode, TFs were allowed to interact with DNA consensus sequences in solution prior to antibody binding and detection, which potentially allowed for TF binding to DNA. Concentration-dependent binding curves are derived from fitting the measured DNA/protein ratio signals over available consensus DNA (in solution) to a single-site binding model (left); binding affinities to each consensus motif is plotted as K_d value with standard errors (right). Specificity and affinity measurements of TFs to DNA sequences for all 11 newly identified TFs, which bound to one or more of the 12 DNA consensus motifs (bcd, D, gt, kni, Kr, Mad, Med, pan, prd, tin, ttk, twi).

To account for the fact that some TFs are known to require multimerization or dimer formation for efficient binding to DNA, we measured saturation binding curves in two distinct modes for 8 TFs: the first mode favoring monomer-DNA binding, in which the target DNA was presented to a surface bound TF (Figure 6.9 A). In a second, the TF was loaded into the chamber, which allowed the formation of homodimers while binding to a target DNA (Figure 6.9 B). Several TFs indeed showed increased binding affinities when allowed to interact in a dimer-like manner (Figure 6.9 B). Binding specificity was more pronounced for zeste (z) when multiple recognition sites were presented, which is also known to require a much longer DNA binding sequence (16 bp) for efficient binding than its consensus sequence would suggest [179]. Interestingly, although almost the same DNA binding motif had been discovered for both gooseberry TFs in a bacterial one-hybrid study (B1H) [180], they do not display the same binding preference. This seems to be supported by the fact that different non-overlapping control elements drive the specific expression of *gsb* and *gsb-n* [181]. The only candidate that showed non-specific binding was bunched (bun), which had been reported to contain a leucine zipper dimerization domain and as such may require other proteins to bind DNA efficiently [182].

6.4.2.2 Assessment of TF integrity

To further estimate how many of the expressed TFs could functionally bind DNA, we randomly selected 8 TFs spanning the entire size range and tested each TF against a library of particular DNA sequences. This De Bruijn library contains all 65,532 8mer sequences computationally segmented into 1,524 double-stranded target DNAs [13]. Labeling the 70 bp oligonucleotides with a Cy5- and a Cy3-fluorophore, respectively, allowed us to harbor all of the 1,524 DNA fragments on one MITOMI device with 768 unit cells by programming each unit cell with two different sequences. ITT mix with the selected TF was loaded into the chambers and allowed to interact with the DNA sequences before trapping the TF-DNA complexes bound to the surface area under the button valve.

Of the 8 TFs tested, 3 TFs bound strongly and specifically to DNA allowing us to determine position weight matrices (PWMs) for these TFs. An additional 2 TFs showed DNA pull-down, but the quality of the data was not sufficient to derive PWMs, and 3 TFs failed to bind to DNA (Figure 6.10). Based on this analysis we estimate that 38-63% of TFs on the array can functionally bind DNA. This is a conservative estimate, as TFs may be folded correctly, but require a co-factor for functional DNA binding. Nevertheless, it compares well with a previous estimate of 35-41% of functional TFs derived from a protein array generated by expression of human proteins in yeast, followed by large-scale purification, and microarraying [8].

Chapter 6. Drosophila transcription factor array

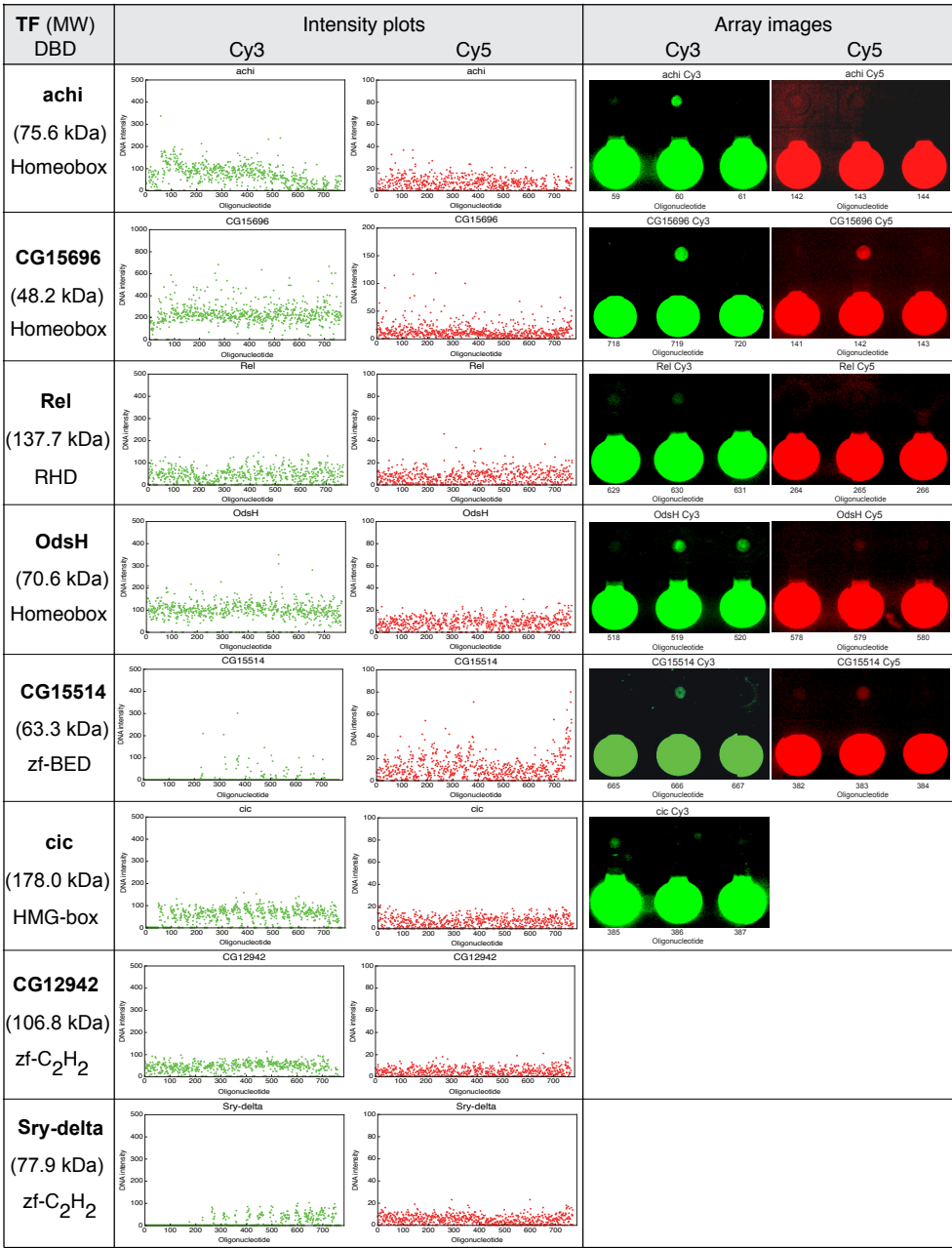


Figure 6.10: Binding ability of randomly chosen TFs was assessed on an oligonucleotide array containing all possible 65,532 8mer sequences (de Bruijn library). The table summarizes plots of measured Cy3 and Cy5 intensities after background correction for each oligonucleotide and a supporting scan image for each channel for detected TF-DNA binding.

We thus could establish that protein arrays generated by *in vitro* transcription-translation perform as well as classical protein arrays, but drastically reduce cost and labor associated with array generation. A similar comparison to Y1H assays is difficult, as parameters such as TF expression, and ability to bind DNA have not been assessed in Y1H assays. Although, a recent Y1H study identified only 26% of control interactions on previously characterized

cis-regulatory modules, the authors claim that this number should be treated with caution because of missing high-confidence protein-DNA interaction to validate the assay [74].

6.4.2.3 Determination of DNA binding motifs

We completed our TF characterization by determining the position weight matrices (PWMs) for 8 TFs, including Kr, prd, CG15696, Rel, achi, and gt, and *de novo* PWMs for the uncharacterized TFs Pdp1 and Sox21a (Figure 6.11). We adopted and streamlined the method of Fordyce *et al.* [13], permitting the same 8mer measurement depth using 768 unit cells instead of the 4,160 unit cell device previously required. The motifs for Kr, prd, CG15696, Rel and achi determined in our analysis were in good agreement with motifs found in the database of *Drosophila* TF DNA-binding specificities (<http://pgfe.umassmed.edu/TFDBS/>). A motif for gt could not be reproduced with confidence, which is reflected by the variety of published motifs [183]. However, applying a multiple-motif model may reveal a secondary model as shown for many TFs (especially with motifs longer than 10 nucleotides) using the Seed-and-Wobble algorithm [80].

Our determined motif and the novel motifs found for Pdp1 and Sox21a reflected known motifs of other members of the bZip (Pdp1) and HMG-box (Sox21a) family of transcription factors. In particular, Pdp1 had shown strong preference for the gt target DNA in the affinity measurements. This preference was supported by a high similarity between the gt consensus sequence and the motif found for Pdp1, with a very high correlation ($r^2=0.98$) between experiments performed on different days.

6.5 Conclusions

Current approaches to study TF-DNA interactions at a systems level for GRN mapping usually involve rather insensitive high-throughput screens, which can be susceptible to a high false positive/negative rate and thus require validation by additional methods. In this project, we developed a high-throughput gene-centered approach for integrated systems-level interaction mapping (*iSLIM*) of TF-DNA interactions. *iSLIM* uses a single microfluidic platform based on the MITOMI principle for qualitative detection of TF-DNA interactions as well as their quantitative characterization. To demonstrate the versatile applicability of *iSLIM* for studying regulatory networks, we generated an *in situ* synthesized protein array, on which we detected protein binding to specific DNA sequences and determined the binding affinity and specificity of detected binding events.

We successfully expressed 423 full-length *Drosophila melanogaster* TFs (93%) *in situ* from DNA templates, which is, to our knowledge, one of the most comprehensive fly TF collections

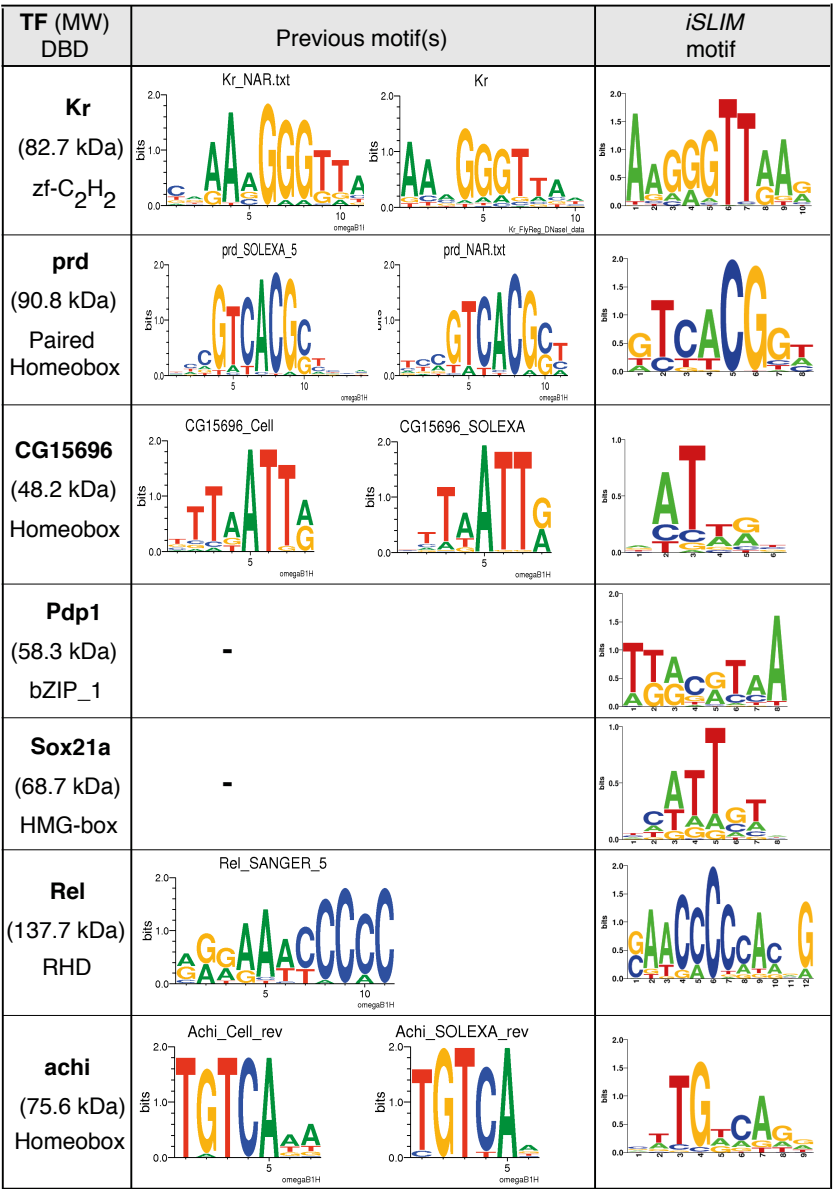


Figure 6.11: Determination of DNA binding motifs by de Bruijn analysis of 7 TFs. The binding specificity of selected TFs was assessed on an oligonucleotide array containing all possible 65,532 8mer sequences (de Bruijn library). The library consisted of 1524 double-stranded oligonucleotides, labeled with Cy3 and Cy5 in equal parts. The motifs found with *iSLIM* platform are shown in comparison to previously known motifs (if any), listed in the database of *Drosophila* TF DNA-binding specificities (<http://pgfe.umassmed.edu/TFDBS/>).

produced on a microfluidic chip. Between 38 and 63% of these TFs are estimated to be functional based on the analysis of binding preferences of 9 randomly selected TFs to 1,524 individual DNA sequences. We further confirmed the TF integrity by simultaneously interrogating our TF array with 12 specific DNA sequences, which allowed us to measure 5,076 (12 DNA x 423 TF) interactions in a single experiment. In measurements with specific DNA sequences known to bind to TFs from a well-characterized GRN we were able to detected at

least 44% of the expected TF-DNA interactions, supporting the previous estimate. However, this is probably a conservative number given that many of these TFs potentially form dimers with other proteins or require co-factors for efficient DNA binding and compares well with other *in vitro* approaches under similar conditions. Moreover, 11 additional interactions, previously unreported within this GRN, were confirmed in affinity and specificity measurements using the same platform, asserting a very low false positive/negative rate for our approach and giving rise to 228 K_d values from binding curves with a total of 5,472 data points. We further demonstrated the versatility of *iSLIM* by determining DNA binding motifs for TFs with known and unknown position weight matrices (PWM).

To conclude, we established a comprehensive methodology for gene-centered mapping of GRNs by rapidly generating a large-scale TF-array from ORF libraries and detecting thousands of TF-DNA interactions in a single experiment. Unlike other gene-centered methods, our approach allows quantitative assessing the expression level of each investigated TF and requires an order of magnitude less time, as a single experiment takes only 1-2 days as opposed to ~2 weeks for Y1H screens. Compared to classical protein arrays we reduce associated costs and labor by circumventing tedious cell-based protein expression, and purification. Moreover, using the same platform we can fully characterize TF-DNA interactions including determination of *de novo* consensus motifs/PWMs and quantitative measurements of binding energy landscapes, which makes it the only method capable of both gene-centered and TF-centered approaches to characterize GRN. In a recent publication, we show that it is even possible to measure the kinetics of TF-DNA interactions in large-scale using a modified MITOMI device [97].

Given these performance advantages over existing methods, our approach will contribute to the large-scale mapping of GRNs and provide a detailed quantitative analysis of the interactions constituting GRNs. Finally, since experiments are conducted *in vitro* and integrated with a next generation detection method, reaction conditions can be precisely controlled enabling the analysis of higher-order interactions, which is not feasible with Y1H methods and may be difficult to achieve using standard protein arrays. *iSLIM* thus provides a universal tool that cannot only complement current methods but has the potential to replace other techniques to systematically map molecular interactions in any given organism, which will aid systems biology in the quantitative characterization of GRNs.

