ENGINEERING DNA GELS FOR CELL-FREE PROTEIN PRODUCTION AND CAPTURE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jason Samuel Kahn

August 2014

ENGINEERING DNA GELS FOR CELL-FREE PROTEIN PRODUCTION AND CAPTURE

Jason Samuel Kahn, Ph.D.

Cornell University 2014

One of the fundamental goals of biological engineering is the harnessing of biological systems to produce desired products. Protein production of exogenous proteins in living cells has long been a staple of molecular biology. However, living biological systems present fundamental limitations as the scientists' desire to produce more complex and varying molecules in cells competes with normal cell processes. Ideally, one can isolate the required cell pathway away from the living system in order to explore the full range of possible molecular permutations allowed by chemistry without the limitations of biology. *In vitro* protein production allows life's central dogma to be performed outside the confines of a cell, creating the possibility of producing toxic proteins and testing full mutation spaces in the DNA-RNA-Protein pathway. This possibility also opens up the need for materials to interface with genes and protein in an *in vitro* platform. Our lab has engineered a DNA gel that interacts with the gene-of-interest to increase protein yields while protecting the gene from degradation. DNA has long been investigated as a genetic material, but only in the pat decade has its vast potential as a generic polymer been elucidated. Beyond its monodispersity, the specificity of binding interactions in Watson-Crick base-pairing

allows a unique level of control over structure and material organization at the nanoscale. In creating networked DNA structures that can incorporate genes into the gel network, we create protein-producing gels. However, further engineering of DNA gels is required in order to produce a system more comparable to the morphology and functionality of a cell, notably the isolation of gene sets and the ability to connect genotype and phenotype when testing protein activity. This thesis work discusses methods to create cell-size DNA gels that possess the ability to both produce and capture protein, practical considerations for DNA manipulation in cell-free protein production, and provides insight into further functionalization of DNA gels to provide more diverse applications in the context of protein engineering.

BIOGRAPHICAL SKETCH

Jason received his B.S. degree in Biological and Environmental Engineering from Cornell University in 2008. As an undergraduate, he had the opportunity to work in a few lab settings, and as a junior he began working with two postdoctoral associates, now professors, Wenlong Cheng and Professor Jianfeng Xu in the laboratory of Professor Dan Luo. As an engineering student, Jason was interested in the control of biological systems and molecules to enable applications outside their typical roles in biomedical applications. In particular, the concept of using DNA as a material and molecular organizer appealed to his desire to explore the potential of biological systems within an engineering context.

Considering the great match between Jason's research interests and the Luo laboratory, he made the decision to pursue my M.S/Ph.D. at Cornell University under the guidance of Professor Luo. Jason also had the opportunity to conduct a portion of his work at the Institute of Bioengineering and Nanotechnology in Singapore under Professor Yi Yan Yang. During his tenure in Professor Luo's lab he explored the use of DNA as a both a genetic and generic material in applications such as nanoparticle organization and crystallization, biosensing, and DNA gel formation and protein production. He will be joining the group of Professor Itamar Willner in the Institute of Chemistry at Hebrew University of Jerusalem from September 2014 as a Lady Davis Fellowship Trust Post-Doctoral Fellow.

To my family and friends

ACKNOWLEDGEMENTS

This thesis represents the culmination of a nearly decade-long journey at Cornell, starting as an undergraduate in the fall of 2004 and continuing with my M.S./Ph.D. work in the fall of 2008. This time may of course be marked by my academic work and research, but even more so by my experiences with people who have provided support, affirmation, and guidance during these years and have played a role in shaping who I am today.

First off, I would like to acknowledge the Department of Energy for not just their financial, but also academic support through the Office of Science Graduate Fellowship. During the program's annual meetings, I had the opportunity to interact with amazing fellows and gain a greater background on how the United States is addressing current and future energy issues across a wide variety of disciplines. Additionally, I would like to thank the National Science Foundation for providing me the chance to conduct a semester of my graduate work at the Institute of Bioengineering and Nanotechnology in Singapore. I did not study abroad for any period as an undergraduate, and this time in Singapore allowed me to grow both inside and outside of the laboratory.

