Development of microfluidic tools and methods for the biochemical characterization of scFv and Cas9 affinity reagents and their applications in molecular detection

THÈSE N^{O} 8614 (2018)

PRÉSENTÉE LE 11 OCTOBRE 2018

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

Ekaterina Emilova PETROVA

acceptée sur proposition du jury:

Prof. A. Radenovic, présidente du jury
Prof. S. Maerkl, directeur de thèse
Prof. S. Reddy, rapporteur
Prof. Y. Shaerli, rapporteuse
Prof. B. Correia, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

I would like to dedicate this thesis to my loving parents …

Abstract

Les progrès récents dans l'ingénierie des protéines, accompagnés par des développements rapides dans les domaines de la biologie synthétique et de la synthétisation de l'ADN, a enclenché une génération de protéines sur mesure. A travers des analyses approfondies sur ces protéines, des applications ont été faites dans les domaines de la médecine, la biotechnologie et la recherche fondamentale. Toutefois, malgré l'augmentation de la cadence de la synthétisation de l'ADN, l'implémentation des protéines synthétique n'est pas encore bien établie. L'une des raisons principales est le manque de technologies à cout réduit et des caractérisation évolutives. Les plateformes microfluidique à haut-débit ont le potentiel de surmonter ces obstacles. Cette thèse décrit l'utilisation des méthodes de pointes dans les domaines de la biologie synthétique acellulaire et de la microfluidique à haut-débit, permettant l'amélioration de la vitesse et des coûts pour analyser non seulement les interactions protéines-protéines mais aussi celles à l'interface ADN-ARN-protéines. En implémentant ces connaissances acquises ceci nous a permis de développer des bio-senseurs acellulaire à base de billes (scFV) ainsi que l'amélioration des méthodes actuelles de détection sans-marqueurs unimoléculaires.

Premièrement, nous avons utilisé une plateforme microfluidique intégrée couplée à un système de transcription/traduction acellulaire in vitro (TXTL) pour caractériser une librairie de petit fragment d'une chaîne unique variable (scFv). La scFv construite retient leur spécificité antigénique quand exprimée dans les TXTL disponibles commercialement. La plateforme microfluidique à haut débit permet de simultanément dépister et sélectionner les scFvs avec des spécificités et affinité définis, évitant des procédures de purifications de protéines et des manipulations de cellules fastidieux et longs. Grâce à leur versatilité, scFvs ont des applications dans différents domaines biologiques tels que la biologie cellulaire et structurale ainsi que dans des applications thérapeutiques et diagnostiques. Basé sur la plateforme de dépistage scFV qui a été développée, nous avons équipé les scFV liants à haute affinités avec des fonctionnalités additionnelles encodées génétiquement. Cette approche a permis la fabrication de bio-senseur sur des billes acellulaire avec une construction de gène rapporteur intégrée pour l'amélioration du signal.

La plateforme microfluidique a été utilisée non seulement pour analyser l'interaction protéine-protéine mais aussi pour étudier les machineries complexes protéiques. Nous avons réussi à reconstruire un système CRISPR-Cas9 in vitro et avons déterminé les valeurs d'affinités absolues à une librairie d'ADN cibles. Cette meilleure compréhension des dépendances pour une fonctionnalité correcte du système synthétisé de novo a trouvé une application dans une méthode de détection unimoléculaire sans-marqueur permettant l'estimation de la charge effective des complexes protéiques testés.

En plus de l'approche in vitro utilisé pour la caractérisation des interactions biomoléculaires, nous avons développé un bio-écran microfluidique qui détecte des petites molécules. Les biologistes synthétiques ont conçu, synthétisé et entièrement caractérisé divers types de parties codées par des gènes et aussi des dispositifs qui peuvent être utilisés pour améliorer les fonctionnalités des cellules. Toutefois, leur caractérisation in vivo est souvent entravée par le manque de technologies de biosécurité de bon marché. Le dispositif décrit ici permet la détection de concentration faible d'arsénite dans l'eau du robinet en utilisant des cellules E. coli modifiées génétiquement. De plus, ceci permet une surveillance environnementale en mesurant les dynamiques temporelles des expressions du bio-rapporteur.

En résumé, tous ces exemples soulignent comment la biologie synthétique acellulaire et la technologie microfluidique à haut débit peuvent faciliter l'application de diverses approches d'ingénierie des protéines.

Mots clés: scFv, CRISPR-Cas9, bio-rapporteur, bio-écran, microfluidiques à haut-débit, expression acellulaire, ingénierie des protéines, biologie synthétique.

Abstract

Recent progress in protein engineering, empowered by the rapidly evolving fields of synthetic biology and DNA synthesis, has enabled the generation of various custom-designed proteins. Their thorough analysis has led to applications in areas including medicine, biotechnology and basic research. However, despite increasing throughput in DNA synthesis, characterization of engineered proteins remains difficult. One of the main reasons is the lack of cost effective and scalable characterization technologies. High-throughput microfluidic platforms have the potential to overcome such hurdles. This thesis describes the development of various state-of-the-art methods in the fields of cell-free synthetic biology and high-throughput microfluidics, vastly improving the speed and cost for probing not only protein-protein, but also DNA-RNA- protein interactions. These methods and biochemical characterizations in turn allowed us to develop a cell-free bead-based scFv biosensor and apply Cas9 in a label-free single molecule detection method.

First, we used an integrated microfluidic platform coupled with an *in vitro* cell-free transcription-translation (TXTL) system to characterize a small single chain variable fragment (scFv) library. The constructed scFvs retained their antigen specificity when expressed in commercially available TXTL. The high-throughput of the microfluidic platform allowed simultaneous screening and selection of scFvs with defined specificity and affinity, obviating tedious and time-consuming cell-handling and protein purification procedures. Owing to their versatility, scFvs have found applications in different biological domains such as structural and cell biology as well as diagnostic and therapeutic applications. Here, based on the scFv screening platform that was developed, we combined high affinity scFv binders with additional genetically-encoded functionalities. This approach allowed the construction of a cell-free bead-based biosensor with an integrated reporter gene construct for signal enhancement.

The microfluidic platform was used not only for protein-protein interaction profiling but also for studying complex protein machineries. We were able to reconstruct an *in vitro* CRISPR-Cas9 system and determine absolute affinity values to a library of DNA targets. Better understanding of the dependencies for proper functionality of the *de novo* synthesized system found application in a label-free single molecular detection method enabling effective

charge estimation of the tested protein complexes. In addition to the *in vitro* approaches used for biomolecular interaction characterization, we also developed a microfluidic biodisplay which senses small molecules.

Synthetic biologists have designed, synthesized, and fully characterized various types of gene encoded parts and devices that can be utilized to create new functions. However, their *in vivo* characterization is often hampered by the lack of cheap biosafety technologies. The device described here allowed detection of low arsenite concentration in tap water using genetically modified *E. coli* cells. Moreover, it enables environmental monitoring. To sum up, all these examples underscore how cell-free synthetic biology together with high-throughput microfluidic technology can facilitate the application of various protein engineering approaches.

Key words: scFv, CRISPR-Cas9, bioreporter, biodisplay, high-throughput microfluidics, cell-free expression, protein engineering, synthetic biology.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Protein engineering

Proteins are one of the major components of cells that allows them to sustain live. Their highly coordinated actions and specificities enable countless elaborate tasks such as DNA replication, RNA synthesis, transcription and translational regulation, immunogenicity, to be performed with explicit accuracy and speed. Despite their complexity, today researchers are capable of engineering their functionalities and utilizing them in various scientific and biotechnology applications.

In the last few decades enormous increase in engineered proteins has been observed. Antibodies with improved biophysical parameters have been constantly generated [88], additional functionalities have been incorporated in their structure improving their stability [112] and therapeutic properties [37]. Enzymes have been customized to retain functionality in ambient conditions or to obtain altered substrate specificity [236–238]. Moreover, the activities of entire protein complexes have been repurposed. DNA polymerases are now capable of synthesizing artificial genetic polymers called XNAs [173]. Another prominent example is CRISPR-Cas9. The system, which native role is to serve as an alternative immune system in prokaryotes, has been adapted in different genome engineering and silencing approaches [50].

While a large amount of native proteins are highly relevant, their application in basic research and biotechnology is often restricted by the high structural complexity, low stability and expression, high toxicity, undesirable biophysical properties such as low affinity and specificity, etc. Adapting proteins for scientific and industrial use has imposed challenges and opportunities for protein engineers.

Protein engineering is a complex process encompassing an iterative cycle of (1) protein design and gene synthesis, (2) protein expression, and (3) protein characterization, which

terminates when desirable functional properties are obtained. A large variety of techniques have been employed over the years to accomplish this elaborate engineering feat. Back in the mid-1980s the first successful attempts of protein engineering were reported. Advances in oligonucleotide synthesis and recombinant DNA technology as well as the advent of PCR enabled scientists to step from answering fundamental questions regarding the structural-functional relationship in proteins [65, 240] toward application-driven approaches. Largely guided by the available crystallographic information, several enzymes were adapted for use in desirable environmental conditions [55, 171]. Next, directed evolution methods were developed in attempt to increase protein engineering efficiency. Here, the protein diversification is introduce by random mutagenesis using either error-prone DNA polymerases working in suboptimal conditions [22, 33, 79] or DNA shuffling based on genomic recombination strategies [212].

However, despite the efficiency of the directed evolution methods [165] the large library sizes that are generated are often exceeding the capacity of conventional selection and screening procedures. In addition, since function is sparsely distributed within the sequence space irrespective of their sizes the synthesized libraries are capable of covering just a small portion of all possible function affiliated sequences. Confronted with all these challenges, the establishment of high-throughput, time and cost effective protein engineering methods has been and is still highly necessary. Today advances in gene synthesis, protein expression and characterization coupled with improvements in computational algorithms, based on the constantly expanding sequence and structure protein datasets, have the potential to greatly enhance the process.

1.2 Protein design methods

Constantly improving next-generation sequencing technologies, bioinformatics and 'big data' computer-aided tools marked the beginning of the 'mega-genomic' era. To date large sequence datasets can be rapidly obtained and analyzed in high-throughput manner, with high fidelity and relatively low cost from any organism under any environmental conditions imaginable. In parallel to the sequence database expansion, the structural database is also constantly growing. These advances are fueling the development of various computational approaches for protein design. By introducing mutated residues onto a high resolution protein structure, molecular modeling programs (Rosetta, YASARA and FoldX) are capable of predicting the most energetically favorable protein variant(s) [117, 235]. Unfortunately the *in silico* models are still lacking in high accuracy and experimental mutagenesis and protein functional validation are still a necessity. As an alternative, a semi-rational design approach

is largely utilized. These rely on integration of prior knowledge of protein sequence, structure and functional data as well as on advanced bioinformatics and machine-learning algorithms.

1.3 DNA synthesis approaches

The expansion of DNA sequencing has been driving the progress in various biology disciplines and in particular gene synthesis. Although still lagging behind in terms of scale and cost, gene synthesis- an indispensable tool in the protein engineering field- is highly required for testing the numerous hypothesis raised by the large metagenomic information available today. It allows researchers to obtain DNA constructs of interest when extraction or modification of naturally occurring sequences is not feasible [15]. It has also accelerated the design-build-test cycle by eliminating time-consuming and labor-intense steps such as cloning and site-directed mutagenesis, especially in cases where large number of vastly modified sequences are required, as for example structure-function relationship and codon optimization studies [119, 180].

Over the last half century a number of *in vitro* gene synthesis strategies have been developed. Their continuous improvement have allowed the generation of DNA construct with varying length and complexity (high GC content, highly repetitive DNA sequences) [21]. The majority of these strategies rely on ligation-based approaches [2, 8, 11, 193] (pool of complementary oligonucleotides spanning the entire sequence of interest are joined together), polymerase cycling assembly (PCA) approaches [21, 213] (pool of oligonucleotides with complementary spanning regions are joined together), or a combination of the two [247]. In addition, recently developed methods such as Gibson assembly and Golden Gate assembly have allowed one-step isothermal assembly of DNA fragments up to several hundred kbp [72]. These approaches have found application not only in gene library construction but also for building entire gene clusters [162].

Currently a large effort is put on the integration of the standard assembly approaches with microarray based oligonucleotide synthesis. This is envisioned to increase the throughput, speed, as well as decrease cost of the gene synthesis process. By using such approaches Quan et al. were able to synthesize a library of a thousand codon-usage LacZ alpha variants using a custom microarray oligo synthesizer [180]. More recently, by applying computational guided design strategies researches were able to synthesize large protein library composed of 6000 ubiquitin variants [217]. To date advances in the microarray synthesis have led to production of longer oligonucleotides (about 200 nt) that have been directly used for small protein or protein domain production and characterization [190]. The development of the DropSynth platform relying on water-in-oil emulsion droplets and magnetic beads oligo

extraction for high-throughput and cost effective gene synthesis has also relied on the use of long, microarray generated DNA strands [176].

Even though DNA synthesis is extremely powerful several challenges still need to be overcome. Although constantly dropping, the cost associated with the process remains high; it was estimated to be about $ 0.07 per nucleotide in 2016 [24]. In addition, the process is also highly error-prone. Upon DNA assembly a mixture of correct and incorrect sequences are generated with mutations arising both from oligonucleotide synthesis and the gene assembly procedures. An established error eradication technique such as site-directed mutagenesis is often used, however, it adds additional cost and decreases the speed of the process. Careful design of the starting building blocks and the use of high-fidelity enzymes, sometimes in combination with mismatch binding and cleavage enzymes [36, 250], have significantly reduced the errors in the range of 1-3 errors per kb.

1.4 *In vitro* protein expression

Conventional strategies for characterizing protein properties and functions rely on long and tedious cell handling and protein purification procedures. Recently, cell-free transcription-translation systems (TXTL) have become a beneficial alternative. Originally developed to study fundamentals behind transcription and translation [156], to date optimized TXTLs enable large-scale protein synthesis in time and cost efficient manner [109, 97]. In addition they serve as a alternative ‘chassis’ for a variety of biotechnology, synthetic biology and diagnostics applications such as: (1) protein synthesis including toxic and non-natural proteins, (2) prototyping regulatory elements and circuits [137, 167], and even (3) the generation of artificial cells [158, 94].

Among all the major benefits of TXTL are:

- decreased experimental timespan due to the elimination of constrains associated with conventional gene-manipulation techniques and cell-handling procedures.
- amenability for miniaturization and high-throughput analysis especially when combined with microfluidic platforms.

Despite all TXTL promising features, challenges still remain. The short TXTL lifetime, due to substrate depletion (ATP, amino acids, etc.) and by-product build up (e.g. inorganic phosphate), often result in low protein yields when compared to cell-based protein synthesis approaches. They require the use of expensive reagents, in particular phosphate chemicals in the form of nucleotides and secondary energy sources. In addition, the use of home-made cell lysates often leads to high batch-to-batch variation. To overcome these limitations, their

composition and techniques for preparation have been continuously refined. Currently a large variety of standardized protocols for cell lysate preparation have been established increasing the TXTL accessibility to larger community [47].

TXTL composition and types

The basic components of TXTL are: (1) a crude cell lysate derived from various types of organisms such as *E. coli*, *B. subtilis*, insect cells (ICE), rabbit reticulocytes (RRL), wheat germ cells (WGE), Chinese hamster ovary cells (CHO), and (2) a supplementary solution composed of all essential amino acids as well as energy-regenerative compounds (nucleotides, energy substrates, cofactors, and salts). Together these two solutions contain all cellular machineries, energy and resources required for the transcription and translation of exogenously added DNA templates. In addition to the conventional TXTLs a commercially available PURExpress system has been developed [196]. It consists of almost one hundred purified components necessary for DNA transcription and translation.

Template construction for TXTL systems

The design and the type of the DNA templates used for protein synthesis have a large impact on protein expression levels. Given the vulnerability of linear templates in crude cell extracts to exonucleases, plasmid DNA is often the preferable choice. However, supplementary factors including gamS protein, DNA binding proteins, chi sequence [136] that inhibit the exonuclease activity have allowed researches to successfully utilize them. An increasing number of commercial and home-made TXTL systems have been developed based on linear template protein expression bypassing tedious cloning procedures and enabling cell-free high-throughput assays.

Proper transcription and translation regulatory elements need to be included into the DNA constructs. In addition to promoter, RBS and terminator, non-obligatory ones such as internal ribosome entry site (IRES), poly-A tail for RNA polymerase destabilization, degradation tags, affinity tags, etc. can be introduced depending on the nature of TXTL systems and the goals of the study.

TXTL systems for antibody production

A remarkable progress in the production of scFv, Fab and even IgG molecules in TXTL has been achieved in the last two decades [5, 192, 248]. The structural stability and functionality of the antibodies are highly dependent on intra- and inter-disulfide bond formation [74]. However, not all organism exploited for TXTL production are ‘fully equipped’ with factors

promoting disulfide bond formation, thus a general strategy for adapting *de novo* antibody synthesis is by supplementing the reaction with redox agents such as glutathione, agents promoting disulfide isomerization and protein folding including protein disulfide isomerase (PDI) and chaperones- DnaK, DnaJ, GroEL and GroES- which increase the solubility of the produced antibodies [110, 192, 249]. In addition, in a more complex approach endoplasmic reticulum (ER) structures can be added, separating the environment into two distinct compartments for protein synthesis and protein folding, mimicking the cell-based protein modification process occurring in periplasmic space in bacteria and ER in eukaryotes. In term of lysate origin various cell types were utilized till now *E. coli* and wheat germ are preferable source in cases when high protein yields are required [109, 110, 129]. Currently significant improvements were made in lysates based on CHO cells to improve their protein production capacity [27, 28, 206] since most successful approaches for production of functional protein have been shown to utilized mammalian cells TXTL systems.

1.5 Protein characterization methods

High-throughput protein characterization has always been of great interest in the field of protein engineering where large protein libraries need to be screened. Over the years a variety of technologies have been developed. Prominent examples are *in vitro* display technologies, protein microarrays as well as recently implemented microfluidic platforms.

In vitro display technologies are an indispensable tool for characterizing protein libraries with diverse sizes and complexities. They enable selection of variants with unique binding properties toward various types of targets including DNA, RNA, proteins and small molecules. During the process the library members are genetically fused to a protein presented on a phage or a cell surface [23, 61, 71, 195, 239]. While the expression of the library members occurs *in vivo*, the iterative selection process is carried out *in vitro*. Depending on the display technology the proteins are either retained by solid-phase phage immobilized or FACs [195]. The enriched population is further screened by methods such as surface plasmon resonance (SPR) or antibody-binding assays, ranging from ELISAs to immunoprecipitation, in order to recover proteins with enhanced binding properties. Despite their robustness, display technologies are often restricted by cost and time concerns.

The first method for protein characterization dependent on cell-free protein synthesis - by in vitro compartmentalization - was established by Griffiths and Tawfik [220] in 1998. During the process water-in-oil emulsions were used to spatially constrain the cell-free expression of a protein coding genes into single compartments. The preserved genotype-phenotype linkage allowed precise evaluation of the protein sequence-functional relationship

[144]. Since then various techniques coupling *in vitro* compartmentalization and microfluidic platforms [100] have been developed making the process even more cost effective and less time-consuming. In parallel, the display technologies have changed dramatically as well. To date cell-free based ribosome display [78] and mRNA display [89] are capable to access the highest potential diversity (10^12- 10^13) while abolishing labour intensive cloning and transformation steps.

Protein microarrays are yet another alternative for high-throughput protein characterization. The fundamental principle was established in 1989 by Ekins [52]. The first generation protein microarrays in which purified proteins are printed on a solid-phase surface is still used for antibody profiling [143], biomarker identification [211], and protein-protein interrogations [252]. However, the requirement for purified proteins immensely increased the cost and labor, has limited its use. As an alternative, second generation protein microarrays have been developed, referred to as self-assembled protein microarrays [183, 184]. Here, protein synthesis is driven in cell-free expression systems flowed across printed protein coding DNA templates, which are increasing the shelf-life of the array. In addition to protein microarrays, DNA microarrays have also found application in rapid high-throughput transcription factor - DNA interaction studies. In this scenario, DNA templates are immobilized on a solid-phase support and afterwards probed with purified or *in vitro* synthesized protein samples.

Owing to miniaturization and parallelization, hybrid microrray and lab-on-chip (LoC) technologies have high potential to enable improved protein profiling in terms of cost, time, throughput, ease-of-operation and analysis.Among the plethora of existing devices, MITOMI (mechanically induced trapping of molecular interactions) [132] and its derivatives are of particular interest. . They are amenable to capture interactions over a large affinity range, allowing the construction of detailed protein binding energy landscape. In general, MITOMI is a two layer polydimethylsiloxane (PDMS) device. It is composed of hundreds of individual unit cells containing a detection area and a back-chamber. Small amount of molecules of interest can be deposited by microprinting onto a glass slide used to program the MITOMI chip. The device operation is facilitated by micromechanical valves that control the solution flow as well as chamber and unit cell segregation. The design also includes a membrane valve, called ‘button’ valve, which enable surface derivatization and trapping of molecular interaction complexes. The MITOMI ‘family’ of devices have been used in various diagnostic applications [63] as well as for detailed biochemical screens of protein-DNA [58, 132], protein-protein [70], and protein-RNA interactions [138]. The use of such devices has enabled biochemical screening as well as regulatory networks analysis and modeling. In addition, it aided the design of novel proteins [21].

The rapid in-depth quantitative binding data exploration enabled by MITOMI is often not feasible when the previously described high-throughput methods are applied. Low affinity and transient protein binders are typically lost when stringent washing conditions are used, even though such binders are known to be highly relevant in biological systems and pharmacology. Adam et. al have shown that in contrast to common beliefs high affinity anti-HER2/neu scFvs possess less tumor targeting specificity and penetration compared to their lower affinity counterparts [1]. Similar results were observed for anti-carcinoembryonic antigen (CEA) [75].

Besides their use in high-throughput protein biochemical interactions mapping microarrays and coupled microfluidics/microarray platforms have been extensively used in diagnostics [82, 63]. In addition, a plethora of other detection tools have been introduced such as bead-based AlphaLISA by Perkin Elmer [16] and SPR-based real-time and label-free immunoassays using the Biacore system [221]. Both of the platforms have increased the detection sensitivity compared to conventional ELISAs, a gold standard for immunodiagnostics. Also they have enabled ease of automation, high throughput and continuous monitoring. Similar to the MITOMI platforms, the multiplexed bead-based Luminex® assays by Life Technologies Corporation immunoassays tremendous potential for immunodiagnostics. The use of magnetic beads in immunoassays have led to truly remarkable applications that takes into account the convenience of handling the magnetic beads/particles [73].

1.6 High-throughput cell profiling

The advent of cell-free synthetic biology as well as the growing plethora of repositories of synthetic biological parts such as the IGEM Registry of Standard Biological Parts (partsregistry.org) have facilitated the rapid design and synthesis of genetic components and networks [53, 54] that can be utilized by living cells in order to explore, improve or repurpose their functionalities [108, 113, 227]. Application-driven approaches have succeeded to implement various types of synthetic parts or even whole devices enabling cells to act as biosensor for toxic metal detection, for example [177].

In addition, plenty of synthetic parts fully characterized *in vitro* are constantly being generated. However, whether their behavior can be recapitulated once placed *in vivo* is an open ended question. Up to date few data is available [153]; one of the reasons behind is the limited amount of tools for high-content cell screening.

Standard approaches for cell culturing and screening are often time-consuming and low-throughput, hampering advances in the field. On the other hand, microtiter-based studies often complemented with expensive robotic equipment have allowed large-scale cell screening but

have also placed cost constrains on their wide-spread use. Microfluidic cell culturing and screening devices provide an interesting alternative. In addition to the economic benefit due to the low reagents consumption, they allow accurate control over the culturing conditions increasing reproducibility. The device operation can be parallelized allowing complex assay procedures to be streamlined and thus maximizing the information gain [134, 241]. The automation by valves/actuators provide an easy way for temporal dynamic analysis which is extremely hard to achieve by utilizing conventional screening methods. And lastly given the closed environment there is reduced contamination risk.

Chapter 2

Coupled microfluidics for *in vitro* synthesis and characterization of a scFv library

2.1 Introduction

The ability to bind vastly diverse antigens with high affinity and specificity have made antibodies an indispensable tool in basic research and in various biotechnology and biomedical applications [104, 121]. As a result, the development of efficient methods for production of antibodies with tailored properties such as increased binding specificity and affinity has always been of great interest.

Hybridoma technology and surface display technologies are among the most widely used approaches for monoclonal antibodies (mAbs) discovery [95, 114, 223, 239]. In general, the processes consist of several distinct steps: (1) antibody repertoire generation, (2) iterative antibody selection, (3) antibody screening, and (4) antibody recovery. However, although robust these technologies are associate with high cost as well as slow and labor intensive procedures.

More recently alternative approaches for antibody discovery have been developed including B cell next-generation sequencing (Ig-seq). Even though bulk Ig-seq is extremely advantageous, given the continuous drop in sequencing cost and broad accessibility, information about endogenous pairing of the variable heavy-chains (VH) and the variable light-chains (VL) is lost. To retain the VH-VL pairing, a variety of strategies have been developed. DeKosky et al. established a pipeline where single B cells deposited in microwells are lysed *in situ*, their mRNA isolated and converted to cDNA that is later used for synthesis

and subsequent sequencing of genetically fused VH-VL constructs [45]. In another study, relative frequency analysis was applied to estimate the VH and VL pairing after performing bulk next-generation antibody sequencing [186]. Although both of these antibody mining strategies allow rapid antibody sequence information gain, functional validation of the *in silico* recovered antibodies is required to determine the accuracy of the bioinformatic antibody identification approaches as well as to pinpoint the most important antibodies generated during an immune response.

ELISA, SPR and ELISPOT are among the most commonly used methods for antibody biochemical characterization. However, due to the often use of stringent washing steps they typically obviate fine-mapping of the antibody binding landscape and as well they are more biased towards recovering high-affinity binders, which may not exhibit the required physiological function [1, 75].

To address these challenges, we developed a pipeline for construction and rapid microfluidics-based synthesis and functional interrogation of recombinant antibodies. Access to antibody sequencing information from B cell Ig-seq studies is valuable resource that can be used for construction of antibody libraries which diversity and size are restricted only by the cost ofthe V-region DNA oligonucletides. Upon library synthesis microfluidic platforms for coupled *in vitro* protein expression and quantitative characterization such as MITOMI [21] can be used for rapid and cost-effective construction of a detailed antibody binding landscape map (Figure 2.1).

To validate our approach, we rapidly synthesized a small rationally designed scFv library against human respiratory syncytial virus (RSV) by utilizing two different gene assembly approaches. All of the variants were successfully *in vitro* expressed and screened on MITOMI. We were able to recover two variants that possess binding affinity similar to motavizumab, an extremely potent anti-RSV neutralizing antibody [140, 245].

2.2 Results

2.2.1 Rational design and gene synthesis of a combinatorial anti-RSV scFv library

The design of our scFv library was aided by previously published data. Wu et. al used an iterative mutagenesis and screening approach to identify affinity matured variants of palivizumab, an FDA approved anti-RSV neutralizing antibody [245]. Sequence analysis of the recovered recombinant antibodies showed that beneficial mutations are located primarily in antibodies complementary determining regions (CDRs).

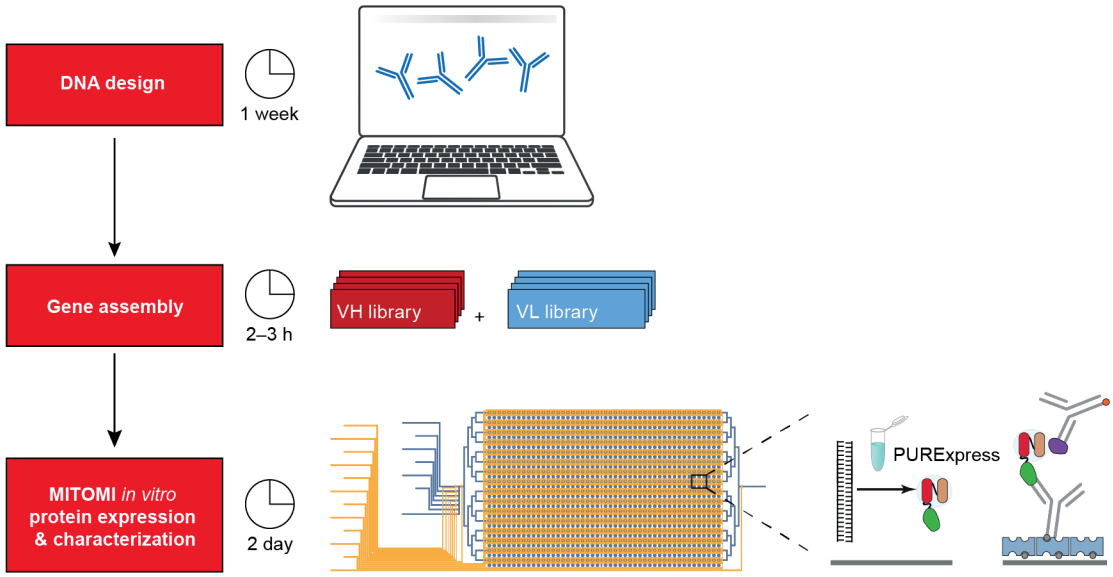

Fig. 2.1 **Overview of microfluidics approach for identifying functionally relevant recombinant antibodies generated based on Ig-seq data.** First, paired V-region sequences identified using bioinformatic and Ig-seq methods can be synthesized and assembled into scFvs. Next, all antibody fragments can be *in vitro* expressed and screened on-chip for precise function.

The DNA sequences of ten palivizumab variants were utilized for the design of a small anti-RSV scFv library. A specialized analysis toolkit (http://www.abysis.org/) was used to identify the VH and VL chains of all retrieved antibodies and to annotate (according to the Kabat numbering scheme) their CDRs and framework regions (FRs). Interestingly, the FRs showed high sequence similarity to VH DP47 and VL DP3 immunoglobulin genes, known to support high scFv expression in bacteria [76]. Four different VH and four different VL sequences were identified. Upon DNA synthesis, a PCR based approach was used to diversify VH FR4 region. Thus, in total we used eight VH and four VL sequences (Figure 2.2a) for the scFv library construction. The fragments were joined in combinatorial fashion resulting in thirty two scFv variants.

The VH and VL sequences were tethered by a flexible peptide linker $(Gly_4Ser)_3$ [229]. To facilitate quantitative scFvs expression and binding measurements as well as to allow scFvs pull-down on solid-phase surfaces a yeGFP tag was fused to the C-terminus of the variants via a rigid Pro-Ala linker. To allow ease of transcription and translation three regulatory elements were incorporated into the scFv design. At the 5' end we placed a T7 promoter followed by a Shine-Dalgarno RBS site and at the 3' end we introduced a T7 terminator. Polymerase cyclic assembly (PCA) and Gibson assembly strategies were utilized for the scFv gene synthesis. For the PCA, DNA constructs with overlapping sequences coding for the VH chain, the VL chain, and the yeGFP tag, were mixed together in a single PCR reaction supplemented with primers amplifying the assembled scFvs. To reduce PCR-based errors high-fidelity DNA polymerases were used: Phusion ($4.4\ e^{-7}$ error rates) and OneTaq (2x improved fidelity compared to Taq). Even though successful, the PCA gave rise to an unspecific product (Figure 2.2b).

In order to improve the quality of the scFv DNA constructs we adapted the Gibson assembly approach. All building blocks together with a linearized pUC19 template were added to a Gibson assembly reaction mix. Sixteen scFv variants were successfully clone (87% efficiency). To take a glimpse at the fidelity of the process seven colonies were Sanger sequenced. No mutations in the coding sequences or the transcription regulatory elements were revealed. In almost all cases adenine deletion or insertion in the long polyA tail was detected which might result from low DNA polymerase processivity or low quality sequence reads. In general, in contrast to the PCA no unspecific product was observed for the Gibson assembled scFvs (Figure 2.2c).

2.2.2 *In vitro* scFv expression assessment

As a prelude to coupled scFv expression and functional characterization on MITOMI, we compared the bench-top *in vitro* scFv expression using the commercial PURExpress and

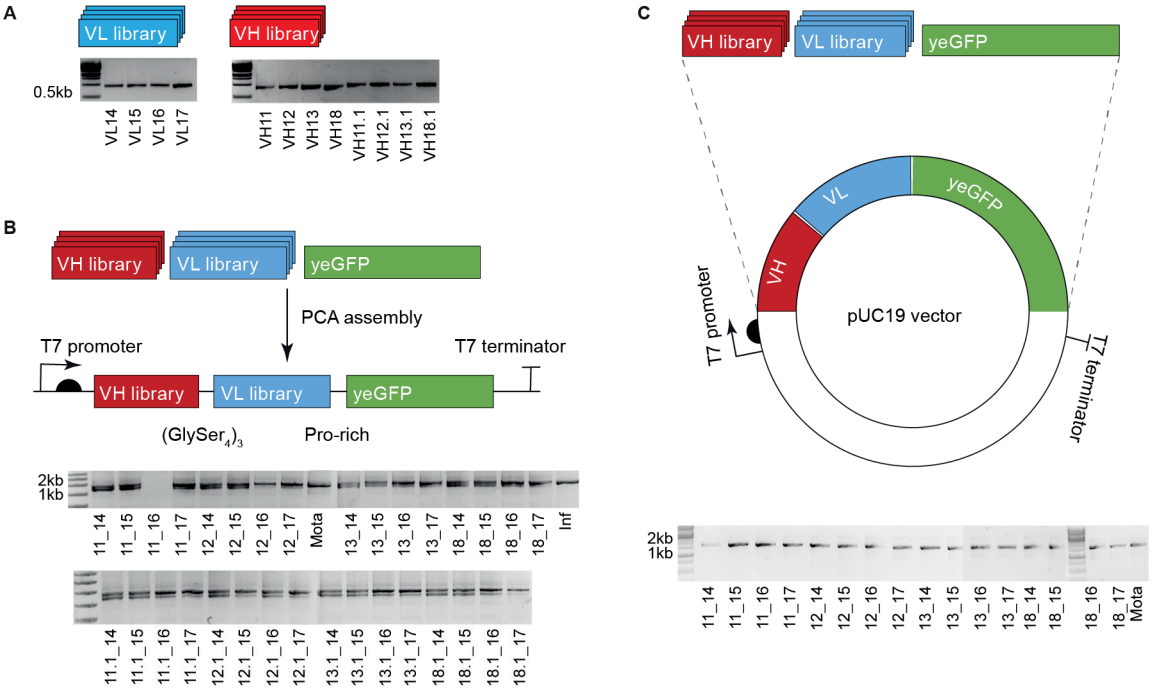

Fig. 2.2 **Gene assembly strategies for scFvs synthesis. (a)** Agarose gel electrophoresis of all VH and VL DNA constructs used in the study. Outline and resulting PCR products from **(b)** the PCA and **(c)** the Gibson assembly strategies.

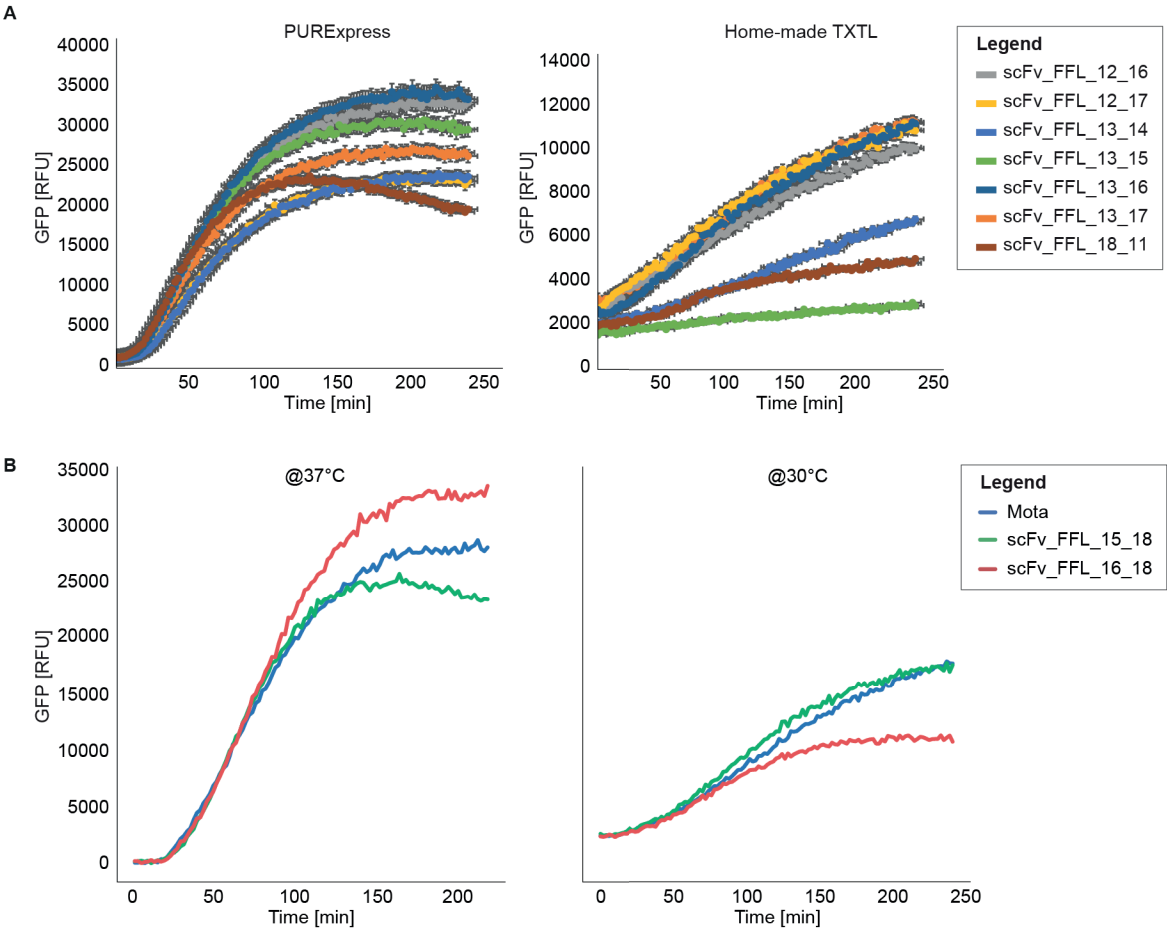

Fig. 2.3 ***In vitro* scFv expression.** Expression profiles of scFvs synthesized using **(a)** the commercial PURExpress and a home-made TXTL system. **(b)** Temperature-dependent scFv expression using the PURExpress.

a home-made TXTL [218]. Given the scFv stability is dependent on intra-chain disulfide bonds we supplemented the reactions with disulfide bond enhancers. In addition, due to DNA exonucleases in the home-made cell lysate, gamS protein was added to the TXTL. Although both the use of both systems resulted in scFv synthesis, higher protein yields were obtained when PURExpress was used and in addition when the reaction were incubated at 37°C (Figure 2.3).