6.6 Future Directions

In order to further assess the overall integrity of each individual TF on the array we will measure binding potential of TFs to genomic DNA, and the binding specificity to specific promoter regions by interrogating the TF array with fragmented DNA sequences covering the entire *Drosophila* genome. This will undoubtedly give us a more reliable estimate of how

many TFs are functional and therefore likely to bind to specific DNA sequences in future sub-network studies on the array.

Given that our proof-of-principle measurement with *iSLIM* on the well-studied even-skipped GRN in *Drosophila melanogaster* already revealed previously unknown network connections with high confidence (low false positive/negative rate), we aim to move from highly characterized sub-networks to less well characterized global networks to link and better understand the regulation in a highly interconnected network. These experiments may also provide more insight into binding site promiscuity, as shown for several consensus motifs in our affinity and specificity measurements. Coupled with association and dissociation rate measurements using the novel kinetic MITOMI device [97] we envision extending the informational content of GRNs significantly. In addition to binary TF-DNA interactions, *iSLIM* could be used to investigate homo- and heterodimer formation, which is known to be required for many TFs to efficiently bind DNA, by implementing co-spotting and multiplexing with different fluorophores. This strategy will also allow systematic studies with a combination of regulatory and co-regulatory factors and will facilitate the feasibility of large-scale protein-protein interaction studies on a proteomic level. In addition, other effects on the binding affinity of TFs, especially CpG methylation of DNA sequences, to enlighten the consequences of epigenetic regulators could easily be implemented in on-chip TF characterization as recently shown [144].

7 General discussion and outlook

On-chip protein synthesis overcomes a lot of the known limitations of conventional protein arrays, notable saving time and reagents as well as reducing the risk of protein degradation during array generation. In this thesis two approaches for on-chip protein synthesis were investigated in order to study protein interactions in large scale using a microfluidic device. This work includes a novel method to purify proteins on-chip from yeast cells as well as a new and improved PCR-based method compatible with large-scale applications for the generation of linear expression-ready DNA templates directly from yeast cell libraries for cell-free protein expression. In addition, a microfluidic method was developed for integrated systems-level interaction mapping (*iSLIM*) and quantitative characterization of transcription factor (TF)-DNA interactions.

While classical protein arrays remain extremely time and resource consuming to produce, this work demonstrates that functional proteins can be retrieved efficiently without the need for tedious bench-top purifications, when coupled to/integrated with a microfluidic platform capable to detect protein interactions. The presented methods are faster than other assays: a single multiplexed experiment takes between 1 and 3 days for cell-free protein synthesis and in-situ protein extraction from cells, respectively, while actual hands-on times are reduced to 1 to 2 hours. In addition, using a microfluidic platform substantially decreases experimental costs by minimizing reagent consumption to a few microliters per experiment. On-chip protein expression from DNA templates runs at a cost of 1-2 cents per synthesized protein. Consequently, one round of TF-DNA interaction screening on our *iSLIM* platform costs ~10 CHF. Protein extraction from deposited yeast cells reduces the expenses by another order of magnitude. A direct cost comparison to other methods is difficult, as screening assays vary largely with respect to experimental procedures and detection mechanisms. In addition to favorable operation times and costs, the problems associated with protein stability for storage and array production are eliminated when producing a protein microarray on demand from either DNA templates or cells.

However, the programmable nature of MITOMI-based devices has been shown to be more than a screening platform: When coupled to an organism-independent DNA library of computationally assembled oligonucleotide sequences, it allows detailed quantitative analysis and full characterization of protein-DNA interactions. At the same time, applying this TF-centered approach validates detected binding events and verifies that MITOMI shows, unlike other interaction screening platforms, a very low false positive rate. Finally, since the method is conducted *in vitro* and integrated with a next generation detection method, reaction conditions can be precisely controlled, enabling the analysis of higher-order interactions, which is not feasible with Y1H methods and may be difficult to achieve using standard protein arrays. Given these performance advantages over existing methods, pairing MITOMI-based detection of molecular interactions with alternatives to conventional protein arrays will contribute to the large-scale mapping of gene regulators networks (GRNs) and provide a detailed quantitative analysis of the interactions constituting GRNs.

7.1 Limitations and suggestions for the improvement of the current approaches

The aim of this work was to investigate different approaches to generate protein arrays to study protein interactions using a microfluidic device. Here we demonstrate the potential of an *in vitro* on-chip cell-based approach and the capacity of an integrated microfluidic method for systems-level TF-DNA interaction mapping and characterization on an *in situ* synthesized protein array. Although performed on a library of yeast ORF-fusion proteins hosted in *Saccharomyces cerevisiae*, the cell-based approach described in chapter 4 can be easily adapted to other protein-fusion cell libraries. It is certainly worthwhile to consider using *E. coli* transformants to facilitate on-chip lysis and increase protein yields. On the other hand, some eukaryotic proteins expressed in bacteria are not functional as a result of missing post-translational modifications or incorrect folding due to differences in the prokaryotic enzymatic machinery. Regardless of the host organism, protein (over-)expression in the wrong spatio-temporal cellular context may be toxic for the cells themselves and trigger a number of modifications.

While a genomically integrated yeast fusion library was considered impractical for generating *in situ* protein arrays from cells, physiological expression levels may be advantageous for the detection of protein interactions to promiscuous binding sites. However, first and foremost the possibilities of a cell-based protein array to detect relevant protein interactions should be assessed to assure array integrity.

7.2. Future applications of microarrays in the field of molecular biology

The second study on cell-free *in situ* protein synthesis using *iSLIM* investigated 60% of all known *Drosophila melanogaster* TFs (454 of 755), of which 93 % (423 of 454) were considered potential candidates for protein interaction mapping after *in situ* protein expression. Arguably, this is one of the most comprehensive fly TF interaction platforms *in vitro*. Nevertheless, to get a complete picture of the fly transcriptome, it is indispensable to include information from potential interaction partners that were not incorporated in the presented work (~44 %).

Currently, most *in vitro* approaches, including those presented in this work, focus on studying one particular protein interaction at a time, which does not reflect the physiological condition, where competing binding partners are abundant and the protein needs to find its specific regulatory site. Moreover, it is well known that many TFs only bind to their target DNA after dimer formation with another protein or association with other co-regulators. Addressing this shortcoming would greatly expand the scope of possible interactions and could be implemented by co-spotting of several proteins or additional co-regulators (e.g. two different expression templates) as described in chapter 4. Furthermore, binding to a pool of fragmented genomic DNA would give a much more realistic picture of functional TFs binding to their regulatory sequence. Multiplexing with different fluorescence labels would increase throughput and narrow down the potential genomic binding sequence.

Finally, all approaches in this thesis are performed *in vitro* in precisely controlled reaction conditions. On the other hand, one of the main limitations of most *in vitro* protein interaction studies is the absence of a cellular environment, which contributes to the complexity of molecular interactions and GRNs. Mapping post-translational modifications, such as phosphorylation or methylation, and other upstream signal transduction pathways is exceedingly important to elucidate contributors to differential gene expression and complete the current understanding of protein interaction networks. The kinetics of these usually enzymatically conveyed modifications could be measured in high-throughput by developing a sensitive on-chip enzyme assay.

7.2 Future applications of microarrays in the field of molecular biology

While post-translational modifications on *in situ* synthesized is relatively difficult to implement, modifications on the DNA level could be more easily pursued. DNA methylation is crucial for epigenetic differences between cell types and pathological states, but its effects on TF binding has not been studied in detail. A recently fabricated microarray of duplex methylated DNA showed the potential of this platform [184] and could be modified to determine the binding preferences and affinities of a TF in the presence of methylated sites.

In this work, the *iSLIM* approach was successfully demonstrated on a TF array consisting of more than 400 *Drosophila* TFs. This approach can be easily adapted to cover all TFs of other organisms, such as mouse or yeast. It would be also interesting to compare the *iSLIM* approach to the human transcriptome study, performed on a classical protein array with ~ 4,200 purified human proteins [8].

The protein-centered and gene-centered approach of *iSLIM* can also be applied to map and characterize other specific sub-networks, such as for proteins involved in ribosome biogenesis. A recent publication on 110 yeast ribosomal protein (RP) promoters uncovered that mutations in certain DNA sequences reduced promoter activity [185]. Measuring binding energy landscapes of RP alone or in combination with other protein against a large DNA library could provide more information on the transcriptional regulation of RP promoters. Furthermore, a gene-centered screen using the TF library could be applied to discover proteins that specifically bind to promoter sequences involved in ribosome biogenesis, which have been much less characterized than RP promoters.

While studies on native promoters provide useful information, defined synthetic promoter libraries allow studying gene regulation systematically. Characterizing the binding properties and determining the binding energy landscapes of each promoter sequence from a recently published library of 128 synthesized eukaryotic promoters [186] will further aid our understanding of gene regulation.

Expanding *in situ* protein synthesis to the entire protein repertoire of an organism on a MITOMI device with higher throughput as described in chapter 3, would establish a comprehensive interaction network on a proteome level. As already outlined in the previous section, screening the proteome array with fragmented genomic DNA will be advantageous to assess the array integrity. In addition to characterizing other sub-networks with selected DNA sequences a library of native and synthetic promoter sequences (see above paragraph) could detect previously unknown proteins with regulatory activity.

Another interesting biological question to be addressed with the *iSLIM* approach is how microRNA (miRNA) binding to mRNA transcripts affects the level gene expression. A recent screen set out to identify 40% of all *Drosophila* genes that are involved in the miRNA pathway [187].

Although in this work we primarily focused on characterizing TFs that associate directly with the genes they regulate, the MITOMI principle has been used to measure other molecular interactions, involving single-stranded RNA and drug molecules [11]. A similar approach could be applied to investigate pathways that affect gene regulation further downstream, for

7.2. Future applications of microarrays in the field of molecular biology

example, to identify genes that are involved in the microRNA (miRNA) pathway and measure how miRNA binding to mRNA transcripts affects gene regulation. Another example is to identify the role of regulatory sequences in untranslated regions or non-coding RNAs using a comprehensive library of cDNA complementary to RNA transcripts.

Mapping of molecular interactions and their full characterization has been long tackled by opposing approaches. However, when coupled to next-generation sequencing and detailed kinetic measurements [97], the microfluidic approach based on the MITOMI principle has the potential to characterize the entire transcriptome of any given organism and contribute substantially to our current understanding of the regulatory network and its underlying mechanisms.

A MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

A.1 Abstract

Gene regulatory networks (GRNs) consist of transcription factors (TFs) that determine the level of gene expression by binding to specific DNA sequences. Mapping all TF-DNA interactions and elucidating their dynamics is a major goal to generate comprehensive models of gene regulatory networks (GRNs). Measuring quantitative binding affinities of large sets of TF-DNA interactions requires the application of novel tools and methods. These tools need to cope with the difficulties related to the facts that TFs tend to be expressed at low levels in vivo, and often form only transient interactions with both DNA and their protein partners. Our approach describes a high-throughput microfluidic platform with a novel detection principle based on the mechanically induced trapping of molecular interactions (MITOMI). MITOMI allows the detection of transient and low-affinity TF-DNA interactions in high-throughput.

A.2 Introduction

Transcription factors (TFs) are proteins that bind to specific DNA sequences and regulate the level of gene expression by either promoting or blocking the transcription of specific genes. These specific TF-DNA interactions are part of a dynamic gene regulatory network (GRN), which is beginning to be understood by generating complex models from data of both experimental and computational methods [188, 189].

The comprehensive characterization of GRNs requires large-scale quantitative measurements of TF-DNA interactions in a high-throughput format. However, conventional experimental methods to study molecular interactions are limited by being either non-quantitative [77, 190] or relatively low-throughput [191]. Generating microarrays of immobilized double-

stranded DNA sequences and probing them with off-chip purified proteins has been widely used for the detection of TF-DNA interactions.

The more recent introduction of microfluidics in the field of protein and DNA microarrays is a promising approach to scale-down and parallelize biological assays and study individual molecular interactions in a miniaturized format. The use of microfluidics platforms has many advantages: samples can be detected with high-precision and sensitivity while at the same time decreasing the amount of consumables, and time needed, as compared to more conventional methods [192]. With the development of multilayer soft lithography for the rapid prototyping of microfluidic systems [193, 194] and microfluidic large-scale integration (MLSI) [114] microfluidic technology has become appealing to the field of biology. These MLSI devices are generally fabricated from elastomeric materials such as PDMS (polydimethylsiloxane) and harbor micron-sized channels with thousands of integrated micromechanical valves. Among the increasing number of applications in biology, microfluidic platforms emerged as powerful screening tools to study molecular interactions, which show the potential of realizing high-throughput and high-precision measurements [36, 12].

With the highly integrated microfluidic device described in this protocol a novel detection method has been established based on the mechanically induced trapping of molecular interactions (MITOMI). MITOMI allows the capture of transient and low-affinity interactions between DNA sequences and TFs at equilibrium, and thus the measurement of absolute binding affinities. In short, picoliter-sized reaction chambers within the array of the microfluidic chip are aligned to spots of dsDNA sequences printed onto an epoxy-coated glass slide using standard DNA microarray instruments. After loading an *in vitro* transcription/translation (ITT) mixture together with genomic DNA coding for the TF, the TF of interest is synthesized on-chip and can bind to freely diffusing target DNA. These binding events are separated by micromechanical valves and thus will be detected independently by MITOMI.

A.3 Materials

A.3.1 Mask & Wafer Fabrication

A.3.1.1 Instruments for mask & wafer fabrication

1. Mask writing on Heidelberg DWL200 laser lithography system (Heidelberg Instruments Mikrotechnik GmbH)
2. Mask and wafer development using DV10 (Süss MicroTec AG)
3. Wafer cleaning with oxygen plasma before processing using Tepla300 (PVA Tepla AG)
4. MA6 Mask Aligner (Süss MicroTec AG) for exposure of wafers
5. Sawatec LMS200, programmable coater for negative resist and Sawatec HP401Z, programmable hot plate for soft bake (Sawatec AG)
6. Süss RC-8 THP, manual coater and hotplate for positive resist (Süss MicroTec AG)

A.3.1.2 Materials for mask & wafer fabrication

Chemicals used in mask and wafer fabrication are from Rockwood Electronic Materials, Gréasque, France, and of Metal-Oxide-Semiconductor (MOS) quality, unless otherwise stated.

1. Masks: square blank 5" Nanofilm SLM 5 (Nanofilm)
2. Silicon wafers (diameter: 100 ± 0.5 mm, thickness: 525 ± 25 μ m, conductivity type: P, dopant: Boron, resistivity range: 0.1-100 Ω cm; Okmetic)
3. Photoresists: AZ9260 positive photoresist (MicroChemicals GmbH); SU-8 negative photoresist GM1060 and GM1040 (Gersteltec)
4. Chrome etching of masks: CR7 consisting of $(\text{NH}_4)_2 \text{Ce}(\text{NO}_3)_6$; HClO_4
5. Developers: MP 351 for mask and AZ 400K for AZ9260 coated wafers (AZ Electronic Materials); PGMEA (1-methoxy-2-propyl-acetate) for manual development of SU-8 wafers;
6. Solvents: Remover 1165 composed of 93% NMP (N-methyl-2-pyrrolidone; Sotrachem Technic) and 7% PGMEA for masks; isopropyl alcohol (IPA); acetone for wafers

A.3.2 MITOMI device fabrication by multilayer soft lithography

1. PDMS (Polydimethylsiloxane) resin: heat curable silicone elastomer (Dow Corning Sylgard 184)
2. TMCS (trimethylchlorosilane) (Sigma-Aldrich)
3. Mixing and degassing of PDMS: Thinky Mixer ARE-250 equipped with adaptor for 100 ml disposable PP beakers (C3 Prozess- und Analysentechnik GmbH)
4. Degassing of PDMS control layer: vacuum desiccator (Fisher Scientific AG)

5. Spin coating of PDMS flow layer: programmable spin coater SCS P6700 (Specialty Coating Systems Inc.)
6. Stereomicroscope, SMZ1500 (Nikon AG)
7. Manual hole punching machine and pin vises, 21 gauge (0.04" OD) (Technical Innovations, Inc.)