The staff and faculty Department of Biological and Environmental Engineering has always been there for support during my undergraduate and graduate careers. I started research in Professor Dan Luo's as an undergraduate, and as my academic advisor and chair of my committee, his guidance has shaped my work at Cornell and

given me insight into successfully making science interesting and engaging to a larger audience. I am grateful to my committee members Professor John March and Professor Ulrich Wiesner for their advice and time whenever I sought their views. I would also like to thank Professor Wenlong Cheng and Professor Jianfeng Xu, who were postdocs in the laboratory when I started as an undergraduate and provided hands-on help as I became more comfortable in the lab setting.

Nothing I can write on paper would begin to express how grateful I am for my family, who has always been a source of encouragement and guidance. To feel you have your family behind you is a blessing that I have been lucky to have through my time at Cornell. I have always appreciated the fact that my parents, my brother, and my sister can be honest with each other, even if we disagree, with the knowledge that everyone is always looking out for the other.

Of course, Ithaca wouldn't have been Ithaca if not for my friends. Within the lab itself, a group of us became very close. I would like to thank Shawn Tan, Natt Kiatwuthinon, Thua Tran, Michael Campolongo, Thomas Derrien, Mark Hartman, Edward Rice, Songming Peng, and Mervin Zhao for the camaraderie during the long hours. In particular, Shawn Tan was always there to act as a sounding board for new ideas. Brenda Marchewka, the student services coordinator for our department, was there to help me navigate my requirements and provide encouragement (and candy) should I wander into her office. Throughout the years I have also had the chance to work with wonderful undergraduate researchers, notably Matthew Mikhail, Amelia Adams and Swati Sureka. It was a joy getting to know them and to have them involved in my projects, and I know they have bright futures ahead of them as they

move onto their careers. Outside of lab, I worked as a Graduate Resident Fellow at Hans Bethe House for two years, and would like to thank Professor Scott MacDonald, Erica Ostermann, Denise Shaw, and my fellow GRF's for the amazing West Campus community of which I loved being a part.

Last but not least, I would like to truly thank all of my friends who over the years have become family to me, and in doing so, have made Ithaca a home. I am not going to go through everybody's names individually, as so many people have defined my time in Ithaca and sometimes it is often the small interactions and gestures that make all the difference, but I would like to think they know who they are. I believe that people do really make a place, and I could not have asked for a greater group of people to define my experience. I will reflect fondly on my time spent at Cornell and in Ithaca, but am excited to keep the connections I have made with so many wonderful people.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH ..... **iii**