2.2.3 scFv functional assessment on MITOMI

Based on a known motavizumab-RSV F crystal structure [140] and using a computation algorithm called Fold from Loops, Correia et. al designed and synthesized peptide, called

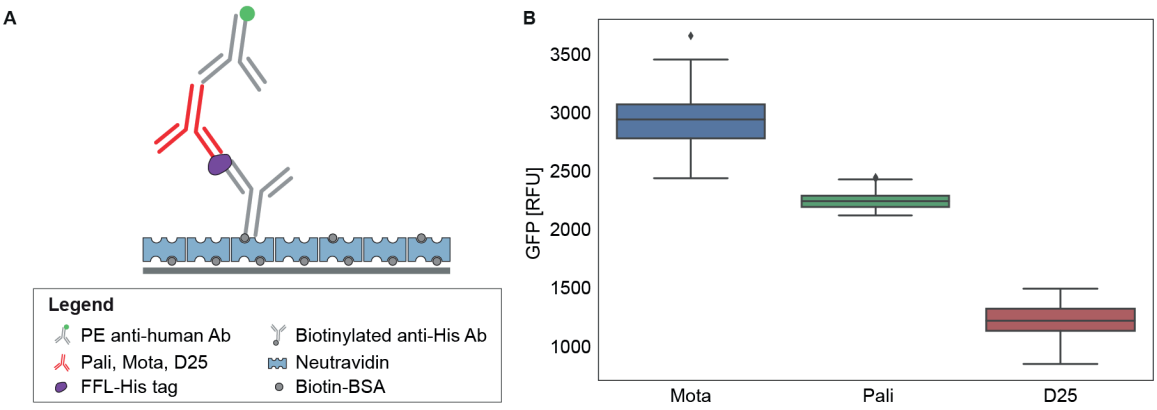

Fig. 2.4 **FFL and anti-RSV IgGs interaction on MITOMI. (a)** Schematic of the on-chip immunoassay for FFL detection. FFL with concentration 2 μM was captured by a biotinylated ani-His antibody. Three different anti-RSV IgGs were used as secondary antibodies. Phycoerythrin labelled anti-human IgG was used as a detection probe. **(b)** Boxplot comparing the relative binding affinities of all three anti-RSV IgGs to FFL.

FFL, that functionally mimics a specific RSV F epitope [43]. The FFL peptide was thoroughly studied. SPR analysis revealed its very high binding affinity to motavizumab (Kd=29 pM), similar to the values measured for RSV F and motavizumab (Kd=34.6 pM) [140]. To allow ease of purification and detection a modified His-tagged FFL version was generated. In all subsequent studies instead of relying on recombinant RSV F proteins we made use of the FFL-His tag peptide.

To explore FFL functionality on MITOMI we measured the interaction between FFL and three different anti-RSV IgGs: motavizumab (Mota), palivizumab (Pali) and D25 (Figure 2.4). The D25 antibody was used as a negative control since it binds a distinct RSV F epitope [222]. Only negligible FFL binding was observed for D25. In addition, both Mota and Pali were capable of FFL recognition. Measurements of relative binding affinities revealed that Mota is stronger FFL binder, as shown before [245].

Next, we tested the scFvs functionality. Two different immunoassay design strategies were explored. In the first approach, Mota and scFv-EGFR (recombinant anti-EGFR antibody) were bench-top expressed and flowed onto a MITOMI functionalized with biotinylated anti-GFP antibody. FFL and fluorescently labelled anti-His antibody were flowed sequentially. As shown in Figure 2.5a FFL was successfully captured by Mota in contrast to scFv-EGFR, confirming the anti-RSV scFv functionality. The second approach in which the scFvs were used as a FFL detection probe consolidated the previous result (Figure 2.5b). For all subsequent experiments we used the first design strategy given the higher signal-to-noise ratio.

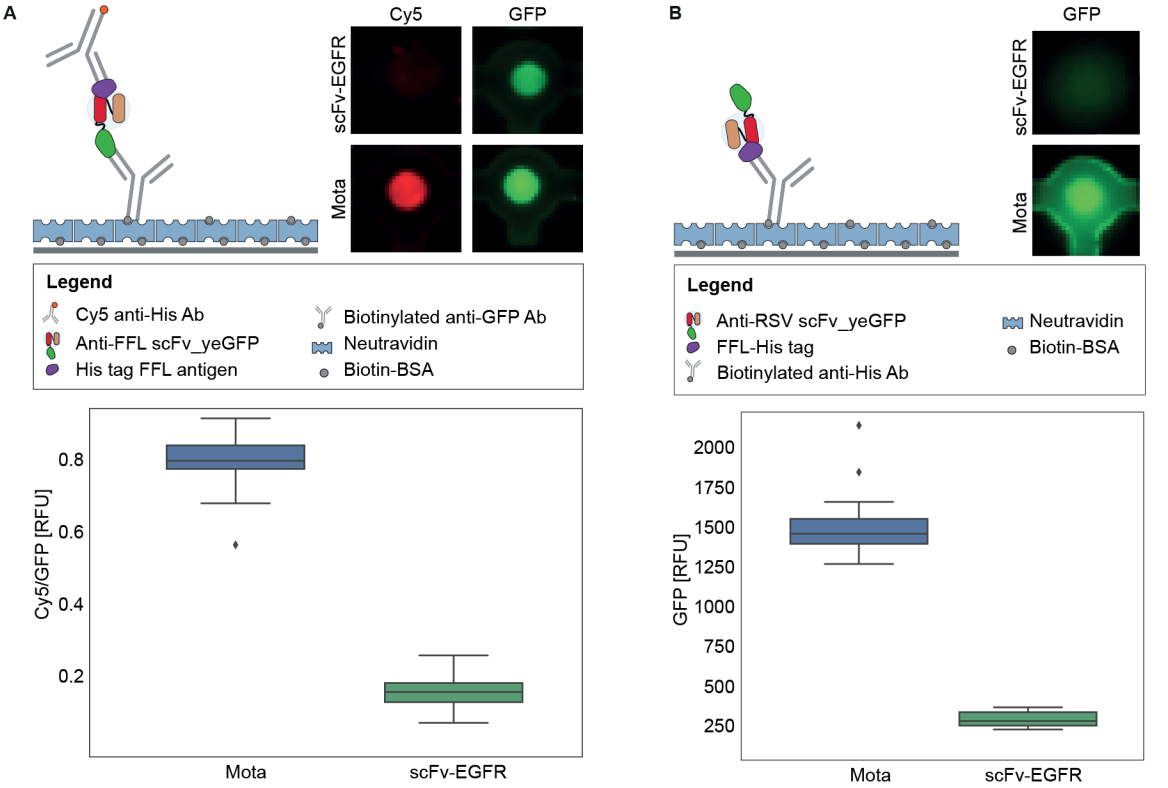

Fig. 2.5 **FFL and anti-RSV scFvs biochemical interrogation on MITOMI.** **(a,b)** Outline of the two immunoassay strategies for 2 μM FFL detection. Fluorescent images showing **(a)** FFL (Cy5) detection when scFvs (GFP) are used as a capture probe and **(b)** FFL detection when scFvs are used as a detection probe. Barplots showing either **(a)** the Cy5/GFP ratio or **(b)** the GFP signal obtained from the scFvs bound to surface immobilized FFL.

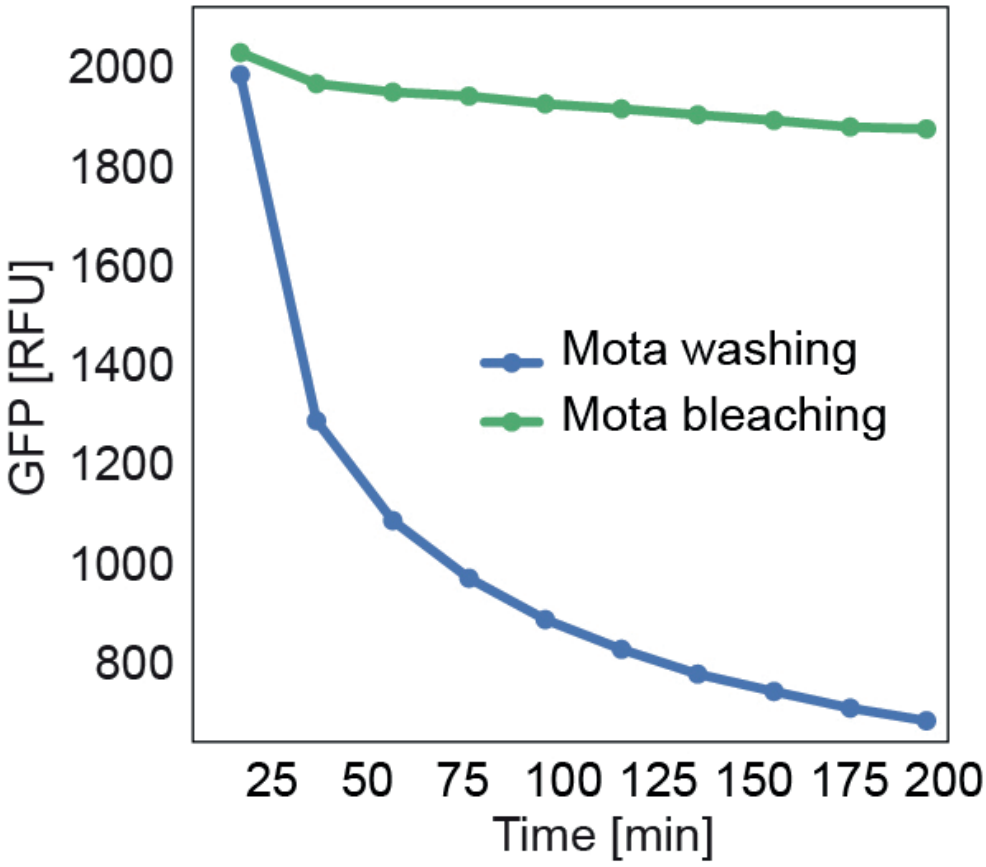

Fig. 2.6 Function mapping the fluorescent signal decrease over time due to Mota bleaching or detachment from surface immobilized FFL.

Although the second approach did not gave rise to highly sensitive measurements, it enabled us to retrieve information about Mota:FFL stability. Continuous washing of the FFL bound scFvs allowed us to determine the Mota:FFL half-life. The calculated koff = $4.06e^{-5}s^{-1}$ only slightly deviated from the one reported from a previous SPR experiment-koff = $3.1e^{-5}s^{-1}$ (Figure 2.6).

Next, we decided to explore if we can perform coupled on-chip scFv expression and functional characterization. Dilution series of DNA templates encoding Mota and scFv-EGFR were microarrayed on an epoxy glass slide used to program a MITOMI chip. Upon surface derivatization and biotinylated anti-GFP antibody immobilization, PURExpress was used for scFv DNA template resuspension and protein expression. After incubation of the

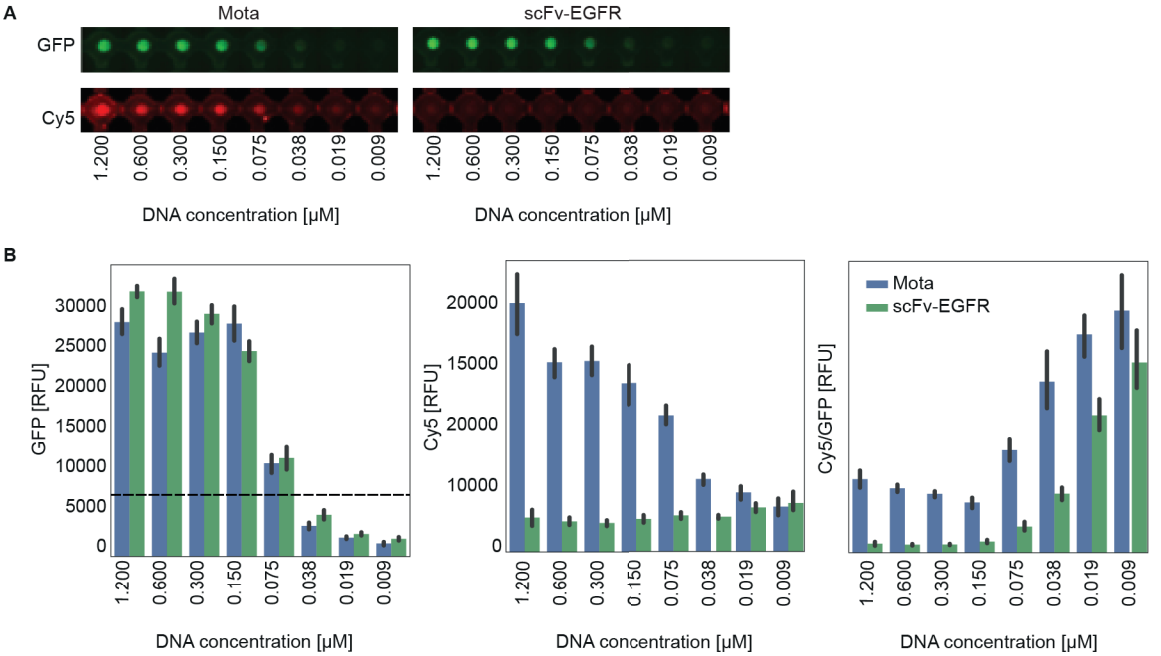

Fig. 2.7 **Integrated scFv synthesis and functional characterization on MITOMI. (a)** Representative fluorescent images showing surface immobilization scFvs (GFP) and bound FFL (Cy5). **(b)** Quantification of the signal obtained from surface immobilized on-chip expressed scFvs, detected FFL and ratio between the captured FFL and the expressed scFvs. 2x dilution series of the scFvs encoding DNA templates were spotted on an epoxy glass slide used for MITOMI programming. The dotted line is showing the set threshold level.

device at 37°C for 2.5-3 h, the expressed scFvs were allowed to diffuse into the detection chambers and bind to the anti-GFP antibody capture probe. Next, FFL was flowed across the immobilized scFvs followed by Cy5-labeled anti-His antibody used as a detection probe. Up to an 8-fold difference in the Cy5 signal was measured between the highly expressed Mota and scFv-EGFR samples. On the other hand, low scFv expression (below 5000 RFU with the preset microscopy setting) led to high Cy5/GFP signal ratio suggesting that fairly high scFv surface passivation is required for proper estimation of the scFv-FFL binding. To obviate such behavior for all proceeding experiments only scFv sample with expression level above certain threshold (5500 RFU) were analyzed (Figure 2.7).

Having successfully expressed Mota and interrogate its interaction with FFL on MITOMI, we sought to conduct functional studies of our anti-RSV scFv library as well as three different Mota scaffold variants.

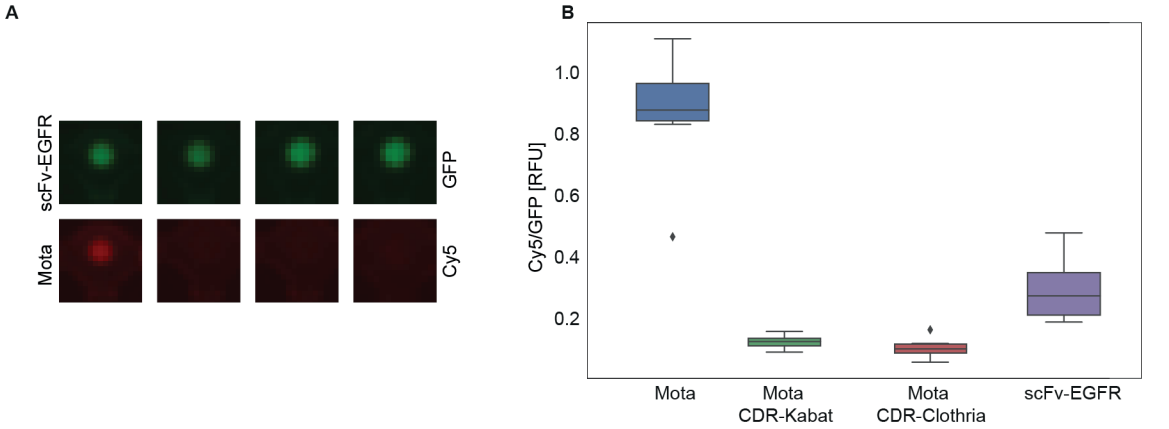

Fig. 2.8 **Screening of scaffold modified anti-RSV scFvs. (a)** Representative fluorescent images showing the surface immobilization of the scFv variants (GFP) and FFL capturing (Cy5) and **(b)** quantitative data. Mota CDRs according to Kabat and Clothria numbering scheme were grafted in anti-EGFR scFv framework.

2.2.4 Screening of anti-RSV variants grafted in anti-EGFR framework

In an attempt to increase the scFv expression we synthesis two additional Mota variants by grafting their CDRs into a scFv-EGFR framework supporting high protein yields. Utilizing the established experimental scheme we measured the binding between the Mota variants and the FFL peptide. Although all constructs reached high expression levels, neither of them were capable of FFL binding. We hypothesized that amino acid residues of Mota framework are either interacting directly with the antigen and thus stabilizing its binding and/or are aiding to specific Mota CDR conformation that is essential for proper scFv-antigen interaction (Figure 2.8).

2.2.5 Screening of anti-RSV affinity variants

Finally, we measured the relative FFL binding across our anti-RSV scFv library. High correlation was observed between the binding affinity of the PCA and Gibson assemble scFv variants. The mutants fell into two distinct regimes. Two of them bound FFL with affinities similar to the one observed for Mota, while for the rest a 0.75 fold decrease was measured. None of the tested scFvs showed improved FFL binding compared to Mota and in general for most we were not able to obtain small affinity differences (Figure 2.9).

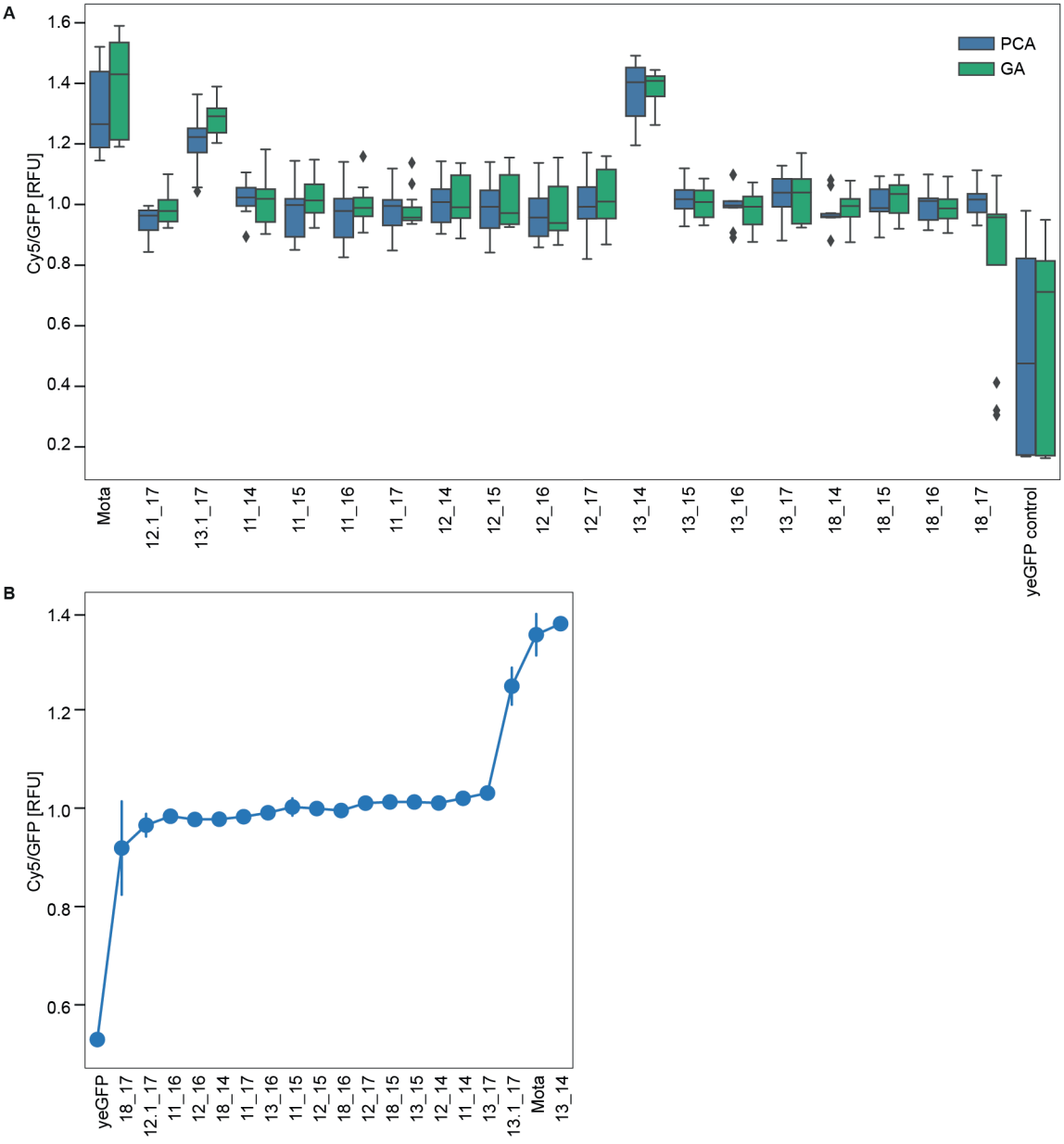

Fig. 2.9 **Coupled microfluidic based expression and screening of a small scFv library.** **(a)** Boxplot showing the anti-RSV scFv library binding affinity landscape. **(b)** Rank ordered anti-RSV scFvs according to their relative affinity binding values.

2.3 Discussion

In this study, we adapted two different strategies for scFv gene assembly. While the PCA required less time (1/2 day) and led to cost reduction ($5.5/rxn), the quality of the scFv constructs was significantly improved with the Gibson assembly. However, as shown the formation of unspecific gene product did not affect the scFv expression and as a result FFL binding measurements. In addition, we also demonstrated that functional anti-RSV scFvs can be synthesized in a simple cell-free reconstructed system such as PURExpress supplemented only with disulfide bond enhancers. This approach differs from the common ones where expensive and complex TXTL systems, often based on mammalian cell lysates, are used [207].

Finally, we were able to streamline the antibody selection and screening process and directly measure the binding properties of the on-chip *in vitro* synthesized scFv variants. This approach eliminated the use of tedious, time-consuming and expensive cell culturing, transformation and protein purification procedures. Overcoming the current technical limitations, caused mainly by variable protein expression, can lead to the establishment of a robust and rapid platform for scFv binding landscape mapping thus aiding to the discovery of physiologically relevant scFvs [1] as well as to the high-throughput characterization of various types of scFv libraries.

2.4 Methods

DNA manipulations

DNA templates encoding scFv assembly parts including VH, VL, Mota, Mota scaffold variants and anti-EGFR were ordered as gBlocks from Integrated DNA technologies (IDT). The sequence was *E. coli* codon optimized. The yeGFP construct was amplified from pKT127-yeGFP plasmid.

Prior to the scFv gene assembly the DNA templates were amplified with primers introducing overlapping regions between the assembly components. The PCR reactions were performed either by OneTaq Master Mix (NEB) or PCR mix containing Phusion High Fidelity DNA polymerase (NEB). The amplicons were purified with PCR purification kit (Qiagen) and their concentrations were determined spectrophotometrically.

scFv gene synthesis

For the PCA assembly, all necessary assembly components were added in equimolar concentration (0.1 μM) to a one-pot PCR reaction consisting of two distinct steps: (1) touch-down PCR(89-72°C), (2) standard PCR. Primers amplifying the assembled scFv products were spiked-in in the beginning of the reaction which was performed by Phusion High Fidelity DNA polymerase.

For the Gibson assembly, pUC19 vector was PCR linearized with primers introducing overlapping regions. Next, DpnI enzyme was supplemented and the reaction was incubated for 1h at 37°C to eliminate plasmid carry-over. Equimolar concentration (0.1 μM) of assembly components together with 5x molar excess of the linearized pUC19 product were mix together with a Gibson Assembly Master Mix (NEB). The reaction was incubated for 1 h at 50°C. The Gibson assembly products were transformed into chemically competent DH5 *alpha E. coli* cells. Next, the cells were plated on 100 μg/ml ampicillin plates and incubated overnight at 37°C. 3-4 colonies of each scFv variant were used for colony PCR performed by OneTaq Master Mix. Cells containing plasmid of interest were inoculated into 0.5 ml LB medium with 100 μg/ml ampicillin into a 96-deepwell plates that were sealed with breath-through foil and incubated for 8h at 37°C. 50 μl of the inoculums were mixed with 50μl 50% glycerol and pipet into a 96-well plate that was stored at -80°C. 0.5 μl of the inoculums were used for a 10 μl colony PCR reaction performed by OneTaq Master Mix. The 0.5 μl of the PCR was used as a starting material for a second PCR reaction that utilized High-Fidelity Phusion DNA polymerase and detergent-free buffer required for formation of uniform spots.

In vitro scFv expression

The bench-top scFv expression was performed in home-made TXTL system utilizing BL21 *E. coli* cells or in PURExpress systems (NEB) according to the manufacture instructions. Both reactions were supplemented with disulfide bond enhancers (NEB) and RNAse inhibitors (Roche).

Microfluidic device and epoxy-coated glass slides preparation

MITOMI devices and epoxy-coated glass slides were performed as described before [21].

DNA arraying and device alignment

The gene assembled scFvs were resuspended in 1% BSA in H2O for visualization during the alignment procedure as well as to prevent covalent linkage between the DNA and the glass substrate. The samples were plated in a 384-well plate (ArrayIt) and spotted onto epoxy-coated glass slides by a QArray2 microarrayer (Genetix) and MP2 split-pins (ArrayIt). Each spot on the array was generated by two consecutive printing programs. Immediately after DNA spotting the arrays were manually aligned to a microfluidic PDMS device with a Nikon SMZ1500 stereoscope and bonded for 4 h at 80°C. The assembled devices were stored in dark at room temperature until used.

Surface chemistry and MITOMI operation

The device was set up for use by actuating the control lines and the flow lines with 15-20 psi and 2.5-3.5 psi of pressure, respectively. Initially, the chamber valves remained closed to prevent spotted DNA resuspension. After all samples were flowed we washed the surface of the chip by PBS for 5 min. The accessible surface area was derivatized by flowing though the chip: (1) 2 mg/ml biotin-BSA (Thermo Fisher Scientific) for 15 min and (2) 0.5 mg/mL Neutravidin (Thermo Fisher Scientific) resuspended in PBS for 15 min. After these steps the 'button' membrane was activated and biotin-BSA solution was flowed for 15 min. Depending on the assay 6.67 nM solution of PBS resuspended of either biotinylated anti-GFP antibody (Abcam) or biotinylated anti-His antibody (Abcam) was flowed for 2 min thus preventing gradient formation. Next, the 'button' was deactivated and the antibody flow was continued for 15 min. In assays where biotinylated anti-His antibody was used 2 μM FFL was flowed for 2 min in 'button' closed regime followed by 5 min in 'button' open regime. After this step the 'button' was closed and PURExpress solution was flowed across the chip for 2 min. The exit valve was closed to build up pressure and the chamber valves were opened allowing dead-end filling and DNA resuspension. Next, the exit valve was opened and the chamber valves closed. PURExpress flow was continued for 5 min after which the sandwich valves were closed allowing unit cell segregation and the chamber valves and 'button' opened. The device was place on a hot plate for 3-4 h at 37°C. During this incubation, transcription and translation occurred and the synthesized scFvs diffused to the detection chamber. The expressed scFv was allowed to bind to equilibrium either to the immobilized FFL or to the anti-GFP capture probe. An image of the chip was obtained (GFP channel). The 'button' was closed and the PBS was flushed for 5min. In assays where biotinylated anti-GFP antibody was used 2 μM FFL was flowed for 2min in 'button' closed regime followed by 5min in 'button' open regime. Next, Cy5 anti-His antibody was flowed for 10 min. At this point

images of the chip were acquired (GFP channel and Cy5 channel). For Mota:FFL stability test, upon antigen-antibody interaction, the chip was multiplexed into two and PBS was flowed continuously in one of the two parts. Time-lapse imaging was performed (GFP channel). The second half of the chip was used to acquire the fluorophore bleaching rates.

Image and data analysis

The microfluidic device was scanned using a fluorescent microscope (Nikon ECLIPSE Ti). The images were obtained with a 40× objective lens and and the following exposure times: GFP channel- 500 ms, Cy5 channel- 250 ms with illumination power set to 100%. All image and data analysis were performed by custom Python scripts.

Chapter 3

Genetically encoded protein biosensor

3.1 Introduction

Various methods have been developed enabling antibody biophysical properties and expression to be improved in cost- and time-efficient manner. In addition, redesigned native antibodies by rebuilding them in different formats as well as construction of antibody chimeras have tremendously increased the scope of their applications. Currently, antibodies and recombinant antibody fragments are widely used as delivery and therapeutic agents [4, 169, 189] as well as high-affinity protein binding probes in various biosensors [35, 161].

Given the importance of antibody-based biosensors in a variety of analytic and diagnostic applications, their development has received significant attention. In general, they are composed of three units: (1) a biosensing/biorecognition element coupled to (2) a transducer and (3) a signal processing unit. The biosensing component acts to recognize, bind and sequester a target analyte from a solution to a solid-phase surface, while the transducer acts to produce a chemical or physical signal. Even though, antibody-based biosensors have successfully met the growing needs of specific molecule detection they possess several well recognized shortcomings. One of the main concerns is the high cost. In addition, the construction of highly sensitive immunosensors rely on the use of antibody chimeras possessing enzymatic moieties often unavailable commercially, imposing the need for utilizing complex chemical procedures. However, even though effective such modifications are often altering the antibody functionality or causing batch-to-batch variability.

Advances in synthetic biology have opened new avenues for *in vitro* biomolecular detection [200]. Over the years the use of cell components with tailored properties implemented in synthetic gene networks have allowed the construction of complex whole-cell biosensors [113, 116, 230]. Different logic functions have been implemented, often based on transcription cascades and signal transduction pathways, allowing defined modulations of the signal

output [34]. More recently, the adoption of freeze dried TXTL systems with embedded synthetic networks have allowed the development of ease-to-use and relatively cheap *in vitro* detection and point-of-care diagnostic tools capable of detection of RNA molecules in picomolar range [167, 168].

Here, by utilizing the current advances in *in vitro* diagnostics and antibody engineering we aimed to develop a scFv-based protein biosensor. As a design framework the conventional ELISA scheme was utilized. However, instead of relying on the use of expensive antibody molecules we substituted them with genetically encoded scFv conjugates possessing effector functionalities such as solid-surface immobilization or binding signal transduction. To convert the binding event into an optical signal a reporter DNA template encoding yeGFP protein was used. scFv transducers possessing DNA binding domains or RNA polymerase domain facilitated protein expression from the reporter DNA in the presence of home-made TXTL system, thus enabling the gain of a fluorescent signal output. We began by establishing a streamlined protocol for scFv modular assembly, *in vitro* protein expression and on-chip functional assessment that gave rise to on-demand production of functional protein-reagents. Next, we used the scFv conjugates for the construction of immunosensors capable of detection of analyte concentration in the nanomolar range (Figure 3.1).

3.2 Results

3.2.1 Design and *in vitro* synthesis of scFv conjugates

To begin, we designed DNA templates encoding scFv chimeras capable of detecting FFL. For antigen recognition we used the anti-FFL 12_17 scFv which was functionally characterized on MITOMI (see Chapter 2). The sequence was appended with regulatory elements including T7 promoter and Shine-Dalgarno RBS site enabling rapid *in vitro* protein expression as well as a linker sequence allowing modular assembly to three different moieties conferring solid-phase immobilization or signal transduction. For scFv conjugate synthesis the PCA strategy was adapted given its simplicity and speed (see Chapter 2). Following the previously established protocol VH, VL and yeGFP templates were premixed with DNA constructs coding for the various functional moieties (Figure 3.2a,b). In our general design scheme the functional domains were fused to the N-terminus of the scFvs via a (Gly4Ser)3 linker. A yeGFP was fused in frame to the C-terminus of the scFV conjugates allowing protein quantification (Figure 3.2c).

First, we constructed scFv fused to an enhanced monomeric avidin (eMa). eMa is an monomeric engineered peptide version of Rhizavidin dimer with a rigid binding pocket stabi-

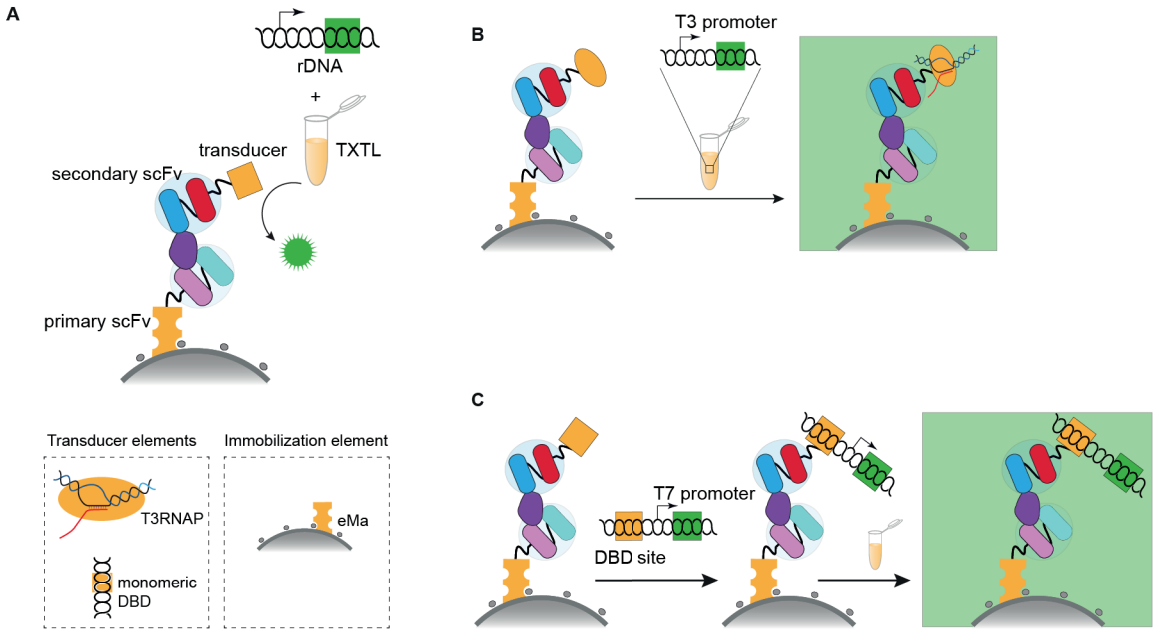

Fig. 3.1 **Design and operation of a genetically encoded protein biosensor.** **(a)** The designed biosensor is composed of two *in vitro* synthetized scFv conjugates enabling solid immobilization and signal transduction. The transducer element acting as a detection antigen probe is also capable of either **(b)** aiding to the expression of a reporter DNA template embedded into TXTL or **(c)** pulling down a reported DNA template which expression is supported entirely by a TXTL.

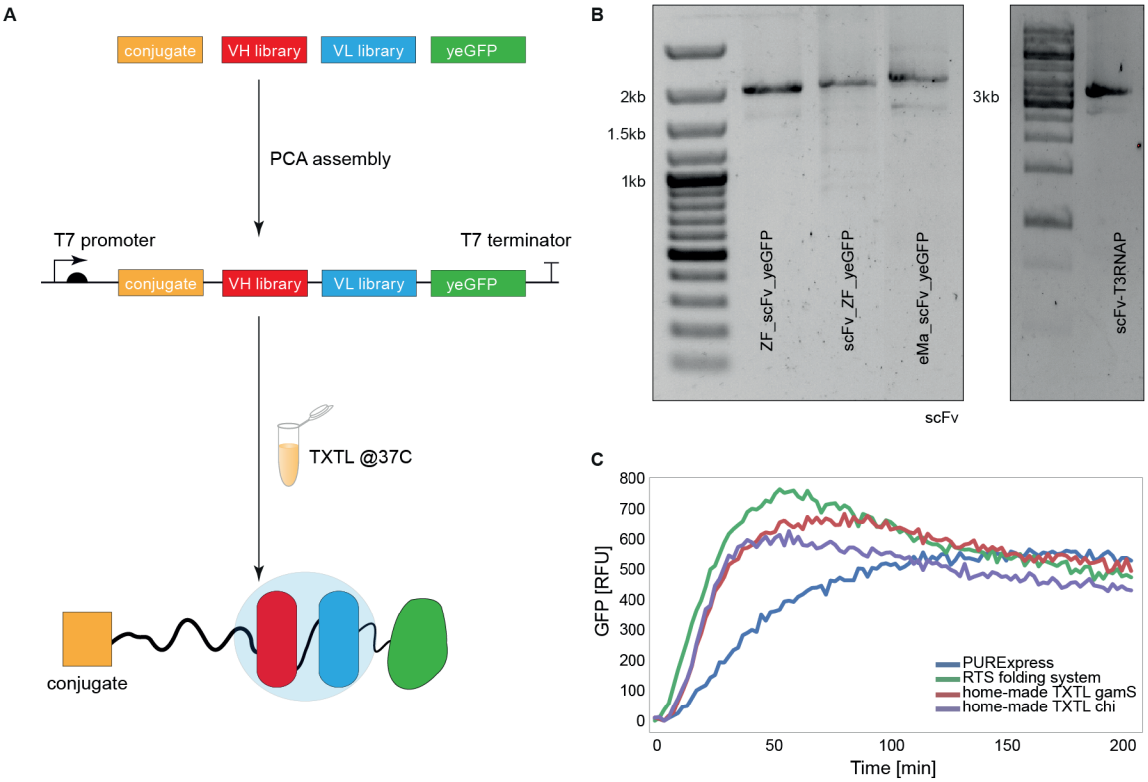

Fig. 3.2 **Construction and *in vitro* expression of scFv conjugates. (a)** Outline of the scFv conjugates construction and protein expression. **(b)** Agarose gel electrophoresis of the gene assembled scFv conjugates. **(c)** *In vitro* expression of the scFv conjugates in different TXTL systems. The volume of the reactions, incubated at 29°C, was set to 5 μl. scFv-ZF construct was used as a DNA template.

lized by a disulfide bond [126]. The integration of immobilization functionality eliminated the use of commercially available biotinylated antibodies often used for surface immobilization. The resulting fusion molecule was used to immobilize the scFv to a biotin monolayer (Figure 3.1a).

Next, we designed and constructed scFv chimeras with signal transduction properties by joining together the scFv domain with a DNA binding domain (DBD). The DBDs, present in many transcription factors, can fold correctly without the assistance of addition factors into functional units that recognize and bind specific DNA sequence. Based on their tertiary structure various DBD families have been identified, the most prominent being zinc-fingers (ZFs), helix-loop-helix, leucine zipper [81, 216]. A plethora of protein-DBD conjugates are widely applied in domains such as gene expression regulation and genome engineering. One of the most widely used DBDs are the ZFs given their intrinsic programmability and free online access to databases storing binding specificity information. Here, we made use of zif BDB (ZF) [21], given its small size, high binding affinity and monomeric mode of DNA recognition. Moreover, we also tested MarA and SoxS, a well-known AraC regulators [51]. To assess whether effector domain conjugation is altering scFv functionality in addition to scFv-ZF construct we also synthesized ZF-scFv. Once bound to FFL the scFv-DBDs were capable of pulling down reporter DNA templates possessing 9 bp long target binding sites (Table 3.1). To obviate yeGFP expression interference the DBD binding site was positioned at the 5' untranslated region, 85 bp away from the T7 promoter. Upon scFv-DBD:DNA binding extensive washing step was performed to remove the excess reporter DNA molecules that was followed by *in vitro* yeGFP expression (Figure 3.1a,c).