A.3.3 Epoxy slide preparation

Chemicals for epoxy-coating of glass slides are from Sigma-Aldrich.

1. Standard (76 x 26 x 1 mm) microscope glass slides (VWR)
2. Milli-Q water
3. Ammonium hydroxide (30%)
4. Hydrogen peroxide (30%)
5. Solvents: acetone, toluene, isopropyl alcohol (IPA)
6. 3-Glycidoxypyl-trimethoxymethylsilane (3-GPS; 97%)
7. Nitrogen gas supply

A.3.4 DNA synthesis

A.3.4.1 Synthesis of linear template DNA

All chemicals used for synthesis of DNA are from Sigma-Aldrich. For DNA primer sequences see Table A.1.

1. DNA primers (Integrated DNA Technologies, IDT)
2. dNTPs (Roche Diagnostics AG)
3. Yeast genomic DNA (Merck Chemicals Ltd.)
4. Polymerase enzyme, Expand High Fidelity PCR system (Roche Diagnostics AG)
5. Elution buffer: 10 mM TrisCl, pH 8.5

A.3.4.2 Synthesis of target DNA

1. Primers, 5'CompCy5, (Integrated DNA Technologies, IDT)
2. dNTP (Roche Diagnostics AG)
3. Klenow fragment (3' → 5' exo-) (Bioconcepts)
4. Dithiothreitol (DTT)
5. Magnesiumchloride (MgCl₂)
6. Buffer: Tris-HCl, pH7.9
7. 0.5% BSA in dH₂O

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

Table A.1: Primer sequences used to generate ITT linear templates and target DNA library.

Name (Comment)	Sequence
5'CompCy5 (Extension primer for target DNA synthesis)	5'-[Cy5]GTCATACCGCCGGA-3'
Target DNA (Design of target DNA library. Variable binding sites of interest are bracketed by constant linker sequences. 3' ends consist of reverse complementary sequence of CompCy5 extension primer)	5'-[5'LINKER]-[BINDING SITE OF INTEREST]-[3'LINKER]-TCCGGCGGTATGAC-3'
Forward-ORF-HIS (Design gene specific sequence to T _m of 60°C. Start codon is underlined. First 5 codons code for 5xHis-tag. Alternatively 5xHis-tag can be added to Reverse_ORF primer)	5'- CTCGAGAATTGCGCCACCATGACCACCACCACC AC-[GENE SPECIFIC]-3'
Reverse-ORF (Design gene specific sequence to T _m of 60°C. Stop codon is underlined)	5'-GTAGCAGCCTGAGTCGTTATTA-[GENE SPECIFIC]-3'
Forward extension T7 promoter sequence and transcription start site is indicated in italic and bold, respectively	5'- GATCTTAAGGCTAGAGTACTAATACGACTCACTA TAG GGAATACAAGCTACTTGTTCTTTTGCCTCGAG AATTC GCCACC-3'
Reverse extension (Poly(A) track is underlined)	5'- CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCC AAG GGGTTATGCTAGTTTTTTTTTTTTTTTTTTTTTTTT TTTTTT GTAGCAGCCTGAGTCG-3'
Forward final	5'-GATCTTAAGGCTAGAGTAC-3'
Reverse final	5'-CAAAAAACCCCTCAAGAC-3'

A.3.5 Microarraying / Spotting

1. QArray2 microarrayer (Genetix GmbH) equipped with a NanoPrint microarray printer printhead and a 946MP3 micro spotting pin (Arrayit Corporation)
2. Bovine serum albumin (BSA; Sigma-Aldrich) re-suspended in DI water to 2 mg/ml
3. Synthesized target DNA templates (see A.3.4.2)
4. Epoxy-coated microscope glass slides (see A.3.3)

A.3.6 Microfluidic control elements**A.3.6.1 Pressure regulation and control**

1. Precision pressure regulator, BelloFram Type 10, 2-25 psi; 1/8" port size (part. no. 960-001-000; Bachofen SA)
2. Bourdon tube pressure gauges, 0-30 psi (0-2.5 bar), G 1/4 male connection (part. no. NG 63-RD23-B4; Kobold Instruments AG)
3. Custom-designed manual manifolds (rectangular metal casing: 14.5x1x1") for control line regulation with 16 detented toggles and barbs for 1/16" tubing ID; 1/4 NPT connection (Pneumadyne Inc.)
4. Fittings to connect regulators to gauges and to luer manifolds (Serto AG): tee union (SO 03021-8), male adaptor union (SO 01121-8-1/8)
5. Tubing: Tygon 1/4" OD x 1/8" ID (Fisher Scientific AG)
6. Polycarbonate luer fittings (Fisher Scientific AG): multi-port luer manifolds for flow inlet regulation (e.g. part. no. Cole Parmer 06464-87); male luer to luer connector (part. no. Cole Parmer 06464-90)

A.3.6.2 Fluidic connections

1. Disposable stainless steel dispensing needles to connect to syringe 23 gauge, 1/2" long, ID 0.33 mm (part. no 560014; I & Peter Gonano)
2. Tubing (Fisher Scientific AG): for liquid/gas; flexible plastic tubing for fluidic connections, Tygon S54HL, ID 0.51 mm;
3. Steel pins for chip-to-tube interface: tube AISI 304 OD/ID x L \varnothing 0.65/0.30 x 8mm, cut, deburred, passivated (Unimed S.A.)

A.3.7 MITOMI**A.3.7.1 Surface Chemistry**

1. Biotinylated bovine serum albumin (Sigma-Aldrich), reconstituted to 2 mg/ml in DI water (referred to as: BSA-biotin)

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

2. Neutravidin (Thermo Scientific Pierce), reconstituted to 0.5 mg/ml in PBS (referred to as: NA/PBS)
3. 1:1 solution of biotinylated Penta-histidine antibody (Qiagen AG) in 2% BSA

A.3.7.2 Transcription Factor Synthesis

1. TNT T7 coupled wheat germ extract mixture (Promega AG)
2. FluoroTect GreenLys-tRNA (Promega AG)
3. Linear template DNA coding for transcription factor (see A.3.4.1)

A.3.8 Data acquisition

1. Modified ArrayWorx scanner (Applied Precision) for detection and softWoRx software
2. Axon GenePix (Molecular Devices) for analysis

A.4 Methods

The entire process involves seven distinct steps that will be described in details (see Figure A.1). Molds for the two-layer microfluidic device are fabricated on silicon wafers patterned from laser-written chrome masks in a class 100 clean-room environment. A set of control and flow molds is used to produce a double-layer device by multilayer soft lithography [36].

Libraries of Cy5-labeled target DNA sequences are synthesized and spotted onto epoxy-coated microscope slides. The DNA arrays are aligned to the microfluidic device containing 768 unit cells and bonded overnight. Assembled chips are mounted on a microscope stage and connected to a pneumatic setup. Each unit cell of the device can be controlled by three individually addressable micromechanical valves, which allow compartmentalization of each unit cell, control of DNA chamber access and the detection area. A circular button membrane is used to mask a precisely defined area during surface derivatization and for the mechanical induced trapping of molecular interactions (MITOMI).

Upon surface derivatization the device is loaded with a mixture of wheat germ in vitro transcription/translation (ITT) extract and the linear DNA template coding for the TF of interest. Spotted target DNA and the synthesized TF localize to an antibody deposited underneath the circular button area. MITOMI is performed by actuating the button membrane and trapping bound complexes in equilibrium. These complexes can subsequently be visualized by scanning the device with a DNA array scanner. Binding affinities are determined by quantifying the detected signals.

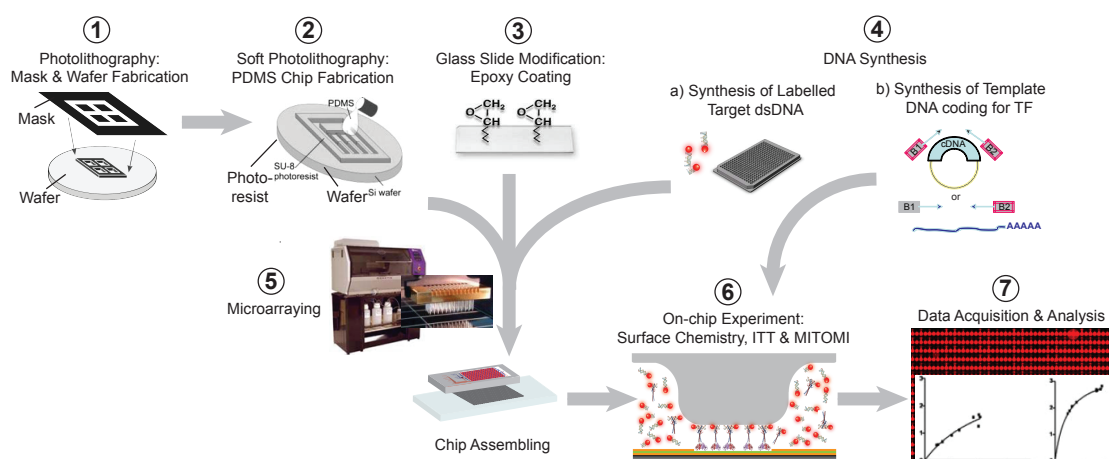


Figure A.1: (1) Molds for the two-layer microfluidic device are produced on silicon wafers reproduced from chrome masks. (2) A double-layer device is fabricated of PDMS (polydimethylsiloxane) by multilayer soft lithography using a control and flow mold as template. (3) Microscope glass slides are surface modified with an epoxysilane coating. (4a) Short, fluorescently labelled target DNA sequences are synthesized and (5) spotted onto the epoxy-coated glass slides using a microarrayer before aligning and bonding them to microfluidic chip generated in step (2). (6) After surface derivatization of the glass slide a mixture is loaded containing wheat germ in vitro transcription/translation (ITT) extract and the synthesized linear DNA template (4b) coding for the TF of interest. On-chip synthesized TFs are pulled down by immobilized antibody pulls down and can bind spotted target DNA sequences. TF-DNA interactions are captured by a mechanism based on mechanically induced trapping of molecular interactions (MITOMI) and (7) binding affinities quantified from detected interactions after scanning.

Note for researchers without clean-room and/or PDMS fabrication facilities: the MITOMI devices may also be obtained directly for a nominal fee from the California Institute of Technology (<http://kni.caltech.edu/foundry/>) and Stanford Microfluidic Foundries (<http://www.stanford.edu/group/foundry/index.html>). Up-to-date protocols and microfluidic design files can be found on our laboratory homepage (<http://microfluidics.epfl.ch>).

A.4.1 Mask & Wafer Fabrication

All processes in this section are performed in a class 100 clean room.

A.4.1.1 Mask fabrication

1. The two layer device is designed in CleWin4 (WieWeb software).
2. Each layer is reproduced as a chrome mask using a Heidelberg DWL200 laser lithography system with a 10 mm writing head and a solid state wavelength stabilized laser diode (max. 110mW at 405 nm).
3. For the development of masks, first the dispenser arm within the DV10 development chamber is purged for 10 sec, then a developer mixture (MP 351:DI water 1:5) is applied twice to the mask (15 sec), agitated for 45 sec and drained, before rinsing and drying (50 sec).

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

4. Developed masks are chrome etched for 110 sec, rinsed, cleaned twice 15 min in 1165 remover bath, quick rinsed and air dried.

A.4.1.2 Flow mold fabrication

1. A 3" silicon wafer is cleaned in a plasma stripper (Tepla 300) with 400 ml/min oxygen gas at 500 W and a frequency of 2.45 GHz for a period of 7 min.
2. A 1-2 μm thin layer of GM1040 negative resist is deposited on the oxygen plasma cleaned silicon wafer by spin coating first for 10 sec at 500 rpm (ramp 100 rpm/sec), then for 46 sec at 1100 rpm (ramp 100 rpm/sec), followed by a short quick spin for 1 sec at 2100 rpm, and finally for 6 sec at 1100 rpm.
3. The pre-coated wafer is baked for 15 min at 65 °C and 15 min at 105 °C with a ramp of 4 °C/min, and allowed to cool down to room temperature.
4. The wafer is exposed for 2 sec in flood exposure mode with an alignment gap of 15 μm , using a lamp with a light intensity of 10 mW/cm² (further settings: WEC type: cont, N₂ purge: no, WEC-offset: off).
5. The exposed wafer is baked on a hotplate for 35 min at 100 °C.
6. Positive resist AZ9260 is spin coated on the pre-coated wafer for 10 sec at 800 rpm, followed by 40 sec at 1800 rpm (ramp 1000 rpm/sec) to yield a substrate height of around 14 μm .
7. The coated wafer is then baked on a hotplate for 6 min at 115 °C.
8. The soft-baked positive resist is allowed to rehydrate for 1 h.
9. The wafer is exposed on a MA6 mask aligner for 2 intervals of 18 sec with 15 sec. waiting time in photolithography (soft contact) mode at 360 mJ/cm² with a light intensity of 10 mW/cm² (broad band spectrum lamp). The alignment gap is set to 15 μm (further settings: WEC type: cont, N₂ purge: no, WEC-offset: off).
10. Following 1 h relaxation time the wafer is developed in a development chamber (DV10) for 8-12 min based on visual observation after each cycle of the following routine with a total time of 4 min: A development mixture (AZ 400K:DI water 1:4) is dispensed and agitated on the wafer in three cycles, drained, rinsed (total time: 3:15 min), and finally dried.
11. In a final step, channels of the flow mold are annealed and rounded at 160 °C for 20 min to create a geometry that allows full valve closure.

A.4.1.3 Control mold fabrication

1. Negative photoresist SU-8 GM1060 is spin coated on an oxygen plasma cleaned silicon wafer (see step 1 in A.4.1.2) first for 10 sec at 500 rpm (ramp 100 rpm/sec), then for 50 sec at 1500 rpm (ramp 100 rpm/sec), followed by a short quick spin for 1 sec at 2500 rpm, and finally for 6 sec at 1500 rpm to yield a height of ~14 μm .
2. The coated wafer is baked for 30 min at 130 °C and 25 min at 30 °C on a hotplate.

3. The wafer is exposed on a MA6 mask aligner for 3 intervals of 20 sec with 15 sec. waiting time (all other settings see step 9 in A.4.1.2).
4. The exposed wafer is developed manually for 2 x 5 min in PGMEA, rinsed with IPA and dried with an air gun.