ACKNOWLEDGEMENTS ..... **v**

TABLE OF CONTENTS ..... **viii**

LIST OF FIGURES ..... **ix**

LIST OF TABLES ..... **xi**

CHAPTER I: INTRODUCTION **1**

SECTION 1.1 - INTRODUCTION TO CELL-FREE PROTEIN PRODUCTION SYSTEMS AND DIRECTED EVOLUTION PLATFORMS 1

SECTION 1.2 - ENGINEERING DNA MATERIALS 5

SECTION 1.3 - CELL-FREE EXPRESSION OF PROTEINS 13

SECTION 1.4 - SIGNIFICANCE OF THIS DISSERTATION 24

CHAPTER 2: SYNTHESIS OF DNA GEL MICRODROPLETS **27**

SECTION 2.1 - MOVING FROM BULK GEL TO MICROGELS 27

SECTION 2.2 - MICROFLUIDIC DEVICE FABRICATION AND TESTING 28

SECTION 2.3 - SUMMARY 39

SECTION 2.4 - MATERIALS AND METHODS 40

CHAPTER 3: PROTEIN EXPRESSION – PLASMID AND X-DNA DESIGN **41**

SECTION 3.1 - PLASMID SELECTION AND LINEARIZATION 41

SECTION 3.2 - SELECTION OF RED FLUORESCENT PROTEIN 47

SECTION 3.3 - GEL FORMATION BASED ON STICKY END SEQUENCE 51

SECTION 3.4 - SUMMARY 62

SECTION 3.5 - MATERIALS AND METHODS 63

CHAPTER 4: GEL FUNCTIONALIZATION **67**

SECTION 4.1 - PRODUCTION OF HOMEMADE E.COLI LYSATE 67

SECTION 4.2 - NTA-FUNCTIONALIZATION OF X-DNA 71

SECTION 4.3 - TESTING OF FUNCTIONAL GROUP PRESENCE AND FUNCTIONALITY 73

SECTION 4.4 - INCORPORATION OF NTA INTO DNA GEL 77

SECTION 4.5 - DNA BIRD NEST FUNCTIONALIZATION 79

SECTION 4.6 - INCREASING THERMOSTABILITY OF DNA-BASED GELS 87

SECTION 4.7 - SUMMARY 104

SECTION 4.8 - MATERIALS AND METHODS 106

CHAPTER 5: CONCLUSION AND FUTURE PERSPECTIVE **111**

BIBLIOGRAPHY **117**

LIST OF FIGURES

Figure 1: Formats of the various DNA topologies explored in the Luo laboratory.....8
Figure 2: Branched DNA structures can be crosslinked to form a hydrogel.....11
Figure 3: Comparison of protein production between P-gel and solution-phase (SPS) expression.....12
Figure 4: A comparison of display technologies.....21
Figure 5: Overview of the microfluidic device construction.....31
Figure 6: Overview of P-gel microdroplet expression.....33
Figure 7: P-gel droplet diameters for 35 μm channel width at 0.8 flow rate ratio.....34
Figure 8: Microfluidic filters at the head of each input and image of the double junction.....34
Figure 9: Breaking of emulsions.....38
Figure 10: DNA microgels settled at the bottom of the extracted aqueous phase.....38
Figure 11: P-gel microdroplet imaging.....39
Figure 12: MluI site addition confirmation.....45
Figure 13: Colony growth showing the expression of red fluorescent proteins.....50
Figure 14: Fluorescence comparison between three red fluorescent proteins.....51
Figure 15: Solutions of (A) dsRed (B) mCherry (C) and tdTomato expressed in cell-free lysate.....51
Figure 16: Views of X-DNA droplets in emulsions and after extraction into bulk aqueous phase.....57
Figure 17: Comparison of bulk gel formation using X-DNA possessing four (A) AatII, (B) ApaI, (C) BspEI, and (D) MluI overhangs.....59
Figure 18: Cell-free expression comparison of GFP in solution using different plasmid linearization sites.....60
Figure 19: Cell-free expression levels of pIVEX2.3d-GFP linearized at different cutting sites.....60
Figure 20: 2.5% agarose gel showing that all X-DNA with different sticky ends form in high yield.....61
Figure 21: Cell-free expression of GFP in linear (MluI-digested) and circular plasmid formats.....62
Figure 22: Four single-stranded oligos are annealed to form X-DNA.....64
Figure 23: Plasmid map of pIVEX2.3d-GFP and mutated MluIpIVEX2.3d-GFP.....65
Figure 24: Agarose-NTA-Ni2+ gel beads incubated with GFP-6His in (A) traditional cell-free lysate and (B) cell-free lysate without DTT.....68
Figure 25: Optical microscopy showing the effects of lysate incubation on the P-gel microdroplets.....70
Figure 26: Chemical structure of maleimido-C3-NTA (MC3N).....72
Figure 27: Overview of the MC3N functionalization procedure.....73
Figure 28: Mass spectroscopy results for MC3N-modified DNA.....74
Figure 29: 2.5% agarose gel showing X-DNA and MC3N-functionalized X-DNA.....76
Figure 30: GFP retained after X-DNA incubation with GFP-His.....77