The design approach was then extended to the construction of scFv chimeras with genetically encoded enzymatic activity. As a functional moiety we used T3 RNA polymerase (T3RNAP). Several design approaches were tested. Remarkably, protein fusion to the T3RNAP C-terminus altered its enzymatic activity, as described previously. Thus, the design was based on a strategy used for Gal4-T7RNAP construction [163]. In the case of scFv-T3RNAP we omitted the yeGFP and fused (-2)T3RNAP construct to the C-terminus of the scFv via a Pro-Ser-Thr-Ser-Arg-Gly-Gly-Pro-Val-Pro linker. Using reporter DNA templates under T3 promoter, constitutive expression of yeGFP was directly induced in small, solution-phase TXTL reactions (Figure 3.1a,b).

Upon assembly, the scFv conjugates were expressed using either the PURExpress or a home-made TXTL, both supplemented with RNAse inhibitors and disulfide bond enhancers. In the case of home-made TXTL, we also added gamS protein enabling protein expression from linear templates (Figure 3.2c).

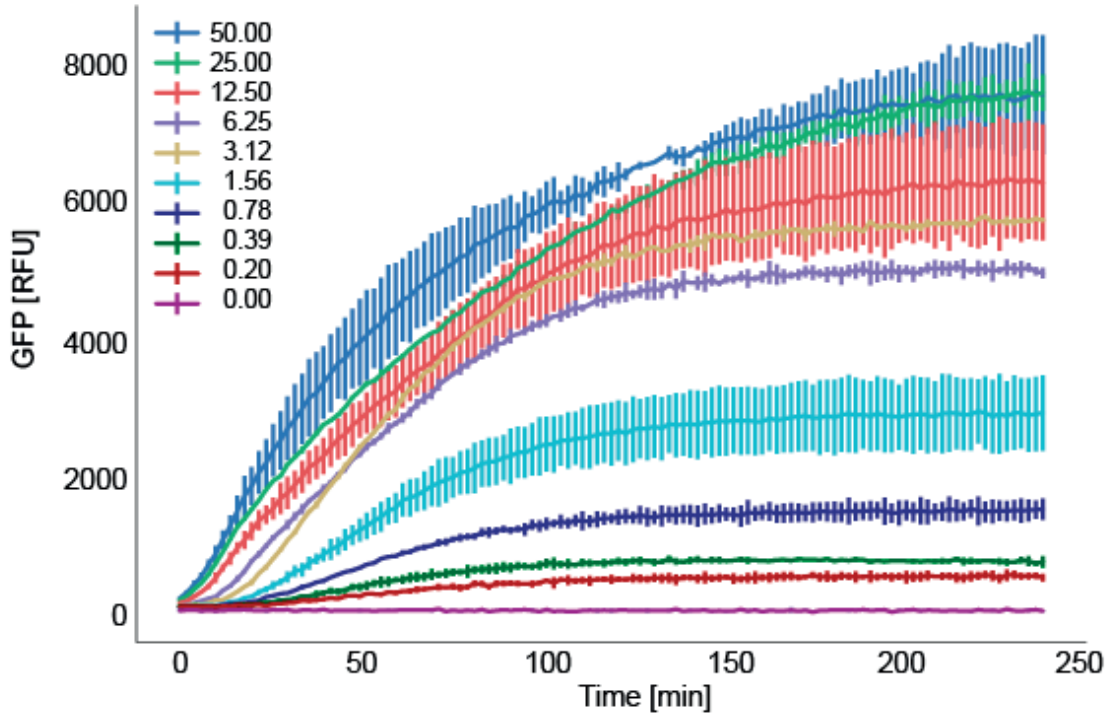

Fig. 3.3 **Reporter DNA titration study.** Reporter DNA template with ZF binding site was titrated in a home-made TXTL. The induced yeGFP expression was monitored for 4 h. The annotated DNA concentrations are in nM. (N=2)

To access the lowest reported DNA concentration required for induction of observable yeGFP expression, we performed a DNA titration study. Dilution series of reporter DNA templates were mixed with TXTL. The reactions were incubated at 37°C and the expression profile was monitored for up to 3h. Even 0.2 nM reporter DNA induced yeGFP expression significantly higher compared to the negative control (Figure 3.3).

3.2.2 Assessment of scFv conjugate functionality

We initially examined DBD-scFv functionality. To test whether the scFv retained its antigen binding properties after the conjugation, we flowed bench-top expressed proteins across a MITOMI chip immobilized with 2 μM FFL. Even though all variants were successfully synthesized, only the ZF-scFv was capable of FFL binding. The observed ZF-scFv and scFv-ZF discrepancy suggested the importance of proper domain orientation for scFv functional preservation (Figure 3.4). Next, we tested ZF domain activity. The same experimental procedure was performed, however, in addition fluorescently labelled reporter templates with cognate ZF binding site were flowed. Binding was observed only in the presence of FFL and

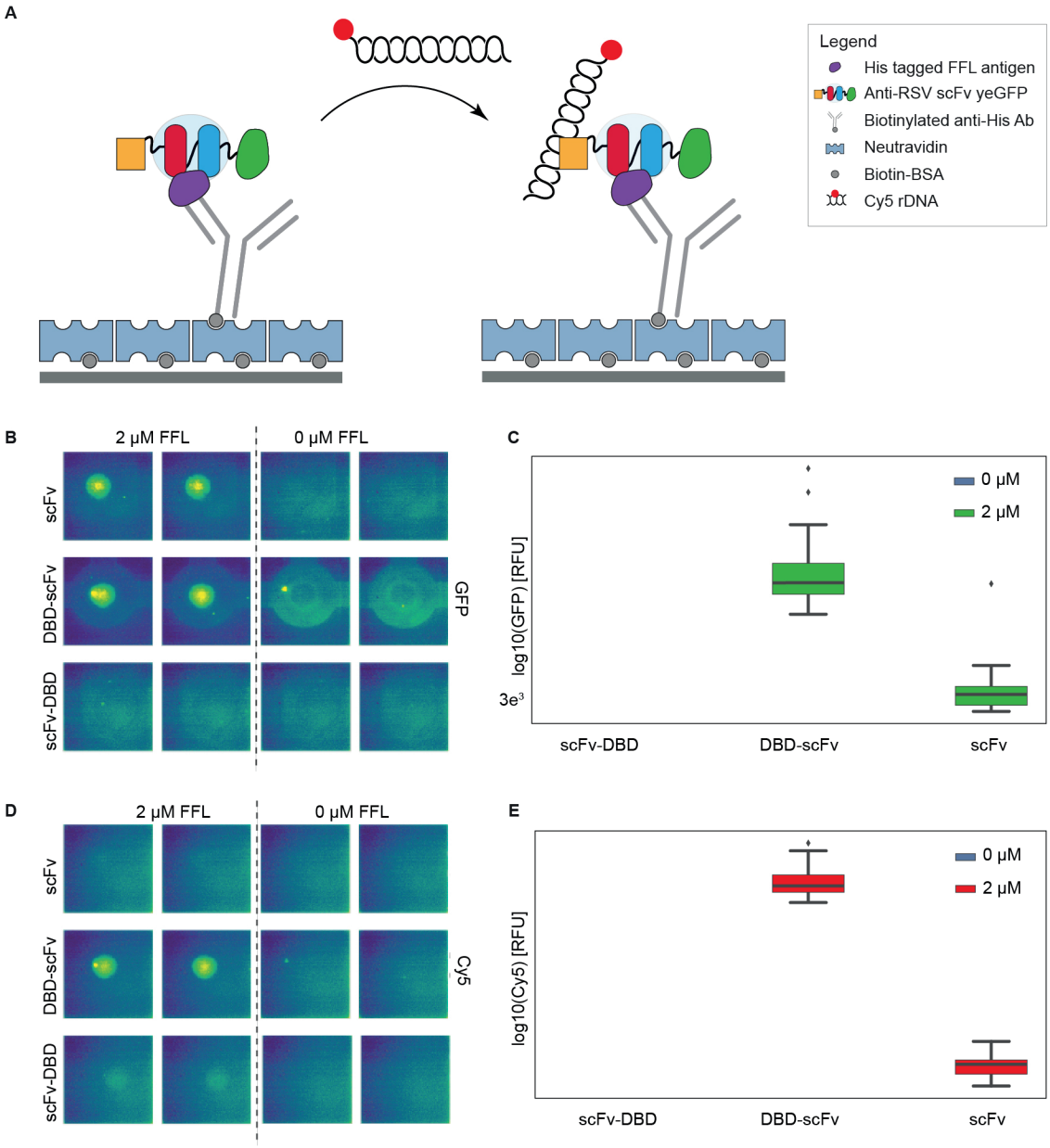

Fig. 3.4 **On-chip interrogation of the DBD-scFv functionality. (a)** General scheme of the assay. scFv variants conjugated to ZF domain are used as detection probes for on-chip surface immobilized FFL. Upon scFv binsing a fluorescent DNA template with ZF binding site is used to access the functionality of the ZF domain. Fluorescent images and quantitative plots showing the interaction **(b)** between FFL:scFv constructs (GFP) and the reporter DNA: scFv construct (Cy5).

functionally active ZF-scFv. Subsequently, we studied the temporal dynamics of cell-free yeGFP expression using pulled-down reporter DNA templates on MITOMI. PURExpress was flowed and incubated at room temperature. Protein expression was induced as quickly as 30 min. A significant difference in the expression levels were observed only in the presence of ZF-scFv and FFL. The maximum fold change occurred around 5-6 h, afterwards the signal reached a plateau (Figure 3.5).

Next, we examined scFv-T3RNAP. When premixed in PURExpress with a DNA template coding for yeGFP under the control of a T3 promoter the conjugate was able to induce yeGFP expression. To test orthogonality to T7 polymerase, which is present in PURExpress, the reporter DNA template was placed in a reaction without scFv-T3RNAP conjugate (Figure 3.6). Only negligible expression was detected. Validation of the scFv domain was performed by a bead-based assay which will be described in the next section.

eMa-scFv was also assayed on-chip. After bench-top expression the protein was flowed across the chip and successfully immobilized via biotin and eMa interrogation. Upon continuous PBS washing only negligible amount of protein was detached from the surface. In addition, we also observed the preserved functionality of the scFv unit which was capable of capturing 2 μM FFL (Figure 3.7).

3.2.3 Development of a complete protein biosensor

Next, we sought to use as a solid-phase support streptavidin coated magnetic beads instead of MITOMI devices. Such an approach has the potential to be broadly accessible without requiring specialized equipment. In addition, readout measurements using a plate reader can be used instead of complex fluorescent microscope setups.

The general experimental scheme was preserved. First, biotinylated anti-His antibodies bound to the surface of neutravidin coated Dynabeads were used as FFL capturing probes. Next, scFv transducer elements were added to the beads allowing the detection of the exposed FFL peptides. In the case of scFv-T3RNAP the beads were resuspended in a TXTL containing reporter DNA templates. The sample was directly placed on a plate reader and the expression of yeGFP measured (Figure 3.8a). In the second scenario in which ZF-scFv was used, reporter DNA templates were added to the beads and incubated for 15 min allowing the DNA molecules to be pull-down by the scFv conjugates. Next, after extensive washing step the beads were placed in TXTL reaction mix inducing yeGFP expression (Figure 3.9a).

Finally, ZF-scFv based biosensor was employed to detect three different concentrations of FFL: 10 μM, 5 μM and 0.5 μM (Figure 3.8b). yeGFP expression which plateau after 3 h of incubation was observed in all tested samples. As expected, the FFL concentration strongly affected the protein synthesis levels. Negligible expression was observed in the absence of

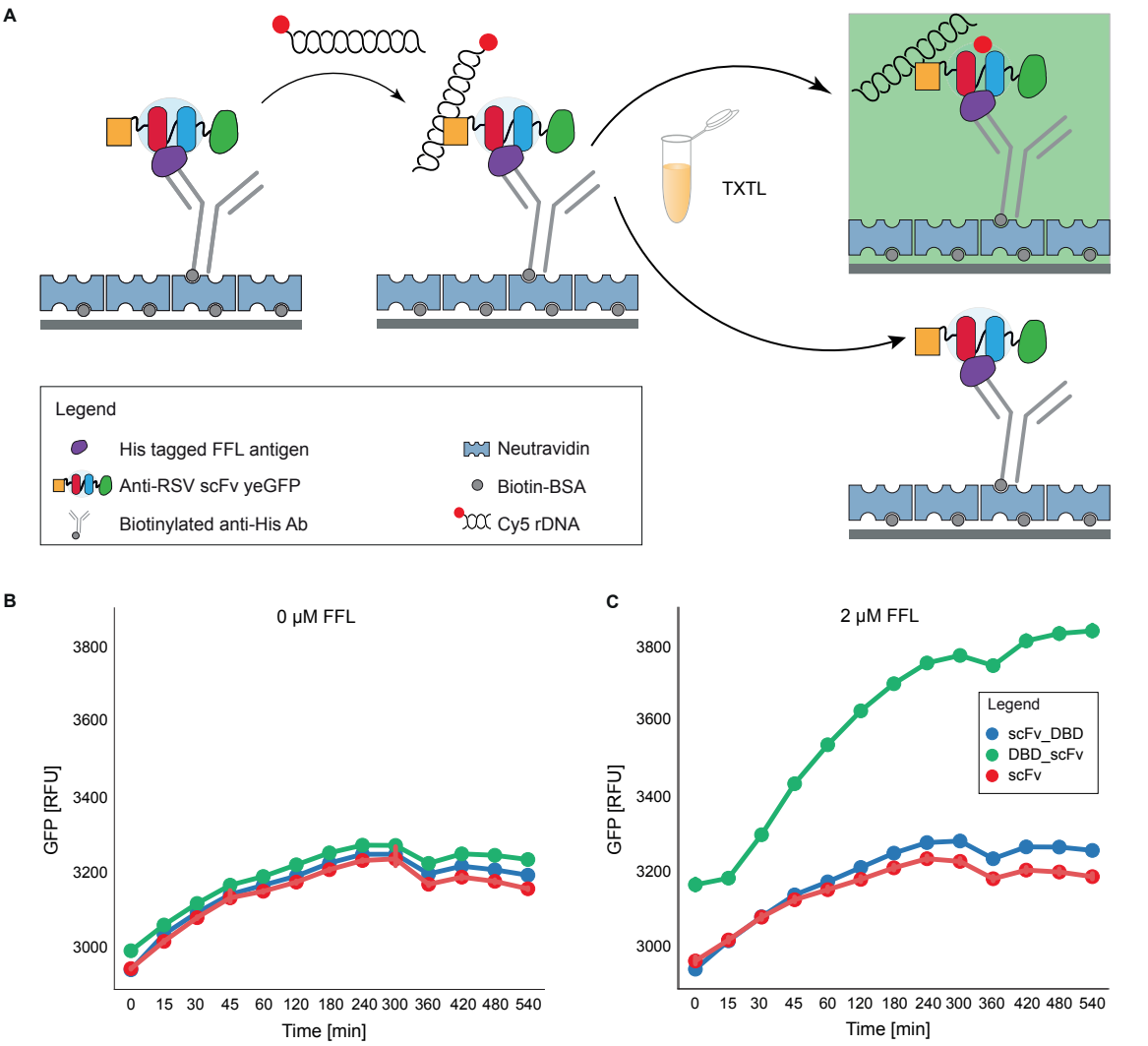

Fig. 3.5 **Temporal dynamics of DBD-scFv driven yeGFP expression.** **(a)** Outline of the experimental procedure. The surface of the microfluidic device was either immobilized with **(b** 0 μM FFL or **(c)** 2 μM FFL. For FFL detection and detection signal enhancement DBD-scFv and reporter DNA construct with ZF cognate site.

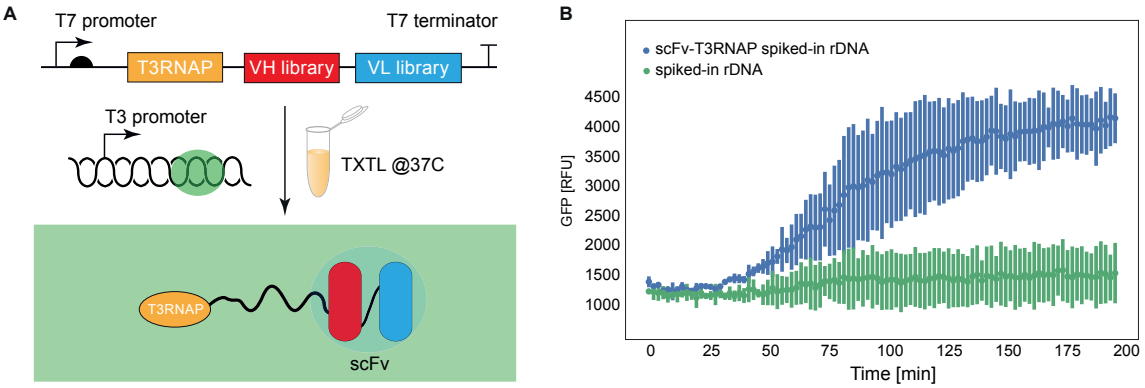

Fig. 3.6 **scFv-T3RNAP activity validation.** **(a)** Outline of the experimental procedure. scFv-T3RNAP construct and a reporter DNA template possessing a T3 promoter are added together to a TXTL. **(b)** Time series of yeGFP expression from a reporter DNA template embedded in PURExpress in the presence or absence of a scFv-T3RNAP. (N=2)

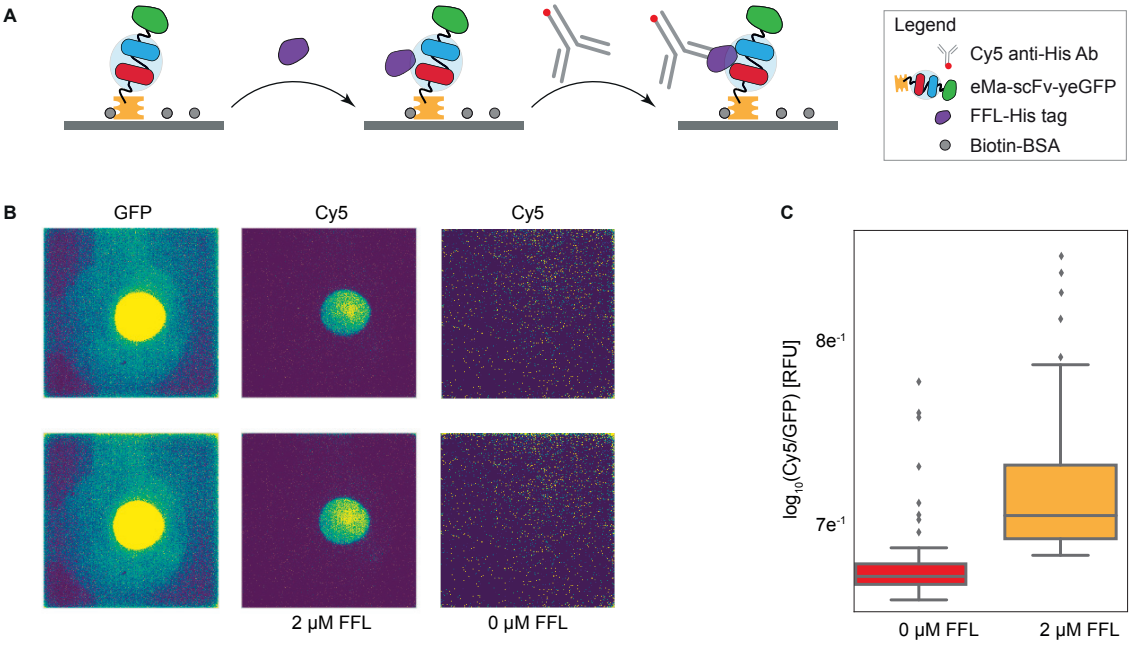

Fig. 3.7 **eMa-scFv activity validation.** **(a)** General scheme for validation of eMa-scFv functionality on MITOMI. **(b)** Representative fluorescent images **c** and quantitative data from eMa-scFv:FFL binding interrogation.

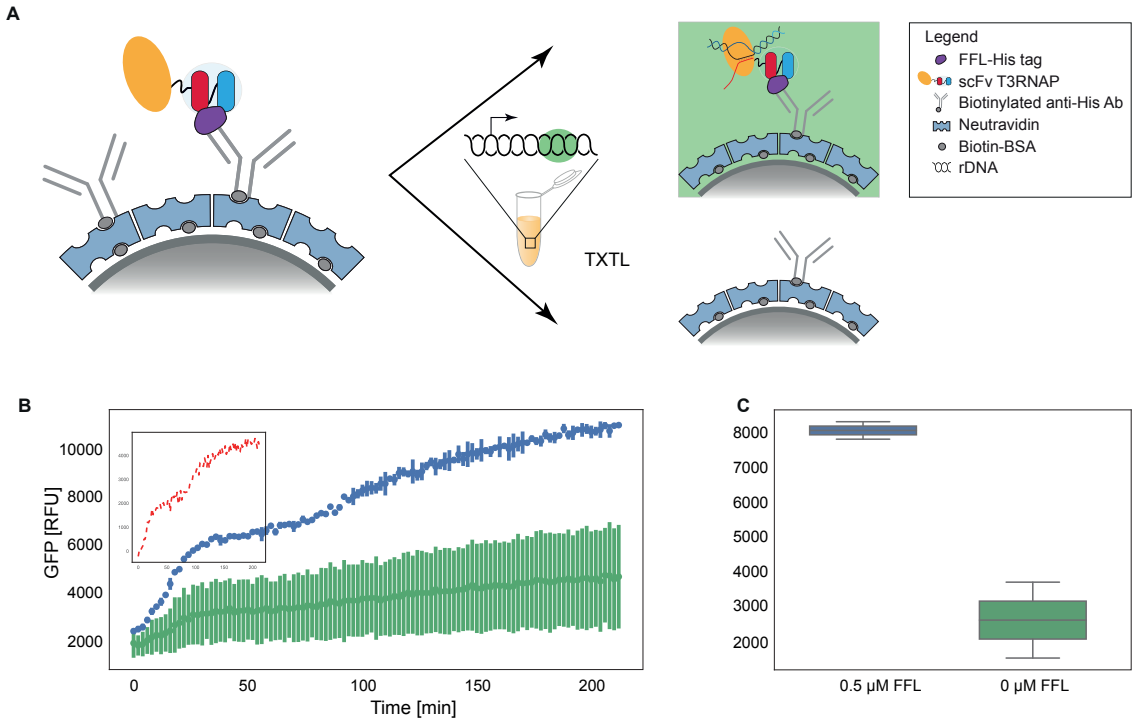

Fig. 3.8 **scFv-T3RNAP protein biosensor. (a)** Outline of the experimental procedure. scFv-T3RNAP transducers are used as detection probes of surface immobilized FFL. In addition they are capable of direct yeGFP expression induction when presented to a TXTL system with embedded reporter DNA template under a T3 promoter. **(b)** yeGFP expression profile when the beads are functionalized with 0 μM or 2 μM FFL. **(c)** Boxplot comparing the averaged yeGFP signal after 4h of incubation in the presence or absence of FFL. (N=2)

FFL. In addition, we titrated FFL and we were able to detect concentrations down to 67.5 nM (Figure 3.8c). Finally, we assayed the scFv-T3RNAP biosensor when sampled with 0.5 μM (Figure 3.9b,c). Almost four fold difference in the detection signal was observed after four hours of incubation. A significant difference was observed between the positive and negative control.

3.3 Discussion

Here, we established a pipeline for construction, *in vitro* synthesis and functional validation of scFv chimeras that can be used as on-demand low-cost alternative to the expensive conventional ELISA reagents. To obviate the need of complex chemical procedures for scFv labeling a modular assembly approach was utilized to genetically fuse an anti-FFL scFv domain to three different u nctional moieties: T3RNAP, ZF and eMa. While T3RNAP and ZF domains allowed signal detection enhancement, eMa enabled solid-phase surface immobilization via biotin interrogation. In addition, we developed biosensor containing scFv transducer elements for the quantitative detection of FFL. Construction and validation of scFv conjugates toward distinct protein epitopes will be a step toward the development of entirely genetically encoded biosensor which has the potential to significantly reduce the cost of the conventional ELISA assays.

3.4 Methods

DNA template synthesis

DNA templates encoding DBDs, eMa and the anti-FFL scFv VH and VL chains were codon optimized for *E.coli* and synthesized as gBlocks by Integrated DNA technologies (IDT). The T3RNAP and yeGFP constructs were amplified from BBa_K346000 part (Registry of Standard Biological Parts) and pKT127-yeGFP plasmid, respectively. Prior to the PCA assembly, all sequences coding for scFv conjugate modules were amplified via PCR with primers containing 25-30 bp flanking overlap regions. Transcription and translation regulatory elements were included in the primers amplifying the sequences encoding the C- and N-termini. To yield linear ready-to-expresss scFv conjugate templates 0.1 μM of each assembly module was added to a PCR reaction that consist of touch-down PCR step (98-72°C, 1 min extension) and standard PCR step. Primers amplifying the final product were supplemented in the beginning of the reaction.

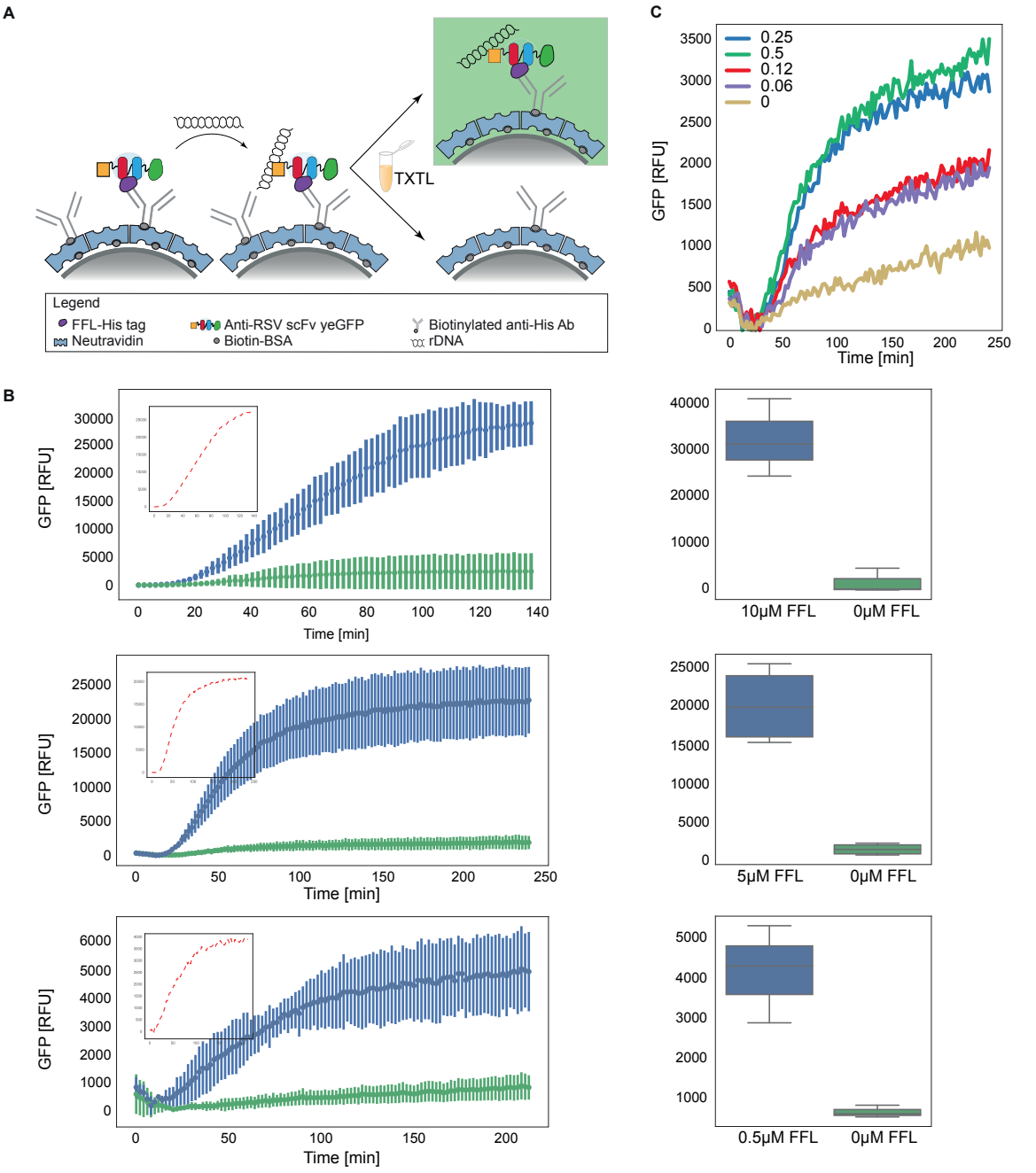

Fig. 3.9 **Detection of FFL in nanomolar range with DBD-scFv bead-based biosensor.]** **(a)** Outline of the experimental setup. ZF-scFv transducer are used to detect bead immobilized FFL and to pull-down reporter DNA template with ZF binding site. Upon washing, the beads are present to TXTL which induced yeGFP expression from the surface immobilized DNA templates. **(b)** yeGFP expression profile different FFL concentration. The boxplots comparing the averaged yeGFP signal after incubating the beads with TXTL for 4 h in the presence or absence of FFL (N=2). **(c)** Nanomolar FFL titration study.

To generate reporter DNA templates yeGFP gene was amplified with primers containing regulatory elements including DBD binding site, promoters, RBS site, and terminators. To quantify the amount of pull-down reporter DNA template by the scFv conjugates during the microfluidics assay a Cy5-labeled 3' primer was used.

All PCR reactions were performed by Phusion polymerase (NEB). All sequence are shown in Table X.

In vitro scFv conjugates expression

scFv conjugates were expressed *in vitro* using commercial PURExpress (NEB) according to the manufacturer instructions or home-made TXTL system [218]. The reactions were prepared on ice and supplemented with Protector RNase Inhibitor (Roche) and disulfide bond enhancers (NEB). 5 μM gamS or 2.5 μM chi DNA were added to the TXTL system to support protein expression from linear DNA construct. Template DNA was added to the reaction mix without purification. Depending on the application the reaction volume varied between 5-40 μl and thus the DNA volume varied between 0.2-1.6 μl. The reactions were centrifuged to the bottom of wells in optical 384-well polystyrene plates (Nunc). While PURExpress reactions were incubated at 37 °C for up to 4 h, the TXTL reactions were performed at 29 °C.

To study reporter DNA concentration-dependent yeGFP expression a dilution series of DNA template in 5 μl home-made TXTL reactions. Prior to use the reporter DNA template was purified by PCR purification kit (Qiagen) and the concentration was determined spectrophotometrically.

The fluorescence of scFv conjugates or reporter DNA template was measured by Biotek SynergyMx microplate reader (excitation 485 ± 9 nm, emission 515 ± 9 nm at a sensitivity of 70) Fluorescence intensity was transformed into concentration unit using a GFP calibration curve acquired by plate reader measurement of 5 μl dilution series of recombinant GFP protein (Biovision) resuspended in PBS.

Microfluidic device fabrication and epoxy glass slide preparation

All procedures for microfluidic device fabrication and epoxy coating of the glass slides were performed as described before [21].

MITOMI setup and operation

The MITOMI devices were operated by actuating the control lines and the flow line pressure at 15-20 psi and 2.5-3.5 psi, respectively. The chamber valves remained closed during the experiments. After each solution flow PBS was flushed for 5 min. The accessible surface area was derivatized by flowing: 1) 2 mg/ml biotin-BSA (Thermo Fisher Scientific) for 15 min and 2) 0.5 mg/mL Neutravidin (Pierce) resuspended in PBS for 15 min. After these steps the 'button' membrane is activated and biotin-BSA solution was flowed again for 15 min. Then a 6.67 nM anti-His antibody solution (Abcam) was flowed across the chip for 2 min, preventing gradient formation. Next the 'button' membrane was opened and the antibody flow was continued for 15 min. 2 μM FFL resuspended in PBS was flowed for 2 min in 'button' closed regime followed by 5 min flow in 'button' open regime. After this steps bench-top expressed scFv transducer elements were flowed across the chip for 5 min. The device was placed on a fluorescent microscope setup and an image is taken (GFP channel, 40x, 250 ms, 100% laser power) to determine relative scFv conjugates amounts when in solution phase. The chip was washed in a 'button' closed regime and an additional image was obtained to determine the relative amount of bound scFv conjugates. Next, reporter DNA template with cognate ZF binding site was flowed for 5 min. Additional images (Cy5 channel, 40x, 250 ms, 100% laser power) were obtained before and after wash with activated 'button' membrane. Finally, PURExpress was flowed for 5 min with a pressurized 'button' membrane. The exit valve was activated for 1 min while the solution is being continuously flowed on-chip in a 'button' open regime to build-up pressure. Finally, the exit valve was opened and the sandwich valves were closed allowing segregation between individual chambers. Images (GFP channel, 40x, 250 ms, 100% laser power) were taken every 5 min during the first two hours and every 30 min during the next 6-12 h to monitor the yeGFP expression profile.

Bead-based assay

MyOne™ Streptavidin T1 beads (Life Technologies) were used for all bead-based assays. Each reaction was set to 10 μl for which 2 μl of beads were resuspended in 1x PBS. Afterwards, the beads were pelleted using a magnetic stand (Invitrogen DynaMag™-Spin) for 30–60 s until the solution is clear. The supernatant was carefully aspirated. The beads were washed three times with 20 μl of 1x PBS. Each washing step involves adding the wash solution, mixing the solution by aspiration until the beads are resuspended, then pelleting the beads and removing the supernatant. The beads were pelleted again and resuspended in a 10 μl mixture of pre-incubated for 15 min FFL and 200 nM biotinylated anti-His antibody (Abcam). Next, the bead reaction was incubated at room temperature for at 15 min with

Table 3.1 **DBDs characteristics**

	size [bp]	DNA target sequence	binding mode
Zif BDB	264	tGAG GTA GTGt	monomer
MarA	381	ATGGGAGGTTTTGCTAAAT	monomer
SoxS	321	ATGGGAGGTTTTGCTAAAT	monomer

gentle agitation. Following the incubation, the beads were pelleted against the magnetic stand, washed three times with 20 μl of 1x PBS, and resuspended in 10 μl pre-incubated solution of bench-top expressed 250 nM scFv transducers and 100 nM reporter DNA template, in the case of ZF-scFv. The mixture was again incubated for 15 min at room temperature with gentle agitation. Next addition washing procedure was performed and the beads were finally resuspended in 10 μl TXTL solution containing reporter DNA template, in the case of scFv-T3RNAP. Finally, the bead solutions were pipetted in optical 384-well polystyrene plates (Nunc). The plate was spin down quickly in order the solutions to be placed at the bottom of the wells, later the used wells were covered with 35 μl of Chill-Out Liquid wax (Biorad) to avoid evaporation. Reactions were performed in a Biotek SynergyMx plate reader at 37°C with constant shaking. Every 2 min for 4 h EGFP fluorescence (excitation 485 ± 9 nm, emission 515 ± 9 nm at a sensitivity of 70) were measured.

Data analysis

All images and acquired data were analyzed by custom Python scripts.

Chapter 4

In vitro biophysical characterization of CRISPR-Cas9

4.1 Introduction

CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats–Cas9-associated) system provides a RNA-mediated adaptive immune response against invading plasmids and bacteriophages in prokaryotes and algae organisms [39, 197, 243]. Its original role has been repurposed and today it is widely used as an RNA-guided DNA targeting tool [12, 197]. The system is composed of a *Streptococcus pyogenes* Cas9 endonuclease which when programmed with a dual-guided RNA or a chimeric single-guided RNA (sgRNA) is capable of binding and cleaving 17-20 bp target DNA site flanked at the 3' end with short nucleotide sequence called protospacer adjacent motif (PAM) [6, 215]. The ease of use and programmability has driven CRISPR-Cas9 widespread adoption in different genome-editing applications [25, 50, 130]. It rapidly became the preferable tool compared to Zinc Finger Nucleases (ZFNs) [226] and Transcription Activator Like Effector Nucleases (TALENs) [149], both of which require cumbersome assembly and characterization. In addition, disruption of the Cas9 endonuclease activity allows the use of the system in various gene activating (CRISPRa) [49, 91, 115] and gene repression (CRISPRi) [49, 91, 179] studies as well as real-time imaging [40, 91, 151].

The specificity of the CRISPR-Cas9 system is of utmost importance for its safe translation in genome modification and gene therapy applications [214]. However, a number of *in vitro* [170] and *in vivo* [44, 92, 120, 133] studies have posed concerns about its off-target binding and cleavage activity that can cause unwanted mutations and chromosomal rearrangements.

Currently, the design of the sgRNA relies on the use of predictive algorithms [9, 145, 146, 253]. However, such computational models are typically based on parameters obtained from biochemical experiments and genome-wide studies. In addition, molecular features playing a role in CRISPR-Cas9 specificity such as guide RNA stability, conformation and binding strength are usually completely neglected given that they still remain largely unexplored. Thus, the development of tools allowing rapid high-throughput CRISPR-Cas9 biochemical characterization was needed. In-depth exploration of the factors determining CRISPR-Cas9 binding specificity in reproducible and cost-effective manner can aid to the optimization of the current *in silico* predictive models and as well as the scope of CRISPR-Cas9 applications.

In this work we developed a high-throughput microfluidic platform for rapid interrogation of the binding and cleavage activity of an *in vitro* reconstructed CRISPR-Cas9 system. Our approach enabled precise control over the concentration of all essential system components-Cas9, sgRNA and DNA targets,- which is a prerequisite for reproducible quantitative measurements. Furthermore, it bypassed expensive, time-consuming and tedious cell-handling and protein purification procedures increasing its scalability and feasibility. We were able to rapidly, within a day, screen two small mutant DNA libraries and infer information about the mutation tolerance levels of the *in vitro* Cas9/sgRNA system and the impact of sgRNA sequence on the Cas9 DNA interaction. Finally, we also obtained insight into the correlation between CRISPR-Cas9 binding and cleavage activity.

4.2 Results

4.2.1 Design and synthesis of cell-free reconstructed CRISPR-Cas9 system

We first established a pipeline for the *in vitro* synthesis of all essential CRISPR-Cas9 components including Cas9 protein, sgRNA, and DNA targets. This approach obviated long and labor-intensive cell-handling and protein purification procedures. In addition, it also took advantage of the MITOMI microfluidic platform for coupled expression and rapid CRISPR-Cas9 biochemical characterization.

To test both binding specificity and cleavage activity we used two types of *Streptococcus pyogenes* Cas9 proteins- a catalytically-active Cas9 and catalytically-dead Cas9 (dCas9) that contains point mutations at the RuvC and HNH cleavage domains [68, 98]. Cas9 genes were cloned under the control of a T7 promoter that facilitated Cas9 *in vitro* synthesis. In addition, a yeGFP-tag was fused to the C-terminus of dCas9 allowing ease of protein concentration quantification as well as pull-down via anti-GFP antibody. The protein yields after 2 h of

incubation with commercial cell-free PURExpress system at 37°C were typically in the range of 0.6-1.4 μM.

To simplify the assay a chimeric sgRNA instead of a dual-guided RNA was used. Evaluation of their activity revealed similar functionality. For the sgRNA synthesis we used an *in vitro* transcription system. To drive the expression of double stranded DNA bearing T7 promoter at the 5' end followed by 20 bp DNA, homologous to the protospacer region, and a long sgRNA DNA scaffold located at the 3' terminus, was used. We designed two sgRNA possessing different secondary structures in order to explore the effect of sgRNA sequence and conformation on the CRISPR-Cas9 binding specificity.