A.4.2 MITOMI device fabrication by multilayer soft lithography

1. The control layer mold is placed in a glass petri dish lined with aluminum foil to facilitate easy removal. Care must be taken that the aluminum foil lining doesn't contain any holes.
2. To generate a hydrophobic surface, both flow and control mold are exposed to vapor deposits of TMCS for 30 min by placing them into a sealable plastic container with 1 ml TMCS filled into a plastic cap. TMCS treatment is repeated for 10 min each time prior to PDMS chip fabrication.
3. For the control layer, 60 g of a 5:1 Sylgard mixture (50 g Part A : 10 g Part B) is prepared, mixed for 1 min at 2000 rpm (~400 x g) and degassed for 2 min at 2200 rpm (~440 x g) in a centrifugal mixer.
4. The mixture is poured onto the control layer mold and degassed in a vacuum desiccator for 10 min.
5. For the flow layer, 10.5 g of a 20:1 Sylgard mixture (10 g Part A : 0.5 g Part B) is prepared, mixed for 1 min at 2000 rpm (~400 x g) and degassed for 2 min at 2200 rpm (~440 x g) in a centrifugal mixer.
6. The mixture is spin coated onto the flow layer with a 15 sec ramp and a 35 sec spin at 2200 rpm.
7. After removing the control layer mold from the vacuum chamber any residual surface bubbles are destroyed by blowing on top of the PDMS layer. Any visible particles on top of the control channel grid are carefully removed using a toothpick.
8. Both layers are cured in an oven for 30 min at 80°C.
9. Following polymerization, both molds are taken from the oven and allowed to cool for 5 min.
10. The control layer is then diced with a scalpel and holes (1-8 and B, S, C, O in Figure A.2 A) are punched at the control input side using a hole puncher or an 21 gauge luer stub.
11. The channel side of the control layer is thoroughly cleaned with Scotch Magic Tape.
12. The cleaned control layer is then aligned to the flow layer on the stereomicroscope.
13. The device is bonded for 90 min at 80°C in an oven.
14. Bonded devices are removed from the oven and allowed to cool for 5 min.
15. Following the outline of the control layer each individual device is cut with a scalpel and peeled off the flow layer.
16. Holes are punched for the sample inlet and outlet (S1-S7 and O in Figure A.2 A) using

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

a hole puncher.

17. The flow channel side is cleaned thoroughly with tape before aligning the device to a spotted glass slide (see A.4.5).
18. The flow mold is cleaned of any residual polymerized PDMS either by peeling off the thin layer of PDMS using a pair of tweezers or by an additional PDMS layer. For the latter, 11 g of a 10:1 Sylgard mixture (10 g Part A : 1 g Part B) is mixed for 1 min at 2000 rpm ($\sim 400 \times g$), degassed for 2 min at 2200 rpm ($\sim 440 \times g$), poured on the flow mold cured in the oven for 30 min at 80°C , and peeled off after cooling down to room temperature. The control mold is cleaned with a nitrogen air gun of any PDMS debris

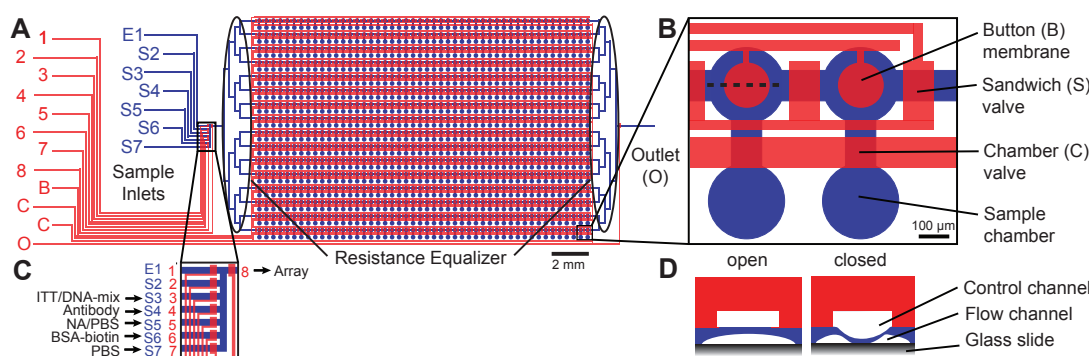


Figure A.2: MITOMI chip: (A) Drawing of a MITOMI device with 768 unit cells in the flow layer (blue) which are controlled (red) by 2388 valves. Resistance equalizers towards solution inlets and outlet ensure equal flow velocities and even derivatization in each row of the channels. (B) The magnified view shows two single unit cells each controlled by 3 different micro-mechanical valves: the chamber in each dumbbell-shaped unit cell hosts a different target DNA sequence and is isolated with the chamber valve during surface modification steps for subsequent pull-down to an immobilized antibody in the detection area under the button membrane. For diffusion of samples to the immobilized antibody chamber valves are opened while sandwich valves between individual unit cells are closed in order to prevent cross-contamination between different samples. (C) Tygon tubing is loaded with different sample solutions and connected to a metal pin that is plugged into a hole at the end of each channel within the sample inlet tree (also see magnified insert of Fig. 2 B). Loading of the device with samples via the inlets (S2-S7) is controlled by opening and closing the corresponding control valves. (D) A cross-section of one of the unit cells is shown to illustrate the detection mechanism based on mechanically induced trapping of molecular interactions (MITOMI). A thin deflectable membrane can be pushed down by actuating the water-filled control channel and consequently physically trap any material under this area in the flow channel.

A.4.3 Glass slide preparation

A.4.3.1 Cleaning procedure

1. All glassware is prepared by rinsing with Milli-Q water.
2. 750 ml Milli-Q water and 150 ml ammonium hydroxide are heated to 80°C in a staining bath.
3. 150 ml hydrogen peroxide is carefully poured to the ammonium solution.
4. Glass slides are added into the staining bath and incubated for 30 min.
5. After removal from the staining bath the glass slides are allowed to cool for 5 min.

6. Glass slides are then rinsed with Milli-Q water in the staining bath.
7. Clean glass slides are dried with nitrogen and stored in a dust free box.

A.4.3.2 Epoxysilane deposition

1. Before epoxysilane deposition all glassware is rinsed with acetone and dried at 80 °C.
2. Cleaned glass slides are incubated for 20 min in 891 ml toluene with 9 ml 3-GPS.
3. After rinsing with fresh toluene to remove unbound 3-GPS the glass slides are dried with nitrogen.
4. Glass slides are baked at 120 °C for 30 min.
5. Following sonication in toluene for 15 min glass slides are rinsed with fresh IPA.
6. Coated glass slides are dried with nitrogen and stored in a dust free box under oxygen free conditions until usage.
7. In case of systematic PDMS chip delamination: Prior to DNA spotting, glass slides are rinsed with toluene and dried with nitrogen.

A.4.4 DNA synthesis

A.4.4.1 Synthesis of linear template DNA

Generation of linear templates from genomic DNA or cDNA (see Note 1) of the TF of interest by a two-step polymerase chain reaction (PCR) method in which the first step amplifies the target sequence and the second step adds required 5'UTR and 3'UTR for efficient ITT.

1. For the first PCR step a mixture of the following components is prepared in a final volume of 50 µl:

1	µM	forward-ORF primer
1	µM	reverse-ORF primer
100	ng	yeast genomic DNA
200	µM	of each dNTP of a nucleotide mix
2.5	units	HiFi polymerase enzyme mixture

2. After initial denaturation for 4 min at 94 °C, the first PCR amplification is performed with 30 cycles as follows:
3. The correct product of this step should be ascertained on a 1% agarose gel.
4. For the second PCR step a mixture of the following components is prepared to yield a final volume of 100 µl:
5. After initial denaturation for 4 min at 94 °C, the second PCR amplification is performed with 10 cycles as follows:

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

(a)	Denaturation	94°C for 4 min
(b)	Annealing	53°C for 60 sec
(c)	Elongation	72°C for 90 sec and finish with a final extension at 72°C for 7 min

5	nM	forward extension
5	nM	reverse extension
1	μl	PCR product (from previous step)
200	μM	of each dNTP of a nucleotide mix
2.5	units	HiFi polymerase enzyme mixture

(a)	Denaturation	94°C for 4 min
(b)	Annealing	53°C for 60 sec
(c)	Elongation	72°C for 90 sec and finish with a final extension at 72°C for 7 min

6. To each reaction 1 μl of final primer mix is added (Forward final + Reverse final; each at 1 μM final concentration) and cycling continued for 30 cycles after 4 min at 94 °C as follows:

(a)	Denaturation	94°C for 4 min
(b)	Annealing	50°C for 60 sec
(c)	Elongation	72°C for 90 sec and finish with a final extension at 72°C for 7 min

7. The final product can be used directly in ITT reactions or purified on spin columns and eluted in 100 μl 10 mM TrisCl (pH 8.5).

A.4.4.2 Synthesis of target DNA

1. Small Cy5 labeled, dsDNA oligos are synthesized by isothermal primer extension in a reaction of a total volume of 30 μ l containing:

6.7	μ M	5'CompCy5
10	μ M	library primer
5	units	Klenow fragment (3'→5' exo-)
1	mM	of each dNTP of a nucleotide mix
1	mM	Dithiothreitol (DTT)
50	mM	NaCl
10	mM	MgCl ₂
10	mM	Tris-HCl, pH 7.9

2. All reactions are incubated at 37°C for 1 h followed by 20 min at 72°C and a final annealing gradient down to 30°C at a rate of 0.1 °C/sec.
3. After the synthesis, 70 μ l of a 0.5% BSA dH₂O solution are added to each reaction.
4. The entire volume is then transferred to a 384 well plate and a 6 fold dilution series prepared with final concentrations of 5'CompCy5 of 2 μ M, 600 nM, 180 nM, 54 nM, 16 nM and 5 nM (see Note 2).

A.4.5 Microarraying / Spotting

Spotting target DNA onto epoxy-coated microscope slides is performed by a QArray2 DNA microarrayer.

1. Before each spotting routine an Eppendorf or Falcon tube filled with DI water and the spotting pins is submerged in a sonicator water bath to clean the pins. During the spotting routine a sterilizing loop (1 sec DI water, followed by 1 sec air drying) between different DNA samples keeps pins clean throughout the spotting procedure.
2. Sample plate(s) of target DNA are placed in an external source plate stacker before starting the spotting routine (see Note 3).
3. The dilution series for each target DNA sequence is spotted as a microarray with a column and row pitch of 373 μ m and 746 μ m, respectively.
4. Spotted arrays are aligned manually to a tape-cleaned PDMS device (see A.4.2) on a Nikon SMZ1500 stereoscope and bonded overnight in an incubator at 40 °C.
5. DNA arrays can be stored in a sealed box protected from light and dust at room temperature for several weeks.

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

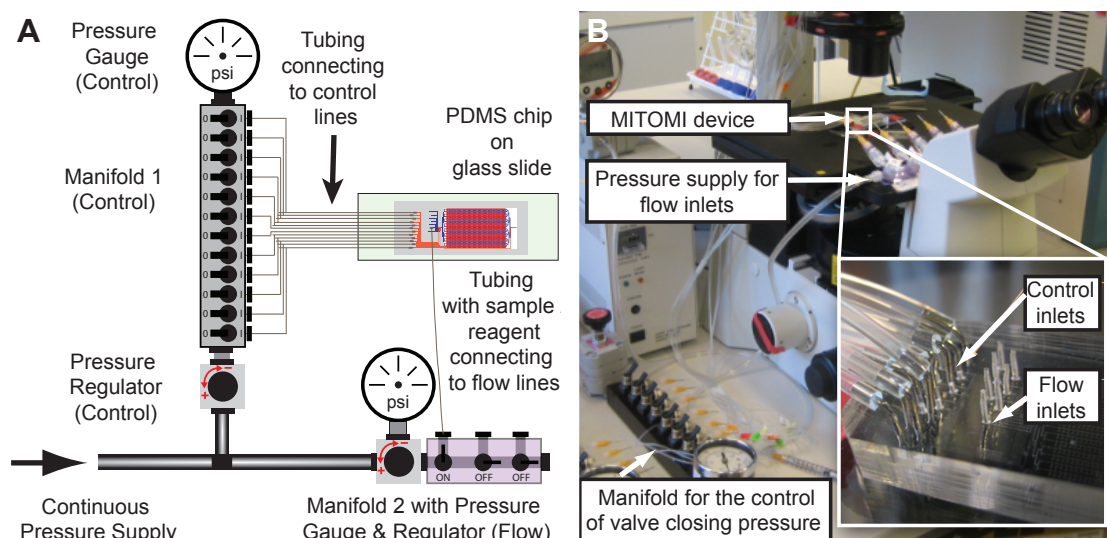


Figure A.3: Setup of the experiment. (A) Schematic of fluidic control set-up using regulated pressure and manually controlled manifolds. (B) Photograph of an assembled MITOMI device placed on the stage of a microscope where flexible tubing (Tygon) is connected via metal pins to the inlets that actuate the control lines on the device (magnified insert). The tubing is filled with DI water displacing the air in the channels when actuated with pneumatic pressure that is controlled with manifolds. Each reagent for the on-chip experiment is filled into a pin-end flexible tube (Tygon) and connected to the flow inlets. The pressure of the flow can be controlled with a gauge.

A.4.6 On-chip experiment (surface chemistry & MITOMI)

A.4.6.1 Mounting MITOMI device to microfluidic control elements

1. The assembled device is mounted on a light microscope and connected to tubing (for details see Figure A.3).
2. Control channels are filled with DI water by actuation with ~5 psi of pneumatic pressure which forces the air from the dead-end channels into the bulk porous silicone. This procedure eliminates subsequent gas transfer into the flow layer upon valve actuation, as well as prevents evaporation of the liquid contained in the flow layer.
3. Devices are actuated with 15 to 20 psi in the control lines and between 5 and 8 psi for the flow line.
4. Upon actuation button membrane and sandwich valves are opened again; chamber valves remain closed during the following initial surface derivatization steps to prevent liquid from entering the sample containing chambers.

A.4.6.2 Surface chemistry

The surface area within the flow channels of the device is modified by depositing layers of BSA-biotin, NA/PBS, and biotinylated antibody onto the epoxy coated glass slide (see Figure A.4). Using a syringe the different sample solutions are loaded into short pieces

(30-40 cm) of clean Tygon tubing which are hooked to a metal pin that is then pushed into the corresponding flow sample inlet holes (see Figure A.2).

1. Tubing with 30 μ l of BSA-biotin is inserted into the sample inlet hole S6. The port on manifold 2 for flow inlet regulation (see Figure A.3) is actuated and valves are opened by switching the corresponding toggles on manifold 1 in the following order: valve 6 controlling sample flow inlet S6, equalizer (1), array inlet (8).
2. Once the air in the channels of the array is displaced with liquid, the outlet valve (O) is opened and the equalizer valve (1) opened.
3. After flowing BSA-biotin for \sim 20 min the array inlet valve (8) and the sample valve (6) are closed again and the port on manifold 2 turned off.
4. The tubing is disconnected from the port of manifold 2.
5. The process of valve and port opening / closing is performed for the following samples.
6. After BSA-biotin derivatization the surface area is washed for 2-3 min with 5 μ l PBS (S7).
7. A 30 μ l solution of NA/PBS (S5) is flowed for \sim 20 min and washed again for 2-3 min with 5 μ l PBS.
8. The button membrane (B) is closed and PBS washing continued for 1 min (\sim 2 μ l) making sure button is closed.
9. The remaining surface area is passivated with 30 μ l BSA-biotin (\sim 20 min) and washed with 5 μ l PBS (2-3 min) while button is actuated.
10. 30 μ l of a 1:1 solution of biotinylated penta-His antibody in 2% BSA is loaded (S4). To ensure that all channels are saturated with antibody solution the button membrane is opened only after flowing 5 μ l.
11. After finishing the antibody deposition (total \sim 20 min), the surface is washed again with 5 μ l PBS for 2-3 min.
12. The surface derivatization procedure is finished with a final 5 μ l PBS washing step (see Note 4).

A.4.6.3 DNA pull-down & detection of interactions on-chip

1. 25 μ l TNT T7 coupled wheat germ extract is prepared and spiked with 1 μ l tRNA^{Lys}-Bodipy-FI and 2 μ l of linear expression ready template coding for the appropriate transcription factor (TF).
2. The mixture is immediately loaded onto the device (S3) and flushed for 10 μ l (around 5-7 min) while the button membrane is closed.
3. Chamber valves (C) are opened and the outlet valve (O) is closed to allow for dead end filling of the chambers with wheat germ extract.
4. Chamber valves are closed and the outlet valve opened again and flushing is continued for an additional 10 μ l (5-7 min).
5. Sandwich valves (S) that separate each unit cell are closed.