Figure 31: A.) MC3N-functionalized and B.) unmodified P-gel micrdoplets incubated with GFP-His.....79
Figure 32: SEM images of Gel1 before (A) and after (B) mechanical breakup of gel linkages to form DNA bird nests.....80
Figure 33: Synthesis of bird nest DNA gel.....83
Figure 34: MC3N-functionalized and non-functionalized DNA bird nests incubated with GFP-6His.....83
Figure 35: DNA bird nests after incubation in lysate containing GFP-His.....86
Figure 36: Comparison of GFP expression levels using bird nest DNA gels.....86
Figure 37: Depiction of longest continuous hybridized strands in ligated X-DNA.....88
Figure 38: Effect of psoralen crosslinking on ligation efficiency.....93
Figure 40: Testing the ability of redesigned X-oligos to form branched X-DNA.....94
Figure 41: X-DNA and psoralen-crosslinked X-DNA stability in SDS-PAGE gel.....98
Figure 42: SDS-PAGE gel comparing different UV exposure times for psoralen crosslinking.....99
Figure 43: 2.5% agarose gel comparing the dehybridization of X-DNA and psoralen-crosslinked X-DNA in nuclease-free water.....100
Figure 44: Melting curve for P-gel and thermostable P-gel microdroplets.....102
Figure 45: Demonstration of thermostability in psoralen-crosslinked gels.....103
Figure 46: Cell-free protein expression using psoralen-crosslinked and non-crosslinked P-gel microdroplets.....104
Figure 47: The functional group 5' Thiol Modifier C6 S-S.....108

LIST OF TABLES

Table 1: Diameter of DNA microdroplets synthesized in microfluidic device.....35
Table 2: Comparison of buffers used for gel ligation..... 36
Table 3: Aqueous inputs for the microfluidic production of P-gel microdroplets.....40
Table 4: Comparison of X-DNA sticky end overhangs..... 56
Table 5: Different four base-pair overhangs used in the X-DNA design..... 56
Table 6: Primer sequences for different aspects of pIVEX2.3d manipulation..... 65
Table 7: Considerations for sequence redesign of X-DNA to enable symmetrical psoralen intercalation and higher ligation efficiency after psoralen crosslinking..... 92
Table 8: Comparison of free energy of hybridization between sequences in the original X-DNA strands and redesigned X-DNA strands.....93
Table 9: Components in a cell-free protein expression reaction mixture..... 107
Table 10: Operating conditions for mass spectrometer..... 109
Table 11: Primers used for DNA bird nest synthesis..... 110
Table 12: Sequences for the original X-DNA and redesigned X-DNA for psoralen crosslinking.....110

CHAPTER I: INTRODUCTION

In this dissertation, I will demonstrate the versatility of DNA-based hydrogels as a platform for cell-free protein production. In the introduction, I will discuss how DNA can be rationally designed and how various applications beyond genetics emerge from the various engineered topologies. Furthermore, I will discuss cell-free expression and how this methodology has led to new capabilities in the testing and evolution of proteins. Next, I will explore the engineering of DNA as a material for hydrogels purposed for cell-free protein production. I will explain the design considerations needed for gel formation and protein expression, as well as how the DNA format allows for methods of functionalization.

Section 1.1 - Introduction to cell-free protein production systems and directed evolution platforms

The engineering of biological systems can take many forms, with systems biology providing vast amounts of information gathered about signaling pathways[1–3] and synthetic biology creating novel pathways or branches from existing networks[4–8]. In the most general sense, biological engineers aim to gather information about systems in order to manipulate them to produce desired products. Much of this work makes use of prokaryotic cells, bypassing the complexities and environmental requirements of eukaryotes.

Biomedical applications most often focus on eukaryotic systems, or the interplay between the prokaryotic realm and its influence on eukaryotic systems. Drug and gene delivery are vast fields, both academically and commercially, but yet the majority of bioengineering in the strictest sense takes place in prokaryotic systems. The presence of a nucleus in eukaryotes greatly complicates the ability to introduce foreign DNA and tap into existing cell pathways. On the other hand, prokaryotes readily take up plasmid DNA and have well-understood mechanisms for the replication of foreign DNA. The lack of a nuclear membrane also means there is direct access to the prokaryotic genome and that the machinery for transcription (the synthesis of RNA from a DNA template) and translation (the synthesis of protein from an RNA template) is present in the same solution volume as the rest of the cell processes.

Viewing the prokaryotic cell, especially the well-studied *Escherichia coli* (*E. coli)*, as a natural compartment in which we can harness the power of biology has led to great excitement when considering the possibilities of bioengineering to address issues ranging from energy to materials production. In this manner, the cell is seen as a factory taking raw, simple molecules to produce more complex chemistries.