For profiling the binding and cleavage activity of the *in vitro* reconstructed Cas9/sgRNA complex two different mutant libraries, which we refer to as CC and L DNA library, were used. Both of them were comprised of short double stranded DNA templates labelled with biotin and/or fluorophores allowing pull-down, when necessary, and DNA quantification. The designed templates were composed of 20 bp protospacer region followed by a 5'-NGG-3' PAM sequence and two flanking regions of random sequences. In addition, all of them possessed a mutation window with different length (from 4 bp to 16 bp) and position at the DNA protospacer region (Figure 4.1).

4.2.2 Initial validation of the cell-free reconstructed CRISPR-Cas9 system

Next, we developed a general strategy for evaluation of the binding specificity of the *in vitro* reconstructed CRISPR-Cas9 system on MITOMI. An overview of the procedure is shown in Figure 4.2a,b. Upon synthesis the DNA library was spotted on a glass slide used to program the microfluidic device; every single spot was compartmentalized within an individual unit cell. A dilution series of each DNA sample was used enabling quantitative measurements of the Cas9/sgRNA DNA interrogation to be obtained. After performing standard surface chemistry, a dCas9/sgRNA complex was flowed across the chip and pulled-down by immobilized anti-GFP antibody. The complex was assembled bench-top by premixing 250 nM dCas9 and 4x excess of sgRNA in a binding buffer. Next, the spotted DNA templates were rehydrated and allowed to interrogate with dCas9/sgRNA exposed on the surface. Finally, after reaching equilibrium the signal from the free DNA molecules in solution was measured, the chip was washed and rescanned to quantify the signal from the on-chip immobilized dCas9 and the capture DNA targets.

Initially, dilution series of two 60 bp Cy5-labeled DNA templates were used for spotting: (1) a consensus DNA target and (2) a random DNA sequence that was used as a negative

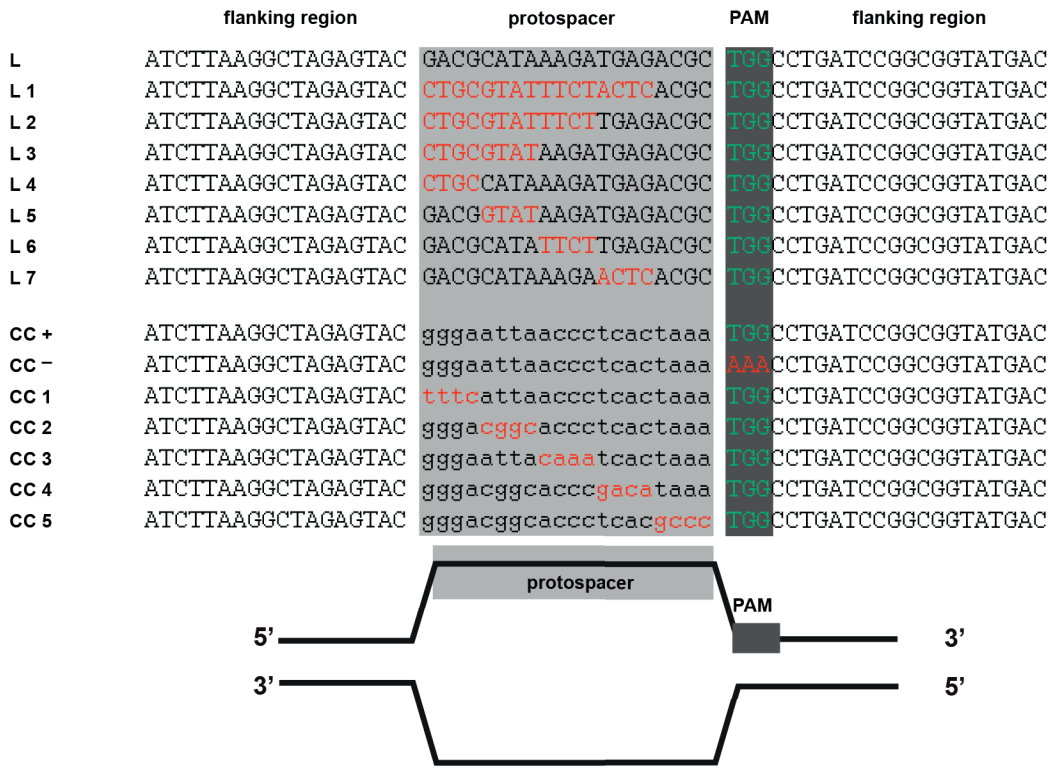

Fig. 4.1 **Sequences of CRISPR-Cas9 DNA libraries** Sequence alignment and annotations of the L and the CC DNA library. The mutation windows are shown in red.

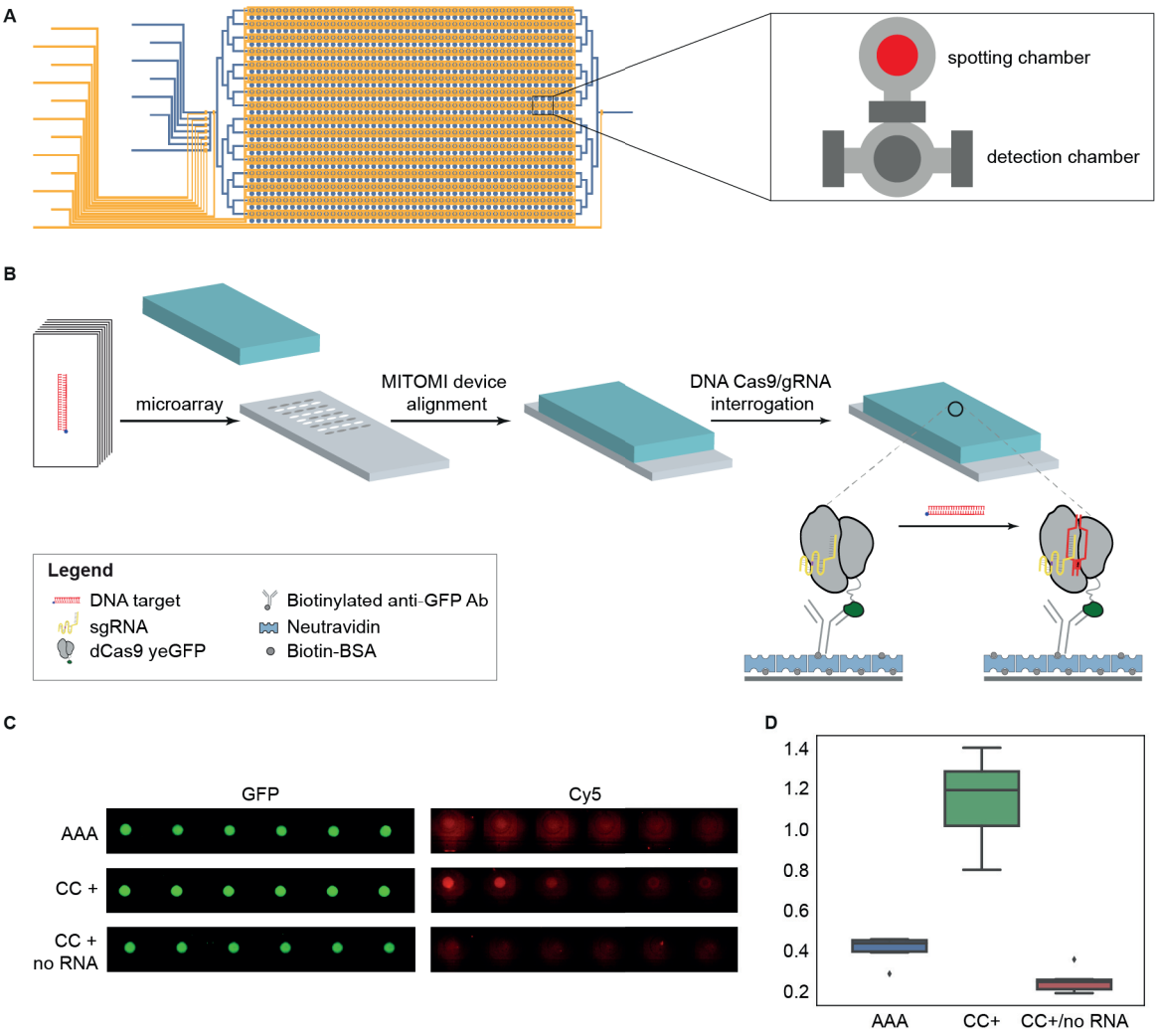

Fig. 4.2 **Cas9/gRNA DNA interrogation on MITOMI. (a)** General scheme of MITOMI 768 device. **(b)** Microarray spotting, device alignment and biochemical assay for studying dCas9/gRNA DNA interrogation. **(c)** Representative images and quantitative data from an assay where a consensus DNA template a random DNA sequence were used to access the Cas9/sgRNA functionality.

control. The chip was multiplexed and dCas9 alone or in preformed complex with sgRNA were flowed. Addition of sgRNA sufficiently enhanced the dCas9 binding to the consensus DNA sequence, proving the role of the sgRNA as an allosteric activator of Cas9 DNA binding (Figure 4.2c,d). On the other hand, only negligible Cas9/sgRNA binding was observed when the random DNA sequence was assayed.

4.2.3 High-throughput analysis of CRISPR-Cas9 biochemical properties

Evaluation of the binding specificity and affinity of CRISPR-Cas9

Next, we measured the Cas9/gRNA CC library interrogation (Figure 4.3a) We observed that four adjacent nucleotide substitutions located at the protospacer 'seed' region (the sequence situated next to the PAM site) as well as complete elimination of the PAM site strongly affected Cas9 binding. Sliding the mutation window away from the 'seed' reconstituted the binding and there was negligible position-dependent difference. To investigate whether this observation holds true for another sgRNA-DNA pair, we tested the L library (Figure 4.3b). Interestingly, the binding affinity toward the consensus DNA target was lower compare to one measured for the CC library. In addition, the binding trends differ from the ones observed before. Preservation of the four base pair adjacent to the PAM was not sufficient to support stable dCas9/sgRNA binding as shown for the CC library. Up to twelve bases from the PAM site were required to be error-free in order the mutations to be tolerated. Mutation windows composed of 8 or 12 base were shown to almost completely eliminating the dCas9/sgRNA binding. The results suggest the importance of sgRNA sequence and for well-controlled CRISPR/Cas9 interrogation.

Evaluation of the cleavage activity of CRISPR-Cas9

Upon binding to the DNA substrate the wild-type Cas9 is undergoing large conformational rearrangements causing DNA hydrolysis. Data from genome-wide sequencing studies have suggested that 'off-target' DNA binding is not always supporting DNA cleavage [225, 246]. To study Cas9/sgRNA DNA binding and cleavage correlation we monitored the enzymatic activity of the CRISPR-Cas9 system on MITOMI. For this, dCas9 variant was exchanged with enzymatically active wild-type Cas9. In addition, the surface functionalization was reverted. Instead of Cas9/gRNA complex immobilized on the surface, a biotinylated dual fluorescently labelled DNA targets were bound. An overview of the experiment procedure is shown in Figure 4.4a. Upon flowing and 2 h incubation of the preformed Cas9/gRNA

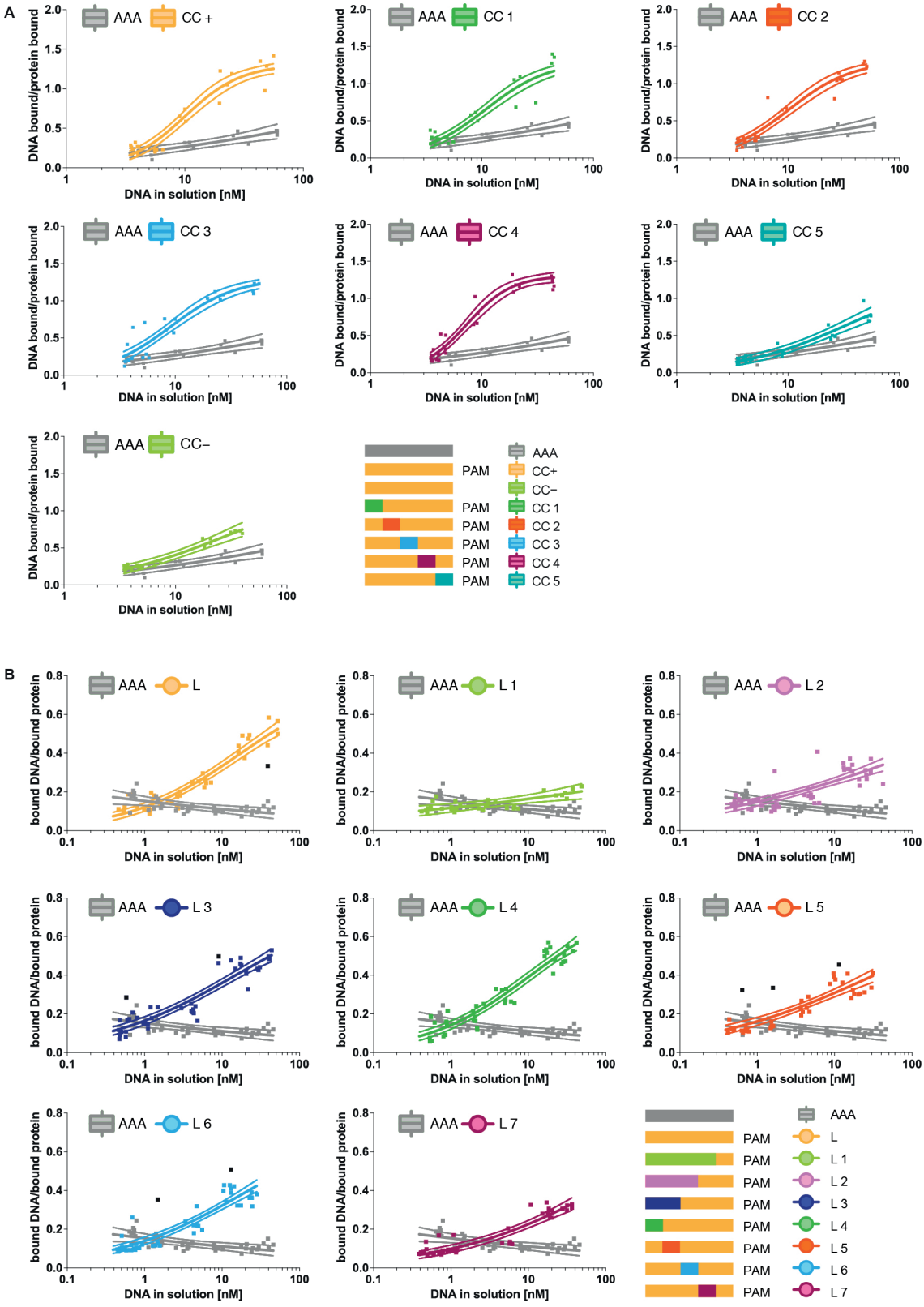

Fig. 4.3 ***In vitro* reconstructed Cas9/sgRNA DNA interrogation on MITOMI]**Binding curves obtained by DNA concentration-dependent measurements of Cas9/gRNA and **(a)** CC DNA library and **(b)** L DNA library.

complex, signal drop in the fluophore intensity of the free (non-biotinylated) DNA strand, an indicator for CRISPR/Cas9 system cleavage activity, was observed. Negligible enzymatic activity was detected when sequences without PAM site or possessing mutation in the region adjacent to the PAM site were used (Figure 4.4b,c). This suggests that stable pairing between the Cas9/sgRNA and the DNA template is required for DNA cleavage. In addition, for those experiments we also made use of guide RNA composed of crRNA and tracrRNA. The results showed similar trends.

4.3 Discussion

Here, we successfully the *in vitro* reconstructed CRISPR-Cas9 system and validated its functionality. By utilizing the MITOMI microfluidic platform we rapidly and in a high-throughput manner probed the binding affinity of two Cas9/sgRNA complexes toward two mutant DNA libraries. The obtained absolute quantitative data allowed us to take a glimpse of their binding profiles. In general, the *in vitro* Cas9/sgRNA system showed high mutation tolerance. The results suggest that the sgRNA sequence and thus conformation structure is strongly affecting the affinity between CRISPR-Cas9 and the DNA targets. In addition, we were also able to study Cas9 cleavage and binding activity toward identical DNA targets for which a high correlation was observed.

Overall, we have shown the promise of MITOMI to be used as a tool for rapid high-throughput biochemical characterization of CRISPR-Cas9 activity, for both binding and cleavage activity. Given the high-content quantitative binding information that can be acquired, the platform can serve as a source for parameters used to model complex genetic regulatory networks improving the CRISPR-Cas9 performance in various synthetic biology applications. In addition, it can aid to the widely used prediction algorithms which mainly rely on qualitative data from poorly controlled *in vivo* studies.Other avenues that can be opened are comparison and characterization of the growing plethora of CRISPR associated proteins, currently 33 different subtypes have been identified. All of them are showing distinct DNA interrogation mechanisms, control of nuclease activity and binding properties, however their integration is slowed down due to the current lack of efficient platform for characterization. The *in vitro* reconstruction allow ease of component recharging of the dCas9/sgRNA system while the MITOMI system flexibility allows multiplex approach for protein variants characterization.

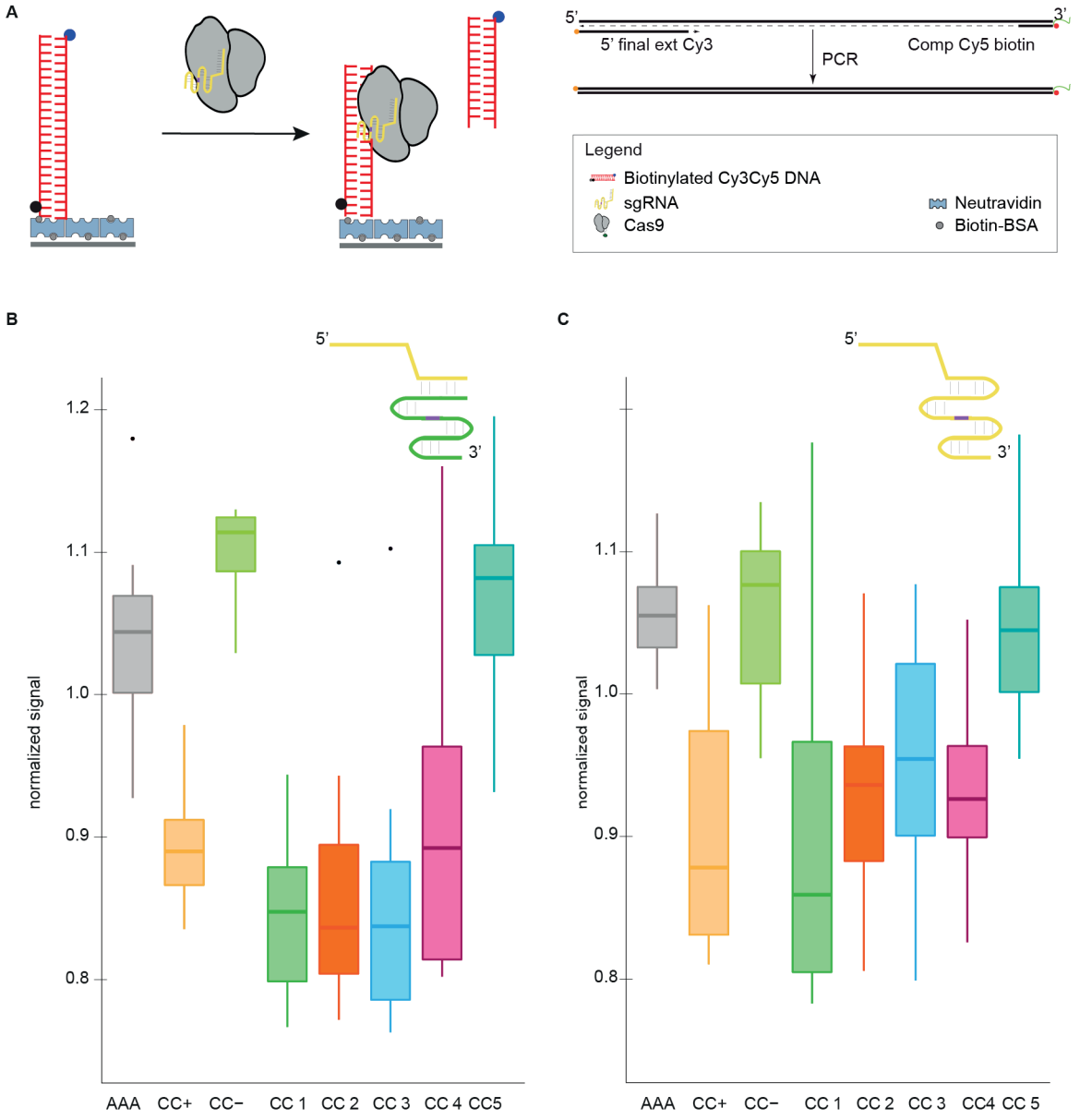

Fig. 4.4 ***In vitro* reconstructed CRISPR-Cas9 cleavage activity on MITOMI. (a)** Outline of the experimental setup and DNA target synthesis. **(b,c)** Boxplots showing the cleavage activity of Cas9 complexed with guide RNA composed of crRNA and tracrRNA **(b)** or sgRNA **(c)**. The normalized signal is representing the ratio between the Cy3/Cy5 ratio before and after CRISPR-Cas9 cleavage. CC DNA templates were used as DNA targets.

4.4 Methods

DNA manipulations

Plasmids pMJ824(pET-T7p_Cas9) and pMJ841(pET-dCas9) were ordered form Addgene. pMJ841 was used as a template for the amplification of the dCas9 gene. The product was subsequently cloned into a previously generated pUC19-T7p_yeGFP plasmid using the Gibson Assembly Master Mix (NEB) according to the manufacturer's instructions. Prior to the Gibson assembly the pUC19-T7p_yeGFP plasmid was linealized, 1 μl DpnI (NEB) was directly added to the PCR reaction which was subsequently incubated for 1 h at 37° to eliminate plasmid carry over. The gibson product was transformed in chemically competent *E.coli* cells (NEB). Upon transformation the 100 μl of the cell inoculums were plated overnight at 37° on 100 μg amplicillin agar plates. Next, 10 μl cloning PCR reactions were performed by using 5 of the grown clones on the agar plates. Taq polymerase was used for the amplification. The clones with proper length was inoculate in LB medium supplemented with 100 μg amplicillin and incubated for 8 h at 37°C. Plasmids were isolated from each inoculated clone and their concentrations were determined with a Nanodrop spectrophotometer. Sanger DNA sequencing revealed that no sequence alterations were introduced during the cloning process.

Double stranded DNA templates coding for the sgRNA molecules were generated via PCR. Equimolar concentrations of a 127 nt long single stranded sgRNA DNA template and sgRNA universal primer 5'AAAAAAAGCACCGACTCG3' were added to a 50 μl PCR reaction, the final concentration of the DNA oligonucleotides was 500 nM. The reaction was carried out at 60°C for 5 cycles to ensure complete primer extension. Subsequently, the product was purified by PCR purification kit (Qiagen) according to manufacturer instructions and stored at -20°C for later use.

The CC and L DNA libraries were prepared either by isothermal Klenow extension by using the Comp Cy5 primer as previously described [21] or by a two-step PCR reaction. The two-step PCR reaction consists of: 1) a PCR reaction in which equimolar concentrations of oligonucleotides encoding the DNA target and universal Comp Cy5 primer were premixed together in a 50 μl reaction to a final concentration of 5 μM, 2) an amplification of the double stranded DNA template, generated during the previous step, by using a 5'-biotinylated primer and the Comp Cy5 primer.

All oligonucleotides used in the study were ordered for IDT. All PCR reactions were performed by Phusion polymerase (NEB) if not otherwise stated.

In vitro synthesis of Cas9 and sgRNA

Both Cas9 and dCas9 were synthesized using the commercial PURExpress kit (NEB) according to manufacturer instructions. The reaction volume was 5 μl, each reaction was supplemented with 0.2 μl RNase inhibitors (Roche) and DNA protein encoding template with 50 ng/μl final concentration. For the Cas9 the *in vitro* transcription and translation was performed on a PCR machine at 37°C for 3 h. In the case of dCas9 the reactions were centrifuged to the bottom of 384-well polystyrene plates (Nunc) and EGFP fluorescence (excitation 485 ± 9 nm, emission 515 ± 9 nm at a sensitivity of 70) was measured every two minutes for 3h by a Biotek SynergyMx plate reader. The quantification of the dCas9 concentration was based on a previously generated GFP calibration curve [154].

sgRNAs were synthesized under sterile conditions by using the MEGAscript T7 kit (Ambion) following the manufacturer instructions. After the RNA expression 1ul Turbo DNAse (Life Technologies) was added directly to the 25 μ.l reactions and incubated for 15 min at 37°C. After the digestion the RNAs were purified by lithium chloride precipitation method. The RNA molecules was resuspended in H2O. The quality of the RNAs was checked by agarose gel electrophoresis and their concentrations were determined spectrophotometrically. Upon this step the samples were store at -20°C for up to two weeks.

In vitro reconstruction of Cas9/sgRNA complex

Bench-top expressed dCas9 and sgRNA were resuspended in binding buffer containing 20 mM Tris pH=7.5, 100 mM KCl, 1 mM DTT, 50 μg/ml heparin, 0.01% Tween, 100 μg/ml BSA, 5% glycerol. The final concentration of the samples was 250 nM and 1000 nM, respectively. The reaction was incubated for 15min at 37C.

For the cleavage assay the binding buffer for Cas9 and sgRNA assembly was modified and contained 20 mM HEPES pH=7.5, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol. The final components concentration as well as incubation conditions were kept the same. In case of tracrRNA and crRNA, the concentrations of the two molecules were 1000 nM each.

Microfluidics device fabrication and epoxy glass slides preparation

MITOMI 768 device fabrication and epoxy glass slides preparation were performed as previously described [21].

DNA arraying

Upon synthesis each DNA target was diluted in 2% BSA in water, allowing visualization during alignment and preventing covalent linkage between the glass slide and the DNA.

Afterwards 2x dilution series was prepared and plated in a 384-well plate (ArrayIT) used for DNA spotting on the glass slides with by MP2 pin. The samples were delivered two times to each spot. Upon DNA arraying the microfluidic device was aligned manually with SMZ1500 (Nikon) stereoscope and bonded overnight at 80°C.

Surface chemistry and MITOMI operation

The device was operated by setting the control lines and the flow line pressure to 15-20 psi and 2.5-3.5 psi, respectively. Initially, the chamber valves remained closed to prevent any liquid from resuspending the spotted DNA samples. The accessible surface area was derivatized by flowing 1) 2 mg/ml biotin-BSA for 20 min followed by PBS washing for 5 min; 2) 0.5 mg/mL Neutravidin (Pierce) resuspended in PSA for 15 min followed by PBS washing for 5 min. Afterwards the 'button' membrane was closed and biotin-BSA solution was flowed for 20 min followed by PBS washing for 5 min. For the binding assays, 6.67 nM solution of biotinylated anti-GFP antibody was flowed for 2 min to prevent gradient formation, the 'button' was opened and the antibody flow was continued for 15 min. After this step the 'button' was closed and the pre-assemble dCas9/sgRNA complex was flowed for 2 min, the 'button' membrane was reopened and the flow was continued for 5min. The device was flushed with PBS for 5min, while the 'button' were closed. Subsequently the chamber valves were opened and the exit valve closed allowing dead volume filling and DNA spot resuspension. Next, the chamber valves were closed and exit valve open and PBS washing was done for 5min. Finally, the sandwich valves were closed, the chamber valves as well as the 'button' was opened allowing DNA diffusion to the detection chamber and DNA Cas9/sgRNA interrogation. The device was place on a hot plate for 3 h at 37°C. After the incubation the chip was scanned by fluorescent microscope (Nikon) to detect the DNA templates in solution. The button and chamber valves were closed and the sandwich valves opened. The chip was washed with PBS for 5 min and imaged again to detect the trapped DNA templates as well the dCas9 signal. For the cleavage assay, instead of anti-GFP antibody flow the PBS was used to resuspended the DNA spots in a way similar to described above. Closing the sandwich valves and opening the 'button' and chamber valves allowed DNA target immobilization via the exposed Neutravidin molecules. After reaching binding equilibrium the 'buttons' were closed and PBS washing was performed for 5 min. An image of the chip was acquired. Afterwards, preassemble Cas9/sgRNA complex was flowed for 5min. The sandwich valves were closed and the button valves open. The chip was incubated for 3 h at 37°C on a hot plate. Finally, the chip was washed with PBS for 5min and addition image was obtained.

Image and data analysis

The microfluidic device was scanned using a fluorescent microarray scanner (ArrayWorx Biochip Reader, Applied Precision, USA). The images were analyzed using microarray image analysis software (GenePix Pro v6.0, Molecular device). Non-linear regression analysis of the data was performed using Prism 5.0 (GraphPad).

Chapter 5

Single Molecule Localisation and Discrimination of DNA-Protein Complexes by Controlled Translocation Through Nanocapillaries

Reprinted with permission from: Roman D. Bulushev*, Sanjin Marion*, Ekaterina Petrova*, Sebastian J. Davis*, Sebastian J. Maerkl, and Aleksandra Radenovic (2016) Nano Lett., 2016, 16 (12), pp 7882–7890 *Equally contributing authors

Using optical tweezers we performed controlled translocation of DNA-protein complexes through nanocapillaries. We used RNA polymerase (RNAP) with two binding sites on a short (7.2 kbp) DNA fragment and a dCas9 protein tailored to have five binding sites on λ-DNA (48.5 kbp). Measured localisation of binding sites showed a shift from the expected positions on the DNA which we explained using both analytical fitting and a stochastic model. From the measured force versus stage curves we extracted the non-equilibrium work done during the translocation of a DNA-protein complex and used it to obtain an estimate of the effective charge of the complex. In combination with conductivity measurements we provided a proof of concept for discrimination between different DNA-protein complexes simultaneous to the localisation of their binding sites.

5.1 Introduction

DNA-protein interactions ubiquitously regulate almost all aspects of cellular function, such as DNA chromosome maintenance, replication, transcriptional regulation, and DNA repair [139]. Most of these interactions occur after complex protein search and binding to a sequence specific DNA target. Because of the broad spectrum of possible interactions it is crucial to gain a better understanding of DNA-protein association, specifically the intricacies of binding.

Over the years, numerous methods have been developed to elucidate the role of DNA-protein interactions in cellular processes and to facilitate the translation of research into biotechnological applications [69]. Eletrophoretic mobility shift assay (EMSA) [67], nuclease footprinting [62], Selex-based approaches [191], protein-binding microarray approaches [32, 132], chromatin immunoprecipitation-based microarray (ChIP), ChiP-seq [13, 101, 159] inhibition of enzymatic degradation are among the most frequent ensemble methods used to reveal sequence-specific protein binding to DNA. While quantitative analysis of thermodynamic and kinetic parameters can be assessed by surface plasmon resonance (SPR) [56], by EMSA, or, more recently, by mechanically induced trapping of molecular interactions (k-MITOMI) [70]. These methods, even though they are successful in probing average interaction characteristics, disregard fine details of DNA-protein complex formation. In addition, most of them employ the use of chemical crosslinking agents, labelling tags and/or complex amplification protocols that can interfere with the proper protein conformation and as a result the proper DNA-protein interaction.

Single-molecule (SM) techniques have emerged to complement these methods and are suitable to characterize rare DNA-protein interactions with high sensitivity and reveal interesting phenomena deriving from the complex mechanisms and inhomogeneous dynamics of DNA-protein interactions.

For example, atomic force microscopy (AFM) [102] and fluorescence resonance energy transfer (FRET) [86] have been used to describe the global structure of DNA-protein complexes and to probe assembly dynamics. Besides localisation, SM techniques such as optical [242] or magnetic [57] tweezers can reveal binding dynamics, elongation rates, and pausing of DNA-protein complexes.

Nanopore-based sensing and force spectroscopy are the latest, label-free, additions to the growing plethora of SM methods used to detect and characterize DNA-protein interactions [30, 90, 175, 181, 199, 204, 203, 215]. In free translocation experiments local analyte characteristics can be directly related to the time-dependent ion conductance. Due to their tunable size, solid-state nanopores are well-suited to detect proteins attached on a long double strand of DNA [17, 205]. Although free translocation experiments allow high throughput

detection, the high speed of translocations and the large distribution of translocation times complicate localisation of protein sites [175, 251]. In addition, nanopore-based sensing is performed in non physiological conditions in a buffer with high salt content ($> 0.4\ M$) allowing for detection only of certain DNA-protein complexes that are capable of sustaining such conditions.

To overcome these hurdles we have recently used a combination of glass nanocapillaries and optical tweezers to detect and characterize different proteins bound to DNA [30]. We were able to elucidate their effective charge and role of the drag force, present due to electroosmotic flow in glass nanocapillaries [122–125]. Compared to free translocations our system allows to isolate and trap single DNA/DNA-protein complexes inside a nanocapillary and translocate them back and forth with a controlled speed in order to obtain biophysical information from force and current data. In addition, experiments in this system are performed in physiological conditions imitating the natural environment for DNA-protein complexes.

This study extends and complements our previous work by focusing on precise protein localisation, charge and size discrimination based on force and current traces, aided by robust analytical and numerical modelling. In this work we made use of two well characterised DNA binding proteins E.coli RNA polymerase (RNAP) and dCas9. RNAP-DNA complex formation is critical for the initiation of gene transcription. On the other hand, dCas9, as a part of the CRISPR-Cas9 complex, has been utilised in various genome engineering studies alongside transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs) [21], etc. One of the biggest advantages of dCas9 is its ease of programmability, dCas9-DNA complexes can be formed at different sites by means of different guide RNA molecules thus allowing us to investigate in details the fine resolution of our method.

Our method recognizes specific or high affinity sites whilst characterising their size and charge and, as any SM force spectroscopy-based technique, could be adapted to monitor DNA-protein interaction dynamics.

5.2 Results

5.2.1 Controlled traslocation of DNA-protein complexes trough nanocapillaries

The set-up used in this work is the same as in our previous study on controlled translocations of DNA-protein complexes [30, 31] and is similar to what has been used with solid state nanopores [106, 107, 204, 224] and nanocapilleries [164, 209]. A simple PDMS fluidic cell allows the separation of two chambers by a glass nanocapillary (schematic in Figure

5.1a). A polystyrene bead, previously coated in DNA-protein complexes, and trapped with optical tweezers, is brought to a distance y from the opening of the nanocapillary. By applying a voltage bias the DNA is driven inside the nanocapillary which is confirmed by a corresponding drop in current and force (Figure 5.1b). Once the DNA is inside the capillary a nanopositioning stage moves the capillary in the opposite direction until the DNA exits. The conversion between stage position z and bead to capillary opening distance y is linear in all our data so we use relative distances in z and y interchangeably.

The controlled translocation of DNA continues smoothly with a change of the stage position z, marked by a flat force profile (Position 1 on Figure 5.1b), until a DNA-protein complex approaches the opening of the capillary (Position 2). Near the opening there is the largest gradient of the electrostatic potential and the largest electroosmotic flow, and thus the largest force on the complex. In thin nanopores, for example solid state nanopores that are close to two dimensional, the potential gradient is well localised. For capillaries it depends on the manufacturing process [123, 30], and causes an electrostatic potential more extended in space of the form $\sim 1/(1+x/\xi)$ with ξ the electrostatic decay length [30, 31] and x growing inside the capillary with the origin at the opening. In our experiments we found $\xi \approx 75$ *nm* by fitting to the tails of the force profiles for bare DNA [31]. When the force exerted by the external potential on the DNA strand is comparable to the force exerted on the DNA-protein complex the protein "jumps" to the other side of the capillary opening (change from position 2 to 3 on Figure 5.1b) [30]. This is identified by a characteristic peak at y_p in both force and current (Position 3 on Figure 5.1b). A nett positively charged complex, as in out case, will jump outside the nanocapillary (positive work) followed by an increase in the length s (coiling) of DNA between the capillary opening and the bead. As the stage position z is increased, the complex stays at the same position until the extension of the DNA returns to the pre-jump level (Position 4 on Figure 5.1b). After the contour length s has become tense again, a change in the stage z also moves the DNA-protein complex. A negatively charged complex first causes a coiling, before jumping when the DNA has accumulated sufficiently resulting in a mirrored jump shape (negative work) [204]. The jump profile in current and force can be used to discriminate bound proteins as is shown later but both can also be used as a tool to localise the binding site.

5.2.2 Localising protein binding sites on DNA

To study localisation of proteins we used dCas9 and RNAP which are known to bind to specific sites along DNA. [41, 148] In contrast to free translocation experiments [181, 203], we carried out our experiments in physiological ionic and pH conditions. All DNA-protein complexes were probed in the same buffer (100 mM KCl, 10 mM HEPES, 5% glycerol, pH

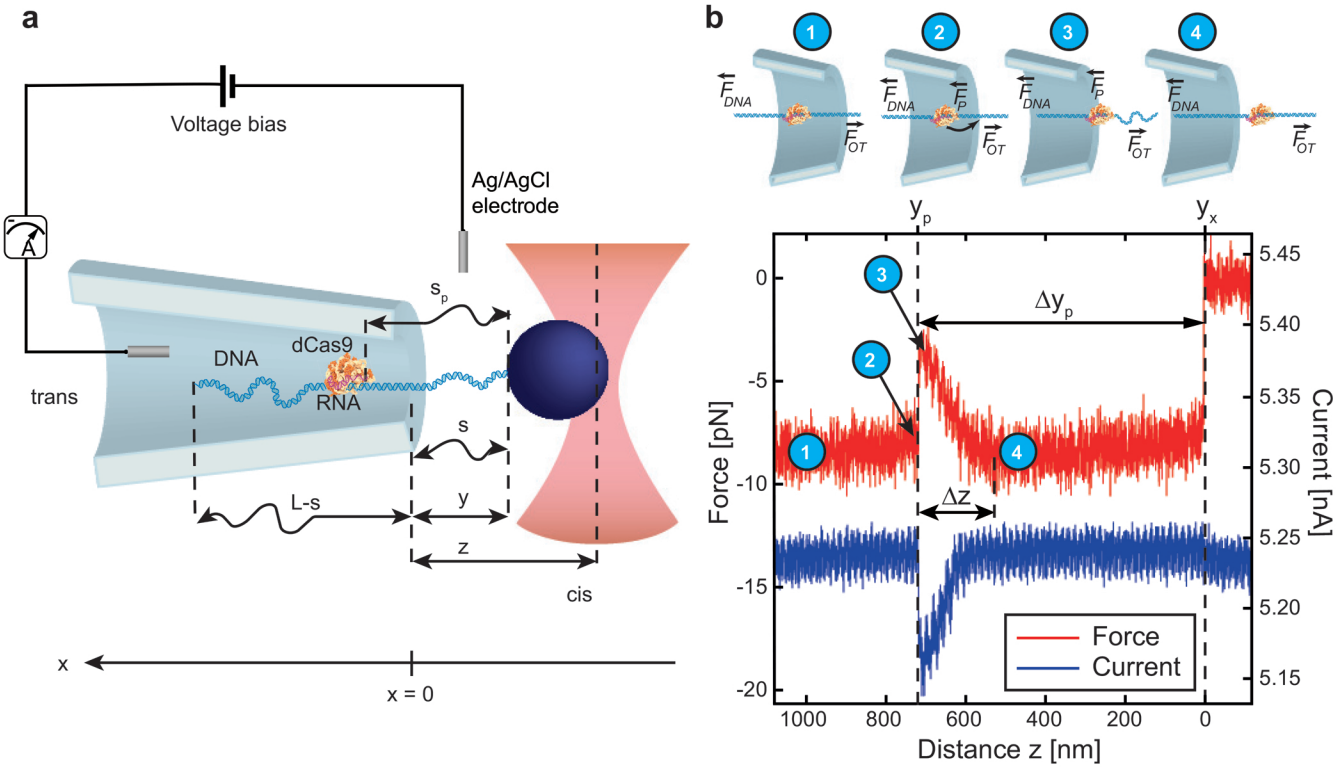

Fig. 5.1 **Set-up and detected signals during controlled translocations of DNA-protein complexes through nanocapillaries (a)** Schematic of the experimental set-up. The distance between the optical trap equilibrium position and the capillary opening (stage distance) is marked as z, the distance between the capillary opening and the surface of the bead is marked as y. The total length of the DNA is L, while s is the length of the DNA contour between the bead and the capillary opening. The bead to protein distance, as measured along the contour, is marked s_p. **(b)** Typical force and current curves are shown as a function of the stage z with the jump location marked as y_p and the DNA exit out of the capillary marked as y_x. The protocol used to move the stage and thus pull out the DNA was a linear increase of z with a speed $v \approx 500\ nm/s$. We measured the relative distance Δy_p between two points y_x and y_p, and the protein jump width is marked as Δz. The force curve is marked at locations corresponding to the individual schematics placed above: 1– before the protein approaches the area of the jump, when the force on the protein becomes comparable to the force on the DNA, 2– location at the moment of the jump when the force on the protein is larger than the force exerted by the DNA on the protein, 3– the moment the protein has finished the jump, and the DNA has coiled, 4– after the jump when the DNA coiling has reduced and the protein is moving with the stage.