Appendix A. MITOMI - A microfluidic platform for in vitro characterization of transcription factor-DNA interactions

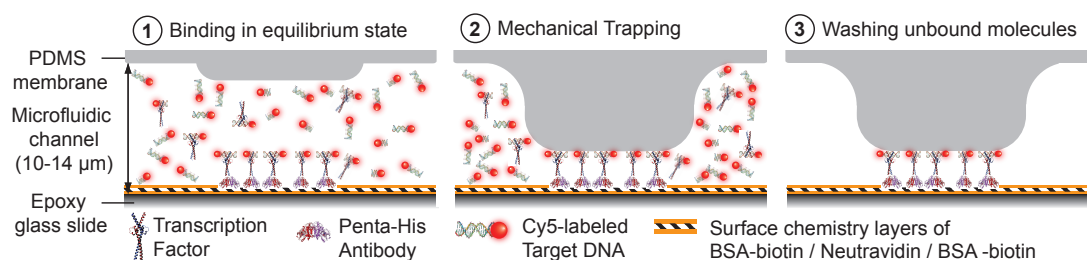


Figure A.4: Schematic of the MITOMI principle. The gray structure at the top of each panel represents the deflectable button membrane that can be brought into contact with the glass surface. (1) His-tagged TFs are localized to immobilized penta-His antibody at the epoxy-coated glass slide. Specific binding between Cy5-labeled target DNA and TFs are at steady state when (2) the button membrane is actuated and brought into contact with the surface. Any molecules in solution are expelled while surface-bound material is mechanically trapped. (3) Unbound material that was not physically protected is washed away, and the remaining molecules are quantified.

6. After ensuring that all sandwich valves are closed the chamber valves and button membranes are opened.
7. The device is incubated for 90 min at room temperature to allow for protein synthesis and diffusion of the samples to the immobilized antibody under the button membrane.
8. After the incubation period the device is scanned on a modified ArrayWoRx microarray scanner.
9. Button membranes are closed to trap bound samples.
10. Chamber valves are closed, sandwich valves opened and the channels washed with 10 μ l PBS (5-7 min).
11. The washed device is scanned once more with closed button membranes to detect the trapped molecules.

A.4.7 Data acquisition & analysis

1. For each experiment two images are analyzed with GenePix3.0 (Molecular Devices): The first image, taken directly after the 60-90 min incubation period before washing, is used to determine the concentration of solution phase or total target DNA concentration (Cy5 channel). The second image, taken after MITOMI and the final PBS wash, is used to determine the concentration of surface bound protein (FITC channel) as well as surface bound target DNA (Cy5 channel).
2. Dissociation equilibrium constants can be calculated for each experiment using a curve fitting software (e.g. Graphpad Prism5 or Mathematica) by performing global nonlinear regression fits using a one site binding model to the data plotted as surface bound target DNA (RFU) divided by surface protein concentration (RFU) (or effectively fractional occupancy) as a function of total target DNA concentration in RFU (see Figure A.5)
3. Relative K_d (RFU^{-1}) must be transformed into absolute K_d (M^{-1}) using a calibration curve previously established by measuring known quantities of 5'CompCy5 primer.

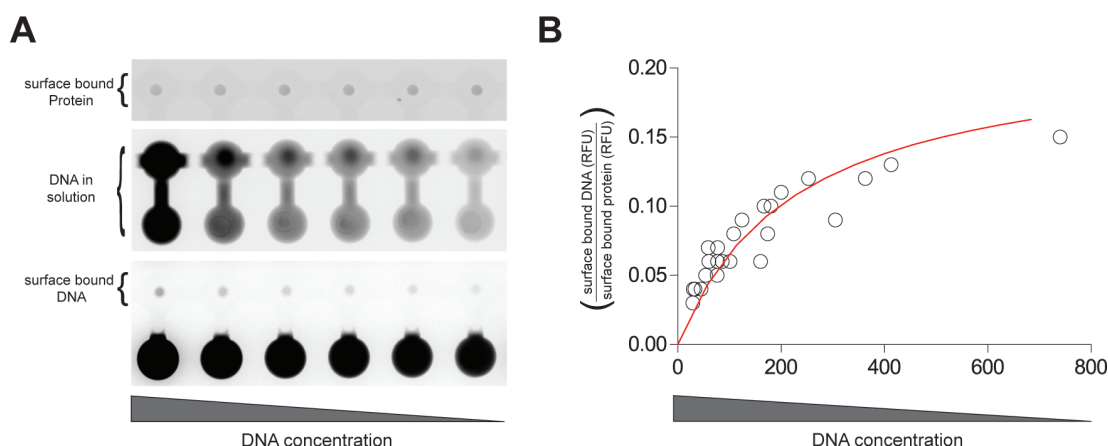


Figure A.5: Steps to analysis of MITOMI experiments. (A) Fluorescence scans of subsequent MITOMI steps. Scan of immobilized, fluorescence labeled protein (Top), solubilized target DNA (Middle), and surface bound target DNA after mechanical trapping and washing step (Bottom). (B) The fraction of surface bound target DNA is plotted against concentration of target DNA in solution. Dissociation constants are determined by performing a nonlinear regression fit using a one site binding model.

4. The change of free Gibbs energy ($\Delta\Delta G$) with $\Delta\Delta G = RT\ln(K_{d,ref})$ is calculated at a temperature of 298 K, where the highest affinity sequence serves as the reference.

A.5 Notes

1. Linear expression templates can also be synthesized from bacterial cDNA clones after lysing them in 2.5 μ l Lyse n'Go buffer (Pierce) at 95°C for 7 min. The lysate serves as template in an Expand High Fidelity PCR reaction (Roche). The first PCR product is then purified using the Qiaquick 96 PCR purification kit (Qiagen) and eluted in 80 μ l of 10 mM TrisCl, pH 8.5.
2. The on-chip DNA concentration can be increased by raising the numbers of repetitive stamps of sample DNA per spot during the spotting routine (multiple returns of spotting pin to the same spot)
3. The humidity inside the spotter is set to 65-80% to prevent the samples in the source plate from evaporating during long spotting routines (> 2 h).
4. An additional passivation step can be included by flowing 5 μ l BSA-biotin for 2-3 min after the antibody immobilization, then closing the button, followed by 2-3 min of 5 μ l BSA-biotin. This additional BSA-biotin passivation step was found to reduce background signal.

B Primer library

B.1 Primers for Drosophila TF-array

Table B.1: Universal primers for amplification of ORF-GST fragment from pF3A-WG-GST vector.

Name, Length, Abbreviation	Sequence (Comment)
5'-forward, 23 bp sp6-for-10down	5'-GTATCCGCTCATGGATCTCGATC-3' (primes 10 bp downstream of SP6 promoter)
3'-reverse, 22 bp sp6-rev-24up	5'-CGGTTTTATGGACAGCAAGCGA-3' (primes 24 bp upstream of T7 terminator)

Table B.2: Gene-specific primer sequences of target DNA.

TF name	Sequence (5'-clamp - TF-consensus motif - 5'-extension-3')
bcd	5'-CGCGGATTAGCTCCGGCGGTATGAC-3'
D	5'-CGCGTCCATTGTTCTCTCCGGCGGTATGAC-3'
gt	5'-CGCGTCCATTGTTCTCTCCGGCGGTATGAC-3'
kni	5'-CGCAAAACTAGAGCAACTCCGGCGGTATGAC-3'
Kr	5'-CGCTAACCCTTTTGCTCCGGCGGTATGAC-3'
Mad	5'-CGCGCTGCCGGCGCGGCCTCCGGCGGTATGAC-3'
Med	5'-CGCAACAGGCGAAACTCCGGCGGTATGAC-3'
pan	5'-CGCCTTTGATCCTCCGGCGGTATGAC-3'
prd	5'-CGCCCAATTTGTCACGCTCTCCGGCGGTATGAC-3'
tin	5'-CGCCTCAAGTGCTCCGGCGGTATGAC-3'
ttk	5'-CGCATTATCCTGGCTCCGGCGGTATGAC-3'
twi	5'-CGCCCCGCATATGTTGCTCCGGCGGTATGAC-3'
5'-CompCy5	5'-[Cy5]GTCATACCGCCGGA-3'

B.2 Primers to generate linear templates from yeast cells

Table B.3: Yeast cell MORF-plasmid (BG1805) collection

Name, Abbreviation (Comment), Length	Sequence
MORF forward extension MORF-WG-SP6-FW (used in 2-step, 1.5-step and single-step PCR), 94 bp	5'-[GATCTTAAGGCTAGAGTAC][ATTAGGTGACACTATAGAAGGGG] [GAATACAAGCTACTTGTTCTTTTGCA][gtacaaaaagcaggcac AAAATG]-3' 5'-[5'-Final-overhang (19bp)][SP6 promoter with <u>overhang</u> (24bp)] [beta-globin(27 bp)][BG1805-plasmid-specific incl. Kozak+Start (24 bp)]-3'
MORF forward extension MORF-WG-T7-FW (used in 2-step, 1.5-step and single-step PCR), 89 bp	5'-[GATCTTAAGGCTAGAGTAC][TAATACGACTCACTATAGG] [GAATACAAGCTACTTGTTCTTTTGCA][gtacaaaaagcaggcac AAAATG]-3' 5'-[5'-Final-overhang (19bp)][T7 promoter (19bp)] [beta-globin (27bp)][BG1805-plasmid-specific incl. Kozak+Start (24 bp)]-3'
MORF reverse extension MORF-WG-RV (used in 2-step, 1.5-step and single-step PCR), 123 bp	5'-[CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTTATGCTAG] [TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT][gtagcagcctgagtcg][TTATTAA] [GCATAATCAGGAACATCGTATG]-3' 5-[T7-terminator incl. 3-Final-overhang (48bp)] [Poly(A)-track][linker sequence (30bp)][2xSTOP (6pb)] [BG1850-plasmid-specific (23bp)]-3'
Gene-specific forward MORF-attB-FW (primes in attB region of BG1805 plasmid), 24 bp	5'-gtacaaaaagcaggcac AAAATG -3' 5'-[BG1805-plasmid-specific + KOZAK + START]-3'
Gene-specific reverse MORF-HA-RV (primes after HA-tag in BG1805 plasmid), 29 bp	5'- TTATTAAGCATAATCAGGAACATCGTATG -3' 5'-[2 x STOP + BG1805-plasmid-specific]-3'
Final forward 5'-Final-FW , 19 bp	5'-GATCTTAAGGCTAGAGTAC-3'
Final reverse 3'-Final-RV , 18 bp	5'-CAAAAAACCCCTCAAGAC-3'

WG primers designed for cell-free protein expression with TNT Wheat Germ ITT kits (Promega)

C Protein Purification

C.1 Buffers

Table C.1: Buffers for protein purification under native conditions.

Buffer	Chemicals
Stock buffer	50 mM NaH ₂ PO ₄ ; 0.3 M NaCl; pH 8.0
Lysis buffer	10 mM imidazole; 50 mM NaH ₂ PO ₄ ; 0.3 M NaCl; pH 8.0
Wash buffer	20 mM imidazole; 50 mM NaH ₂ PO ₄ ; 0.3 M NaCl; pH 8.0
Elution buffer	250 mM imidazole; 50 mM NaH ₂ PO ₄ ; 0.3 M NaCl; pH 8.0

Add 0.05% Tween 20 (5 ml of a 10% Tween 20 stock solution). All buffers adjusted to pH 8.0 using NaOH.

Table C.2: Buffers for protein purification under denaturing conditions.

Buffer	Chemicals
Stock buffer	0.1 M NaH ₂ PO ₄ ; 0.01 M Tris-Cl; pH 8.0
Lysis buffer	8 M urea; 0.1 M NaH ₂ PO ₄ ; 0.01 M Tris-Cl; pH 8.0
Wash buffer	8 M urea; 0.1 M NaH ₂ PO ₄ ; 0.01 M Tris-Cl; pH 6.3
Elution buffer 1	8 M urea; 0.1 M NaH ₂ PO ₄ ; 0.01 M Tris-Cl; pH 5.9
Elution buffer 2	8 M urea; 0.1 M NaH ₂ PO ₄ ; 0.01 M Tris-Cl; pH 4.5

All buffers adjusted immediately prior to use using NaOH (pH 8.0) and HCl (all other pH).

C.2 SDS-PAGE

C.2.1 SDS gels

SDS gels were casted using the following protocol for 13.5% gels.

Running gel solution (pH 8.8)

6.0	ml	40% Acrylamide/Bisacrylamide (37.5 : 1 cross-linker ratio)
7.5	ml	1 M Tris/HCl, pH 8.8
0.2	ml	10% SDS
0.1	ml	10% APS
6.1	ml	DI water
20.0	ml	total
10	μl	TEMED (added when ready to cast)

Ammonium persulfate was prepared freshly, and stored at 4°C for max. 1 week.
Acrylamide/bisacrylamide solutions were kept in the dark.

Stacking Gel Solution (pH 6.8)

10	ml	40% Acrylamide/Bisacrylamide (37.5 : 1 cross-linker ratio)
1.25	ml	1 M Tris/HCl, pH 6.8
0.1	ml	10% SDS
50	μl	10% APS
7.6	ml	DI water
10.0	ml	total
5	μl	TEMED (added when ready to cast)

Appendix C. Protein Purification

C.2.2 SDS PAGE - Buffer System (Laemmli)

Sample Buffer (SDS Reducing Buffer) - added to protein pellet

3.55	ml	deionized water
0.625	ml	1 M Tris-HCl, pH 6.8 (6 g Tris base in 60 ml DI water)
3.15	ml	glycerol
2.0	ml	10% (w/v) SDS (10 g/100 ml)
0.2	ml	0.5% (w/v) Bromophenol Blue
<hr/>		
9.5	ml	total

50 μ l β -mercaptoethanol (BME) was added to 950 μ l sample buffer prior to use. Sample (protein pellet) was diluted at least 1:2 with sample buffer and heated at 95°C for 4 minutes.

10x Electrode (Running) Buffer, pH 8.3 (1 L)

30.3 g Tris base

144 g glycine

10 g SDS

1 L DI water

Kept at 4°C. If precipitation occurred, warmed up to room temperature before use. 50 ml of 10x stock were added to 450 ml deionized water for each electrophoresis run.

Coomassie staining solution

50% MeOH

10% Acetic Acid

0.05% Brilliant Blue R-250 (dissolved in methanol)

Destaining solution 50% MeOH

40% milli-Q water

10% Acetic Acid

D Antibodies

D.1 Primary antibodies

Table D.1: Primary antibodies (biotinylated) for on-chip protein pulldown used in chapter 4

antibody name (Epitope)	supplier (art. no.)	isotype	used for yeast collection	on-chip dilution (conc.)	Western blot dilution (conc.)
anti- GFP antibody (Biotin)	abcam (ab6658)	IgG goat polyclonal	GFP	1:100 (10 μ g/ml)	
anti- GST antibody (Biotin)	abcam (ab71283)	IgG goat polyclonal	GST	1:100 (10 μ g/ml)	1:2,000 (0.5 μ g/ml)
anti- protein A antibody (Biotin)	abcam (ab18598)	IgY chicken polyclonal	MORF	1:20 (50 μ g/ml)	1:2,000 (0.4 μ g/ml)
anti- HA antibody (Biotin)	abcam (ab26228)	IgG rabbit polyclonal	MORF	1:100 (10 μ g/ml)	
Penta-His Biotin Conjugate	Qiagen (34440)	IgG1 mouse monoclonal	MORF	1:50 (4 μ g/ml)	1:2,000 (0.1 μ g/ml)

D.2 Secondary antibodies

Table D.2: Secondary antibodies (fluorescence-labeled) for on-chip protein detection used in chapter 4

antibody name (Epitope)	supplier (art. no.)	isotype	used for yeast collection	on-chip dilution (conc.)
anti- GST antibody (FITC)	abcam (ab71283)	IgG goat polyclonal	GST	
anti- GST antibody (Hilyte Fluor 647)	abcam (ab64370)	IgG2a mouse monoclonal	GST	
Penta-His Alexa Fluor 488 Conj.	Qiagen (35310)	IgG1 mouse monoclonal	MORF GST	1:200 (1 μ g/ml)
Penta-His Alexa Fluor 555 Conj.	Qiagen (35350)	IgG1 mouse monoclonal	MORF GST	1:200 (1 μ g/ml)
Penta-His Alexa Fluor 647 Conj.	Qiagen (35370)	IgG1 mouse monoclonal	MORF GST	1:200 (1 μ g/ml)

E Plasmid Purification Protocols

All plasmid purification steps from *E. coli* overnight cultures in LB medium were carried out at room temperature using Qiagen plasmid purification kits.