The study and engineering of the cell as a living factory has also opened up the possibility of further refining biological processes to more directly serve the needs of the researcher. A biological engineer aiming for more control of protein production or to utilize the cell to select for more desirable protein mutants can deconstruct the cell into its simplest components[5,6] in order to focus on the specific pathway required for protein expression or synthesis of a molecule of choice[9,10]. In this manner, changes can

be introduced without restrictions imposed by a living cell – namely, the requirement to maintain a living, dynamic system. Yields of desired target molecules can be increased over *in vivo* production as there are no cells that require maintenance of processes extraneous to the desired task, and the energy supply and raw materials can be refocused towards the production of a protein of interest. Isolation and purification of the proteins is also less time-consuming, as the proteins are now in free solution rather than in the confines of cells that must be lysed.

In considering the format of such systems, it is straightforward to picture how cell-free protein production occurs – introduction of an expression plasmid into a cell-lysate solution will induce expression of the gene. However, one can greatly expand the use of these systems by not only breaking them down into their basic components, but also engineering materials to mimic natural processes in this well-controlled environment.

There has been much work devoted to studying cell-free protein expression, but very little applies to the manner and state in which the DNA is presented to the cell's enzymatic machinery outside a cell-based system. The accepted convention for cell-free expression is solution-phase expression, whereby the DNA for the protein of interest is simply added to the system either in a plasmid or linear gene-cassette format. Circular plasmid has been shown to have significantly higher expression than linearized gene, due significantly to the actions of nucleases that degrade DNA; a linear DNA strand possesses free ends that exonucleases readily attack. However, in considering DNA organization within a living cell, DNA is not truly a free-floating entity.

DNA within the cell contains a level of organization that is not present through simple introduction of a plasmid to a solution. Even within prokaryotes, which do not possess a nuclear structure or the level of DNA organization found in eukaryotes, the DNA is normally found within a high-density region in the single large compartment of the cell. Though this is not true within all prokaryotes, as certain organisms have shown the presence of multiple linear or circular chromosomes[11], it is true for most and importantly for the prokaryotic organism *Escherichia coli* (*E. coli*). In addition to the lack of a nucleus, most prokaryotes also do not possess histones, which are the proteins associated with wrapping eukaryotic DNA to condense the incredibly long genome into a small enough volume to fit into the nucleus. Rather, supercoiling provides another mechanism to package DNA. Regardless of the specific method of packing, DNA within cells is most often stored and accessed in a region of extremely high DNA density, whether it is in an actual nucleus of the eukaryotic cells or within the less defined 'nucleoid' of a prokaryote.

Though some forms of DNA immobilization have been used in cell-free systems, mostly in the form of gene immobilization, none have taken the steps to utilize DNA materials as a bulk organizer and scaffold. The ability to link gene-encoding DNA into a larger material provides a means of organization and further functionality as the scaffold can be chemically modified, notably through thiol and click chemistry[12,13]. Such modification was performed in our lab to form DNAsomes, which form spontaneously through hydrophobic interactions from lipid-functionalized DNA[14].

Section 1.2 – Engineering DNA materials

The structure of DNA was discovered and detailed in the 1950's, but the enthusiasm surrounding this discovery was focused on the incredible potential this discovery held for genetics rather than the fundamental chemical and mechanical aspects of the polymer in chemical and material applications – and rightly so. However, if we remove ourselves from the realm of basic biology and genetics and view DNA as a generic polymer, it truly stands alone. There are a number of reasons DNA is a unique polymer for engineering materials at the nano- and microscale, and I will discuss how different DNA topologies can be created and used for different applications.

A single strand of DNA is comprised of a sequence of four monomer units – guanine (G), cytosine (C), adenine (A), and thymine (T). These nucleotides, or bases, can be encoded specifically into any sequence, creating complex sequences that are not readily achievable with other polymers; the number of distinct sequence permutations afforded by an N base strand is 4^N. This would not be of great importance and use, however, if DNA existed only as a single polymer strand. Rather, its natural form is a double helix structure, with two strands bound to each other through hydrogen bonding (hybridization). In this manner, polymers can be synthesized to interact with each other in a controllable manner through sequence design.