7.5) barring the addition of 0.01% TWEEN to RNAP to prevent sticking of the protein to the capillary. Applied voltage bias varied between 150-200 mV to optimize the capture rate during bead approach and signal to noise ratio of the current measurement. In the case of dCas9 we took advantage of the single guide RNA technique to engineer five distinct sites along the 16500 nm (48.5 kbp) λ-DNA, whereas for RNAP we probed two distinct sites along a shorter, 2450 nm (7.2 kbp) DNA (see Methods). In the case of dCas9 we performed experiments both with a single RNA guide present in the mixture as well as with several (two and three). The presence of multiple binding site possibilities did not affect the localisation of each individual site and shows the potential of using our method to perform multiplexed localisation. Indeed in some cases we were able to observe multiple binding events on the same DNA. Measured protein binding locations for all sites are shown in Figure 5.2a and 5.2b as histograms. The binding site position $\Delta y_p = y_x - y_p$ was determined as the distance between the DNA exit out of the capillary (y_x) and the position of the protein jump inside the capillary (y_p) (Figure 5.1b). For each binding site studied we obtained a histogram which included both specific and unspecific events with clustering seen at positions close to the theoretically predicted ones. We eliminated all events which deviated strongly from these clusters as non-specific binding. With larger statistics this method may be able to better assess non specific binding of proteins and determine if any differences are present in size or charge which would hint at variation in binding interactions between specific and non specific binding. Non-specific binding has been seen both in vivo and in vitro [246, 120, 38, 160] and results, for example in the case of dCas9, from mismatching between the RNA guide sequence and the DNA, bulge formations, or transient binding [215]. The obtained localisation has an average standard deviation of 97 nm and 49 nm respectively for dCas9 and RNAP, with a smaller error near the free end of DNA and thus a smaller error for shorter DNAs, as seen with RNAP. We assume the error comes from a wide distribution of drag forces on the DNA strand, the stochastic nature of the process, unspecific interactions not taken into account by our statistics, friction between the DNA and the capillary (if the pulling occurs off axis), and a minor contribution also comes from the different voltages used. To verify that there is no dependence on the translocation protocol we performed a reverse protocol (threading the complex into the capillary) but found no difference in the localisation value which we attribute to a small or negligible hysteresis [204, 30].

Localisation shift and analytical modelling

Although we localised protein binding sites, both thanks to force and current traces (Fig. 5.3a), the positions were shifted in comparison to the expected locations (Fig. 5.2a,b). For positions close to the free end of the DNA strand, the shift is small, or even negligible

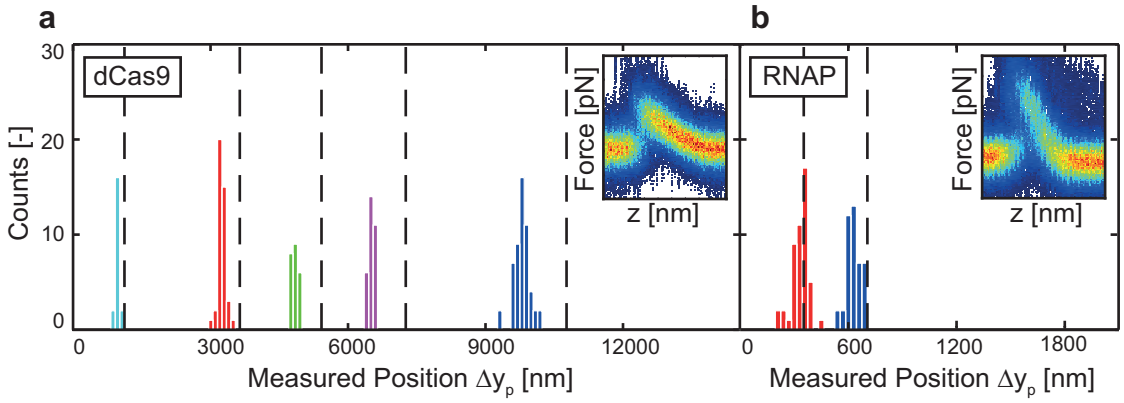

Fig. 5.2 **Localisation of protein binding sites on DNA in physiological conditions (a)** Localisation histograms obtained from force curves for five dCas9 sites along λ-DNA done in 100 mM KCl, 10 mM HEPES, 5% glycerol, pH 7.5. Black, vertical, dashed lines represent expected (theoretical) binding sites at 1100, 3567, 5315, 7118, and 10557 nm from the free end. Different coloured histograms represent different binding sites with only specific events shown. In total 17 different capillaries were used across the five sites and the number of single complexes used is 20, 42, 24, 31, and 54 for each site respectively. **(b)** Localisation of two RNAP binding sites along a 2448 nm long DNA in 100 mM KCl, 10 mM HEPES, 5% glycerol, 0.01% TWEEN, pH 7.5. Black, vertical, dashed lines are theoretical binding sites at 395, and 753 nm from the free end. In total 4 different capillaries were used across the two sites and the number of events is 49 and 43 for each site respectively. Insets in panel a (b) represent the density plot of 60 (122) averaged force curves for dCas9 at site 1100 nm (RNAP at site 800 nm) which were shifted to the same DNA force level and normalized to a probability of 1 for each z. Force curves also include events without an equal DNA base level before and after the jump that were not used for later non-equilibrium work analysis.

in the case of one RNAP binding site, and it grows as the expected position is closer to the tethered end of DNA. Fig. 5.3a shows how the difference between the measured and expected positions for dCas9 and RNAP (inset) $\Delta_{shift} = (L - s_p) - \Delta y_p$ depends on the theoretical location of the complex. This shift can be explained by taking into account that the experimentally measured distance Δy_p is not the same as its position s_p along the DNA contour. At any distance y between the capillary opening and the bead, the DNA is extended to a contour length s such that the extension $\mu(s) = y/s < 1$. The shift Δ_{shift} is due to an early exit of the DNA strand out of the capillary (the DNA strand's exit length s_x is shorter than the full length L of the DNA strand), caused by fluctuations, as well as the DNA being extended by the pulling force. For each contour length s there is a corresponding y such that

$$\Delta y_p = s_x \mu(s_x) - s_p \mu(s_p), \tag{5.1}$$

with s_x the contour length corresponding to the DNA exit location y_x. The effect on the localisation shift is that it makes the apparent protein position smaller than expected.

The extension of the DNA molecule, $\mu(s) = y/s$, which is required to explain the shift, can be obtained analytically in the strong stretching regime [135] which corresponds to the range of forces on DNA measured in typical experiments [31, 232]. It is useful to compare how the results change between nanopores and nanocapillaries, as both can be used for localisation. The force on the DNA can be written as $F_{DNA} = \lambda_{DNA} \Delta V f(s)$ where $f(s)$ is a general functional dependence of the potential on the length of DNA inside the capillary, such that $f(s) = 1$ for nanopores and $f(s) = 1 - \frac{1}{1+\frac{L-s}{\xi}}$ for nanocapilleries [31, 30]. λ_{DNA} is the effective linear charge density of DNA reduced by screening and drag. We assume that the extending force comes from both the force on the DNA and the force exerted by the complex, where, if it is small, we can continue in the strong stretching regime. Approximating that the protein jump is located at the point where the force is largest we can write for the extending force $F_{ext} = F_{DNA} - F_p(s = s_p) = F_{DNA} - q^* V/\xi$. The extension in the high stretching regime is then

$$\mu(s) = \frac{y}{s} = 1 - \frac{1}{\sqrt{4 L_p F_{ext}/k_B T}}, \tag{5.2}$$

with $L_p = 50\ nm$ the persistence length of DNA. Fig. 5.3a shows the analytically obtained fit (from (5.1) and (5.2)) to the shift for dCas9 assuming a nanocapillary geometry with $\xi = 75\ nm$ and $V = 175\ mV$. The fit results in $\lambda_{DNA} = -0.06\ e/bp$, $L - s_x = 200nm$ and the effective DNA-protein complex charge of $q^* = 5e$. The obtained λ_{DNA} is larger than the experimental mean of $-0.04\ e/bp$ (Fig. 5.3d) which we attribute to the strong stretching regime approximations used. Our analytical formulas can predict the shift in the range of

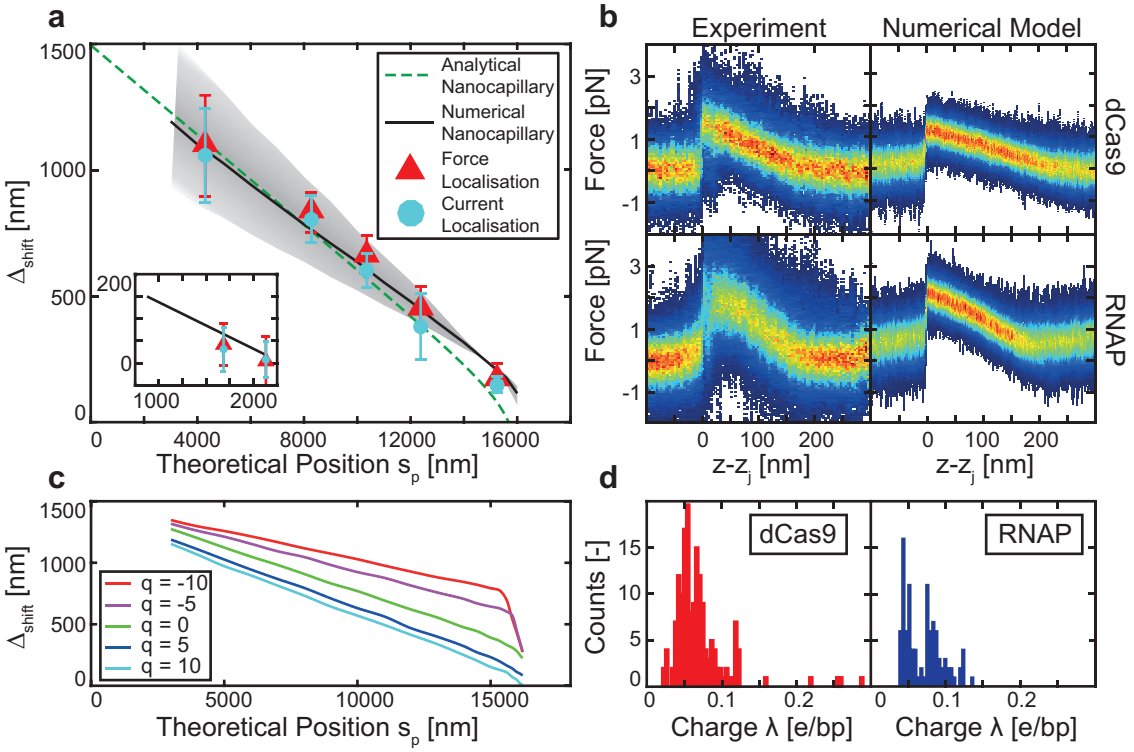

Fig. 5.3 **Protein binding site localisation shift (a)** Difference between measured and expected location $\Delta_{shift} = (L - s_p) - \Delta y_p$ for dCas9 binding sites on λ-DNA. Red triangles correspond to localisation shift obtained from force curves, and blue circles from current curves, all for 5 tailored binding sites on λ-DNA. Green dashed line represents the best fit obtained with the analytically derived nanocapillary shift formula. Full black line represents the numerically obtained fit with $\xi = 75$ nm, while shaded area shows the variation of the numerical fit curve if λ_{DNA} varied from $-0.02\ e/bp$ (upper border) to $-0.08\ e/bp$ (lower border) as in panel d. Inset shows the same for two RNAP binding sites on a short 2448 nm DNA with an analytical fit with the same parameters as in the full panel except $\xi = 10$ nm (most likely due to boundary effects). **(b)** Comparison of 47 and 110 force versus stage position curves plotted as density plots for dCas9 and RNAP respectively (left column) with numerically obtained plots (right column). The sites shown correspond to the same as in the insets of Fig. 5.2 (i.e 1100 nm and 800 nm for dCas9 and RNAP respectively). Before averaging, all curves where shifted so that the jump position is at $z = 0$ and all noisy curves, without a well defined DNA force level before and after the jump, were rejected. **(c)** Plot of localisation shift obtained from the numerical stochastic model using different values of the effective charge of the complex for the same parameters as the numerical fit in panel a. Charge $q = 0$ corresponds to no protein, i.e. bare DNA localisation shift. **(d)** Histograms of measured DNA linear charge density λ_{DNA} for dCas9 and RNAP attributed to changes in the electroosmotic flow induced drag force. The histograms correspond to a total of 10 (2) nanocapillaries used for dCas9 (RNAP) localisation.

small charges, but in order to better fit the shift, and thus accurately predict the localisation, a more precise modelling of how the complex charge affects translocation is necessary.

Numerical modelling scheme

To better explain the influence of the DNA-protein complex charge on the shift, we used a stochastic modelling scheme [93] previously implemented to explain controled translocation events in both nanopores [204] and nanocapillaries [30]. The model is based on two coupled Langevin equations for two state variables of the system, the bead position away from equilibrium at the stage position z, denoted r, and the length of the contour between the bead and the capillary opening s. We solve two Langevin equations with an external force determined by the total free energy of the system $G(r,s)$ which has contributions from the charged DNA worm like chain, optical trap, and a point protein with an effective charge q^* (coming from both electrostatics and electroosmotic flow induced Stokes drag [30]). Using our stochastic model we can obtain numerical fits to the experimentally determined shift. Fig. 5.3a shows the fit obtained with parameters $\lambda_{DNA} = -0.04\ e/bp$, $V = 200$ mV, $\xi = 75$ nm and an effective charge of the complex $q^* = 10e$ for dCas9 and the same except $\lambda_{DNA} = -0.02\ e/bp$ and $q^* = 12e$ for RNAP. λ_{DNA} values in the fit were taken as the means of the experimentally measured ones. We can also predict the general shape of the force curves using our simplified cone geometry (Fig. 5.3b).

Figure 5.3c shows that complex protein charge influences the shift for both positively and negatively charged bound proteins. The shaded area in Figure 5.3a shows the numerical fit variation if the linear charge density of DNA changes from its smallest to its largest experimentally measured value. This can easily account for the large standard deviations seen on the localisation shift as explained by the wide distribution of effective linear charge densities of DNA $\lambda_{DNA} = \lambda^0_{DNA} - \lambda_{DRAG}$ seen in the DNA force levels (see Fig. 5.3d). This is caused by a variability of electroosmotic flow induced drag on the DNA (λ_{DRAG}) in nanocapillaries [123, 122, 125, 124, 30]. In spite of the drag force variability we can account for the experimentally determined localisation shift and thus localise proteins in a comparable/better range as other SM methods, while being conducted in physiological conditions. In practice the electroosmotic flow can be controled by coating capillaries with lipids [85] or polymers [84, 202] thus reducing the experimental error.

5.2.3 Discrimination of DNA-bound proteins using non-equilibrium work and conductivity

In order to simultaneously discriminate DNA-protein complexes whilst localising, two methods, able to detect differences in charge or size, were used. The first is based on non-equilibrium work analysis of the force curves during the jump [30, 204], the second uses the characteristic of nanocapillaries to obtain current versus stage curves and uses them to differentiate proteins using conductivity drops [208].

The stochastic protein jumps being non-equilibrium events, work analysis can be utilized in order to extract equilibrium quantities used for protein discrimination. Our experiment starts in an equilibrium state at time t_A and at stage position z_A and ends in an equilibrium state at time t_B and stage position z_B. The work done by a variation of the control parameter z is then

$$\tilde{W}_i = \int_{t_A}^{t_B} v \frac{\partial H}{\partial z} dt = \int_{z_A}^{z_B} F_{ot}(z) dz, \tag{5.3}$$

with $v = dz/dt \approx 500\ nm/s$ the speed with which the stage is moving the capillary, and H the hamiltonian of the system. This is connected to an integral of the measured optical tweezers force over the stage from the state at z_A to the state at z_B, where the protein event is wholly located in between these two stage positions. The work W_i done in an individual protein jump curve is computed as the area under the peak of the force-distance curve, from the point it starts to change from the base DNA level until the point it returns to this level. This work is corrected by subtracting the average force on the DNA $W_i = \tilde{W}_i - F_{DNA}\Delta z$. The measured works W_i obtained from force curves can be inserted into the non-equilibrium work equation first shown by Jarzysnki [96],

$$\left\langle \exp\left(-\frac{W_i}{k_B T}\right)\right\rangle = \exp\left(-\frac{\Delta G_{AB}}{k_B T}\right), \tag{5.4}$$

and connected to the free energy difference between the state before and after the jump ΔG_{AB}. With knowledge of our set-up geometry and parameters, we can connect free energy differences to physical properties of the DNA-protein complexes. Assuming that the change of the worm like chain free energy before and after the jump is negligible, that the force profile returns to the same value after the jump, and that changes in stage and contour length are approximately equal $y_B - y_A \approx s_B - s_A$, the difference in free energies before and after the jump can be simplified to

$$\Delta G_{AB} = q^*\, V \cdot \frac{\Delta z}{\xi + \Delta z}, \tag{5.5}$$

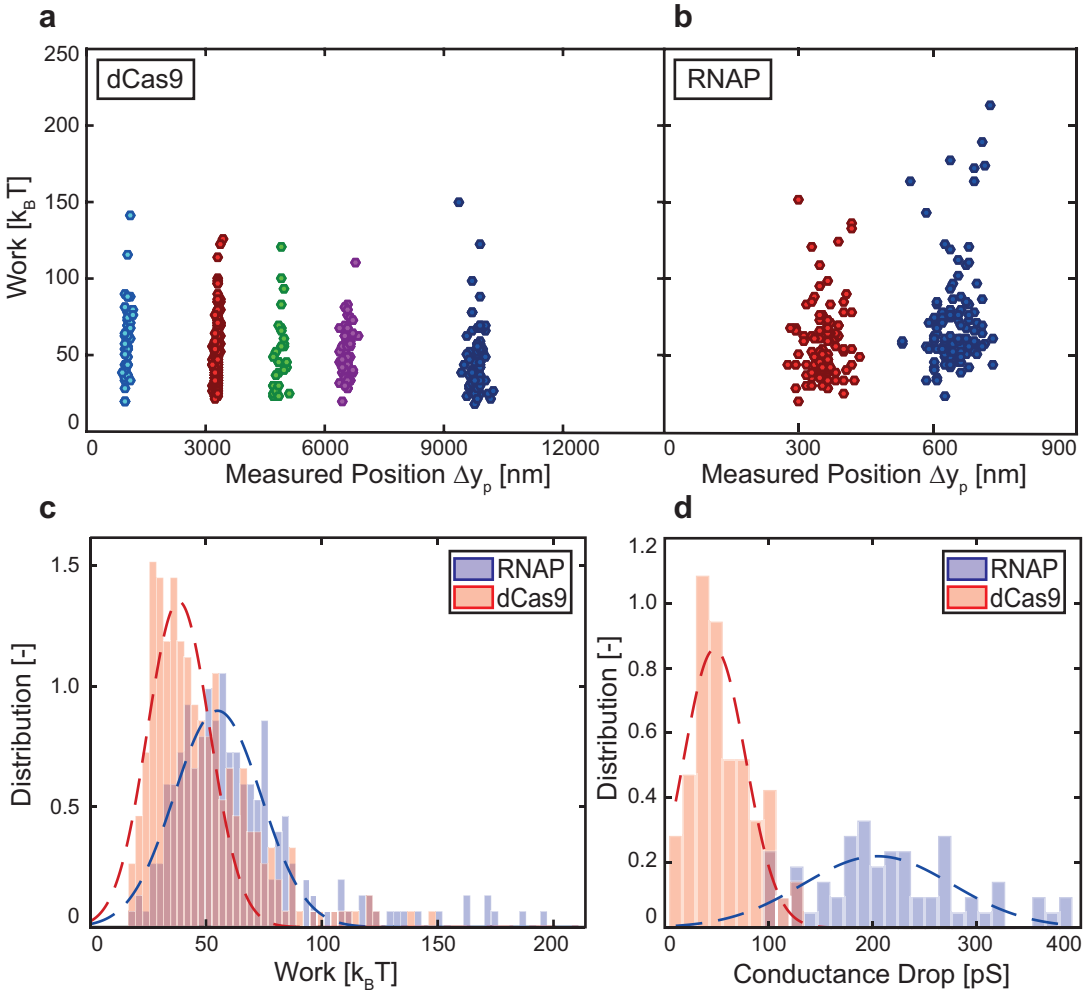

Fig. 5.4 **Analysis of works invested into protein jumps and conductance drops** **(a)** Scatter plots of measured works W_i for five different dCas9 binding sites (as in Figure 5.2a). **(b)** Scatter plots of measured works W_i for two different RNAP binding sites (as in Figure 5.2b). Here the scale of the x-axis is ten times larger than in panel (a). **(c)** Comparison of jump event works obtained for dCas9 and RNAP. All binding site locations have been put together for all measurements done at $V = 200$ mV. Buffer conditions (100 mM KCl, 10 mM HEPES, 5% glycerol, pH 7.5) were the same for all dCas9 sites and with added 0.01% TWEEN for RNAP. From these distributions ΔG_{AB} values were computed using the Jarzysnki equality to be 31.7 and 37.9 k_BT for dCas9 and RNAP respectively. **(d)** Comparison of conductance changes for dCas9 and RNAP obtained from the maximal change of the conductance during a protein translocation relative to the DNA level.

where q^* is the effective charge of the complex, V the voltage at which the experiment was carried out, and Δz the width of the jump event.

Using the Jarzynski equality we can determine the effective charges from our non-equilibrium measurements for discrimination. Work analysis for both dCas9 and RNAP results in a wide clustering (Figure 5.4a and b) due to the non-equilibrium and stochastic nature of the translocation events as well as a wide distribution of electroosmotic flow induced drag. Assuming that there is no difference between binding sites, we can group all the work values obtained at the same voltage for both dCas9 and RNAP (Figure 5.4c). Although similar, it is possible to discern two different DNA-protein complex distributions. In order to quantify the difference, we can apply the Jarzynski equality (5.4) to the obtained distributions. Using the free energy relation (5.5) with a determined event width $\Delta z = 220$ and 170 nm, for dCas9 and RNAP respectively, and an electrostatic decay length $\xi = 75$ nm it is possible to obtain a value for the effective charge of the complex. For dCas9 we obtain $q^*_{Cas9} = 3.1e$ and for RNAP $q^*_{RNAP} = 3.4e$. First we must note the difference between these charge values ($\sim 3e$) and the charge values obtained in the stochastic fit to the localization shift ($\sim 10e$). As we have a wide distribution of drag forces, most likely due to different capillaries used, and thus different flow velocity distributions, we will have a distribution of effective charges q^* due to drag going into the work analysis. Because the Jarzynski equality is biased towards small works due to the factor e^{-W_i/k_BT}, we will preferentially obtain contributions from smaller charge values, i.e. works, and the distribution will be skewed to show the lower bound for the complex effective charge. Both RNAP and dCas9 (with GFP attached, see Methods) having similar hydrodynamic radii (~ 5 nm [99, 188]) we expect them to feel a similar drag force which would explain the similarity of the determined effective charge. The determined effective charges will also have a contribution from the bare electrostatic charge, but we expect a minor contribution as drag overcomes electrostatic charge [30]. In spite of the similarities of dCas9 and RNAP, the distributions of work values and effective charge from the Jarzynski equality support discrimination.

To complement the small difference in effective charges of RNAP and dCas9, we can additionally discriminate them by comparing the current drops obtained simultaneously with force peaks for both proteins (Figure 5.4d). In a buffer of same ionic strength, RNAP exibits a larger current drop than dCas9 and enables a clearer discrimination between the proteins than the non-equilibrium work analysis. The origin of the conductance drop is known to depend strongly on salt conditions [201] and can be attributed to several different scenarios (current blockage [111], molecular friction [105], or a change in the advective current from current polarization [30]). A combination of conductance changes and non-equilibrium work analysis thus enables us to discriminate two proteins even if they have a similar charge

or shape. In practice one could use either one, the other, or both methods to discriminate proteins while tailoring the sensitivity to either charge or size.

5.3 Discussion

We have demonstrated the viability of single molecule force spectroscopy using optical tweezers and glass nanocapillaries as a sensing tool for discrimination and localisation of DNA-protein complexes on two proteins: RNAP and dCas9. We detected the proteins shifted from their expected positions but were able to explain and correct for it by using both analytical and numerical modelling and localize them to within an accuracy of 50 nm.

In the context of the applicability of this technique, we note the nanopore limit of (5.5). For thin nanopores ($< 5nm$), like graphene [3, 48, 228, 233] and MoS_2 [127], the free energy difference for a jump simplifies, under the same assumptions, to $\Delta F = q\,V$, with q as the bare electrostatic charge, making the analysis of experiments much simpler as the dependence on geometry is negligible. However this comes at the cost of a worse signal to noise ratio for current measurements and harder experiments since combining optical tweezers with nanopores is significantly more complex than with nanocapillaries. In addition nanocapillaries are more sensitive to the size of the proteins due to the electroosmotic flow induced drag that can be reduced and controlled with lipid or polymer surface coating [84, 85, 202], thus allowing this technique to be tailored to be either sensitive to charge or size of the proteins, enabling discrimination whilst simultaneously localising them.

We believe that further implementation of optical tweezers combined with nanocapillaries is robust enough to precisely measure binding sites of previously uncharacterised proteins, which make up a large part of proteins known to bind to DNA, whilst simultaneously inferring their apparent hydrodynamic radius and charge. It could also study differences between different binding conformations, for example, binding of RNAP in the initiation stage versus the elongation stage, or Cas9 versus dCas9. Not only does our method recognize specific or high affinity sites whilst characterising their size and charge, but, with enough statistics, could be used to infer about low affinity or non-specific binding sites. The use of the technique to detect multiple proteins bound to a single DNA (multiplexed detection) could also be put to use in detecting the presence of tandem repeats, such as the expansion of a CGG repeat occuring in Fragile X syndrome [60].

5.4 Methods

Combination of nanocapillaries and optical tweezers

Quartz capillaries with a 0.4 mm outer diameter and 0.3 mm inner diameter (Hilgenberg) were pulled using a P-2000 laser-assisted puller (Sutter Instrument) with a program containing 2 lines (Table 5.1). The pulled capillaries had a diameter of 200-300 nm. They were shrunk to diameters of 43-58 nm under SEM accordingly to [210] using a 500 pA beam current and 3 kV accelerating voltage. Only nanocapillaries with a circular, symmetric opening were used in this work. The shrunken nanocapillaries were embedded in a PDMS fluidic cell consisting of two chambers as in [30, 31] and treated by oxygen plasma at $\approx$ 50 W for 2 minutes. Afterwards they were filled with a buffer, containing 100 mM KCl, 10 mM HEPES, pH 7.5, 5 % glycerol filtered through an Anotop 25 filter (Watman) beforehand. The buffer used for experiments with RNAP also contained 0.01 % TWEEN to reduce proteins sticking to the glass walls. The capillaries were used in the setup combining optical tweezers and nanocapillaries described in [30, 31]. The stiffness of the optical trap estimated using the power spectrum density method was in the range of 60 -120 pN/μm [Neuman and Block].

Line	Heat	Filament	Velocity	Delay	Pull
1	600	4	10	145	0
2	600	4	10	145	140

Table 5.1 **Laser puller program.** A program used to pull capillaries to diameters of 200-300 nm.

Formation of bead-DNA-protein complexes

In order to obtain bead-DNA-protein complexes first DNA-protein complexes were formed and then they were incubated for 15 min at 37°C, 250 rpm with 10 μL ($6.7 \cdot 10^6$) 3 μm streptavidin-coated polystyrene beads (Bangs Laboratories) suspended in the experimental buffer. Afterwards the beads were centrifuged for 5 min, 4°C, 1000 g and the solution was exchanged to the experimental buffer. The formed complexes were stored on ice during experiments. For dCas9, fresh complexes were prepared every day, whereas for RNAP the same complexes could be used for up to two days due to their stability. The experiments were performed at room temperature. After 45-60 min the buffer in the cis chamber was exchanged to a new one and new beads-DNA-protein complexes were introduced.

In the case of dCas9, first, 140 nM of guide RNA was mixed with $\approx$ 7 nM of dCas9 in the buffer containing 20mM HEPES, 100mM KCl, 5mM $MgCl_2$, 1mM DTT, 5% glycerol and in the presence of 0.5 Units of RNase inhibitors (Roche) for 15 min at 37°C, 250 rpm.

To form a complex of RNAP on DNA a T7A1 promoter was used. On this promoter, in the absence of UTP in the solution, RNAP stalls in solution after 20 bp [118]. A 7.3 kb ds DNA fragment containing a biotin tag on one end was obtained by PCR of the plasmid pRL574 kindly provided by Landick's lab. 50 nM RNAP (NEB) was mixed with $\approx$ 500 pM of DNA with a single T7A1 promoter, 500 μM ApU (TriLink), 100 μM of ATP, CTP and GTP (Invitrogen) in a buffer, containing 25 mM Tris/HCl, pH 8.0, 100 mM KCl, 4 mM MgCl2, 1 mM DTT, 3 % glycerol, 0.15 mg/ml BSA for 30 min at 37°, 250 rpm.

Preparation of dCas9

A DNA encoding the catalytically inactive dCas9 (Cas9(D10A/H840A)) version from S. *pyogenes* was PCR amplified from the pMJ841 plasmid (Addgene). The resulting PCR product was fused at the 3' end with yeGFP construct and cloned by Gibson assembly into a custom T7 RNAP expression vector whose sequence was verified by DNA sequencing (Microsynth). The generated vector was used for the in vitro synthesis of dCas9 yeGFP tagged protein by PURExpress In Vitro Protein Synthesis Kit (NEB) according to the manufacturer's conditions. The protein expression levels were monitored over time using a spectrophotometer (SynergyMx, Biotek) and the final dCas9 yeGFP concentration was quantified based on the GFP calibration curve.

Design and preparation of single guide RNAs(sgRNAs)

λ-DNA was screened for the presence of PAM motifs (5'X20NGG3') separated by more than 5000 bp. Six sgRNAs (single guide RNAs) were designed bearing complementarity to the 20 bp 5' adjacent PAM motif sites. sgRNAs secondary structure predictions were performed by Mfold online software. sgRNAs were prepared by in vitro transcription of dsDNA templates carrying a T7 promoter sequence. Transcription templates were generated by PCR amplification of ssDNA templates (IDT)containing the T7 binding site, 20 bp sequence complementary to the DNA target site and sgRNA scaffold sequence. The sgRNAs were synthetized by in vitro transcription using the MEGAshortscript T7 Transcription kit (Ambion) according to the manufacturer's conditions. The sgRNAs were treated with Turbo DNAse (Ambion), purified by lithium chloride precipitation method and resuspended in nuclease-free water supplemented with RNAse inhibitors (Promega). The quality of the sgRNA was assessed prior to use by agarose gel electrophoresis.

Stochastic model

We used a stochastic modelling scheme [93] previously implemented to explain controled translocation events in both nanopores [204] and nanocapillaries [30]. The model is based on two coupled Langevin equations for two state variables in the system, the bead position $r = y + \rho$ away from equilibrium at the stage position z with $\rho = 1.5\ \mu m$ the bead radius, and the length of the contour located between the bead and the capillary opening s. We solve two Langevin equations with an external force determined by the total free energy of the system $G(r,s)$. The free energy contains contributions from the optical trap , the free energy of a charged DNA molecule in the nanocapillary $G_{DNA}(s)$, the free energy contribution from a protein bound to the DNA at position s_p $G_p(s)$, and the worm like chain (WLC) free energy of DNA $G_{wlc}(r,s)$ which couples the equations. We numerically solve the two coupled equations, while slowly varying the stage position z with a speed $v \approx 500\ nm/s$ from the DNA being almost completely inside the capillary until it exits. In order to determine any numerical parameter, we make ~ 100 averages of pulling protocols with random starting conditions.

Chapter 6

A Microfluidic Biodisplay

Reprinted with permission from: Francesca Volpetti*, Ekaterina Petrova* and Sebastian J. Maerkl (2017) ACS Synth. Biol., 2017, 6 (11), pp 1979–1987 *Equally contributing authors

Synthetically engineered cells are powerful and potentially useful biosensors, but it remains difficult to deploy such systems due to practical difficulties and biosafety concerns. To overcome these hurdles, we developed a microfluidic device that serves as an interface between an engineered cellular system, environment, and user. We created a biodisplay consisting of 768 individually programmable biopixels and demonstrated that it can perform multiplexed, continuous environmental monitoring. The biodisplay detected 10 µg/l sodium arsenite in tap water using a research grade fluorescent microscope, and reported arsenic contamination down to 20 µg/l with an easy to interpret "skull and crossbones" symbol detectable with a low-cost USB microscope or by eye. The biodisplay was designed to prevent release of chemical or biological material to avoid environmental contamination. The microfluidic biodisplay thus provides a practical solution for the deployment and application of engineered cellular systems.

6.1 Introduction

With the advent of synthetic biology, a number of biological sensors have been engineered capable of monitoring the environment [20, 227]. Although such synthetic biological systems are in principle powerful sensors and information processing units, they often lack direct applicability because suitable interfaces between the environment, the user and the engineered

biological system do not exist [77, 219]. The lack of such interfaces results in problems related to safely deploying genetically modified organisms (GMOs), while making it possible for them to interact with the environment and user [244]. Aside from safety concerns, interfaces can solve practical problems such as keeping the engineered biological system alive for extended periods of time, automatic and frequent sampling of the environment, and facilitating readout.

Microfluidics originated as a tool to enable analytical measurements in chemistry and biology [234]. In the last decades, microfluidic devices have found a plethora of applications in biology [234] spanning high-throughput screening [131], cell-based assays [18, 157], and molecular diagnostics [63]. The use of microfluidic devices increases throughput, reduces cost, and can enable novel measurements [64]. In most instances the purpose of microfluidics is to enable or conduct analytical measurements of molecules or cells. Few examples depart from this dogmatic application of microfluidics and instead employ microfluidic devices as soft robots capable of movement and camouflage [147], or as microfluidic games [187]. One recent example demonstrated how engineered biological systems can be deployed by combining bacterial sensors with passive microfluidic channels in a wearable hydrogel-elastomer hybrid device capable of sensing and reporting the presence of diacetylphlorogluicinol (DAPG), isopropyl beta-D-thiogalactopyranoside (IPTG), acyl homoserine lactones (AHL) and rhamnose [127].

We previously developed a method to culture and interrogate 1152 S. *cerevisiae* strains on a microfluidic device. Each strain was cultured in a dedicated microchemostat and was interrogated optically to provide single-cell, phenotypic information on each strain. We applied this platform to the comprehensive, single-cell analysis of yeast proteome dynamics [46] and a detailed characterization and modeling of transcriptional regulation in yeast [182]. Here we show that this approach can be used to generate a bacterial biodisplay. Unlike previous bacterial biodisplays [177], each of the 768 pixels can be independently programmed with a specific bacterial strain or clone. We used the biodisplay to characterize the response of several engineered bacterial strains to small molecule analytes for extended periods of time. To demonstrate environmental monitoring, we programmed the biodisplay with a "skull and crossbones" symbol using a bacterial arsenic sensor strain. The symbol is displayed when arsenic contaminated water is sensed by the biodisplay. We also demonstrated a multiplexed biodisplay that reports on the presence or absence of two small molecules: arabinose and arsenic. The use of an easy-to interpret symbolic display drastically simplifies analysis, and is readily understood by a layman without any special hardware or analysis requirements. To enable field-deployment of our biodisplay we showed that readout can be conducted using an off-the-shelf low-cost USB microscope, a cellphone, or direct visual inspection. Finally,

to increase shelf life and facilitate shipping/deployment of the biodisplay we programmed the device with *B. subtilis* spores, which allowed storage of the display at 80°C for 1 month without any detectable decrease in viability or sensing functionality.

6.2 Results

6.2.1 Biodisplay Programming, Culturing, Sampling and Readout

The biodisplay has a resolution of 48 × 16 for a total of 768 programmable pixels with a density of 64 pixels per inch (Figure 6.1a). Each pixel consists of a 250 μm square cell chamber, two side channels for medium and sample introduction, sandwich valves to isolate adjacent pixels and a neck valve to isolate the pixel from the side channels. Each pixel of the biodisplay can be programmed with a different bacterial strain. Standard microarray spotting was used to array bacterial cells or spores on an epoxy-coated glass slide. *E. coli* was spotted in medium containing 10% glycerol to preserve cell viability, as we observed that dried *E. coli* spots failed to regrow after spotting. *B. subtilis* spores could be spotted in water without addition of glycerol. The bacterial array was then aligned to the microfluidic chip and bonded at 37°C for 1 h.