E.1 Plasmid DNA Purification using the QIAprep Spin Miniprep Kit

All centrifugation steps were carried out at room temperature with 14,680 rpm in a conventional table-top microcentrifuge (Eppendorf 5424).

1. Pellet 1-5 ml bacterial overnight culture by centrifugation at 10,000 rpm for 3 min at room temperature (15-25°C).
2. Resuspend pelleted bacterial cells in 250 µl Buffer P1 (with RNase A solution added) and transfer to a microcentrifuge tube.
3. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4-6 times until the solution becomes clear. Do not allow the lysis reaction to proceed for more than 5 min. If using LyseBlue (1 to 1000 in buffer P1) reagent, the solution will turn blue.
4. Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4-6 times. If using LyseBlue reagent, the solution will turn colorless.
5. Centrifuge for 10 min.
6. Apply the supernatant from step 5 to the QIAprep spin column by decanting or pipetting. Centrifuge for 30-60 s and discard the flow-through.
7. Wash the QIAprep spin column by adding 0.75 ml Buffer PE (with 96% ethanol added before use). Centrifuge for 30-60 s and discard the flow-through. Transfer the QIAprep spin column to the collection tube.
8. Centrifuge for 1 min to remove residual wash buffer.
9. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris-Cl, pH 8.5) or water to the center of the QIAprep spin column, let stand for 1 min, and centrifuge for 1 min.

E.2 Plasmid DNA Purification using the QIAprep 96 Turbo Miniprep Kit

1. Resuspend pelleted bacterial cells in 250 μ l Buffer P1 and transfer to the flat bottom block (if cells were not harvested in this block) provided with the kit.
2. Add 250 μ l Buffer P2 to each sample. Dry the top of the flat-bottom block with a paper towel, seal the block with the tape provided, gently invert the block 4-6 times to mix, and incubate at room temperature for 5 min. It is important to mix gently by inverting the block. Do not shake vigorously, as this will result in shearing of genomic DNA. If necessary, continue inverting the block until the solution becomes viscous and slightly clear.
3. Remove the tape from the block. Add 350 μ l Buffer N3 to each sample, dry the top of the flat-bottom block with a paper towel, and seal the block with a new tape sheet. Gently invert the block 4-6 times.
4. Remove the tape from the block. Pipet the lysates from step 3 (850 μ l per well) into the wells of the TurboFilter plate. Unused wells of the TurboFilter plate should be sealed with tape. Apply vacuum until all samples have passed through.
5. Switch off vacuum and ventilate the QIAvac 96 slowly. Discard the TurboFilter plate. Transfer the QIAprep plate containing the cleared lysates to the top plate of the manifold. Seal any unused wells of the QIAprep plate with tape.
6. Replace plate holder in the base with waste tray. Place the top plate squarely over the base, making sure that the QIAprep plate is seated securely. Apply vacuum. The flow-through is collected in the waste tray.
7. Switch off vacuum. Wash QIAprep plate by adding 0.9 ml of Buffer PE to each well and applying vacuum. Repeat once.
8. After Buffer PE has been drawn through all wells, apply maximum vacuum for an additional 10 min to dry the membrane. Important: This step removes residual Buffer PE from the membrane. The removal is only effective when maximum vacuum is used.
9. Switch off vacuum, and ventilate the QIAvac 96 slowly. Lift the top plate from the base (not the QIAprep plate from the top plate), vigorously tap the top plate on a stack of absorbent paper until no drops come out, and blot the nozzles of the QIAprep plate with clean absorbent paper to remove residual ethanol from the buffer PE.
10. Replace waste tray with an empty blue collection microtube rack (provided with the QIAvac 96). Place a 96-well microplate directly on the rack. Place the top plate back on the base, making sure that the QIAprep plate is positioned securely.
11. To elute DNA, add 100 μ l of Buffer EB (10 mM Tris-Cl, pH 8.5) or water to the center of each well of the QIAprep plate, let stand for 1 min, and apply maximum vacuum for 5 min. Switch off vacuum and ventilate QIAvac 96 slowly. For increased DNA concentration, an elution volume of 75 μ l can be used.



Bibliography

Bibliography

- [1] Furney, S. J., Higgins, D. G., Ouzounis, C. A. & Lopez-Bigas, N. Structural and functional properties of genes involved in human cancer. *BMC Genomics* **7**, 3 (2006).
- [2] Boyadjiev, S. A. & Jabs, E. W. Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clinical genetics* **57**, 253–266 (2000).
- [3] Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
- [4] Lopez-Bigas, N., De, S. & Teichmann, S. A. Functional protein divergence in the evolution of Homo sapiens. *Genome Biology* **9**, R33 (2008).
- [5] MacBeath, G. & Schreiber, S. L. Printing Proteins as Microarrays for High-Throughput Function Determination. *Science* **289**, 1–4 (2000).
- [6] Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
- [7] Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
- [8] Hu, S. *et al.* Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610–622 (2009).
- [9] Ramachandran, N. *et al.* Next-generation high-density self-assembling functional protein arrays. *Nature methods* **5**, 535–538 (2008).
- [10] Gerber, D., Maerkl, S. J. & Quake, S. R. An in vitro microfluidic approach to generating protein-interaction networks. *Nature methods* **6**, 71–74 (2009).
- [11] Einav, S. *et al.* Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nature Biotechnology* **26**, 1019–1027 (2008).

- [12] Maerkl, S. J. & Quake, S. R. Experimental determination of the evolvability of a transcription factor. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18650–18655 (2009).
- [13] Fordyce, P. M. *et al.* De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature Biotechnology* **28**, 970–975 (2010).
- [14] Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- [15] Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4569–4574 (2001).
- [16] Deane, C. M., Salwiński, L., Xenarios, I. & Eisenberg, D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Molecular & cellular proteomics : MCP* **1**, 349–356 (2002).
- [17] Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nature methods* **7**, 667–8– author reply 668 (2010).
- [18] Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
- [19] Ramaswamy, S. & Golub, T. R. DNA microarrays in clinical oncology. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 1932–1941 (2002).
- [20] Bulyk, M. L. DNA microarray technologies for measuring protein-DNA interactions. *Current Opinion in Biotechnology* **17**, 422–430 (2006).
- [21] Silzel, J. W., Cercek, B., Dodson, C., Tsay, T. & Obremski, R. J. Mass-sensing, multianalyte microarray immunoassay with imaging detection. *Clinical chemistry* **44**, 2036–2043 (1998).
- [22] Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genetics* **26**, 283–289 (2000).
- [23] Huang, R. P. Detection of multiple proteins in an antibody-based protein microarray system. *Journal of immunological methods* **255**, 1–13 (2001).
- [24] Ramachandran, N., Srivastava, S. & LaBaer, J. Applications of protein microarrays for biomarker discovery. *Proteomics. Clinical applications* **2**, 1444–1459 (2008).

Bibliography

- [25] Joos, T. O. *et al.* A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* **21**, 2641–2650 (2000).
- [26] Bacarese-Hamilton, T., Bistoni, F. & Crisanti, A. Protein microarrays: from serodiagnosis to whole proteome scale analysis of the immune response against pathogenic microorganisms. *BioTechniques Suppl*, 24–29 (2002).
- [27] Paweletz, C. P. *et al.* Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* **20**, 1981–1989 (2001).
- [28] Ptacek, J. *et al.* Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679–684 (2005).
- [29] Stiffler, M. A., Grantcharova, V. P., Sevecka, M. & MacBeath, G. Uncovering quantitative protein interaction networks for mouse PDZ domains using protein microarrays. *Journal of the American Chemical Society* **128**, 5913–5922 (2006).
- [30] Pace, C. N., Treviño, S., Prabhakaran, E. & Scholtz, J. M. Protein structure, stability and solubility in water and other solvents. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1225–34– discussion 1234–5 (2004).
- [31] Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M. Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nature Biotechnology* **17**, 573–577 (1999).
- [32] He, M. & Taussig, M. J. Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Research* 1–6 (2001).
- [33] Ramachandran, N. *et al.* Self-Assembling Protein Microarrays. *Science* **305**, 86–90 (2004).
- [34] He, M. *et al.* Printing protein arrays from DNA arrays. *Nature methods* **5**, 175–177 (2008).
- [35] Chatterjee, D. K. *et al.* Protein microarray on-demand: a novel protein microarray system. *PLoS ONE* **3**, e3265 (2008).
- [36] Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
- [37] Jolma, A. & Taipale, J. Methods for Analysis of Transcription Factor DNA-Binding Specificity In Vitro. *Sub-cellular biochemistry* **52**, 155–173 (2011).

- [38] Galas, D. J. & Schmitz, A. DNAase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Research* **5**, 1–14 (1978).
- [39] Tullius, T. D., Dombroski, B. A., Churchill, M. E. & Kam, L. Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. *Methods in enzymology* **155**, 537–558 (1987).
- [40] Brenowitz, M., Senear, D. F. & Kingston, R. E. DNase I footprint analysis of protein-DNA binding. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.] Chapter 12*, Unit 12.4 (2001).
- [41] Connaghan-Jones, K. D., Moody, A. D. & Bain, D. L. Quantitative DNase footprint titration: a tool for analyzing the energetics of protein-DNA interactions. *Nature Protocols* **3**, 900–914 (2008).
- [42] Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Research* **9**, 3047–3060 (1981).
- [43] Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* **9**, 6505–6525 (1981).
- [44] Smith, A. J. P. & Humphries, S. E. Characterization of DNA-binding proteins using multiplexed competitor EMSA. *Journal of Molecular Biology* **385**, 714–717 (2009).
- [45] Stenger, D., Grussem, W. & Baginsky, S. Mass spectrometric identification of RNA binding proteins from dried EMSA gels. *Journal of Proteome Research* **3**, 662–664 (2004).
- [46] Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
- [47] Wright, W. E., Binder, M. & Funk, W. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Molecular and Cellular Biology* **11**, 4104–4110 (1991).
- [48] Blackwell, T. K. & Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104–1110 (1990).
- [49] Jolma, A. *et al.* Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research* **20**, 861–873 (2010).
- [50] Zykovich, A., Korf, I. & Segal, D. J. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Research* **37**, e151 (2009).

Bibliography

- [51] Slattery, M. *et al.* Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
- [52] Horak, C. & Snyder, M. ChIP-chip: A genomic approach for identifying transcription factor binding sites. *Methods in enzymology* **350**, 469–483 (2002).
- [53] Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nature Genetics* **28**, 327–334 (2001).
- [54] Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- [55] Lee, T. I. *et al.* Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science* 1–6 (2002).
- [56] Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
- [57] Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Research* **36**, 5221–5231 (2008).
- [58] van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature Biotechnology* **18**, 424–428 (2000).
- [59] Liu, X., Noll, D. M., Lieb, J. D. & Clarke, N. D. DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Research* **15**, 421–427 (2005).
- [60] Rhee, H. S. & Pugh, B. F. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell* **147**, 1408–1419 (2011).
- [61] Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
- [62] Truax, A. D. & Greer, S. F. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods in molecular biology (Clifton, N.J.)* **809**, 175–188 (2012).
- [63] Li, J. J. & Herskowitz, I. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* **262**, 1870–1874 (1993).
- [64] Walhout, A. J. & Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297–306 (2001).

- [65] Deplancke, B., Dupuy, D., Vidal, M. & Walhout, A. J. M. A gateway-compatible yeast one-hybrid system. *Genome Research* **14**, 2093–2101 (2004).
- [66] Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7382–7387 (2000).
- [67] Meng, X., Brodsky, M. H. & Wolfe, S. A. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature Biotechnology* **23**, 988–994 (2005).
- [68] Noyes, M. B. *et al.* A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Research* **36**, 2547–2560 (2008).
- [69] Deplancke, B. *et al.* A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**, 1193–1205 (2006).
- [70] Reece-Hoyes, J. S. *et al.* Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nature methods* **8**, 1059–1064 (2011).
- [71] Reece-Hoyes, J. S. *et al.* Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. *Nature methods* **8**, 1050–1052 (2011).
- [72] Gaudinier, A. *et al.* Enhanced Y1H assays for *Arabidopsis*. *Nature methods* **8**, 1053–1055 (2011).
- [73] Stormo, G. D. & Zhao, Y. Determining the specificity of protein-DNA interactions. *Nature Reviews Genetics* **11**, 751–760 (2010).
- [74] Hens, K. *et al.* Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nature methods* **8**, 1065–1070 (2011).
- [75] Christensen, R. G. *et al.* A modified bacterial one-hybrid system yields improved quantitative models of transcription factor specificity. *Nucleic Acids Research* **39**, e83 (2011).
- [76] Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nature Genetics* **36**, 1331–1339 (2004).
- [77] Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnology* **24**, 1429–1435 (2006).
- [78] Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).

Bibliography

- [79] Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Research* **19**, 556–566 (2009).
- [80] Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
- [81] Siggers, T., Duyzend, M. H., Reddy, J., Khan, S. & Bulyk, M. L. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Molecular Systems Biology* **7**, 555 (2011).
- [82] Warren, C. L. *et al.* Defining the sequence-recognition profile of DNA-binding molecules. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 867–872 (2006).
- [83] Puckett, J. W. *et al.* Quantitative microarray profiling of DNA-binding molecules. *Journal of the American Chemical Society* **129**, 12310–12319 (2007).
- [84] Nutiu, R. *et al.* Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology* **29**, 659–664 (2011).
- [85] Liedberg, B., Nylander, C. & Lunstrom, I. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators* **4**, 299–304 (1983).
- [86] Arda, H. E. & Walhout, A. J. M. Gene-centered regulatory networks. *Briefings in Functional Genomics* **9**, 4–12 (2010).
- [87] Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**, 669–680 (2009).
- [88] Zhao, Y., Granas, D. & Stormo, G. D. Inferring binding energies from selected binding sites. *PLoS Computational Biology* **5**, e1000590 (2009).
- [89] Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E. & Boeke, J. D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 10315–10320 (1996).
- [90] Deplancke, B. Experimental advances in the characterization of metazoan gene regulatory networks. *Briefings in functional genomics & proteomics* **8**, 12–27 (2009).
- [91] Fried, M. G. & Bromberg, J. L. Factors that affect the stability of protein-DNA complexes during gel electrophoresis. *Electrophoresis* **18**, 6–11 (1997).
- [92] Tullius, T. D. Physical studies of protein-DNA complexes by footprinting. *Annual review of biophysics and biophysical chemistry* **18**, 213–237 (1989).