Specific hydrogen bonding between strands occurs through Watson-Crick base-pairing, which is defined by the complementary base-pairing of adenine-thymine

(A-T) and guanine-cytosine (G-C). While adenine-thymine base pairing is conducted through two hydrogen bonds, the cytosine-guanine interaction occurs through three hydrogen bonds and thus provides a stronger associative force. This is an important consideration in DNA sequence design, as the degree of association between strands in a DNA duplex determines at what conditions, namely temperature, the strands begin to dissociate[15,16]. Furthermore, when a variety of different strands are being used in an experiment, the relative strength of interactions between strands becomes extremely important, as bases in a single-stranded DNA can also interact amongst themselves to form secondary structures that may interfere with the intended application.

The precise control of DNA length capable in synthesis, both naturally and synthetically, is a highly appealing aspect of DNA as a polymer for material applications and as a molecular organizer at the nano- and microscales. The monodispersity of biological polymerization processes, or 'living' polymerization, is extremely high when compared to other, bulk methods of polymerization, such as free radical or cationic/anionic addition[17]. This high degree of monodispersity allows for the controlled use of DNA at both small length scales, down to a few bases in length, and enormous aspect ratios when considering the approximately 2 nm diameter of double-stranded DNA. The stiffness of DNA also changes depending on whether it is being used in a single- or double-stranded format. The persistence length of a single strand DNA is approximately 2 nm while that of double-stranded is closer to 50 nm (≈150 base pairs)[18]. This broad difference allows us to design branched systems and to consider each arm as 'stiff', which gives excellent control over resulting topology.

Our laboratory has created DNA-based materials over a range of applications using DNA in linear, branched, and network formats, as shown in Figure 1. Used in its more natural, linear format, DNA can be used as a coating ligand for nanoparticle assemblies. In coating nanoparticles with single-stranded DNA, particles interactions can be modified as the DNA seeks out complementary sequences coating other particle sets. Groups have recently used such capabilities to specifically encode crystalline structures with nanoparticles of different materials[19–23], but our lab has focused mainly on the use of DNA to control gold nanoparticle crystallization conditions and spacing in solution[24], on surfaces[25], within free-standing sheets[26], and at the water-air interface[27,28]. Even when complementary base-pairing is not utilized, DNA is an excellent coating ligand based on the precise control of length and monodispersity[28]. The inherent negative charge of DNA means that densely coated particles will not interact strongly at low salt concentrations, and while aggregation may occur at high salt concentration, crystallization occurs in a much smaller space of conditions. Monovalent and divalent salts will have also have vastly different effects on the system, as divalent salt possess the ability to 'link' strands together to increase effective interactions between particles[28].