After assembling and bonding the biodisplay, cells were allowed to recover overnight using an automated culturing routine (Figure 6.1b). A medium reservoir was attached to the device and the waste outlet was connected to a receptacle containing 70% aq. EtOH to immediately sterilize and contain the device outflow. First, the entire device was loaded with medium by out gas priming [80]. Once the device was fully primed with medium, cells were cultured with the following three-step culturing routine: (i) flowing medium through the side channels (10 min), (ii) allowing medium to diffuse into the biopixels (45 min) and (iii) flowing lysis buffer through the side channels (10 min). Whenever fluid was flowed through the side channels, the biopixel chambers were sealed off from the side channels using a set of valves. In step (ii) the biopixel valves were opened to allow diffusion of molecules in and out of the biopixel chamber. During this step, individual pixels were isolated from one another by closed sandwich valves. The lysis segment of the culturing routine prevented biofilm and microcolony formation outside of the biopixels [10]. By actively preventing accumulation and growth of cells in areas other than the biopixels we avoided device clogging and allowed experiments to proceed for several days; the longest experiment conducted here lasted 1 week. Culturing was performed at 37 °C using a thermo glass plate (Okolab). Culturing at 37 °C increased growth rate, leading to faster accumulation of biomass, but was otherwise not required and incubation at room temperature was also possible. After overnight

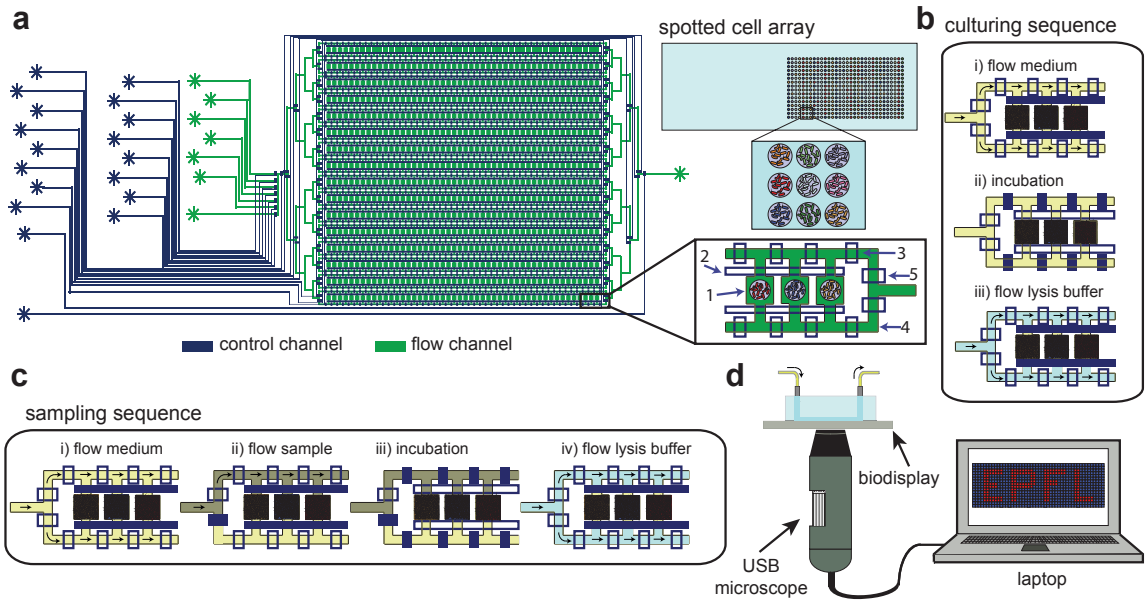

Fig. 6.1 **Biodisplay schematic and use.** **(a)** Design of the two-layer microfluidic device (control channels in blue and flow channels in green). The device is composed of 16 rows and 48 columns for a total of 768 biopixels. Each pixel consists of a chamber (1), chamber valve (2), sandwich valves (3) and side channels (4). The side channels in turn can be specifically addressed using the side channel valves (5). A bacterial cell array is spotted on a glass slide and aligned with the PDMS chip. **(b)** The culturing sequence consists of 3 steps: (i) flowing medium with the chambers closed (10 min), (ii) incubation with the sandwich valves closed and chamber valves open (45 min), and (iii) flowing lysis buffer with the chamber closed (10 min). **(c)** The sampling sequence consists of 4 steps: (i) flowing the medium in both side channels (10 min), (ii) flowing the sample in one side channel (10 min), (iii) incubation with both medium and sample (45 min), and (iv) flowing lysis buffer (10 min). The sequence is than repeated continuously for the extent of the experiment. **(d)** The biodisplay can be imaged using a standard research grade epi-fluorescent microscope, a USB fluorescent microscope, a cellphone camera, or by eye.

culturing, each bacterial spot had outgrown, and populated its dedicated biopixel. We did not observe any contamination resulting from microarraying. Cells occasionally populated empty biopixels during culturing, which was avoided by spotting a non fluorescent *E. coli* strain in otherwise empty biopixels.

Once the biodisplay was online, we switched from the culturing routine to a sampling routine that enabled frequent testing of a liquid sample (Figure 6.1c). The sampling routine was identical to the culturing routine, with an additional step added prior to diffusive mixing of the pixel and side channels. After flowing medium through both side channels, the sample was flowed through one of the side channels. During the diffusion step, both medium and sample were allowed to diffuse into the biopixel chambers. This process allowed us to aspirate a water sample directly, rather than having to premix it manually with the medium solution. By using a dedicated sampling port, it was possible to automatically sample a source for extended periods of time in 75 min intervals without user interference.

We employed a standard research grade epi-fluorescent microscope (Nikon) and a hand-held, low-cost USB fluorescent microscope (Dino-Lite) for device readout and quantitation (Figure 6.1d). The standard epi-fluorescent microscope equipped with a motorized stage was programmed to image each biopixel individually whereas the hand-held USB microscope could image either the entire display at once or acquire several subsections. Both approaches were used to acquire time-course and end point measurements. We also demonstrated that the biodisplay could be observed directly by eye using a LED flashlight and an emission filter, and images could be acquired using a standard cellphone camera.

6.2.2 Multiplexed characterization of bacterial strains

We tested the biodisplay device by culturing and characterizing nine *E. coli* strains engineered to induce GFP or RFP expression in response to arabinose (Table S 6.1). These strains were obtained from the iGEM registry of standard biological parts and were generated by various student teams. The fluorescent protein genes were under the regulation of a pBAD promoter or derivatives thereof. We programmed the biodisplay with the nine strains, each strain filling 2 columns of the device for a total of 32 pixels per strain. Strains were induced with arabinose after overnight culturing at 37°C. Consequent culturing and induction were performed at room temperature (23-24°C) (Figure 6.2a) or 37°C (Figure S6.5). Room temperature induction was used in order to assess the performance of the biodisplay and cell-based sensors at ambient temperature in order to reduce platform complexity by eliminating the need for a temperature control element.

We designed the biodisplay device so that it was possible to specifically address four sets of four rows with a different solution. We used this modality to characterize each strain

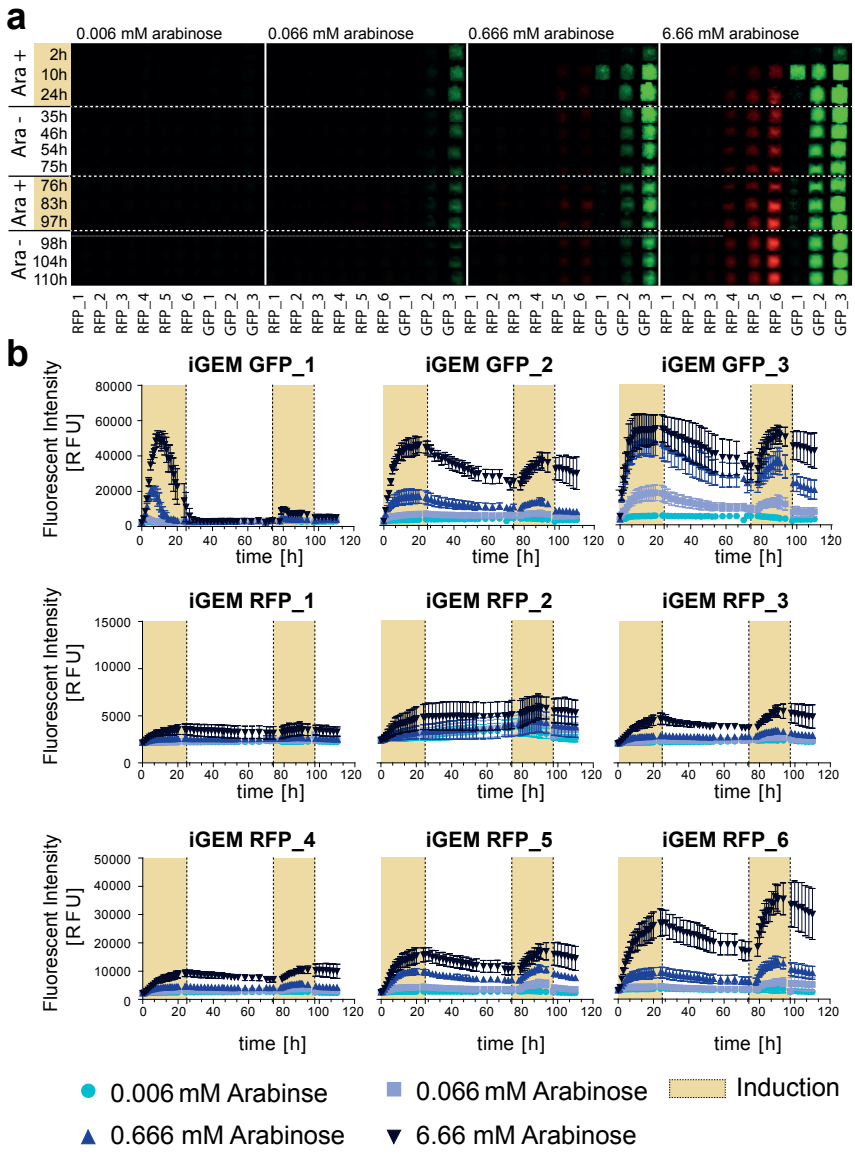

Fig. 6.2 **Using the biodisplay for strain characterization.** **(a)** Fluorescence kymograms of nine arabinose-sensing *E. coli* strains. Shown is a representative biopixel for each strain and condition tested. The strains were induced twice for 24 h with 0.006 mM, 0.066 mM, 0.666 mM and 6.66 mM arabinose. **(b)** Quantitative analysis of fluorescence pixel intensities for each strain under four arabinose concentrations. The yellow background denotes the two induction periods. Each data point represents the average of eight pixels with error bars representing to the std. deviation of the means.

against four arabinose concentrations: 0.006 mM, 0.066 mM, 0.666 mM, 6.66 mM (0.0001%, 0.001%, 0.01% and 0.1% w/v) arabinose in lysogeny broth (LB) medium. Altogether, we tested 36 strain – inducer combinations by measuring nine strains in response to four arabinose concentrations with eight technical repeats for each combination. The biodisplay was placed on an automated research-grade fluorescent microscope and imaged over time.

We performed two consecutive 24 hour inductions, with the second induction starting 51 hours after the end of the first (Figure 6.2). Upon induction we observed a measurable signal in all strains. All three of the GFP strains showed a strong response to 6.66 mM arabinose while only strain #3 induced highly in response to 0.666 mM. Strains #1 and #2 induced to low levels in response to 0.666 mM and failed to induce to appreciable levels at lower arabinose concentrations, whereas strain #3 also induced in the presence of 0.066 mM arabinose. These results indicate that strains #2 and #3 could be used in combination to distinguish three different inducer concentrations ranging over 3 orders of magnitude. All RFP strains induced expression to differing degrees in response to 6.66 mM arabinose and in the cases of RFP strains #2-6 also to 0.666 mM arabinose. All strains, with the exception of strain GFP_1 slowly decreased in intensity during the 51 hour de-induction period. Strain GFP_1 exhibited a transient expression profile with signal beginning to decrease during the induction phase. GFP_1 is also the only strain that failed to induce during the second induction when cultured at room temperature but induced well in the second induction when cultured at 37ºC (Figure S6.5). The transient expression profile can be attributed to the DH5-alpha background used and the presence of a LVA protease tag which is not present in any of the other strains tested.

Using the same approach we tested nine engineered arsenic responsive bacterial strains [142, 231]. We chose arsenic responsive strains for use in a biodisplay for low-cost, continuous environmental arsenic monitoring. After overnight growth, we flowed LB medium supplemented with 0 µg/l, 10 µg/l, 100 µg/l and 500 µg/l sodium-arsenite for 24 hours and quantitated GFP expression over time. The experiment was performed at room temperature. Strain EDI_as_4 in particular showed high expression of GFP and detected sodium-arsenite down to 10 µg/l (Figure S6.6). We thus chose this strain for all further arsenic sensing applications.

This series of experiments demonstrated that our biodisplay is a viable high-throughput platform for long-term culturing and characterization of bacterial strains. Theoretically, up to 768 strains can be characterized in parallel on this platform under dynamically changing media conditions [46], which is otherwise difficult to achieve using existing standard microplate batch cultures or low-throughput microfluidic devices.

6.2.3 Biodisplay

High levels of arsenic in groundwater are common in a number of countries including China, India, and the USA. Not only is arsenic a concern in water supplies, high arsenic levels have also been found in a number of food sources [128], particularly rice [87, 141]. In Bangladesh, an estimated 20–45 million people are affected by arsenic levels that exceed the national standard of 50 µg/L and the WHO/EPA standard of 10 µg/L. Worldwide it is estimated that over 200 million people are exposed to unsafe arsenic levels of 10 µg/L or above [150]. Arsenic also tops the ATSDR substance priority list, which lists those substances "determined to pose the most significant and potential threat to human health". Our biodisplay could serve as a continuous environmental water monitor that is low-cost, deployable in resource limited settings, and easy to interpret.

We programmed our biodisplay with an arsenic sensing *E. coli* strain (EDI_as_4) spelling "As", the elemental symbol for arsenic, and a "skull and crossbones" symbol commonly used to warn of toxic substances (Figure 6.3a). All pixels not programmed with the arsenic sensing strain were programmed with a nonfluorescent strain to prevent invasion and cross-contamination by the arsenic sensing strain. After overnight growth, sodium-arsenite in tap water was sampled every 75 min by the device using the sampling routine described above. The biodisplay was kept at room temperature for the entire duration of the sampling routine. Using the research grade fluorescent microscope for readout, the biodisplay detected as little as 10 µg/L sodium-arsenite, and 20 µg/L sodium-arsenite was detected with the hand-held low-cost fluorescent USB microscope (Figure S6.7). Fluorescent signal and the appearance of the "As" and "skull and crossbones" symbols was detected after 10 h by the research grade microscope (Figure S6.8). After 24 h a strong fluorescent signal had developed that could be easily imaged with a low-cost hand-held USB microscope. A negative control biodisplay, which was run under identical conditions but exposed to tap water without added arsenic showed no detectable fluorescent signal. We also performed an experiment demonstrating that the biodisplay is able to sense arsenic containing water after 1 day of continuous operation. After overnight growth, arsenic-free tap water was sensed by the biodisplay for 24 h. After the 24 h period, the arsenic-free water source was replaced with water containing 50 µg/L sodium-arsenite and an increase in fluorescent signal was observed over time (Figure S6.9).

Given the large number of programmable pixels and high pixel density on our biodisplay we explored the possibility of generating a biodisplay that could sense and report the presence or absence of multiple substances in a sample. We selected the same arsenic sensing strain as above and the iGEM GFP_3 strain for arabinose sensing. The arsenic sensing strain was spotted in the form of a "skull and crossbones" pattern whereas the arabinose sensing strain was spotted in a "smiley" pattern. We performed four separate experiments exposing

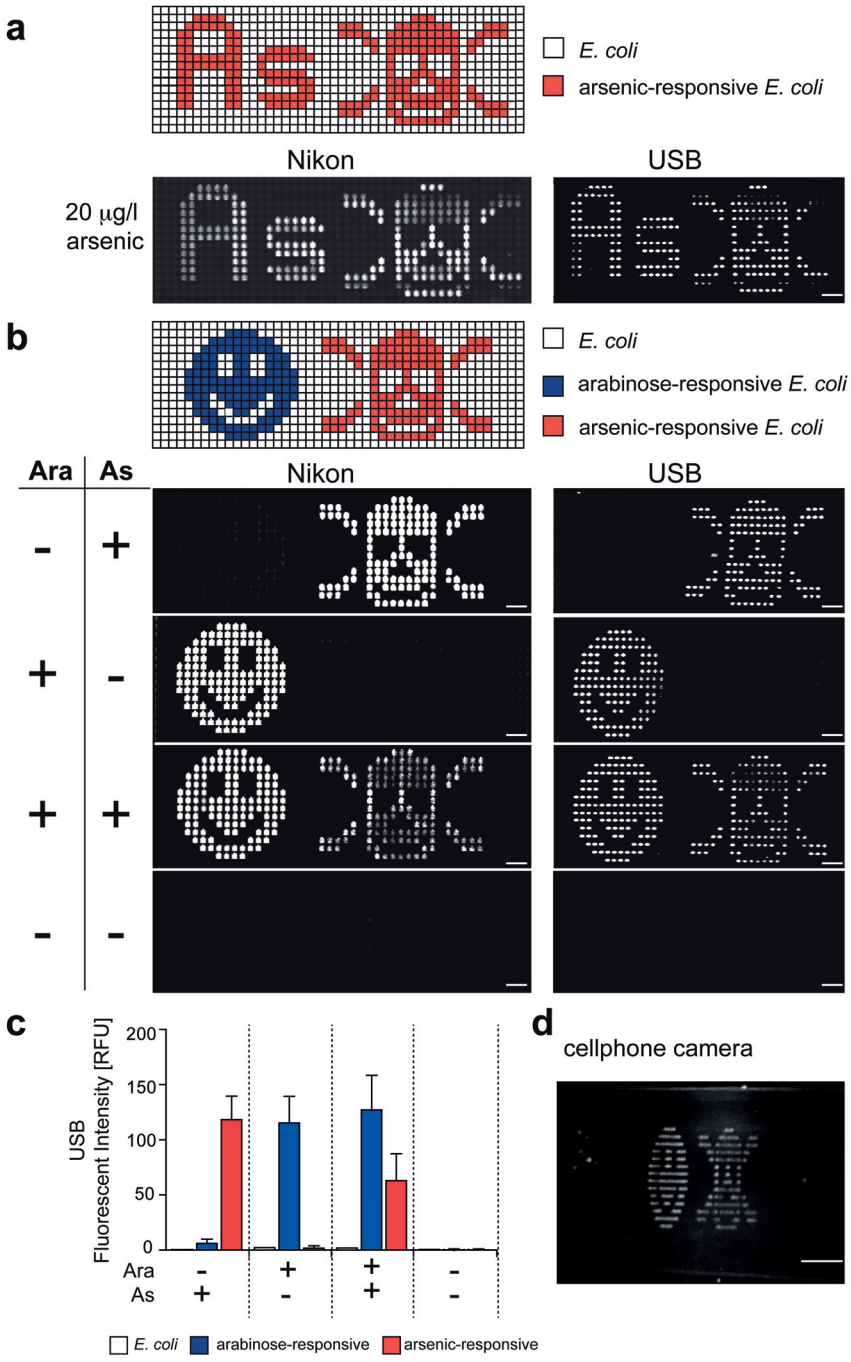

Fig. 6.3 **Biodisplay. (a)** Arsenic-responsive *E. coli* was spotted according to the indicated pattern, spelling "As" and forming a "skull and crossbones" symbol. Twenty µg/L of sodium-arsenite in tap water was sampled and images were acquired using an epifluorescent and USB microscope after 24 h. **(b)** Multiplexed detection of arabinose (blue) and arsenic (red) by two *E. coli* strains that were spotted to display a "smiley" and a "skull and crossbones" symbol, respectively. Epifluorescent and USB microscope images of the chip were acquired after flowing all combinations of 6.66 mM arabinose and 100 µg/L sodium-arsenite in tap water in four independent experiments. Scale bars: 750 µm **(c)** Quantitation of GFP intensity after 24 h of induction, measured using the USB microscope. **(d)** The device exposed to both arsenic and arabinose, was imaged using a mobile phone. The image was obtained using the illumination with blue LEDs and a band-pass filter, placed in front of the cellphone camera. Scale bar: 3.5 mm.

a new biodisplay to one of the four possible combinations of arsenic and arabinose. After sampling for 24 h, the biodisplay accurately reported the presence/absence of these two small molecules (Figure 6.3b). We note that the arsenic-sensing strain used (EDI_as_4) was engineered to be tunable by arabinose, explaining the difference in intensity of the "skull and crossbones" symbol between experiments sampling arsenic and arabinose versus arsenic alone. [231] The patterns could be readily imaged using a research grade and USB fluorescent microscope. Because of the ease with which the symbols could be visualized we also tested whether the device could be imaged directly with a cellphone using a blue LED flashlight for illumination and a fluorescence emission filter held in front of the cellphone camera. Both symbols could readily be imaged with this simple setup (Figure 6.3d). The symbols could also be seen by eye using the LED flashlight and emission filter. Imaging was performed in an illuminated laboratory during daylight hours, with a small curtain shielding the device from direct sunlight (Figure 6.3d, Figure S6.10).

6.2.4 Spores Enable Long-Term Storage

In order to develop a biodisplay that can withstand long-term storage and shipping without requiring a cold chain or special conditions, we explored the use of bacterial spores. Bacterial spores are well-known for their incredible robustness to adverse environments [194, 7] and thus would be ideal for programming a long-term storage biodisplay.

We generated spores from two *B. subtilis* strains, each containing a genomically integrated Phy-spank promoter driving the expression of either mCherry or GFP. The spores were spotted onto glass slides together with *B. subtilis* and *E. coli* cells. The slides were aligned and bonded to a microfluidic device. These biodisplays were then stored for 1 day or one month at 40 and 80°C. After the storage period we introduced LB medium (Figure 6.4a,b). Neither *E. coli* nor *B. subtilis* cells could be grown after any of the storage condition tested. Only when spotted in the presence of glycerol and without an extended storage period could *E. coli* and *B. subtilis* cells be grown on-chip. *B. subtilis* spores, on the other hand, germinated after all storage conditions, notably even after incubation at 80 °C for one month. Germination was observed as early as 1.5 h after the start of the culturing routine. To test whether a spore-based biodisplay retains sensing capability after storage at 80 °C for one month, we induced reporter expression with 1 μM IPTG and observed a rapid increase of fluorescence within 90 min. *B. subtilis* biopixels not exposed to IPTG remained nonfluorescent (Figure 6.4c). To decrease the risk of environmental biological and chemical release, the entire spore biodisplay with attached tygon tubing was embedded in PDMS (Figure S6.11a). No adverse effects on spore germination and induction were observed (Figure S6.11b,c). Increasing the flow pressure from 20 to 140 kPa as well as the actuation pressure of the control valves from

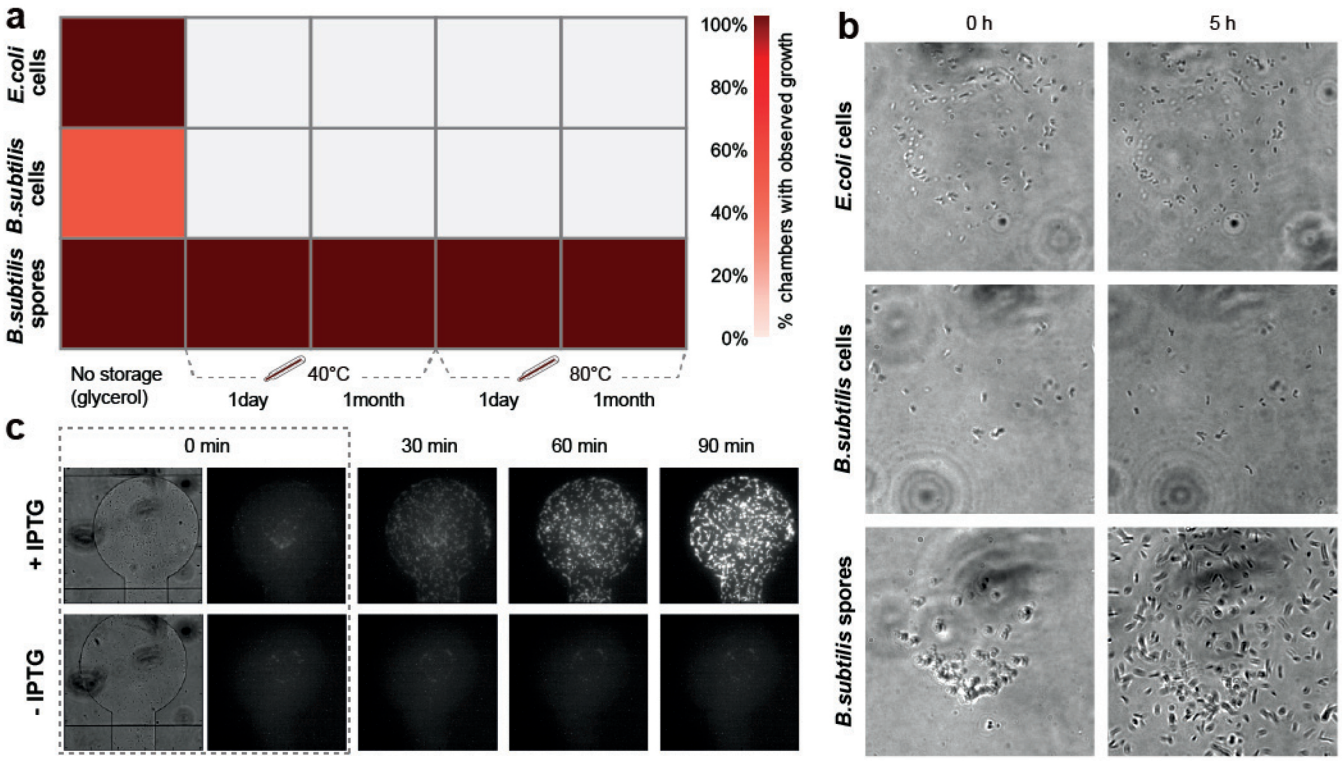

Fig. 6.4 **Spore biodisplay.** **(a)** Each table entry represents the number of chambers with observed growth divided by the total number of spotted chambers for a specific temperature and time condition. **(b)** Bright-field images of a magnified view of the central part of representative pixels after storage at 80°C for 1 month. Images were acquired immediately after spots were resuspended in medium and after a 5 h incubation at 37°C. **(c)** Time-lapse images of germinated Phy-spank-mCherry *B. subtilis* spores after storage at 80°C for one month. Cells were induced with 1 µM IPTG.

70 to 172 kPa did not lead to device failure, demonstrating its improved tolerance to higher pressures.

6.3 Discussion

We developed a biohardware device by combining engineered bacterial strains with a microfluidic chip operated under the control of an automated system. The resulting biodisplay consists of 768 independently programmable biopixels and we showed that it could culture bacterial strains for extended periods of time. Continuous culturing with the capability of dynamically changing media makes the biodisplay a useful platform for high-throughput bacteria cell analysis and may find application in high-throughput screening and characterization of synthetically engineered bacterial strains.

Biological systems are powerful sensors, but challenging to deploy for several reasons. First, engineered biological systems generally need to be cultivated under fairly well controlled conditions, and often require the presence of antibiotics in order to maintain the synthetic network. Second, a major concern limiting the applicability of synthetic systems is the possible escape of genetically modified cells into the environment. Not only is the escape of GMOs a concern, but also the release of genetic material, which could be taken up by environmental bacteria through horizontal gene transfer. Our biodisplay solves both issues by providing a platform that enables long-term culturing, while decreasing the risk of release of genetic and chemical material into the environment. The small scale of the device requires low quantities of medium and antibiotics: 2 mL for a 6 day experiment. The device outflow is sterilized and collected to prevent environmental contamination with biological and chemical components.

The ability to culture bacterial sensors on the biodisplay for extended periods of time with frequent readout makes the biodisplay a potentially useful environmental monitoring tool. We showed that our biodisplay can detect arabinose and sodium-arsenite in tap water using bacterial strains previously engineered by other research groups [231, 142] and student teams. Several heavy metal sensors have already been engineered, including arsenic [103, 243], mercury [243] and lead sensors [19]. Information processing systems also exist, including genetic toggle switches [66], logic functions [155], band-pass filters [14], and event counters [59]. By combining event counters or toggle switches with sensors it is possible to implement memory, which could be useful in instances when the biodisplay can be monitored only intermittently, but a transient contamination should be logged and reported. The use of band-pass filters would enable the development of a biodisplay that reports the concentration of an analyte without requiring quantitative analysis of the reported signal. The biodisplay

can also reduce the complexity of engineered strains, since functions can potentially be distributed among several strains [198]. For example, to generate a sensor that senses and reports the presence/absence of three different substances would not require incorporation of all three functions into a single strain but could instead be implemented in three individual strains. It may also be possible to develop chemiluminescent biodisplays or color biodisplays using chromogenic proteins. Communication between pixels as previously demonstrated using hydrogen peroxide [177] could be used to couple sensing biopixels, logic biopixels, and reporter biopixels.

Unlike previous biopixel arrays [177] where each pixel contained the same bacterial strain, each pixel on our display can be specifically programmed with a different strain, drastically increasing the multiplexing capacity and information content of our biodisplay with a maximum of 768 unique strains, each of which could contain multiple, genetically encoded functions. A previous demonstration of integrating an *E. coli* arsenic biosensor in a microfluidic device was limited to a single culture, employed a complex scheme that decoupled culturing from measurement, and used a research-grade fluorescent microscope for readout [29]. A scheme called InfoBiology was used for the transmission of information using arrays of bacterial cells spotted on microtiter sized agar plates [166]. 144 colonies of a handful of strains were spotted in this scheme, but lacked integration with a microfluidic device for cell culturing and environmental sampling.

We previously developed a low-cost, portable hardware system that contains all necessary components for microfluidic device operation, making the biodisplay readily deployable in resource limited settings [174]. By using bacterial spores the biodisplay can be stored and shipped under ambient conditions and is viable for extended periods of time. The material cost of the biodisplay itself is negligible at less than 1 USD per device. In this particular example a biohardware platform applied to continuous environmental monitoring of arsenic levels in water could replace a USD 26000 instrument with an annual running cost of over USD 3000 (Ova 7000, Modern Water Monitoring Limited). The relatively slow response time of our biodisplay is acceptable in this and other applications, where continued chronic exposure of a substance is to be avoided. For time critical applications the response time of the biosensors would need to be improved, for example by developing phosphorylation based sensors and reporters rather than transcriptional based genetic networks [178].

Programmable biopixel displays drastically simplify device readout and interpretation. Instead of reporting the presence of arsenic using a display-wide pattern of signal oscillations [177], our biodisplay generates an easy to interpret “skull and crossbones” symbol. We showed that this concept can be easily multiplexed using different symbols to report the presence of multiple small molecules. The number of biopixels is scalable with previous

work on two layer polydimethylsiloxane (PDMS) chips having demonstrated up to 4160 unit cells per device [64, 58]. Immediate optical readout without specialized equipment combined with the fact that computing and decision-making steps are conducted by the biodisplay itself make it low-cost, easy to deploy, and easy to interpret. Biohardware hybrid devices therefore combine the advantages of biological systems with the advantages of mechanical, electrical and optical systems. The ability to engineer biological systems on the molecular level and combine them with novel hardware opens new opportunities for how biological, mechanical, electrical and optical systems are integrated and the types of functions they can perform.

6.4 Methods

Materials

All bacterial strains and plasmids used in the study are listed in Table S1. Cells for arsenic detection were a gift from Baojun Wang (University of Edinburgh): pBW103 ParsR-Amp30C (Addgene plasmid #78638), pBW300 ParsR-Amp32T (Addgene plasmid #78652), pBW102 ParsR-Amp32C (Addgene plasmid #78637) and pBW101 ParsR-gfp (Addgene plasmid #78636). LB medium was purchased from Applichem Panreac (A0954). l-(+)-Arabinose (A3256–100G), 0.05 M (35000–1L-R) and chloramphenicol (23275) were purchase from Sigma. Kanamycin sulfate (T832.1) was bought from Carl Roth. Isopropyl-beta-d-thiogalactoside for spore induction was purchased from Roche. Schaeffer and Fulton Spore Stain Kit was bought from Sigma-Aldrich.

Device Fabrication

Device molds were made using standard photolithography techniques. The mold for the control layer was made using GM 1070 SU-8 photoresist (Gersteltec Sarl, Switzerland) with a height of 30 μm. The flow layer was made with AZ9260 photoresist (Gersteltec Sarl, Switzerland) with a thickness of 14 μm. A Suss MJB4 single side mask aligner was used to expose the wafers. The flow mold was baked for 2 h at 135°C to anneal the flow channels. The molds were then treated with chlorodimethylsilane (DMCS) prior to being exposed to PDMS (Sylgard 184). The double layer microfluidic device were made by using multilayer soft lithography.(46) For the flow layer, PDMS was prepared with a ratio of 20:1 (part A:B) and spin coated with a speed of 3000 rpm for 1 min, in order to get a thickness in the range of 20–40 μm. The control mold was placed in a Petri dish and coated with a 5:1 ratio of PDMS, in order to get a 0.5 cm thick device. Both layers were baked for 30 min at 80°C. Then the control layer was cut and peeled from the control layer mold and placed on top of

the flow layer. Alignment of the two layers was performed by eye using a stereomicroscope; markers, placed around the main structure of the device aided the alignment process. After alignment the two PDMS layers were baked for an additional 1.5 h at 80°C. The chip was then aligned to a spotted glass slide.

Cell Culture, Transformation and Spore Formation

E. coli and *B. subtilis* cells were cultured in LB medium at 37°C and 200 rpm with appropriate antibiotics, when required. All plasmid transformations and *E. coli* studies were performed in DH5-alpha cells. *B. subtilis* sporulation was performed as described previously [83]. For spore enrichment, a three day old *B. subtilis* cell suspension was centrifuge for 1 min at 15000 rpm, the pellet was washed three times with water. Schaeffer and Fulton spore staining was performed to verify the presence of spores.

Cell Arraying

Overnight cultures of *E. coli* and *B. subtilis* were centrifuged for 5 min at 3000 rpm. For the *E. coli* display the cell pellet was resuspended in 100 µL LB with 10% glycerol and appropriate antibiotics. For the spore biodisplay spores were resuspended in water. Cell suspensions and *B. subtilis* spore solutions were plated in conical, polypropylene 96-well plates. The samples were spotted with a 0.7 nL delivery volume pin (946MP2B Arrayit) on a glass slide by using a microarray robot (QArray2, Genetix). Glass slides were coated with epoxysilane (3-Glycidoxypropyl-dimethoxymethylsilane 97% AC216545000 Acros organic). The spotting parameters were as follows: 100 ms inking time, 10 ms stamping time, max number of stamps per ink 30. A wash procedure was included that consisted of washing the spotting pin with 70% ethanol for 2 s, washing with water for 2 s followed by pin drying. Depending on the spotting pattern and number of samples used the spotting time varied between 10 to 20 min.

Cell-Display Culturing

The *E. coli* array was aligned to a PDMS chip and incubated for 1 h at 37°C. The spore-array was aligned to a MITOMI PDMS chip [132] and incubated depending on the experimental procedure as follows: 40 °C for 1 day, 40 °C for 1 month, 80°C for 1 day, or 80°C for 1 month. After the incubation step the devices were placed on a temperature-controlled glass plate (H401-NIKON-TI_SR_GLASS/H401_T_CONTROLLER; Okolab) at 37°C. Tygon tubes were filled with deionized water and connected to the inlets of the control lines,

pressure was applied in order to actuate the valves. For the *E. coli* display 68.9 and 13.8 kPa were used, for the control and flow lines, respectively. For the spore-display 103.4 kPa and 24.1–27.5 kPa were used instead. For the spore display we used a 768 unit cell MITOMI chip [132], but performed the same culturing routine as on the biodisplay device. For the cell-culturing and sampling routines a custom-written LabVIEW program was used. During the culturing, LB medium with appropriate antibiotics and lysis buffer containing 30 mM of NaOH (06203–1KG Sigma-Aldrich) and 12% SDS (L3771–100G Sigma) was used. Depending on the experimental procedure LB medium or tap water was supplemented with arabinose, sodium-arsenite, or IPTG. The level of arsenic in the tap water was measured by a colorimetric arsenic test kit (MQuant Arsenic Test, Sigma-Aldrich) with a LOD of 5 µg/L. Arsenic levels in our tap water were below the detection limit, and based on publically available data are likely below 2 µg/L [172].

Imaging

Image acquisition was performed on a Nikon ECLIPSE Ti automated microscope equipped with a LED Fluorescence Excitation System and a Hamamatsu ORCA-Flash 4.0 camera controlled by NIS Elements. Images were taken at 40× magnification (SPlan Fluor, ELWD 40x/0.60, WD 3.6–2.8, Nikon) in fluorescent and bright field mode. Images of all biopixel units were stitched together using either the Grid/Collection Stitching plugin in Fiji or a custom-written Python script. Fluorescent measurements were performed using Genepix software. A FITC USB fluorescent microscope (AM4113T-GFBW, Dino-Lite) was used to acquire fluorescent images. Images of 9 subsections of the device were taken, using lower magnification (10x), and stitched. Pictures were also taken using a cellphone camera. A band-pass filter, centered at 530 nm with a 40 nm bandwidth, was placed in front of the camera of the mobile phone, and the LEDs of the FITC USB microscope were used for illumination of the biodisplay.

6.5 Supplementary information

Identifier	Name	Host strain	Resistance	Inducer	Reporter
BBa_I13517	iGEM RFP_1	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP

BBa_K1333301	iGEM RFP_2	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP
BBa_K1333300	iGEM RFP_3	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP
BBa_K577004	iGEM RFP_4	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP
BBa_K577882	iGEM RFP_5	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP
BBa_I13516	iGEM RFP_6	E. coli DH5 alpha	Chloramphenicol	Arabinose	RFP
BBa_K750000	iGEM GFP_1	E. coli DH5 alpha	Chloramphenicol	Arabinose	GFP
BBa_K584000	iGEM GFP_2	E. coli DH5 alpha	Chloramphenicol	Arabinose	GFP
BBa_K577881	iGEM GFP_3	E. coli DH5 alpha	Chloramphenicol	Arabinose	GFP
pBW103ParsR_Amp 30C (78638)	EDI_as_1	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pBW102ParsR-Amp32C (78637)	EDI_as_2	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pBW103ParsR-Amp30C (78638)	EDI_as_3	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pBW300ParsR-Amp32T (78652)	EDI_as_4	E. coli DH5 alpha	Kanamycin	Arsenic	GFP

pIIUN gfp	UNIL_1	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pAAUN gfp	UNIL_2	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pVUN gfp	UNIL_3	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
10pLtet0UN gfp	UNIL_4	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
pPR arsR abs gfp	UNIL_5	E. coli DH5 alpha	Kanamycin	Arsenic	GFP
BGSC 3A40	B. subtilis subsp. subtilis amyE::Physpank-mCherry	B. subtilis subsp. subtilis		IPTG	mCherry
BGSC 3A39	B. subtilis subsp. subtilis amyE::Physpank-GFP	B. subtilis subsp. subtilis		IPTG	GFP

Table 6.1 **List of plasmids and strains used in this work**

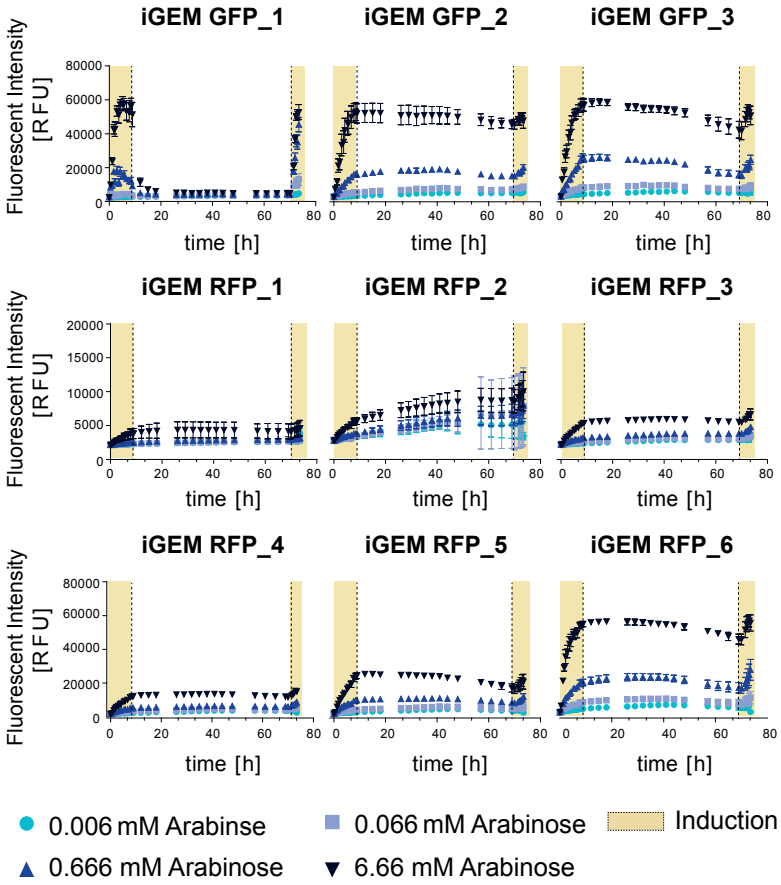

Fig. 6.5 **Biodisplay for strain characterization at 37°C.** Fluorescent intensity of GFP or RFP expression of nine different arabinose sensitive *E. coli* strains. The cells were induced twice with 4 different concentrations of arabinose (0.006 mM, 0.066 mM, 0.666 mM and 6.66 mM). The yellow background indicates the induction intervals. The microfluidic chip was heated to 37°C during the experiment.