- [93] Dai, S. M., Chen, H. H., Chang, C., Riggs, A. D. & Flanagan, S. D. Ligation-mediated PCR for quantitative in vivo footprinting. *Nature Biotechnology* **18**, 1108–1111 (2000).
- [94] Fägerstam, L. G., Frostell-Karlsson, A., Karlsson, R., Persson, B. & Rönnerberg, I. Biospecific interaction analysis using surface plasmon resonance detection applied to kinetic, binding site and concentration analysis. *Journal of chromatography* **597**, 397–410 (1992).
- [95] Carlson, C. D. *et al.* Specificity landscapes of DNA binding molecules elucidate biological function. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4544–4549 (2010).
- [96] Shultzaberger, R. K., Maerkl, S. J., Kirsch, J. F. & Eisen, M. B. Probing the informational and regulatory plasticity of a transcription factor DNA-binding domain. *PLoS Genetics* **8**, e1002614 (2012).
- [97] Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences of the United States of America* (2012).
- [98] Déjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic Loci. *Cell* **136**, 175–186 (2009).
- [99] Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature methods* **6**, 283–289 (2009).
- [100] Sabo, P. J. *et al.* Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nature methods* **3**, 511–518 (2006).
- [101] Crawford, G. E. *et al.* DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature methods* **3**, 503–509 (2006).
- [102] Pabo, C. O. & Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annual review of biochemistry* **61**, 1053–1095 (1992).
- [103] Schneider, T. D. & Stephens, R. M. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Research* **18**, 6097–6100 (1990).
- [104] Benos, P. V., Bulyk, M. L. & Stormo, G. D. Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Research* **30**, 4442–4451 (2002).
- [105] Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4**, 393–411 (2009).
- [106] Zhao, Y. & Stormo, G. D. Quantitative analysis demonstrates most transcription factors require only simple models of specificity. *Nature Biotechnology* **29**, 480–483 (2011).

Bibliography

- [107] Morris, Q., Bulyk, M. L. & Hughes, T. R. Jury remains out on simple models of transcription factor specificity. *Nature Biotechnology* **29**, 483–484 (2011).
- [108] Wunderlich, Z. & Mirny, L. A. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in genetics : TIG* **25**, 434–440 (2009).
- [109] Harrison, D. J. *et al.* Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* **261**, 895–897 (1993).
- [110] Manz, A., Graber, N. & Widmer, H. M. Miniaturized total chemical analysis systems: a novel concept for chemical sensing. *Sensors and actuators B: Chemical* **1**, 244–248 (1990).
- [111] Woolley, A. T. & Mathies, R. A. Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 11348–11352 (1994).
- [112] Unger, M. A., Chou, H. P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).
- [113] Marcus, J. S., Anderson, W. F. & Quake, S. R. Microfluidic single-cell mRNA isolation and analysis. *Analytical Chemistry* **78**, 3084–3089 (2006).
- [114] Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic Large-Scale Integration. *Science* **298**, 1–5 (2002).
- [115] Hansen, C. L., Skordalakes, E., Berger, J. M. & Quake, S. R. 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* 1–6 (2002).
- [116] Lagally, E. T., Emrich, C. A. & Mathies, R. A. Fully integrated PCR-capillary electrophoresis microsystem for DNA analysis. *Lab on a Chip* **1**, 102–107 (2001).
- [117] Liu, J., Enzelberger, M. & Quake, S. A nanoliter rotary device for polymerase chain reaction. *Electrophoresis* **23**, 1531–1536 (2002).
- [118] Ottesen, E. A., Hong, J. W., Quake, S. R. & Leadbetter, J. R. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**, 1464–1467 (2006).
- [119] Marcus, J. S., Anderson, W. F. & Quake, S. R. Parallel Picoliter RT-PCR Assays Using Microfluidics. *Analytical Chemistry* **78**, 956–958 (2006).
- [120] Kobel, S. A. *et al.* Automated analysis of single stem cells in microfluidic traps. *Lab on a Chip* (2012).

- [121] Wu, A. R. *et al.* Automated microfluidic chromatin immunoprecipitation from 2,000 cells. *Lab on a Chip* **9**, 1365–1370 (2009).
- [122] Wu, A. R. *et al.* High throughput automated chromatin immunoprecipitation as a platform for drug screening and antibody validation. *Lab on a Chip* **12**, 2190–2198 (2012).
- [123] Liu, W.-T. *et al.* Microfluidic device as a new platform for immunofluorescent detection of viruses. *Lab on a Chip* **5**, 1327–1330 (2005).
- [124] Emrich, C. A., Medintz, I. L., Chu, W. K. & Mathies, R. A. Microfabricated Two-Dimensional Electrophoresis Device for Differential Protein Expression Profiling. *Analytical Chemistry* **79**, 7360–7366 (2007).
- [125] Valero, A. *et al.* Gene transfer and protein dynamics in stem cells using single cell electroporation in a microfluidic device. *Lab on a Chip* **8**, 62 (2007).
- [126] Hong, J. W., Studer, V., Hang, G., Anderson, W. F. & Quake, S. R. A nanoliter-scale nucleic acid processor with parallel architecture. *Nature Biotechnology* **22**, 435–439 (2004).
- [127] Ashton, R., Padala, C. & Kane, R. S. Microfluidic separation of DNA. *Current Opinion in Biotechnology* **14**, 497–504 (2003).
- [128] Yu, H. *et al.* A simple, disposable microfluidic device for rapid protein concentration and purification via direct-printing. *Lab on a Chip* **8**, 1496–1501 (2008).
- [129] Huang, L. R., Cox, E. C., Austin, R. H. & Sturm, J. C. Continuous particle separation through deterministic lateral displacement. *Science* **304**, 987–990 (2004).
- [130] Liu, C., Cui, D. & Li, H. A hard-soft microfluidic-based biosensor flow cell for SPR imaging application. *Biosensors and Bioelectronics* **26**, 255–261 (2010).
- [131] Ghosh, T., Williams, L. & Mastrangelo, C. H. Label-free detection of protein binding with multisine SPR microchips. *Lab on a Chip* **11**, 4194–4199 (2011).
- [132] Rondelez, Y. *et al.* Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nature Biotechnology* **23**, 361–365 (2005).
- [133] Bessette, P. H., Hu, X., Soh, H. T. & Daugherty, P. S. Microfluidic library screening for mapping antibody epitopes. *Analytical Chemistry* **79**, 2174–2178 (2007).
- [134] Lee, P. J., Hung, P. J., Rao, V. M. & Lee, L. P. Nanoliter scale microbioreactor array for quantitative cell biology. *Biotechnology and Bioengineering* **94**, 5–14 (2006).

Bibliography

- [135] Rohde, C. B., Zeng, F., Gonzalez-Rubio, R., Angel, M. & Yanik, M. F. Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13891–13895 (2007).
- [136] Kobel, S., Valero, A., Latt, J., Renaud, P. & Lutolf, M. Optimization of microfluidic single cell trapping for long-term on-chip culture. *Lab on a Chip* **10**, 857–863 (2010).
- [137] Di Carlo, D., Aghdam, N. & Lee, L. P. Single-cell enzyme concentrations, kinetics, and inhibition analysis using high-density hydrodynamic cell isolation arrays. *Analytical Chemistry* **78**, 4925–4930 (2006).
- [138] Fu, A. Y., Spence, C., Scherer, A., Arnold, F. H. & Quake, S. R. A microfabricated fluorescence-activated cell sorter. *Nature Biotechnology* 1–3 (1999).
- [139] Braschler, T. *et al.* Continuous separation of cells by balanced dielectrophoretic forces at multiple frequencies. *Lab on a Chip* **8**, 280–286 (2008).
- [140] Bichsel, C. A. *et al.* Diagnostic microchip to assay 3D colony-growth potential of captured circulating tumor cells. *Lab on a Chip* **12**, 2313–2316 (2012).
- [141] Gervais, L., de Rooij, N. & Delamarche, E. Microfluidic chips for point-of-care immunodiagnosics. *Advanced materials* **23**, H151–76 (2011).
- [142] Lovchik, R. D., Kaigala, G. V., Georgiadis, M. & Delamarche, E. Micro-immunohistochemistry using a microfluidic probe. *Lab on a Chip* **12**, 1040–1043 (2012).
- [143] Dimov, I. K. *et al.* Stand-alone self-powered integrated microfluidic blood analysis system (SIMBAS). *Lab on a Chip* **11**, 845–850 (2011).
- [144] Raghav, S. K. *et al.* Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBP-beta and KAISO. *Molecular Cell* **46**, 335–350 (2012).
- [145] Schröter, C. *et al.* Topology and dynamics of the zebrafish segmentation clock core circuit. *PLoS Biology* **10**, e1001364 (2012).
- [146] Araci, I. E. & Quake, S. R. Microfluidic very large scale integration (mVLSI) with integrated micromechanical valves. *Lab on a Chip* **12**, 2803–2806 (2012).
- [147] Goldmann, T. & Gonzalez, J. S. DNA-printing: utilization of a standard inkjet printer for the transfer of nucleic acids to solid supports. *Journal of Biochemical and Biophysical Methods* **42**, 105–110 (2000).
- [148] Lausted, C. *et al.* POSaM: a fast, flexible, open-source, inkjet oligonucleotide synthesizer and microarrayer. *Genome Biology* **5**, R58 (2004).

-
- [149] Auburn, R. P. *et al.* Robotic spotting of cDNA and oligonucleotide microarrays. *Trends in Biotechnology* **23**, 374–379 (2005).
- [150] Piner, R., Zhu, J., Xu, F., Hong, S. & Mirkin, C. "Dip-Pen" nanolithography. *Science* **283**, 661–663 (1999).
- [151] Fidalgo, L. M. & Maerkl, S. J. A software-programmable microfluidic device for automated biology. *Lab on a Chip* **11**, 1612–1619 (2011).
- [152] Petranovic, D., Tyo, K., Vemuri, G. N. & Nielsen, J. Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism. *FEMS yeast research* **10**, 1046–1059 (2010).
- [153] Bharadwaj, P., Martins, R. & Macreadie, I. Yeast as a model for studying Alzheimer's disease. *FEMS yeast research* **10**, 961–969 (2010).
- [154] Miller-Fleming, L., Giorgini, F. & Outeiro, T. F. Yeast as a model for studying human neurodegenerative disorders. *Biotechnology journal* **3**, 325–338 (2008).
- [155] Rinaldi, T., Dallabona, C., Ferrero, I., Frontali, L. & Bolotin-Fukuhara, M. Mitochondrial diseases and the role of the yeast models. *FEMS yeast research* **10**, 1006–1022 (2010).
- [156] Barros, M. H., da Cunha, F. M., Oliveira, G. A., Tahara, E. B. & Kowaltowski, A. J. Yeast as a model to study mitochondrial mechanisms in ageing. *Mechanisms of ageing and development* **131**, 494–502 (2010).
- [157] Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- [158] Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- [159] Gelperin, D. M. *et al.* Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development* **19**, 2816–2826 (2005).
- [160] Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Molecular Cell* **21**, 319–330 (2006).
- [161] Newman, J. R. S. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
- [162] Rockel, S., Geertz, M. & Maerkl, S. J. MITOMI: a microfluidic platform for in vitro characterization of transcription factor-DNA interaction. *Methods in molecular biology (Clifton, N.J.)* **786**, 97–114 (2012).

Bibliography

- [163] Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 1–7 (2002).
- [164] Morita, E. H., Sawasaki, T., Tanaka, R., Endo, Y. & Kohno, T. A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein science : a publication of the Protein Society* **12**, 1216–1221 (2003).
- [165] He, B. Z., Holloway, A. K., Maerkl, S. J. & Kreitman, M. Does positive selection drive transcription factor binding site turnover? A test with *Drosophila* cis-regulatory modules. *PLoS Genetics* **7**, e1002053 (2011).
- [166] Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnology advances* (2011).
- [167] Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751–755 (2001).
- [168] Netzer, W. J. & Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343–349 (1997).
- [169] Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14652–14657 (2002).
- [170] Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- [171] Fortini, M. E. & Bonini, N. M. Modeling human neurodegenerative diseases in *Drosophila*: on a wing and a prayer. *Trends in genetics : TIG* **16**, 161–167 (2000).
- [172] Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research* **11**, 1114–1125 (2001).
- [173] Wilczynski, B. & Furlong, E. E. M. Challenges for modeling global gene regulatory networks during development: Insights from *Drosophila*. *Developmental Biology* **340**, 161–169 (2010).
- [174] Philippakis, A. A., Qureshi, A. M., Berger, M. F. & Bulyk, M. L. Design of compact, universal DNA microarrays for protein binding microarray experiments. *Journal of computational biology : a journal of computational molecular cell biology* **15**, 655–665 (2008).
- [175] Foat, B. C., Morozov, A. V. & Bussemaker, H. J. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* **22**, e141–9 (2006).

- [176] Zhou, X. & O'Shea, E. K. Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. *Molecular Cell* **42**, 826–836 (2011).
- [177] Yin, Z., Xu, X. L. & Frasch, M. Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* **124**, 4971–4982 (1997).
- [178] Kim, J., Johnson, K., Chen, H. J., Carroll, S. & Laughon, A. Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature* **388**, 304–308 (1997).
- [179] Chen, J. D. & Pirrotta, V. Multimerization of the Drosophila zeste protein is required for efficient DNA binding. *The EMBO Journal* **12**, 2075–2083 (1993).
- [180] Noyes, M. B. *et al.* Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell* **133**, 1277–1289 (2008).
- [181] Li, X. & Noll, M. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. *The EMBO Journal* **13**, 400–406 (1994).
- [182] Treisman, J. E., Lai, Z. C. & Rubin, G. M. Short sighted acts in the decapentaplegic pathway in Drosophila eye development and has homology to a mouse TGF-beta-responsive gene. *Development* **121**, 2835–2845 (1995).
- [183] Schroeder, M. D., Greer, C. & Gaul, U. How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. *Development* **138**, 3067–3078 (2011).
- [184] Warren, C. L. *et al.* Fabrication of duplex DNA microarrays incorporating methyl-5-cytosine. *Lab on a Chip* **12**, 376–380 (2012).
- [185] Zeevi, D. *et al.* Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Research* **21**, 2114–2128 (2011).
- [186] Rajkumar, A. S. & Maerkl, S. J. Rapid Synthesis of Defined Eukaryotic Promoter Libraries. *ACS Synthetic Biology* 120620090752001 (2012).
- [187] Pressman, S., Reinke, C. A., Wang, X. & Carthew, R. W. A Systematic Genetic Screen to Dissect the MicroRNA Pathway in Drosophila. *G3 (Bethesda, Md.)* **2**, 437–448 (2012).
- [188] Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325**, 429–432 (2009).

Bibliography

- [189] Beer, M. A. & Tavazoie, S. Predicting Gene Expression from Sequence. *Cell* 1–14 (2004).
- [190] Bulyk, M. L., Huang, X., Choo, Y. & Church, G. M. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7158–7163 (2001).
- [191] Majka, J. & Speck, C. Analysis of protein–DNA interactions using surface plasmon resonance. *Analytics of Protein–DNA Interactions* 13–36 (2007).
- [192] Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).
- [193] Duffy, D. C., McDonald, J. C., Schueller, O. J. & Whitesides, G. M. Rapid Prototyping of Microfluidic Systems in Poly(dimethylsiloxane). *Analytical Chemistry* **70**, 4974–4984 (1998).
- [194] Xia, Y. & Whitesides, G. M. Soft Lithography. *Angewandte Chemie (International ed. in English)* **37**, 550–575 (1998).

Glossary

Antibody Protein produced by certain cells of the immune system as response to a foreign substance (antigen). Since an antibody binds specifically to only one antigen, engineered or isolated antibodies can be used to detect or label other proteins or cells.

cDNA (complementary DNA) The DNA complement of an mRNA sequence, synthesized using reverse transcriptase.

Cell The smallest structural unit of living matter capable of functioning independently; a microscopic mass of protoplasm surrounded by a semi-permeable membrane, usually including one or more nuclei and various non-living products, capable - either alone or by interacting with other cells - of performing all the fundamental functions of life.

Chip A semiconductor device cut from a silicon wafer. Refers also to a microfluidic device.

ChIP as an abbreviation stands for chromatin immunoprecipitation and is technique used to determine specific interactions between proteins and DNA *in vivo* (inside a cell) by fixing proteins and associated chromatin in the cell lysate and purification of the DNA-protein complex with a specific antibody.