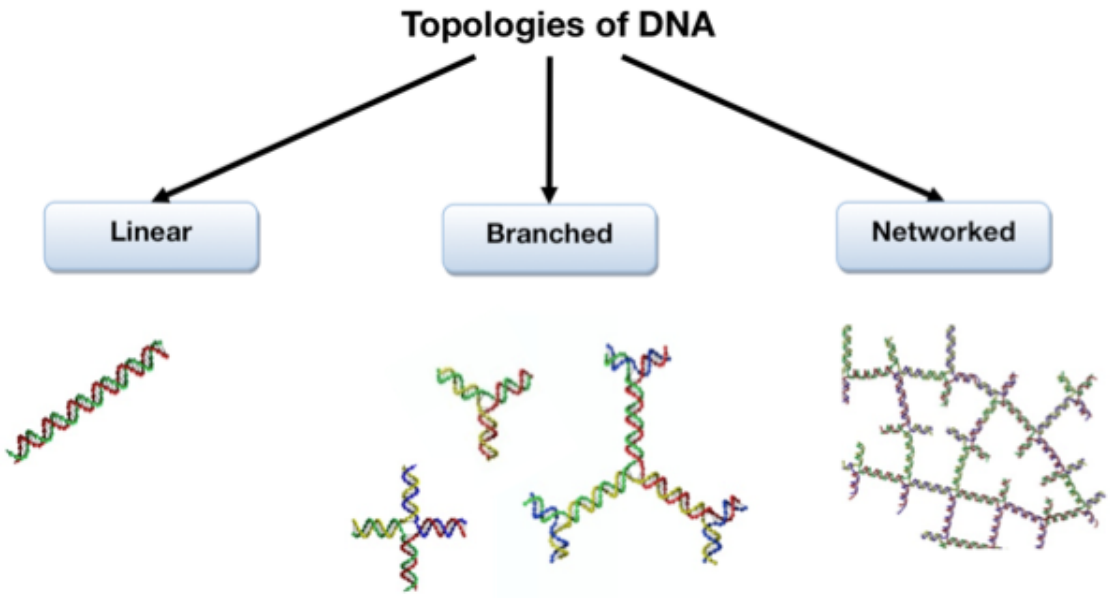


Figure 1: Formats of the various DNA topologies explored in the Luo laboratory.

As an engineered branched molecule, DNA structures present multiple arms that can be modified based on the application. Use of DNA in dendritic structures is particularly useful based on the extensive array of biochemical techniques that can interface with the nucleic acids[29]. The original pioneering work performed by Seeman and colleagues to produce branched DNA was based on the transient four-armed Holliday junctions formed during genetic recombination[30]. By constraining sequence symmetry, the first synthetic four-armed juncture was created and has led to the synthesis of a number of different branched morphologies[31–36] as well as catenanes and knots[37,38]. Though early synthesis attempts were characterized by high conformational flexibility, crossover hybridization utilizing double, triple, and paranemic architectures led to more rigid and controllable structures.

Using the crossover hybridization technique, our lab has synthesized up to six-generation DNA dendrimers[39], and used such structures for the creation of a DNA barcode system[40]. Using green and red fluorescent moieties, branched DNA containing different ratios of these dyes were used for the multiplexed detection of various diseases[40]. Further functionalization of branched DNA that took advantage of the ability to carry different functional moieties on different arms yielded ABC (Anisotropic, Branched, and Crosslinkable) monomers[41]. The branches of this molecule were modified with fluorescent dyes, a UV-crosslinkable group, and a sticky end that was complementary to a target DNA or RNA. In the presence of the target, the ABC monomers were brought together by hybridization with the target, allowing for UV crosslinking into larger DNA aggregations observable under an optical or fluorescent microscope. The ability to organize multiple functional groups on a branched molecule has applications across many fields, including drug delivery[42,43], molecular sensing[44], and DNA logic systems[45].

Networked DNA structures can be formed through a number of mechanisms, but the most direct route is through the linking of branched DNA. Our lab synthesized X-, Y-, and T-DNA that were subsequently used as crosslinkable monomers and covalently linked by DNA ligase (Figure 2)[39,41,46,47]. In order for the branched structures to recognize other DNA monomers in solution, each arm possesses a palindromic sticky end that allows for hybridization. Though our lab covalently ligates these DNA units together, ligation is not required for bulk gelation; longer sticky ends can be used to ensure hybridization is strong enough at room temperature to maintain the gel morphology[48]. However, the direction and application of this thesis work

necessitates the use of both short sticky ends *and* ligation. The production of discrete gels in solution is needed in conjunction with the ability to crosslink DNA encoding for proteins into a growing gel network. As the method of recognition between monomers is based on sticky end sequences, other DNA and RNA displaying this sticky end can also be ligated into the network when digested by selected restriction enzymes. This is the unique aspect of DNA as compared to other polymers – the interactions that induce formation of the material are of the same type that allows interaction with exogenous DNA and RNA targets. This ability naturally makes DNA-based materials a great platform for sensing and capture applications, as our work on DNA nanobarcodes has demonstrated[40,47].