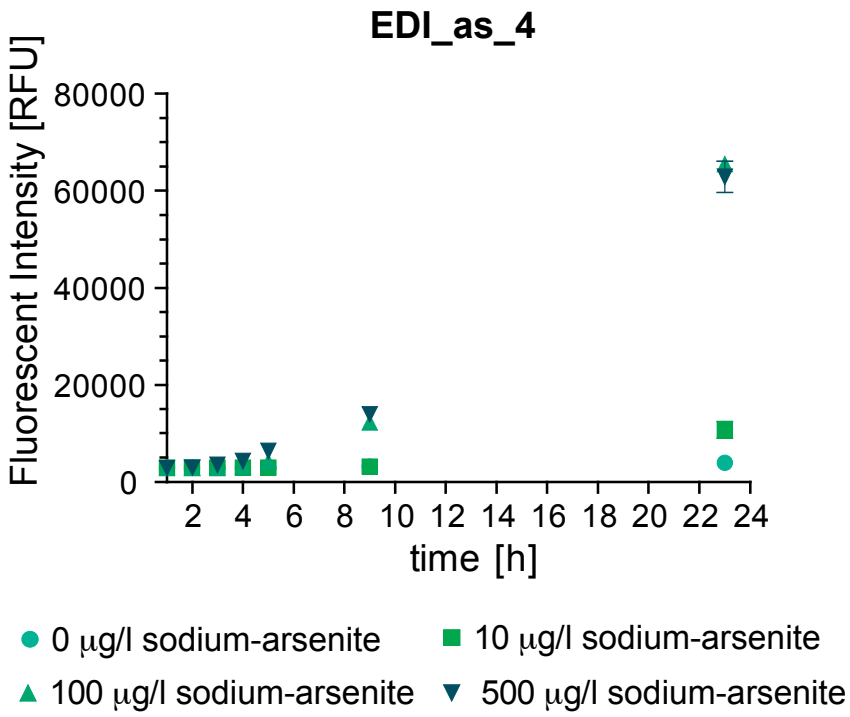

Fig. 6.6 **Arsenic-responsive *E. coli*.** EDI_as_4 was spotted on the chip and 4 different concentrations of sodium-arsenite in LB were sampled for 24h. GFP expression was monitored over time using an epifluorescent microscope.

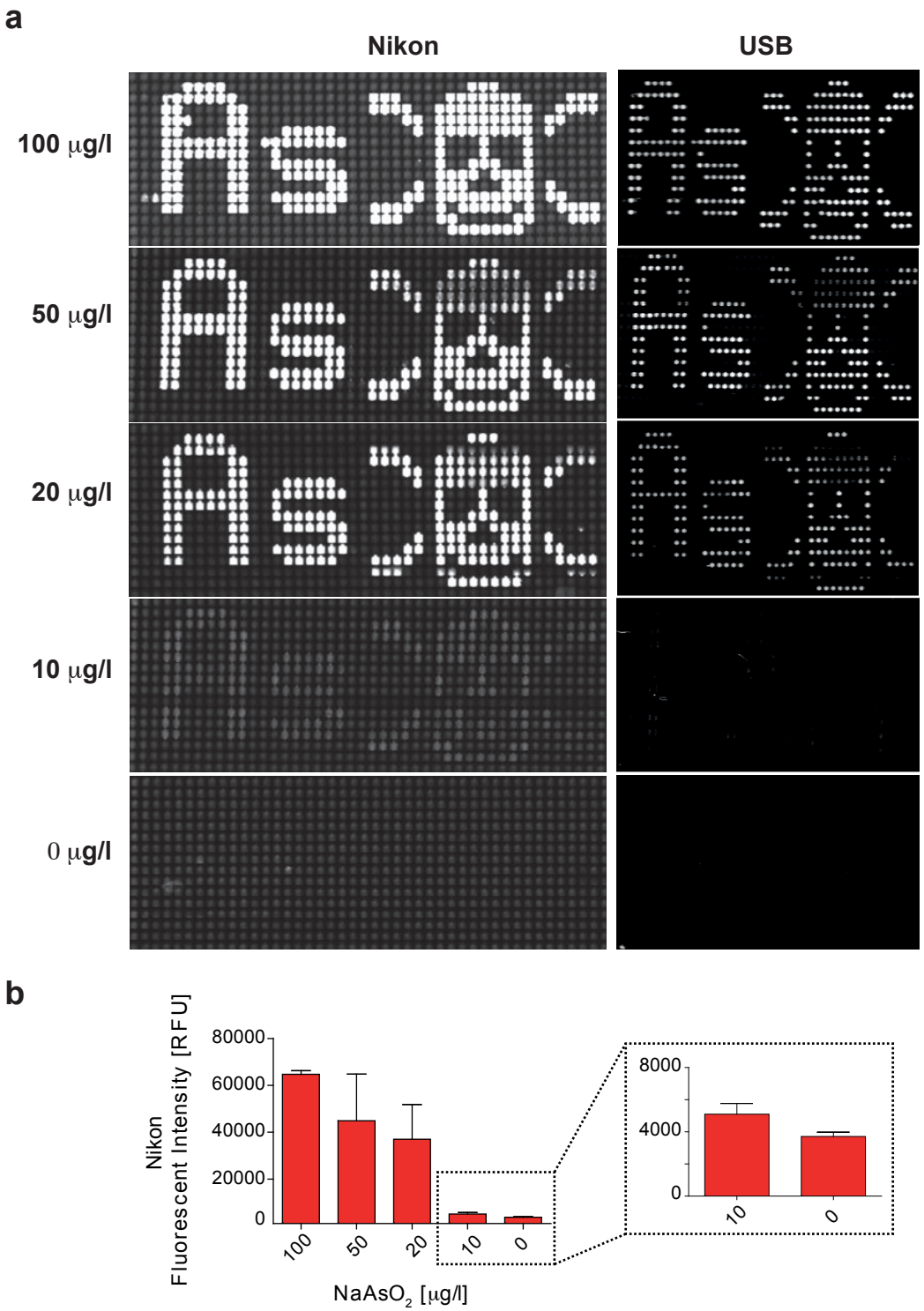

Fig. 6.7 **Arsenic biodisplay (a)** Fluorescent images taken using an epifluorescent and USB microscope after 24 hours of induction with different concentrations of sodium-arsenite (100 µg/L, 50 µg/L, 20 µg/L,10 µg/Land 0 µg/L). Each concentration was sampled in different devices. **(b)** Quantitation of the fluorescent signal, after 24 h of induction and measured with the epifluorescent microscope.

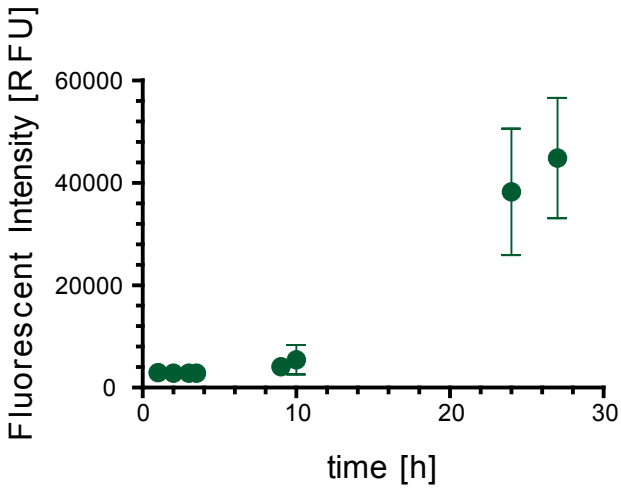

Fig. 6.8 **Arsenic biodisplay over time.** Arsenic-responsive *E. coli* was spotted in a "As" and a "skull and crossbones" pattern. 20 µg/L of sodium-arsenite in tap water was sampled and GFP expression monitored over time using an epifluorescent microscope.

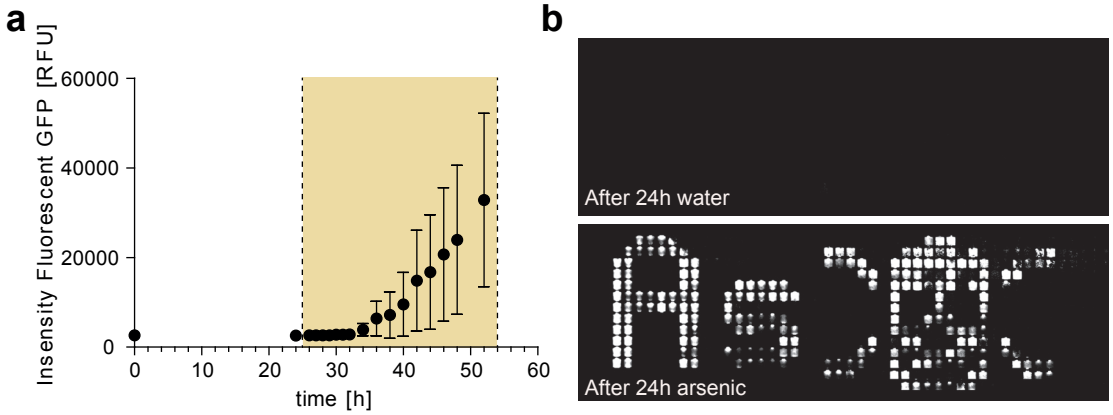

Fig. 6.9 **Delayed arsenite sensing. (a)** Arsenic-responsive *E. coli* was spotted in a "As" and a "skull and crossbones" pattern. After sampling water for 24 h, water containing 50 µg/L of sodium-arsenite was sampled. The GFP signal was monitored using an epifluorescent microscope. **(b)** Fluorescent images of the device after 24 h of water sampling (top) and others 24 h of water containing sodium-arsenite.

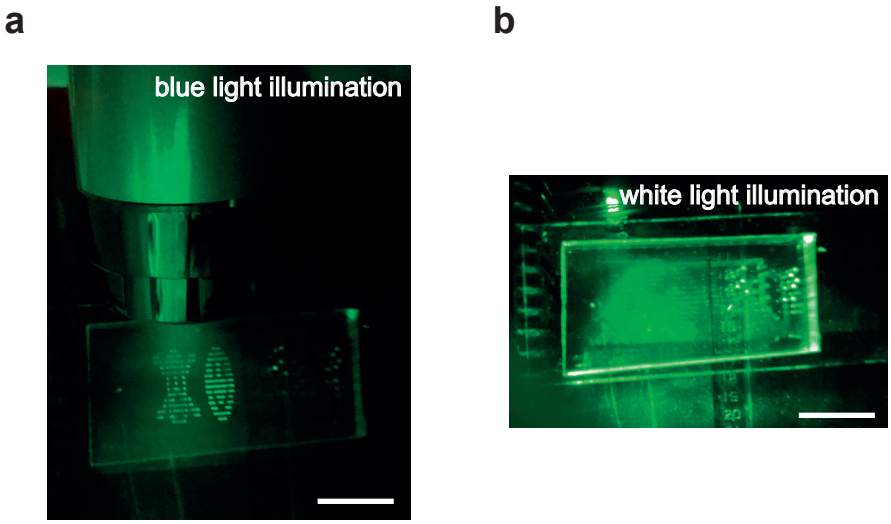

Fig. 6.10 **Cellphone image acquisition. (a)** The microfluidic chip was illuminated using the LEDs of a USB microscope (excitation at 480 nm), a band-pass filter centered at 530 nm with a 40 nm bandwidth was placed in front of the camera of the cellphone. **(b)** The same device was illuminated by white light and imaged using the cellphone and emission filter as in **(a)**. Scale bar 1cm.

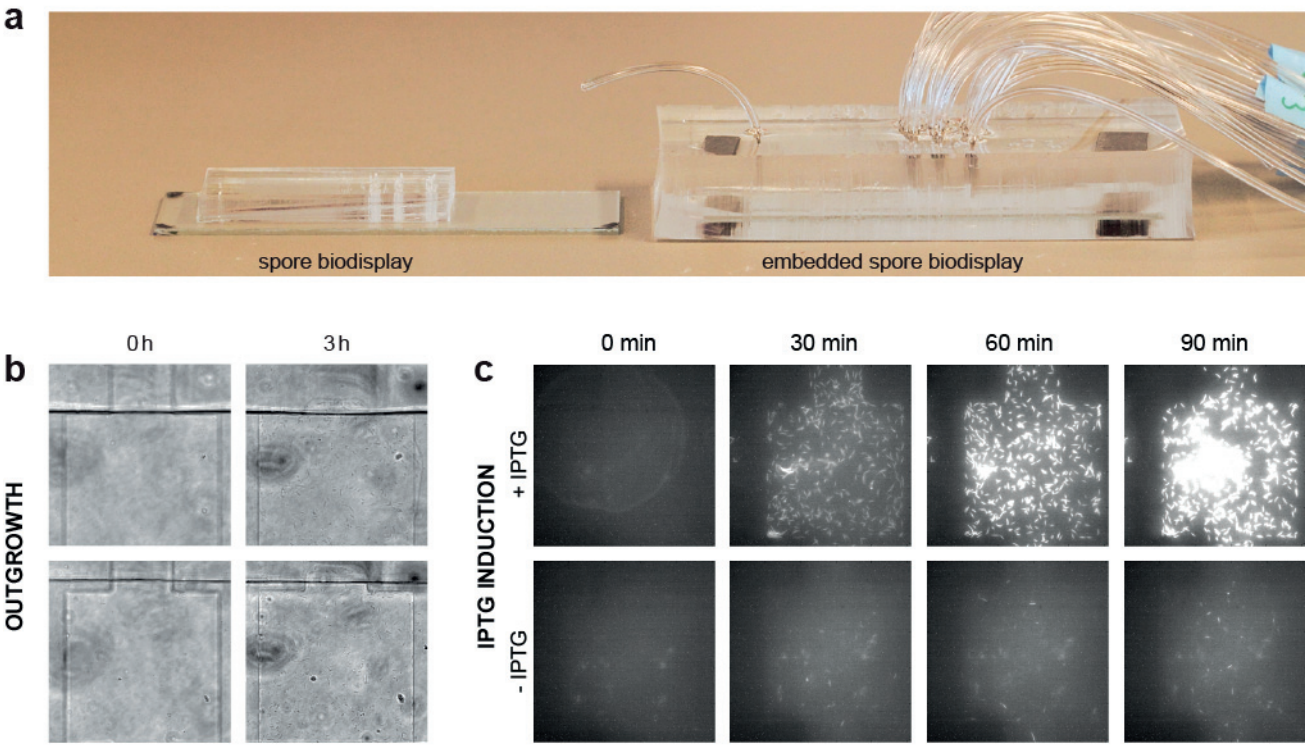

Fig. 6.11 **Embedded spore biodisplay.** **(a)** Photo comparing the spore biodisplay and the embedded PDMS spore biodisplay. **(b)** Bright-field images of magnified center part of spotted chambers showing spore germination. Images are acquired immediately after spot resuspension with medium and 3 h incubation at 37°C. **(c)** Time-lapse images of germinated Phy-spank-mCherry *B. subtilis* spores after induction with 1 µM IPTG.

Chapter 7

Conclusions and Outlook

Conventional methods for protein characterization are often limited in terms of throughput and cost. In the thesis we utilized several state-of-art technologies in the field of cell-free synthetic biology and microfluidics allowing parallel and relatively cheap profiling of custom proteins and protein complexes.

First, we developed methods for combinatorial assembly of anti-RSV scFv variants and their rapid characterization by coupled synthesis and biophysical characterization on a high-throughput microfluidic platform (see Chapter 2). The synthesized proteins retained their functional properties when expressed in PURExpress. We were able to identify two scFv variants possessing higher binding affinity compared to the rest of the library members. However, we did not succeed in measuring small differences between the variants which may be due to small library size. Another major obstacle was the observed variability in the protein expression levels. Optimization of the spotting procedure as well as the expression system composition may help to resolve the problem. Despite all, the developed platform has the potential to outperform current high-throughput display methods for scFvs screening and selection by enabling direct mapping of their binding landscape in cost- and time-effective manner.

Characterizing the binding properties of the scFv library aided to the design of genetically-encoded scFv-conjugates possessing additional functionalities. The synthesized scFv chimeras were used for the development of a genetically encoded protein biosensor (see Chapter 3). The proposed method has the potential to eliminate the used of expensive reagents used in conventional ELISA assays.

In addition to the synthesis of individual biological molecules we were able to develop a method for *in vitro* reconstruction of the CRISPR-Cas9 system. Our cell-free assembly approach together with high-throughput microfluidic platform allowed rapid screening of the Cas9/sgRNA binding and cleavage properties. The fast identification of Cas9 orthologous

as well as structural-guided engineering of the current widely utilized CRISPR-associated proteins have expanded greatly the CRISPR-Cas9 toolbox [26, 42, 185]. However, the molecular mechanism behind all these new systems are largely unexplored. The developed platform can be used for in-depth protein characterization, aid to the functional improvement of the novel candidates and thus can enable their faster implementation into genome and epigenome applications.

Finally, we developed a high-throughput microfluidic platform supplemented with automated valve actuators allowing continuous bacteria cell culturing (see Chapter 5). In contrast to conventional high-throughput cell culturing methods relying on microtiter plate experiments and expensive robotic equipment, the microfluidic biodisplay has made the process relatively cheap and amenable to track temporal expression dynamics. For the study we used cell bioreporters capable of sensing small molecules such as arabinose, arsenite and IPTG. The platform was programmed with microarray of printed *E. coli* cells or *B. subtilis* spores. This increased sufficiently its scalability and also allowed us to generate custom-defined patterns simplifying the interpretation of the readout generated by the bacteria response. To prove the feasibility of engineering in-field biodisplay we employed *B. subtilis* allowing long storage at ambient temperatures as well as transportation. Unfortunately, compared to *E. coli* currently there are limited number of genetically engineered *B. subtilis* bioreporter stain. Furthermore, the platform can be easily implemented with low-cost, portable hardware system for microfluidics device operaton, previously developed in the lab [174]. The scalability and the architecture of the device promise its further use in various studies including cell response to molecular cues that vary in space and time as well as cell-extracellular matrix interactions in terms of mechanical force and biochemical signaling.

References

[1] Adams, G. P., Schier, R., McCall, A. M., Simmons, H. H., Horak, E. M., Alpaugh, R. K., Marks, J. D., and Weiner, L. M. (2001). High affinity restricts the localization and tumor penetration of single-chain fv antibody molecules. *Cancer research*, 61(12):4750–4755.

[2] Agarwal, K., Büchi, H., Caruthers, M., Gupta, N., Khorana, H. G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U., Van de Sande, J., et al. (1970). Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature*, 227(5253):27.

[3] Aksimentiev, A. (2010). Deciphering ionic current signatures of DNA transport through a nanopore. *Nanoscale*, 2(4):468–483.

[4] Alderson, Ralph F and Toki, Brian E and Roberge, Martin and Geng, Wei and Basler, Joshua and Chin, Regina and Liu, Amy and Ueda, Roanna and Hodges, Douglas and Escandon, Enrique and others (2006). Characterization of a CC49-Based Single-Chain Fragment- β-Lactamase Fusion Protein for Antibody-Directed Enzyme Prodrug Therapy (ADEPT). *Bioconjugate chemistry*, 17(2):410–418.

[5] Ali, M., Suzuki, H., Fukuba, T., Jiang, X., Nakano, H., and Yamane, T. (2005). Improvements in the cell-free production of functional antibodies using cell extract from protease-deficient Escherichia coli mutant. *Journal of bioscience and bioengineering*, 99(2):181–186.

[6] Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519):569.

[7] Atrih, A. and Foster, S. J. (1999). The role of peptidoglycan structure and structural dynamics during endospore dormancy and germination. *Antonie van Leeuwenhoek*, 75(4):299–307.

[8] Au, L.-C., Yang, F.-Y., Yang, W.-J., Lo, S.-H., and Kao, C.-F. (1998). Gene synthesis by a lcr-based approach: High-level production of leptin-l54 using synthetic gene in Escherichia coli. *Biochemical and biophysical research communications*, 248(1):200–203.

[9] Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 30(10):1473–1475.

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

[11] Bang, D. and Church, G. M. (2008). Gene synthesis by circular assembly amplification. *Nature methods*, 5(1):37.

[12] Barrangou, R. and Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nature biotechnology*, 34(9):933.

[13] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837.

[14] Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130.

[15] Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., and Voigt, C. A. (2009). Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society*, 131(18):6508–6515.

[16] Beaudet, L., Rodriguez-Suarez, R., Venne, M.-H., Caron, M., Bédard, J., Brechler, V., Parent, S., and Bielefeld-Sévigny, M. (2008). AlphaLISA immunoassays: the no-wash alternative to ELISAs for research and drug discovery. *Nature Methods*, 5(12).

[17] Bell, N. A. and Keyser, U. F. (2015). Specific protein detection using designed DNA carriers and nanopores. *Journal of the American Chemical Society*, 137(5):2035–2041.

[18] Bennett, M. R. and Hasty, J. (2009). Microfluidic devices for measuring gene network dynamics in single cells. *Nature Reviews Genetics*, 10(9):628.

[19] Bereza-Malcolm, L., Aracic, S., and Franks, A. E. (2016). Development and application of a synthetically-derived lead biosensor construct for use in gram-negative bacteria. *Sensors*, 16(12):2174.

[20] Bereza-Malcolm, L. T., Mann, G., and Franks, A. E. (2014). Environmental sensing of heavy metals through whole cell microbial biosensors: a synthetic biology approach. *ACS synthetic biology*, 4(5):535–546.

[21] Blackburn, M. C., Petrova, E., Correia, B. E., and Maerkl, S. J. (2015). Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. *Nucleic acids research*, 44(7):e68–e68.

[22] Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences*, 97(20):10701–10705.

[23] Boder, E. T. and Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology*, 15(6):553.

[24] Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V. W., Holt, L., et al. (2016). The genome project-write. *Science*, 353(6295):126–127.

[25] Bolukbasi, M. F., Gupta, A., and Wolfe, S. A. (2016). Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nature methods*, 13(1):41.

[26] Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P., and Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular cell*, 56(2):333–339.

[27] Brödel, A. K., Sonnabend, A., and Kubick, S. (2014). Cell-free protein expression based on extracts from CHO cells. *Biotechnology and bioengineering*, 111(1):25–36.

[28] Brödel, A. K., Sonnabend, A., Roberts, L. O., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2013). Ires-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems. *PLoS One*, 8(12):e82234.

[29] Buffi, N., Beggah, S., Truffer, F., Geiser, M., Van Lintel, H., Renaud, P., and van der Meer, J. R. (2016). An automated microreactor for semi-continuous biosensor measurements. *Lab on a Chip*, 16(8):1383–1392.

[30] Bulushev, R. D., Marion, S., and Radenovic, A. (2015). Relevance of the Drag Force during Controlled Translocation of a DNA–Protein Complex through a Glass Nanocapillary. *Nano letters*, 15(10):7118–7125.

[31] Bulushev, R. D., Steinbock, L. J., Khlybov, S., Steinbock, J. F., Keyser, U. F., and Radenovic, A. (2014). Measurement of the position-dependent electrophoretic force on DNA in a glass nanocapillary. *Nano letters*, 14(11):6606–6613.

[32] Bulyk, M. L., Huang, X., Choo, Y., and Church, G. M. (2001). Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proceedings of the National Academy of Sciences*, 98(13):7158–7163.

[33] Cadwell, R. C. and Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome research*, 2(1):28–33.

[34] Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology*, 12(5):381.

[35] Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Fernö, M., Olsson, H., and Borrebaeck, C. A. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *European journal of cancer*, 44(3):472–480.

[36] Carr, P. A., Park, J. S., Lee, Y.-J., Yu, T., Zhang, S., and Jacobson, J. M. (2004). Protein-mediated error correction for de novo DNA synthesis. *Nucleic acids research*, 32(20):e162–e162.

[37] Carter, P. J. and Lazar, G. A. (2017). Next generation antibody drugs: pursuit of the 'high-hanging fruit'. *Nature Reviews Drug Discovery*.

[38] Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T. M., Dostie, J., and Pelletier, J. (2014). Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PloS one*, 9(10):e109213.

[39] Chari, R., Mali, P., Moosburner, M., and Church, G. M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature methods*, 12(9):823.

[40] Chen, B., Guan, J., and Huang, B. (2016). Imaging specific genomic dna in living cells. *Annual review of biophysics*, 45:1–23.

[41] Cho, S. W., Kim, S., Kim, J. M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31(3):230.

[42] Chylinski, K., Makarova, K. S., Charpentier, E., and Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, 42(10):6091–6105.

[43] Correia, B. E., Bates, J. T., Loomis, R. J., Baneyx, G., Carrico, C., Jardine, J. G., Rupert, P., Correnti, C., Kalyuzhniy, O., Vittal, V., et al. (2014). Proof of principle for epitope-focused vaccine design. *Nature*, 507(7491):201.

[44] Cradick, T. J., Fine, E. J., Antico, C. J., and Bao, G. (2013). CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic acids research*, 41(20):9584–9592.

[45] DeKosky, B. J., Ippolito, G. C., Deschner, R. P., Lavinder, J. J., Wine, Y., Rawlings, B. M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S. F., et al. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature biotechnology*, 31(2):166.

[46] Dénervaud, N., Becker, J., Delgado-Gonzalo, R., Damay, P., Rajkumar, A. S., Unser, M., Shore, D., Naef, F., and Maerkl, S. J. (2013). A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proceedings of the National Academy of Sciences*, 110(39):15842–15847.

[47] Didovyk, A., Tonooka, T., Tsimring, L., and Hasty, J. (2017). Rapid and scalable preparation of bacterial lysates for cell-free gene expression. *ACS synthetic biology*, 6(12):2198–2208.

[48] Dimitrov, V., Mirsaidov, U., Wang, D., Sorsch, T., Mansfield, W., Miner, J., Klemens, F., Cirelli, R., Yemenicioglu, S., and Timp, G. (2010). Nanopores in solid-state membranes engineered for single molecule detection. *Nanotechnology*, 21(6):065502.

[49] Dominguez, A. A., Lim, W. A., and Qi, L. S. (2016). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature reviews Molecular cell biology*, 17(1):5.

[50] Doudna, J. A. and Charpentier, E. (2014). The new frontier of genome engineering with crispr-cas9. *Science*, 346(6213):1258096.

[51] Duval, V. and Lister, I. M. (2013). MarA, SoxS and Rob of Escherichia coli–Global regulators of multidrug resistance, virulence and stress response. *International journal of biotechnology for wellness industries*, 2(3):101.

[52] Ekins, R. P. (1989). Multi-analyte immunoassay. *Journal of pharmaceutical and biomedical analysis*, 7(2):155–168.

[53] Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335.

[54] Endy, D. (2005). Foundations for engineering biology. *Nature*, 438(7067):449.

[55] Estell, D. A., Graycar, T. P., and Wells, J. A. (1985). Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *Journal of Biological Chemistry*, 260(11):6518–6521.

[56] Fägerstam, L. G., Frostell-Karlsson, Å., Karlsson, R., Persson, B., and Rönnberg, I. (1992). Biospecific interaction analysis using surface plasmon resonance detection applied to kinetic, binding site and concentration analysis. *Journal of Chromatography A*, 597(1-2):397–410.

[57] Fan, J., Leroux-Coyau, M., Savery, N. J., and Strick, T. R. (2016). Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. *Nature*, 536(7615):234.

[58] Fordyce, P. M., Gerber, D., Tran, D., Zheng, J., Li, H., DeRisi, J. L., and Quake, S. R. (2010). De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature biotechnology*, 28(9):970.

[59] Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., and Collins, J. J. (2009). Synthetic gene networks that count. *science*, 324(5931):1199–1202.

[60] Fu, Y.-H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkert, A. J., Holden, J. J., Fenwick Jr, R. G., Warren, S. T., et al. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, 67(6):1047–1058.

[61] Gai, S. A. and Wittrup, K. D. (2007). Yeast surface display for protein engineering and characterization. *Current opinion in structural biology*, 17(4):467–473.

[62] Galas, D. J. and Schmitz, A. (1978). DNAase footprinting a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research*, 5(9):3157–3170.

[63] Garcia-Cordero, J. L. and Maerkl, S. J. (2014). A 1024-sample serum analyzer chip for cancer diagnostics. *Lab on a Chip*, 14(15):2642–2650.

[64] Garcia-Cordero, J. L. and Maerkl, S. J. (2016). Mechanically Induced Trapping of Molecular Interactions and Its Applications. *Journal of laboratory automation*, 21(3):356–367.

[65] Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S., and Rutter, W. J. (1985). Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase a does not play a crucial role in catalysis. *Nature*, 317(6037):551.

[66] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339.

[67] Garner, M. M. and Revzin, A. (1981). 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(13):3047–3060.

[68] Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific dna cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, 109(39):E2579–E2586.

[69] Geertz, M. and Maerkl, S. J. (2010). Experimental strategies for studying transcription factor–DNA binding specificities. *Briefings in functional genomics*, 9(5-6):362–373.

[70] Geertz, M., Shore, D., and Maerkl, S. J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences*, 109(41):16540–16545.

[71] Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L., and Curtiss III, R. (1997). Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature biotechnology*, 15(1):29–34.

[72] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6(5):343.

[73] Gijs, M. A. (2004). Magnetic bead handling on-chip: new opportunities for analytical applications. *Microfluidics and Nanofluidics*, 1(1):22–40.

[74] Glockshuber, R., Schmidt, T., and Plueckthun, A. (1992). The disulfide bonds in antibody variable domains: effects on stability, folding in vitro, and functional expression in Escherichia coli. *Biochemistry*, 31(5):1270–1279.

[75] Graff, C. P., Chester, K., Begent, R., and Wittrup, K. D. (2004). Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 C. *Protein Engineering Design and Selection*, 17(4):293–304.

[76] Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I., Waterhouse, P., Crosby, W. L., Kontermann, R., Jones, P., Low, N., and Allison, T. a. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *The EMBO journal*, 13(14):3245–3260.

[77] Gulati, S., Rouilly, V., Niu, X., Chappell, J., Kitney, R. I., Edel, J. B., Freemont, P. S., et al. (2009). Opportunities for microfluidic technologies in synthetic biology. *Journal of The Royal Society Interface*, page rsif20090083.

[78] Hanes, J. and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences*, 94(10):4937–4942.

[79] Hanes, J., Schaffitzel, C., Knappik, A., and Plückthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature biotechnology*, 18(12):1287.

[80] Hansen, C. L., Skordalakes, E., Berger, J. M., and Quake, S. R. (2002). A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proceedings of the National Academy of Sciences*, 99(26):16531–16536.

[81] Harrison, S. C. (1991). A structural taxonomy of dna-binding domains. *Nature*, 353(6346):715–719.

[82] Hartmann, M., Roeraade, J., Stoll, D., Templin, M. F., and Joos, T. O. (2009). Protein microarrays for diagnostic assays. *Analytical and bioanalytical chemistry*, 393(5):1407–1416.

[83] Harwood, C. R. and Cutting, S. M. (1990). *Molecular biological methods for Bacillus*. Wiley.

[84] He, H., Xu, X., and Jin, Y. (2014). Wet-chemical enzymatic preparation and characterization of ultrathin gold-decorated single glass nanopore. *Analytical chemistry*, 86(10):4815–4821.

[85] Hernández-Ainsa, S., Muus, C., Bell, N. A., Steinbock, L. J., Thacker, V. V., and Keyser, U. F. (2013). Lipid-coated nanocapillaries for DNA sensing. *Analyst*, 138(1):104–106.

[86] Hillisch, A., Lorenz, M., and Diekmann, S. (2001). Recent advances in FRET: distance determination in protein–DNA complexes. *Current opinion in structural biology*, 11(2):201–207.

[87] Hojsak, I., Braegger, C., Bronsky, J., Campoy, C., Colomb, V., Decsi, T., Domellöf, M., Fewtrell, M., Mis, N. F., Mihatsch, W., et al. (2015). Arsenic in rice: a cause for concern. *Journal of pediatric gastroenterology and nutrition*, 60(1):142–145.

[88] Holliger, P. and Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature biotechnology*, 23(9):1126.

[89] Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nature biotechnology*, 23(9):1105–1116.

[90] Hornblower, B., Coombs, A., Whitaker, R. D., Kolomeisky, A., Picone, S. J., Meller, A., and Akeson, M. (2007). Single-molecule analysis of DNA-protein complexes using nanopores. *Nature Methods*, 4(4):315.

[91] Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262–1278.

[92] Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., et al. (2013). Dna targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9):827.

[93] Hummer, G. and Szabo, A. (2003). Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical journal*, 85(1):5–15.

[94] Ishikawa, K., Sato, K., Shima, Y., Urabe, I., and Yomo, T. (2004). Expression of a cascading genetic network within liposomes. *FEBS letters*, 576(3):387–390.

[95] Jakobovits, A., Amado, R. G., Yang, X., Roskos, L., and Schwab, G. (2007). From xenomouse technology to panitumumab, the first fully human antibody product from transgenic mice. *Nature biotechnology*, 25(10):1134.

[96] Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690.

[97] Jewett, M. C. and Swartz, J. R. (2004). Rapid Expression and Purification of 100 nmol Quantities of Active Protein Using Cell-Free Protein Synthesis. *Biotechnology progress*, 20(1):102–109.

[98] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, page 1225829.

[99] Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343(6176):1247997.

[100] Joensson, H. N., Samuels, M. L., Brouzes, E. R., Medkova, M., Uhlén, M., Link, D. R., and Andersson-Svahn, H. (2009). Detection and Analysis of Low-Abundance Cell-Surface Biomarkers Using Enzymatic Amplification in Microfluidic Droplets. *Angewandte Chemie International Edition*, 48(14):2518–2521.

[101] Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316(5830):1497–1502.

[102] Kasas, S., Thomson, N. H., Smith, B. L., Hansma, H. G., Zhu, X., Guthold, M., Bustamante, C., Kool, E. T., Kashlev, M., and Hansma, P. K. (1997). Escherichia coli RNA polymerase activity observed using atomic force microscopy. *Biochemistry*, 36(3):461–468.

[103] Kaur, H., Kumar, R., Babu, J. N., and Mittal, S. (2015). Advances in arsenic biosensor development–A comprehensive review. *Biosensors and Bioelectronics*, 63:533–545.

[104] Keller, M. A. and Stiehm, E. R. (2000). Passive immunity in prevention and treatment of infectious diseases. *Clinical microbiology reviews*, 13(4):602–614.

[105] Kesselheim, S., Müller, W., and Holm, C. (2014). Origin of current blockades in nanopore translocation experiments. *Physical review letters*, 112(1):018101.

[106] Keyser, U., Van der Does, J., Dekker, C., and Dekker, N. (2006a). Optical tweezers for force measurements on DNA in nanopores. *Review of Scientific Instruments*, 77(10):105105.

[107] Keyser, U. F., Koeleman, B. N., Van Dorp, S., Krapf, D., Smeets, R. M., Lemay, S. G., Dekker, N. H., and Dekker, C. (2006b). Direct force measurements on DNA in a solid-state nanopore. *Nature Physics*, 2(7):473.

[108] Khalil, A. S. and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367.

[109] Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS letters*, 442(1):15–19.

[110] Kim, D.-M. and Swartz, J. R. (2004). Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. *Biotechnology and bioengineering*, 85(2):122–129.

[111] Kim, S. C., Kannam, S. K., Harrer, S., Downton, M. T., Moore, S., and Wagner, J. M. (2014). Geometric dependence of the conductance drop in a nanopore due to a particle. *Physical Review E*, 89(4):042702.

[112] Kiyoshi, M., Caaveiro, J. M., Tada, M., Tamura, H., Tanaka, T., Terao, Y., Morante, K., Harazono, A., Hashii, N., Shibata, H., et al. (2018). Assessing the Heterogeneity of the Fc-Glycan of a Therapeutic Antibody Using an engineered FcγReceptor IIIa-Immobilized Column. *Scientific reports*, 8(1):3955.

[113] Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T. S., Cantor, C. R., and Collins, J. J. (2004). Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22):8414–8419.

[114] Köhler, G. and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 256(5517):495–497.

[115] Komor, A. C., Badran, A. H., and Liu, D. R. (2017). Crispr-based technologies for the manipulation of eukaryotic genomes. *Cell*, 168(1-2):20–36.

[116] Kotula, J. W., Kerns, S. J., Shaket, L. A., Siraj, L., Collins, J. J., Way, J. C., and Silver, P. A. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proceedings of the National Academy of Sciences*, 111(13):4838–4843.

[117] Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Structure, Function, and Bioinformatics*, 47(3):393–402.

[118] Krummel, B. and Chamberlin, M. J. (1989). RNA chain initiation by Escherichia coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry*, 28(19):7829–7842.

[119] Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *science*, 324(5924):255–258.

[120] Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology*, 32(7):677.

[121] Lambert, J. S. and Stiehm, E. R. (1993). Passive immunity in the prevention of maternal—fetal transmission of human immunodeficiency virus infection. *Annals of the New York Academy of Sciences*, 693(1):186–193.

[122] Laohakunakorn, N., Ghosal, S., Otto, O., Misiunas, K., and Keyser, U. F. (2013a). Dna interactions in crowded nanopores. *Nano letters*, 13(6):2798–2802.

[123] Laohakunakorn, N., Gollnick, B., Moreno-Herrero, F., Aarts, D. G., Dullens, R. P., Ghosal, S., and Keyser, U. F. (2013b). A landau–squire nanojet. *Nano letters*, 13(11):5141–5146.

[124] Laohakunakorn, N. and Keyser, U. F. (2015). Electroosmotic flow rectification in conical nanopores. *Nanotechnology*, 26(27):275202.

[125] Laohakunakorn, N., Thacker, V. V., Muthukumar, M., and Keyser, U. F. (2014). Electroosmotic flow reversal outside glass nanopores. *Nano letters*, 15(1):695–702.

[126] Lee, J. M., Kim, J. A., Yen, T.-C., Lee, I. H., Ahn, B., Lee, Y., Hsieh, C.-L., Kim, H. M., and Jung, Y. (2016). A Rhizavidin Monomer with Nearly Multimeric Avidin-Like Binding Stability Against Biotin Conjugates. *Angewandte Chemie International Edition*, 55(10):3393–3397.