DamID (DNA adenine methyltransferase identification) Methylation-based chromatin profiling method, in which DNA binding sites are identified by expressing the DNA-binding protein as a fusion protein with DNA methyltransferase, which transfers a methyl group from S-adenosyl-methionine (SAM). Since this process does not occur naturally in eukaryotes, adenine methylation in a fusion protein of interest implies that the region is near a binding site.

DBD DNA binding domains are protein domains containing a specific motif that recognizes a specific DNA sequence. All transcription factors contain at least one DBD. **Diffusion** The movement of molecules from a region of higher concentration to a region of lower concentration.

Bibliography

DNA (deoxyribonucleic acid) The chain-like molecule, that stores and transmits genetic information, is composed of two strands running anti-parallel to each other, which are twisted into a double helical conformation.

DNA footprinting is a laboratory technique used to study where and how proteins and other ligands bind to DNA.

DNA polymerase Enzyme that catalyzes the synthesis of double-stranded DNA, using single-stranded DNA as a template.

Drosophila melanogaster The full name of a species of fruit fly that is widely used in genetic research as a eukaryotic model organism. Of the nearly 300 disease-causing genes in the human genome, more than half have an analogous gene in the Drosophila genome.

E. coli (*Escherichia coli*) Bacterium inhabiting the colon of many animal species, including human. *E. coli* is widely used as a model of cell biochemical function, and as a host for cloning DNA.

Electrophoresis A method of separating molecules by their movement within an electric field. The medium used is usually an agarose or polyacrylamide gel through which the molecules move.

Eukaryote Organisms that have their DNA in a nucleus enclosed by a membrane.

Exon Part of a eukaryotic gene that is transcribed and translated into a protein.

Gene Conceptually, the unit of heredity transmitted from generation to generation during sexual or asexual reproduction. More generally, the term is used in relation to the transmission and inheritance of particular identifiable traits. Since the molecular revolution, it is now known that a gene is a segment of nucleic acid that encodes peptide or RNA.

Gene expression The process by which a gene produces RNA and protein, and hence exerts its effects on the phenotype of an organism.

Genome The complete set of genetic material (genes and non-coding sequences) of an organism, virus or organelle, encoded in DNA or in RNA (some viruses) and inherited in its entirety through a set of chromosomes from the parents.

Genomics Is a discipline that uses molecular characterization to understand the struc-

ture, function and evolution of genes through.

Genetics (from Greek "genesis" meaning "origin") is the science of genes and heredity.

GRN (gene regulatory network) describes the collection of direct and indirect molecular interactions between oligonucleotide sequences (DNA and RNA) and other molecules in a cell, which govern the transcription rate of DNA.

Induction Process that triggers downstream effects such as transcription of a specific gene or synthesis of a specific protein through a physical or chemical stimuli.

In silico coined in analogy to commonly used Latin phrases in biology (*in vitro*, *in vivo*, *in situ*), meaning "performed on a computer or by computer simulation" as opposed to in biological experiments. In silico experiments became an essential part within the emerging field of systems biology to handle the growing amount of sequence and interaction databases.

In situ (from Latin meaning "in position") In biology and bioengineering it refers to the examination of an organism, cell, tissue, phenomenon or reaction, where it occurs without moving it to another location.

In vitro Cells growing in culture as opposed to in an organism (*in vivo*).

In vivo In an intact cell or organism (antonym to *in vitro*).

K_d abbreviation for dissociation constant, which characterizes the strength of binding (affinity) between molecules at equilibrium and is the inverse of the association constant K_m .

Klenow fragment A truncated form of the enzyme DNA polymerase I from *E. coli*, which lost its 5' → 3' exonuclease activity and is therefore used frequently for the production of synthetic DNA.

K_m abbreviation for association constant or Michaelis constant that describes the substrate concentration at its half maximum reaction rate in a Michaelis-Menten kinetic enzymatic model. As an inverse of the dissociation constant K_d , the K_m value is smaller for tighter binding of enzymes to the substrates.

Kozak (consensus) sequence nucleotide sequence in eukaryotic mRNA with the consensus, which is recognized by the ribosome as the start site for protein translation.

Lab-on-a-chip Device that integrates one or several laboratory functions on a single chip of

Bibliography

only millimeters to a few square centimeters in size.

Major groove As the twin helical strands, which form the DNA backbone, are not directly opposite to each other, the width between the strands, or grooves, are not identical. The major groove is 10 Å wider than the minor groove and allows better accessibility of binding proteins, such as TFs.

Mask Typically a transparent substrate with a pattern on its surface defined in an opaque material (chrome metal or ink) used in photolithography.

Microfabrication Term to describe the fabrication and patterning of structures in the micrometer range. Adapted from the semiconductor industry, these methodologies are increasingly used for bioengineering applications.

Microfluidics Fluidics in micrometers-sized channels that are characterized by a laminar flow regime.

Microtechnology The fabrication and application of materials, structures and systems with micron or submicron-scale features.

Microwell array Array of miniaturized wells with diameters and a depths of a few tens to hundreds micrometers.

Minor groove see major groove

mRNA (messenger RNA) The RNA transcript of a protein-encoding gene. The information encoded in the mRNA molecule is translated into a polypeptide of specific amino acid sequence by the ribosomes. In eukaryotes, mRNAs transfer genetic information from the DNA to ribosomes, where it is translated into protein.

ORF The open reading frame is a DNA sequence without a stop codon, which often used to predict the presence a gene or protein-coding regions.

PBM A protein-binding microarray is a high-throughput method to detect protein interactions and determine their function. Based on the DNA microarray technology for genomics studies, DNA oligonucleotides are immobilized to a solid support (generally a glass slide) and interrogated with proteins for a proteomics approach.

PCR Polymerase chain reactions are a widespread molecular biology procedure that allows the production of multiple copies (amplification) of a specific DNA sequence, provided

that the base pair sequence of each end of the target is known. It involves multiple cycles of DNA denaturation, primer annealing, and strand extension, and requires a thermostable DNA polymerase, deoxyribonucleotides, and specific oligonucleotides (primers).

PDMS Poly(dimethylsiloxane). An elastomeric silicone polymer that is commercially available and has properties that make it well suited to applications in microbiology.

Photolithography A process used to transfer a pattern from a mask onto a thin film of photosensitive polymer (photoresist) and then onto the surface of a substrate. Photolithography is commonly used in semiconductor fabrication to fabricate integrated circuits.

Photoresist A photoreactive polymer that undergoes chemical changes that lead to changes in physical properties (such as solubility) after exposure to ultraviolet light.

Plasmid Non-chromosomal, double-stranded DNA molecule that can replicate itself autonomously in a host cell and which are naturally found in many bacteria. Plasmids vary in sizes between 1 and 1000 kbp and often carry an antibiotic resistance genes are frequently located on plasmids. Because of their simple modes of modification and transfer between bacterial cells, plasmids are particularly important as vectors for genetic engineering.

Primer A short DNA fragment that anneals to a template of single-stranded DNA or RNA and thereby provides a starting point from which DNA polymerase extends a new DNA strand to produce a duplex molecule.

Promoter A specific region of DNA located near a gene, recognized by RNA polymerase for the initiation of gene transcription.

Proteome The entire set of proteins expressed by a genome particular cell type, tissue or organism, including all modifications. The term "proteome" is a mix of the words "protein" and "genome" and reflects the interplay between material containing genetic information and proteins on a systems level.

Proteomics The entire set of proteins expressed by a genome particular cell type, tissue or organism. Coined in 1997 in analogy to "genomics" it generally refers to large-scale protein analysis.

PWM In biology a position weight matrix or position-specific scoring matrix is a representation of DNA binding or consensus sequence that a protein binds to and is usually visualized as a sequence logo.

Bibliography

RNA Ribonucleic acid. An organic acid composed of repeating nucleotide units of adenine, guanine, cytosine and uracil, whose ribose components are linked by phospho-diester bonds. RNA is the only information-carrying material in some viruses.

RNA polymerase is an enzyme that initiates the transcription the genetic information from double-stranded DNA onto a single strand of RNA through binding to a specific region of DNA, known as promoter.

Saccharomyces cerevisiae is a species of yeast often used as a simple eukaryotic model organism in molecular and cell biology, which is naturally found on ripe fruits.

SDS-PAGE Abbreviation for sodium dodecyl sulfate polyacrylamide gel electrophoresis, a technique used to separate proteins according to their electrophoretic mobility, a function of the length and charge of the polypeptide chain, which migrate through an inert matrix (polyacrylamide gel) as a result of an imposed electric field and reside at a position that can be correlated to their size.

SELEX stands for Systematic Evolution of Ligands by EXponential enrichment and is also referred to a *in vitro selection* or *in vitro evolution*. A combinatorial technique used in molecular biology to select for oligonucleotide sequences that bind specifically and with high affinity to target ligands.

Sequence logo Graphical representation of nucleotide sequences (in one strand of DNA or RNA) to which proteins bind with a high probability (often referred to as binding motif). Also used to display amino acids in protein sequences.

Soft lithography A set of techniques that makes microstructures by printing, moulding and embossing using a patterned, elastomeric stamp or mould, and/or a polymeric substrate.

Spin coating A process for depositing uniform layers of polymer on a substrate. Rotating the substrate at a high speed spreads the material uniformly over the surface. The viscosity of the material and the rotational velocity of the substrate control the thickness of the layer of material; surface tension flattens the surface of the spun film.

TMCS Trimethylchlorosilane, used as anti-adhesive in microfabrication.

Terminator Specific DNA sequence downstream of a genes coding region that is recognized by RNA polymerase as a signal to stop transcription.

Transcription Process of RNA synthesis from a DNA template via the enzyme RNA poly-

merase, in which the genetic information is copied from a double-stranded (DNA) to a single-stranded oligonucleotide (RNA), enabling the transport to the place of translation.

Transcription factor A protein that binds to specific DNA sequences and regulates the transcription of genes by either activation or repression.

Translation Final process of protein synthesis, in which genetic information in the form of nucleotides on mRNA is translated or decoded into specific amino acids, which are carried by the mediator molecule tRNA and chained together into a specific amino acid sequence when the mRNA passes through the ribosome. The growing chain of amino acids forms a polypeptide and finally folds into an active protein.

tRNA each transfer RNA carries a specific amino acids to the ribosome during protein translation, where it recognizes a specific codon in the mRNA.

UTR In molecular genetics, the untranslated region refers to either the 5'UTR or the 3'UTR, which are sections flanking the 5' or 3' side, respectively, of a coding sequence (CDS) on a mRNA strand.

Vector A plasmid or phage that is used to deliver selected foreign DNA for cloning and in gene transfer.

Wafer Thin disk usually made of pure silicon that is used as substrate for microfabrication.

References:

A. Zaid, H.G. Hughes, E. Porceddu and F. Nicholas. Glossary of Biotechnology for Food and Agriculture - A Revised and Augmented Edition of the Glossary of Biotechnology and Genetic Engineering. Food and Agriculture Organization of the United Nations, Rome, 2001 (<http://www.fao.org/DOCREP/004/Y2775E/y2775e00.htm#Contents>)

Douglas B Weibel, Willow R Diluzio, and George M Whitesides. Microfabrication meets microbiology. *Nat Rev Microbiol*, 5(3):209-18, 2007.

Sylvie Rockel

Avenue Fantaisie 4, CH - 1006 Lausanne | + 41 76 274 95 94 | sylvie.rockel@gmx.de
Date of birth: 06.06.1981 | Marital status: Single | Nationality: German

Education

2008 - 2012	PhD in Bioengineering, group Prof. Maerkl, EPF Lausanne, Switzerland
2006 - 2008	M.Sc. Bioprocess Engineering, TU Dresden, Germany
2005	Master Program „Biotechnology & Business“, RMIT, Melbourne, Australia
2001 - 2004	B.Sc. Process Engineering, TU Dresden, Germany

Professional Experience

Since 2008	Institute of Bioengineering, EPFL, Switzerland, doctoral assistant
2007 - 2008	Faculty of Bioprocess Engineering, TU Dresden, Germany, research assistant
2006 - 2007	Institute of Bioengineering and Nanotechnology (IBN), Singapore, research scholar (scholarship: IBN Youth Research Program)
2006	Biotechnological Centre (Biotec), Dresden, research assistant
2005 - 2006	Murdoch Childrens Research Institute (MCRI), Melbourne, Australia, research scholar (scholarship: German academic exchange service (DAAD))
2004 - 2005	The Paper Technology Specialists, Heidenau, Germany, research assistant

Skills

Microfluidics:	Design & fabrication of multilayer microfluidic devices; microarrays (DNA, protein and cells)		
Biology:	Cell culture, stem cell differentiation, PCR, protein purification, chemical labelling		
Imaging:	Confocal, Fluorescence Correlation Spectroscopy (FCS), Atomic Force Microscopy (AFM)		
Processing:	Fermentation (batch, fed-batch, up to 80 L reactors) and downstream process procedures		
Computational:	Adobe Illustrator & Photoshop, CleWin, GenePix, OriginPro, Mathematica, MatLab (basics), Prism GraphPad, ImageJ, Berkeley Madonna (basics), softWoRx, MS Office, LaTeX		
Languages:	German native	English (C2) full proficiency	French (B2) limited working proficiency Spanish (A1) beginner

Publications & Conferences

Publications

- Rockel, S., Geertz, M., Hens, K., Deplancke, B. and Maerkl, S. J. *iSLIM*: a comprehensive approach to mapping and characterizing gene regulatory networks. *Nucleic Acid Research*. (2012)
- Rockel, S., Geertz, M., and Maerkl, S. J. MITOMI: A Microfluidic Platform for In Vitro Characterization of Transcription Factor-DNA Interaction. *Methods in Molecular Biology*, 2012.786:97-114. **(Appendix A)**
- Geertz, M., Rockel, S. and Maerkl, S. J. A high-throughput microfluidic method for generating and characterizing transcription factor mutant libraries. *Methods in Molecular Biology*, 2011.813:107-23.

Conference Contributions

- Rockel, S., Geertz, M., Hens, K., Deplancke, B. and Maerkl, S. J. *iSLIM*: a comprehensive microfluidic approach to characterize gene regulatory networks. *Microfluidics 2012*, EMBL, Heidelberg, Germany. Poster presentation.
- Rockel, S., Geertz, M., Hens, K., Deplancke, B. and Maerkl, S. J. Generating a large-scale Drosophila transcription factor microarray. *1st International SystemsX.ch Conference on Systems Biology*, 2011, Basel, Switzerland. Invited short talk and poster.
- Rockel, S. and Maerkl, S. J. Large-scale protein interaction mapping using MITOMI. 435. *WE-Heraeus-Conference, Physics of Biological Function*, 2009, Bad Honnef, Germany. Poster presentation.
- Rockel, S. and Maerkl, S. J. Large-scale protein interaction mapping using MITOMI. *NCCR Frontiers in Genetics Annual Meeting*, 2008, Saas-Fee, Switzerland. Poster presentation.
- Rockel, S. and Kraut, R. S. *Sächsischer Biotechnologietag*, 2007, Congress Center Dresden, Germany, Poster presentation.
- Chauhan, A., Rockel, S., Farlie, F. and Meara, J.G. Tissue engineering of bone using a chick chorio-allantoic membrane graft model. *The Royal Australasian College of Surgeons Annual Scientific Congress*, 2006, Sydney, Australia, Abstract.

Associations & Societies

Since 2009 EPFL Toastmasters Club (public speaking skills)

Since 2009 ACIDE (Association of doctoral students at EPFL), organization of group social events

Since 2007 Netzwerk Bioverfahrenstechnik (Bioprocess Engineering Network in Germany)

2005-2006 RMIT Outdoor Club & RMIT Underwater Club