[127] Liu, X., Tang, T.-C., Tham, E., Yuk, H., Lin, S., Lu, T. K., and Zhao, X. (2017). Stretchable living materials and devices with hydrogel–elastomer hybrids hosting programmed cells. *Proceedings of the National Academy of Sciences*, 114(9):2200–2205.

[128] Lynch, H. N., Greenberg, G. I., Pollock, M. C., and Lewis, A. S. (2014). A comprehensive evaluation of inorganic arsenic in food and considerations for dietary intake analyses. *Science of the total environment*, 496:299–313.

[129] Madin, K., Sawasaki, T., Ogasawara, T., and Endo, Y. (2000). A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proceedings of the National Academy of Sciences*, 97(2):559–564.

[130] Maeder, M. L. and Gersbach, C. A. (2016). Genome-editing technologies for gene and cell therapy. *Molecular Therapy*, 24(3):430–446.

[131] Maerkl, S. J. (2009). Integration column: Microfluidic high-throughput screening. *Integrative Biology*, 1(1):19–29.

[132] Maerkl, S. J. and Quake, S. R. (2007). A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315(5809):233–237.

[133] Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., and Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9):833.

[134] Mark, D., Haeberle, S., Roth, G., Von Stetten, F., and Zengerle, R. (2010). Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications. In *Microfluidics Based Microsystems*, pages 305–376. Springer.

[135] Marko, J. F. and Siggia, E. D. (1995). Stretching Dna. *Macromolecules*, 28(26):8759–8770.

[136] Marshall, R., Maxwell, C. S., Collins, S. P., Beisel, C. L., and Noireaux, V. (2017). Short DNA containing χ sites enhances DNA stability and gene expression in E. coli cell-free transcription-translation systems. *Biotechnology and bioengineering*.

[137] Marshall, R., Maxwell, C. S., Collins, S. P., Jacobsen, T., Luo, M. L., Begemann, M. B., Gray, B. N., January, E., Singer, A., He, Y., et al. (2018). Rapid and scalable characterization of CRISPR technologies using an E. coli cell-free transcription-translation system. *Molecular cell*, 69(1):146–157.

[138] Martin, L., Meier, M., Lyons, S. M., Sit, R. V., Marzluff, W. F., Quake, S. R., and Chang, H. Y. (2012). Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nature methods*, 9(12):1192.

[139] McGhee, J. D. and von Hippel, P. H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *Journal of molecular biology*, 86(2):469–489.

[140] McLellan, J. S., Chen, M., Kim, A., Yang, Y., Graham, B. S., and Kwong, P. D. (2010). Structural basis of respiratory syncytial virus neutralization by motavizumab. *Nature Structural and Molecular Biology*, 17(2):248.

[141] Meharg, A. A., Deacon, C., Campbell, R. C., Carey, A.-M., Williams, P. N., Feldmann, J., and Raab, A. (2008). Inorganic arsenic levels in rice milk exceed EU and US drinking water standards. *Journal of Environmental Monitoring*, 10(4):428–431.

[142] Merulla, D., Hatzimanikatis, V., and Meer, J. R. (2013). Tunable reporter signal production in feedback-uncoupled arsenic bioreporters. *Microbial biotechnology*, 6(5):503–514.

[143] Michaud, G. A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P. F., and Schweitzer, B. I. (2003). Analyzing antibody specificity with whole proteome microarrays. *Nature biotechnology*, 21(12):1509.

[144] Miller, O. J., Bernath, K., Agresti, J. J., Amitai, G., Kelly, B. T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D. S., and Griffiths, A. D. (2006). Directed evolution by in vitro compartmentalization. *Nature methods*, 3(7):561.

[145] Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., and Valen, E. (2014). Chopchop: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research*, 42(W1):W401–W407.

[146] Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K., and Giraldez, A. J. (2015). Crisprscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature methods*, 12(10):982.

[147] Morin, S. A., Shepherd, R. F., Kwok, S. W., Stokes, A. A., Nemiroski, A., and Whitesides, G. M. (2012). Camouflage and display for soft machines. *Science*, 337(6096):828–832.

[148] Mulligan, M. E., Hawley, D. K., Entriken, R., and McClure, W. R. (1984). Escherichia coli promoter sequences predict in vitro RNA polymerase selectivity.

[149] Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*, 39(21):9283–9293.

[150] Naujokas, M. F., Anderson, B., Ahsan, H., Aposhian, H. V., Graziano, J. H., Thompson, C., and Suk, W. A. (2013). The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. *Environmental health perspectives*, 121(3):295.

[151] Nelles, D. A., Fang, M. Y., O'Connell, M. R., Xu, J. L., Markmiller, S. J., Doudna, J. A., and Yeo, G. W. (2016). Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell*, 165(2):488–496.

[Neuman and Block] Neuman, K. and Block, S. Optical trapping. *Rev. Sci. Instrum.*

[153] Niederholtmeyer, H., Sun, Z. Z., Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M., and Maerkl, S. J. (2015). Rapid cell-free forward engineering of novel genetic ring oscillators. *Elife*, 4.

[154] Niederholtmeyer, H., Xu, L., and Maerkl, S. J. (2012). Real-time mRNA measurement during an in vitro transcription and translation reaction using binary probes. *ACS synthetic biology*, 2(8):411–417.

[155] Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016). Genetic circuit design automation. *Science*, 352(6281):aac7341.

[156] Nirenberg, M. W. and Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10):1588–1602.

[157] Nobs, J.-B. and Maerkl, S. J. (2014). Long-term single cell analysis of S. pombe on a microfluidic microchemostat array. *PloS one*, 9(4):e93466.

[158] Noireaux, V. and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the national academy of sciences of the United States of America*, 101(51):17669–17674.

[159] Nutiu, R., Friedman, R. C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L., Schroth, G. P., and Burge, C. B. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature biotechnology*, 29(7):659.

[160] O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F., and Segal, D. J. (2015). A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic acids research*, 43(6):3389–3404.

[161] O'Reilly, E. J., Conroy, P. J., Hearty, S., Keyes, T. E., O'Kennedy, R., Forster, R. J., and Dennany, L. (2015). Electrochemiluminescence platform for the detection of C-reactive proteins: application of recombinant antibody technology to cardiac biomarker detection. *RSC Advances*, 5(83):67874–67877.

[162] Oßwald, C., Zipf, G., Schmidt, G., Maier, J., Bernauer, H. S., Muller, R., and Wenzel, S. C. (2012). Modular construction of a functional artificial epothilone polyketide pathway. *ACS synthetic biology*, 3(10):759–772.

[163] Ostrander, E. A., Benedetti, P., and Wang, J. C. (1990). Template supercoiling by a chimera of yeast GAL4 protein and phage T7 RNA polymerase. *Science*, 249(4974):1261–1265.

[164] Otto, O., Steinbock, L., Wong, D., Gornall, J., and Keyser, U. (2011). Note: Direct force and ionic-current measurements on DNA in a nanocapillary. *Review of Scientific Instruments*, 82(8):086102.

[165] Packer, M. S. and Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7):379.

[166] Palacios, M. A., Benito-Peña, E., Manesse, M., Mazzeo, A. D., LaFratta, C. N., Whitesides, G. M., and Walt, D. R. (2011). InfoBiology by printed arrays of microorganism colonies for timed and on-demand release of messages. *Proceedings of the National Academy of Sciences*, 108(40):16510–16514.

[167] Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., and Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell*, 159(4):940–954.

[168] Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., et al. (2016). Rapid, low-cost detection of zika virus using programmable biomolecular components. *Cell*, 165(5):1255–1266.

[169] Pastan, I., Hassan, R., FitzGerald, D. J., and Kreitman, R. J. (2007). Immunotoxin treatment of cancer. *Annu. Rev. Med.*, 58:221–237.

[170] Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., and Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, 31(9):839.

[171] Perry, L. J. and Wetzel, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science*, 226(4674):555–557.

[172] Pfeifer, H.-R. (2011). Arsenic in the different environmental compartments of Switzerland: an updated inventory. *Metals and Related Substances in Drinking Water*, page 250.

[173] Pinheiro, V. B. and Holliger, P. (2014). Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends in biotechnology*, 32(6):321–328.

[174] Piraino, F., Volpetti, F., Watson, C., and Maerkl, S. J. (2016). A digital–analog microfluidic platform for patient-centric multiplexed biomarker diagnostics of ultralow volume samples. *ACS nano*, 10(1):1699–1710.

[175] Plesa, C., Ruitenberg, J. W., Witteveen, M. J., and Dekker, C. (2015). Detection of individual proteins bound along DNA using solid-state nanopores. *Nano letters*, 15(5):3153–3158.

[176] Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D., and Kosuri, S. (2018). Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science*, 359(6373):343–347.

[177] Prindle, A., Samayoa, P., Razinkov, I., Danino, T., Tsimring, L. S., and Hasty, J. (2012). A sensing array of radically coupled genetic 'biopixels'. *Nature*, 481(7379):39.

[178] Purnick, P. E. and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nature reviews Molecular cell biology*, 10(6):410.

[179] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183.

[180] Quan, J., Saaem, I., Tang, N., Ma, S., Negre, N., Gong, H., White, K. P., and Tian, J. (2011). Parallel on-chip gene synthesis and application to optimization of protein expression. *Nature biotechnology*, 29(5):449.

[181] Raillon, C., Cousin, P., Traversi, F., Garcia-Cordero, E., Hernandez, N., and Radenovic, A. (2012). Nanopore detection of single molecule RNAP–DNA transcription complex. *Nano letters*, 12(3):1157–1164.

[182] Rajkumar, A. S., Dénervaud, N., and Maerkl, S. J. (2013). Mapping the fine structure of a eukaryotic promoter input-output function. *Nature genetics*, 45(10):1207.

[183] Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., Walter, J. C., and LaBaer, J. (2004). Self-assembling protein microarrays. *Science*, 305(5680):86–90.

[184] Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., Hu, Y., and LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nature methods*, 5(6):535.

[185] Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, 520(7546):186.

[186] Reddy, S. T., Ge, X., Miklos, A. E., Hughes, R. A., Kang, S. H., Hoi, K. H., Chrysostomou, C., Hunicke-Smith, S. P., Iverson, B. L., Tucker, P. W., et al. (2010). Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature biotechnology*, 28(9):965.

[187] Riedel-Kruse, I. H., Chung, A. M., Dura, B., Hamilton, A. L., and Lee, B. C. (2011). Design, engineering and utility of biotic games. *Lab on a Chip*, 11(1):14–22.

[188] Rivetti, C., Guthold, M., and Bustamante, C. (1999). Wrapping of DNA around the E. coli RNA polymerase open promoter complex. *The EMBO journal*, 18(16):4464–4475.

[189] Robinson, M. K., Shaller, C., Garmestani, K., Plascjak, P. S., Hodge, K. M., Yuan, Q.-A., Marks, J. D., Waldmann, T. A., Brechbiel, M. W., and Adams, G. P. (2008). Effective treatment of established human breast tumor xenografts in immunodeficient mice with a single dose of the α-emitting radioisotope astatine-211 conjugated to anti-HER2/neu diabodies. *Clinical Cancer Research*, 14(3):875–882.

[190] Rocklin, G. J., Chidyausiku, T. M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., Carter, L., Ravichandran, R., Mulligan, V. K., Chevalier, A., et al. (2017). Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*, 357(6347):168–175.

[191] Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J., Mermod, N., and Bucher, P. (2002). High-throughput SELEX–SAGE method for quantitative modeling of transcription-factor binding sites. *Nature biotechnology*, 20(8):831.

[192] Ryabova, L. A., Desplancq, D., Spirin, A. S., and Plückthun, A. (1997). Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. *Nature biotechnology*, 15(1):79.

[193] Sekiya, T., Contreras, R., Takeya, T., and Khorana, H. G. (1979). Total synthesis of a tyrosine suppressor transfer RNA gene. XVII. Transcription, in vitro, of the synthetic gene and processing of the primary transcript to transfer RNA. *Journal of Biological Chemistry*, 254(13):5802–5816.

[194] Setlow, P. (2006). Spores of Bacillus subtilis: their resistance to and killing by radiation, heat and chemicals. *Journal of applied microbiology*, 101(3):514–525.

[195] Sheehan, J. and Marasco, W. A. (2015). Phage and yeast display. In *Antibodies for Infectious Diseases*, pages 105–127. American Society of Microbiology.

[196] Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature biotechnology*, 19(8):751.

[197] Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology*, 15(3):169.

[198] Shong, J., Diaz, M. R. J., and Collins, C. H. (2012). Towards synthetic microbial consortia for bioprocessing. *Current Opinion in Biotechnology*, 23(5):798–802.

[199] Sischka, A., Spiering, A., Khaksar, M., Laxa, M., König, J., Dietz, K.-J., and Anselmetti, D. (2010). Dynamic translocation of ligand-complexed DNA through solid-state nanopores with optical tweezers. *Journal of Physics: Condensed Matter*, 22(45):454121.

[200] Slomovic, S., Pardee, K., and Collins, J. J. (2015). Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences*, 112(47):14429–14435.

[201] Smeets, R. M., Keyser, U. F., Krapf, D., Wu, M.-Y., Dekker, N. H., and Dekker, C. (2006). Salt dependence of ion transport and DNA translocation through solid-state nanopores. *Nano letters*, 6(1):89–95.

[202] Sola, L. and Chiari, M. (2012). Modulation of electroosmotic flow in capillary electrophoresis using functional polymer coatings. *Journal of Chromatography A*, 1270:324–329.

[203] Soni, G. V. and Dekker, C. (2012). Detection of nucleosomal substructures using solid-state nanopores. *Nano letters*, 12(6):3180–3186.

[204] Spiering, A., Getfert, S., Sischka, A., Reimann, P., and Anselmetti, D. (2011). Nanopore translocation dynamics of a single DNA-bound protein. *Nano letters*, 11(7):2978–2982.

[205] Squires, A., Atas, E., and Meller, A. (2015). Nanopore sensing of individual transcription factors bound to DNA. *Scientific reports*, 5:11643.

[206] Stech, M., Hust, M., Schulze, C., Dübel, S., and Kubick, S. (2014). Cell-free eukaryotic systems for the production, engineering, and modification of scfv antibody fragments. *Engineering in life sciences*, 14(4):387–398.

[207] Stech, M., Nikolaeva, O., Thoring, L., Stöcklein, W., Wüstenhagen, D., Hust, M., Dübel, S., and Kubick, S. (2017). Cell-free synthesis of functional antibodies using a coupled in vitro transcription-translation system based on CHO cell lysates. *Scientific reports*, 7(1):12030.

[208] Steinbock, L., Krishnan, S., Bulushev, R., Borgeaud, S., Blokesch, M., Feletti, L., and Radenovic, A. (2014). Probing the size of proteins with glass nanopores. *Nanoscale*, 6(23):14380–14387.

[209] Steinbock, L., Otto, O., Skarstam, D., Jahn, S., Chimerel, C., Gornall, J., and Keyser, U. (2010). Probing DNA with micro-and nanocapillaries and optical tweezers. *Journal of Physics: Condensed Matter*, 22(45):454113.

[210] Steinbock, L., Steinbock, J., and Radenovic, A. (2013). Controllable shrinking and shaping of glass nanocapillaries under electron irradiation. *Nano letters*, 13(4):1717–1723.

[211] Steller, S., Angenendt, P., Cahill, D. J., Heuberger, S., Lehrach, H., and Kreutzberger, J. (2005). Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections. *Proteomics*, 5(8):2048–2055.

[212] Stemmer, W. P. (1994). Dna shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences*, 91(22):10747–10751.

[213] Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 164(1):49–53.

[214] Sternberg, S. H. and Doudna, J. A. (2015). Expanding the biologist's toolkit with CRISPR-Cas9. *Molecular cell*, 58(4):568–574.

[215] Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). Dna interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490):62.

[216] Struhl, K. (1990). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. In *Proteins: Form and Function*, pages 259–266. Elsevier.

[217] Sun, M. G., Seo, M.-H., Nim, S., Corbi-Verge, C., and Kim, P. M. (2016). Protein engineering by highly parallel screening of computationally designed variants. *Science advances*, 2(7):e1600692.

[218] Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of visualized experiments: JoVE*, (79).

[219] Szita, N., Polizzi, K., Jaccard, N., and Baganz, F. (2010). Microfluidic approaches for systems and synthetic biology. *Current opinion in biotechnology*, 21(4):517–523.

[220] Tawfik, D. S. and Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nature biotechnology*, 16(7):652.

[221] Thillaivinayagalingam, P., Gommeaux, J., McLoughlin, M., Collins, D., and Newcombe, A. R. (2010). Biopharmaceutical production: Applications of surface plasmon resonance biosensors. *Journal of Chromatography B*, 878(2):149–153.

[222] Tian, D., Battles, M. B., Moin, S. M., Chen, M., Modjarrad, K., Kumar, A., Kanekiyo, M., Graepel, K. W., Taher, N. M., Hotard, A. L., et al. (2017). Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein. *Nature communications*, 8(1):1877.

[223] Traggiai, E., Becker, S., Subbarao, K., Kolesnikova, L., Uematsu, Y., Gismondo, M. R., Murphy, B. R., Rappuoli, R., and Lanzavecchia, A. (2004). An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. *Nature medicine*, 10(8):871–875.

[224] Trepagnier, E. H., Radenovic, A., Sivak, D., Geissler, P., and Liphardt, J. (2007). Controlling DNA capture and propagation through artificial nanopores. *Nano letters*, 7(9):2824–2830.

[225] Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, 33(2):187.

[226] Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435(7042):646.

[227] Van Der Meer, J. R. and Belkin, S. (2010). Where microbiology meets microengineering: design and applications of reporter bacteria. *Nature Reviews Microbiology*, 8(7):511.

[228] Venkatesan, B. M., Estrada, D., Banerjee, S., Jin, X., Dorgan, V. E., Bae, M.-H., Aluru, N. R., Pop, E., and Bashir, R. (2011). Stacked graphene-Al2O3 nanopore sensors for sensitive detection of DNA and DNA–protein complexes. *ACS nano*, 6(1):441–450.

[229] Völkel, T., Korn, T., Bach, M., Müller, R., and Kontermann, R. E. (2001). Optimized linker sequences for the expression of monomeric and dimeric bispecific single-chain diabodies. *Protein engineering*, 14(10):815–823.

[230] Volpetti, F., Petrova, E., and Maerkl, S. J. (2017). A microfluidic biodisplay. *ACS synthetic biology*, 6(11):1979–1987.

[231] Wang, B., Barahona, M., and Buck, M. (2014). Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks. *Nucleic acids research*, 42(14):9484–9492.

[232] Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophysical journal*, 72(3):1335–1346.

[233] Wanunu, M., Dadosh, T., Ray, V., Jin, J., McReynolds, L., and Drndić, M. (2010). Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nature nanotechnology*, 5(11):807.

[234] Whitesides, G. M. (2006). The origins and the future of microfluidics. *Nature*, 442(7101):368.

[235] Wijma, H. J., Floor, R. J., Jekel, P. A., Baker, D., Marrink, S. J., and Janssen, D. B. (2014). Computationally designed libraries for rapid enzyme stabilization. *Protein Engineering, Design and Selection*, 27(2):49–58.

[236] Wilks, H. M., Halsall, D. J., Atkinson, T., Chia, W. N., Clarke, A. R., and Holbrook, J. J. (1990). Designs for a broad substrate specificity keto acid dehydrogenase. *Biochemistry*, 29(37):8587–8591.

[237] Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., and Holbrook, J. J. (1988). A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science*, 242(4885):1541–1544.

[238] Wilks, H. M., Moreton, K. M., Halsall, D. J., Hart, K. W., Sessions, R. D., Clarke, A. R., and Holbrook, J. J. (1992). Design of a specific phenyllactate dehydrogenase by peptide loop exchange on the Bacillus stearothermophilus lactate dehydrogenase framework. *Biochemistry*, 31(34):7802–7806.

[239] Willats, W. G. (2002). Phage display: practicalities and prospects. *Plant molecular biology*, 50(6):837–854.

[240] Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M., and Smith, M. (1982). Redesigning enzyme structure by site-directed mutagenesis: tyrosyl trna synthetase and atp binding. *Nature*, 299(5885):756.

[241] Woodruff, K. and Maerkl, S. J. (2017). Microfluidic Module for Real-Time Generation of Complex Multimolecule Temporal Concentration Profiles. *Analytical chemistry*, 90(1):696–701.

[242] Woodside, M. T., Anthony, P. C., Behnke-Parks, W. M., Larizadeh, K., Herschlag, D., and Block, S. M. (2006). Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001–1004.

[243] Wright, A. V., Nuñez, J. K., and Doudna, J. A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, 164(1-2):29–44.

[244] Wright, O., Stan, G.-B., and Ellis, T. (2013). Building-in biosafety for synthetic biology. *Microbiology*, 159(7):1221–1235.

[245] Wu, H., Pfarr, D. S., Johnson, S., Brewah, Y. A., Woods, R. M., Patel, N. K., White, W. I., Young, J. F., and Kiener, P. A. (2007). Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *Journal of molecular biology*, 368(3):652–665.

[246] Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Cheng, A. W., Trevino, A. E., Konermann, S., Chen, S., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology*, 32(7):670.

[247] Xiong, A.-S., Peng, R.-H., Zhuang, J., Gao, F., Li, Y., Cheng, Z.-M., and Yao, Q.-H. (2008). Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS microbiology reviews*, 32(3):522–540.

[248] Yin, G., Garces, E. D., Yang, J., Zhang, J., Tran, C., Steiner, A. R., Roos, C., Bajad, S., Hudak, S., Penta, K., et al. (2012). Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. In *MAbs*, volume 4, pages 217–225. Taylor & Francis.

[249] Yin, G. and Swartz, J. R. (2004). Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnology and bioengineering*, 86(2):188–195.

[250] Young, L. and Dong, Q. (2004). Two-step total gene synthesis method. *Nucleic acids research*, 32(7):e59–e59.

[251] Yu, J.-S., Lim, M.-C., Huynh, D. T. N., Kim, H.-J., Kim, H.-M., Kim, Y.-R., and Kim, K.-B. (2015). Identifying the location of a single protein along the DNA Strand Using Solid-State Nanopores. *ACS nano*, 9(5):5289–5298.

[252] Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., et al. (2001). Global analysis of protein activities using proteome chips. *science*, 293(5537):2101–2105.

[253] Zhu, L. J., Holmes, B. R., Aronin, N., and Brodsky, M. H. (2014). Crisprseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PloS one*, 9(9):e108424.

Appendix A

Oligonucleotide and gBlocks used in the studies

VH11	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAAATGCAGGTGACCCTGAGGGAGAGCGGC CCCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGC ACCTTCAGCGGCTTCAGCCTGAGCACCGCCGGCATGAGCGTG GGCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTG GCCGACATCTGGTGGGACGACAAGAAGGACTACAACCCCAG CCTGAAGAGCAGGCTGACCATCAGCAAGGACACCAGCAAGA ACCAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACA CCGCCACCTACTACTGCGCCAGGAGCATGATCACCAACTGGT ACTTCGACGTGTGGGGCCA
VH12	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAAatgCAGGTGACCCTGAGGGAGAGCGGCC CCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGCA CCTTCAGCGGCTTCAGCCTGAGCACCAGCGGCATGAGCGTGG GCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTGG CCGACATCTGGTGGGACGACAAGAAGGACTACAACCCCAGCC TGAAGAGCAGGCTGACCATCAGCAAGGACACCAGCAAGAAC CAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACACC GCCACCTACTACTGCGCCAGGAGCATGATCACCAACTTCTACT TCGACGTGTGGGGCCA

VH18	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAatgCAGGTGACCCTGAGGGAGAGCGGCC CCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGCA CCTTCAGCGGCTTCAGCCTGAGCACCAGCGGCATGAGCGTGG GCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTGG CCGACATCTGGTGGGACGACAAGAAGGACTACAACCCCAGCC TGAAGAGCAGGCTGACCATCAGCAAGGACACCAGCAAGAAC CAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACACC GCCACCTACTACTGCGCCAGGAGCATGATCACCAACTGGTAC TTCGACGTGTGGGGCCA
VH13	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAatgCAGGTGACCCTGAGGGAGAGCGGCC CCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGCA CCTTCAGCGGCTTCAGCCTGAGCACCGCCGGCATGAGCGTGG GCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTGG CCGACATCTGGTGGGACGACAAGAAGGACTACAACCCCAGCC TGAAGAGCAGGCTGACCATCAGCAAGGACACCAGCAAGAAC CAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACACC GCCACCTACTACTGCGCCAGGAGCATGATCACCAACTTCTACT TCGACGTGTGGGGCCA
VL14	GGGCACCACCGTGACCGTGAGCAGCGGTGGCGGAGGCTCCG GCGGTGGTGGATCTGGAGGTGGTGGTAGCGACATCCAGATGA CCCAGAGCCCCAGCACATTGTCTGCTAGTGTCGGCGACCGGG TGACGATAACGTGTTCCGCGTCAAGCTCAGTGGGATACATGC ACTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTGA TCTACGACACCTTCAAGCTGGCCAGCGGCGTGCCCAGCAGGT TCAGCGGCAGCGGCAGCGGCACCGAGTTCACCCTGACCATCA GCAGCCTGCAGCCCGACGACTTCGCCACCTACTACTGCTTCCA GGGCAGCTTCTACCCCTTCACCTTCGGCGGCGGCACCAAGGT GGAGATCAAGCCAGCACCTGCTCCA

VL15	GGGCACCACCGTGACCGTGAGCAGCGGTGGCGGAGGCTCCG GCGGTGGTGGATCTGGAGGTGGTGGTAGCGACATCCAGATGA CCCAGAGCCCCAGCACCCTGAGCGCCAGCGTGGGCGACAGG GTGACCATCACCTGCAGCGCCAGCAGCAGCGTGGGCTACATG CACTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTG ATCTACGACACCAGCAAGCTGGCCAGCGGCGTGCCCAGCAGG TTCAGCGGCAGCGGCAGCGGCACCGAGTTCACCCTGACCATC AGCAGCCTGCAGCCCGACGACTTCGCCACCTACTACTGCTTCC AGGGCAGCGGCTACCCCTTCACCTTCGGCGGCGGCACCAAGG TGGAGATCAAGCCAGCACCTGCTCCA
VL16	GGGCACCACCGTGACCGTGAGCAGCGGTGGCGGAGGCTCCG GCGGTGGTGGATCTGGAGGTGGTGGTAGCGACATCCAGATGA CCCAGAGCCCCAGCACCCTGAGCGCCAGCGTGGGCGACAGG GTGACCATCACCTGCAAGTGCCAGCTGAGCGTGGGCTACATG CACTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTG ATCTACGACACCAGCAAGCTGGCCAGCGGCGTGCCCAGCAGG TTCAGCGGCAGCGGCAGCGGCACCGAGTTCACCCTGACCATC AGCAGCCTGCAGCCCGACGACTTCGCCACCTACTACTGCTTCC AGGGCAGCGGCTACCCCTTCACCTTCGGCGGCGGCACCAAGC TGGAGATCAAGCCAGCACCTGCTCCA
VL17	GGGCACCACCGTGACCGTGAGCAGCGGTGGCGGAGGCTCCG GCGGTGGTGGATCTGGAGGTGGTGGTAGCGACATCCAGATGA CCCAGAGCCCCAGCACACTCTCTGCCTCTGTAGGGGACCGCG TGACCATCACCTGCTCAGCAAGCTCCAGTGTGGGCTACATGC ACTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTGA TCTACGACACCAGCAAGCTGGCCAGCGGCGTGCCCAGCAGGT TCAGCGGCAGCGGCAGCGGCACCGAGTTCACCCTGACCATCA GCAGCCTGCAGCCCGACGACTTCGCCACCTACTACTGCTTCCA GGGCAGCTTCTACCCCTTCACCTTCGGCGGCGGCACCAAGGT GGAGATCAAGCCAGCACCTGCTCCA

Table A.1 VH and VL DNA templates used for scFv gene assembly

Mota CDRs according to Clothria numbering scheme grafted in scFv-EGFR scaffold	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAATGGAGGTGCAGCTGGTGGAGAGCGGC GGCGGCCTGGTGAGGCCCGGCGGCAGCCTGAAGCTGAGCTGC GCCGCCAGCGGCTTCGCCTTCAGCACCTTCAGCGGCTTCAGCC TGAGCACCGCCGGCATGAGCGTGGGCTGGGTGAGGCAGACCC CCGAGAAGAGGCTGGAGTGGGTGGCCGACATCTGGTGGGAC GACAAGAAGCACAGGTTCACCATCAGCAGGGACAACGCCGA GAACACCCTGTACCTGCAGATGAGCAGCCTGAAGAGCGAGGA CACCGCCATCTACTACTGCGCTTCTGCCAGGGACATGATCTTC AACTTCTACTTCGACGTGTGGGGCCAGGGCACCCTGGTGACC GTGAGCGCCGGTGGCGGAGGCTCCGGCGGTGGTGGATCTGGA GGTGGTGGTAGCGACGTGGTGATGACCCAGTCTCCCCTCAGT CTTCCAGTGAGCCTTGGGGATCAGGCGACCATATCATGTTCA GCAAGTAGCCGCGTGGGCTACATGCACTGGTACCTGCAGAAG GCCGGCCAGAGCCCCAAGCTGCTGATCTACTACGACACCAGC AAGCTGGCCAGCGGCGTGCCCGACAGGTTCAGCGGCAGCGGC AGCGGCACCGACTTCACCCTGAAGATCAGCAGGGTGGAGGCC GAGGACCTGGGCGTGTACTTCTGCTTCCAGGGCAGCGGCTAC CCCTTCACCTTCGGCGGCGGCACCAACCTGGAGATCAAGAGG GCCGACGCCGCCCCAGCGCCAGCGCCA

Mota CDRs according to Kabat numbering scheme grafted in scFv-EGFR scaffold	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAATGGAGGTGCAGCTGGTGGAGAGCGGC GGCGGCCTGGTGAGGCCCGGCGGCAGCCTGAAGCTGAGCTGC GCCGCCAGCGGCTTCGCCTTCAGCACCGCCGGCATGAGCGTG GGCTGGGTGAGGCAGACCCCCGAGAAGAGGCTGGAGTGGGT GGCCGACATCTGGTGGGACGACAAGAAGCACTACAACCCCA GCCTGAAGGACAGGTTCACCATCAGCAGGGACAACGCCGAG AACACCCTGTACCTGCAGATGAGCAGCCTGAAGAGCGAGGAC ACCGCCATCTACTACTGCGCCAGCGACATGATCTTCAACTTCT ACTTCGACGTGTGGGGCCAGGGCACCCTGGTGACCGTGAGCG CCGGTGGCGGAGGCTCCGGCGGTGGTGGATCTGGAGGTGGTG GTAGCGACGTGGTGATGACCCAGTCACCCCTCAGTCTTCCGGT CAGCCTTGGTGACCAGGCAACCATTAGCTGCTCTGCATCATCT CGGGTGGGCTACATGCACTGGTACCTGCAGAAGGCCGGCCAG AGCCCCAAGCTGCTGATCTACGACACCAGCAAGCTGGCCAGC GGCGTGCCCGACAGGTTCAGCGGCAGCGGCAGCGGCACCGA CTTCACCCTGAAGATCAGCAGGGTGGAGGCCGAGGACCTGGG CGTGTACTTCTGCTTCCAGGGCAGCGGCTACCCCTTCACCTTC GGCGGCGGCACCAACCTGGAGATCAAGAGGGCCGACGCCGC CCCAGCGCCAGCGCCA

scFv-EGFR	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAatgCAGGTGACCCTGAGGGAGAGCGGCC CCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGCA CCTTCAGCGGCTTCAGCCTGAGCACCAGCGGCATGAGCGTGG GCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTGG CCGACATCTGGTGGGACGACAAGAAGGACTACAACCCCAGCC TGAAGAGCAGGCTGACCATCAGCAAGGACACCAGCAAGAAC CAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACACC GCCACCTACTACTGCGCCAGGAGCATGATCACCAACTGGTAC TTCGACGTGTGGGGCGCCGGCACCACCGTGACCGTGAGCAGC GGTGGCGGAGGCTCCGGCGGTGGTGGATCTGGAGGTGGTGGT AGCGACATCCAGATGACCCAGAGCCCCAGCACCCTGAGCGCC AGCGTGGGCGACAGGGTGACCATCACCTGCAAGTGCCAGCTG AGCGTGGGCTACATGCACTGGTACCAGCAGAAGCCCGGCAAG GCCCCCAAGCTGCTGATCTACGACACCAGCAAGCTGGCCAGC GGCGTGCCCAGCAGGTTCAGCGGCAGCGGCAGCGGCACCGA GTTCACCCTGACCATCAGCAGCCTGCAGCCCGACGACTTCGC CACCTACTACTGCTTCCAGGGCAGCGGCTACCCCTTCACCTTC GGCGGCGGCACCAAGCTGGAGATCAAGccagcacctgctcca

Mota	GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGA CCACAACGGTTTCCCTCTAGAGATAATTTTGTTTAACTTAAGA AGGAGGAAAAAAAAATGCAGGTGACCCTGAGGGAGAGCGGC CCCGCCCTGGTGAAGCCCACCCAGACCCTGACCCTGACCTGC ACCTTCAGCGGCTTCAGCCTGAGCACCGCCGGCATGAGCGTG GGCTGGATCAGGCAGCCCCCCGGCAAGGCCCTGGAGTGGCTG GCCGACATCTGGTGGGACGACAAGAAGCACTACAACCCCAGC CTGAAGGACAGGCTGACCATCAGCAAGGACACCAGCAAGAA CCAGGTGGTGCTGAAGGTGACCAACATGGACCCCGCCGACAC CGCCACCTACTACTGCGCCAGGGACATGATCTTCAACTTCTAC TTCGACGTGTGGGGCCAGGGCACCACCGTGACCGTGAGCAGC GGTGGCGGAGGCTCCGGCGGTGGTGGATCTGGAGGTGGTGGT AGCGACATCCAGATGACCCAGAGCCCCAGCACCCTGAGCGCC AGCGTGGGCGACAGGGTGACCATCACCTGCAGCGCCAGCAGC AGGGTGGGCTACATGCACTGGTACCAGCAGAAGCCCGGCAAG GCCCCCAAGCTGCTGATCTACGACACCAGCAAGCTGGCCAGC GGCGTGCCCAGCAGGTTCAGCGGCAGCGGCAGCGGCACCGA GTTCACCCTGACCATCAGCAGCCTGCAGCCCGACGACTTCGC CACCTACTACTGCTTCCAGGGCAGCGGCTACCCCTTCACCTTC GGCGGCGGCACCAAGGTGGAGATCAAGCCAGCACCTGCTCCA

Table A.2 Sequence of Mota scaffold variants

Ekaterina Petrova

PhD in Bioengineering

Address Avenue de Montchoisi 51,
Lausanne, 1006
Switzerland

E-mail ekaterina.epetrova@gmail.com

EDUCATION

May 2012 – present

École Polytechnique Fédérale de Lausanne (EPFL), PhD in Biotechnology and Bioengineering

PhD Thesis: 'Development of microfluidic tools and methods for the biochemical characterization of scFv and Cas9 affinity reagents and their applications in molecular detection.'

Supervisor Prof. Sebastian Maerkl

Sep 2010 – May 2012

Karolinska Institutet, MSc in Biomedicine

Master thesis: 'Spatial control of membrane receptor function using ligand nanocalipers.'

Supervisor Prof. Ana Teixeira

- evaluated mammalian cell behaviour in response to the spatially organized microenvironmental cues by DNA origami nanoparticles.

Master project: 'Design and evaluation of hyperbranched polymer based nanoparticles for cancer therapy.' *Supervisor Prof. Andreas Nystrom*

- studied hyperbranched polymer based nanoparticles cytotoxicity and major routes of internalization in breast cancer cells.

Sep 2006 – Jun 2010

Sofia University 'St. Kliment Ohridski', BSc in Molecular Biology

In parallel to my Bachelor studies, I was volunteering in the National Genetic Laboratory in the University Hospital of Obstetrics and Gynaecology in Sofia under the supervison of Prof. Alexey Savov. I helped in:

- genotyping and studying inherited neuropathies (Charcot-Marie-Tooth disease, Ducheene muscular dystrophy, etc.).
- a clinical trial of EGFR inhibitors for patients with non-small-cell lung cancer and colon carcinomas.

PROFESSIONAL EXPERIENCE

Apr 2013 – present

PhD Student, EPFL

- treamlined the assembly, synthesis and characterization of small-chain antibody fragments (scFvs) by intergeted microfluidic platform
- developed a genetically encoded scFv-based protein biosensor
- in vitro reconstructed CRISPR-Cas9 system and explore its properties on a microfluidic device
- developed microfluidic biodisplay platform for bacteria screening and small molecule detection
- supervised more than 6 Bachelor, Master and intern students

Apr 2013 – Nov 2017

Graduate Teaching Assistant for iGEM (International Genetically Engineered Machine competition), EPFL

- guided and supervised the students
  - EPFL 2013 team qualified for World Championship Jamboree at MIT (USA) for their project on smart drug delivery system.
  - EPFL 2014 team won a gold medal at the World Championship Jamboree at MIT (USA) for their project on development of a microfluidics platform 'BioPad' based on bacteria mechanosensors.
  - EPFL 2017 team won a gold medal and an award at the World Championship Jamboree at MIT for their 'Aptasense' project on the developement of an microfluidic aptamer-based biosensor.

Jun 2012 – Dec 2012

Research Assistant, Karolinska Institutet

After graduation, I continued to work on my Master's thesis project which resulted in a multidisciplinary oriented paper.

SKILLS

- molecular cloning, CRISPR-Cas9 technology, gene assembly strategies
- cell handling and culturing (bacteria and mammalian cells)
- DNA, RNA and protein purification; in vitro protein expression
- microfluidics (soft lithography fabrication, operation and surface functionalization), deposition of DNA, protein and cells by microarray spotters (ArrayIt)
- microfluidic-based protein-protein and protein-DNA interrogations , bead-based assays for detection of biomolecular interactions, conventional ELISA, SDS-PAGE, Western blot
- fluorescent and confocal microscopy, microplate reader, image processing

SOFTWARE

- Python, R (Basic level), ImageJ, Imaris (Basic), Abobe Suite (Illustrator, Photosop), MS Office

LANGUAGES

- English (Proficiency), Bulgarian (Mother tongue), Spanish (Basic)

CERTIFICATES & COURSES

- Certificates: FELASA B
- PhD courses: Reliability of MEMS, BioMEMS, Data analysis and model classification, Immunology and bioengineering applications

PUBLICATIONS

- 2017 — Volpetti F*, Petrova E*, Maerkl SJ. (2017) A Microfluidic Biodisplay. ACS Synth Biol.; 6(11):1979-1987
- 2016 — Bulushev RD*, Marion S*, Petrova E*, Davis SJ*, Maerkl SJ, Radenovic A. (2016) Single Molecule Localization and Discrimination of DNA-Protein Complexes by Controlled Translocation Through Nanocapillaries. Nano Lett., 16(12):7882-7890
- 2016 — Blackburn MC, Petrova E, Correia BE, Maerkl SJ. (2016) Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. Nucleic Acids Res, 44(7):e68
- 2014 — Shaw A, Lundin V, Petrova E, Fördős F, Benson E, Al-Amin A, Herland A, Blokzijl A, Högberg B, Teixeira AI. (2014) Spatial control of membrane receptor function using ligand nanocalipers. Nat Methods, 11(8):841-6

INTERESTS

- I enjoy various sport activities including: swimming, skiing, paddling, windsurfing, squash.