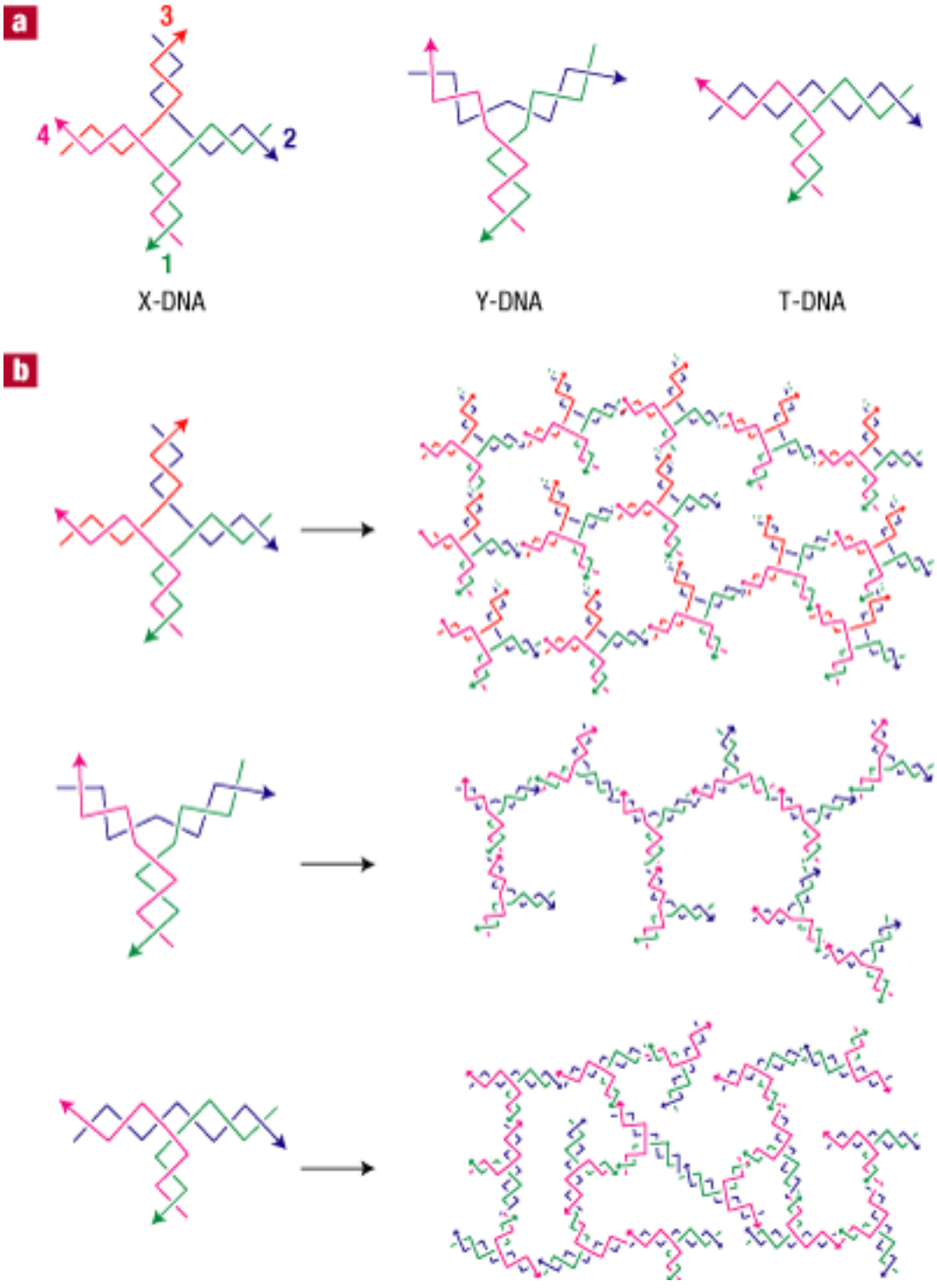


Figure 2: Different branched DNA structures can be crosslinked to form a hydrogel. Note the Y- and T- shaped DNA utilize three different single-stranded DNA sequences while X-DNA utilizes four[46].

This ability to naturally crosslink DNA encoding for proteins of interest into a larger DNA network led to the formation of the our lab's gel known as P-gel, or protein-producing gel[49,50]. When placed into cell-free lysate with the enzymatic machinery necessary for transcription and translation, the protein yield associated with the crosslinked gene was significantly higher than if the same amount of gene had been placed in a conventional setup, consisting of free genes in solution (Figure 3). This phenomenon is discussed further in Chapter 2, and is the basis for using the DNA gel format as a platform for developing further cell-free protein production technology.

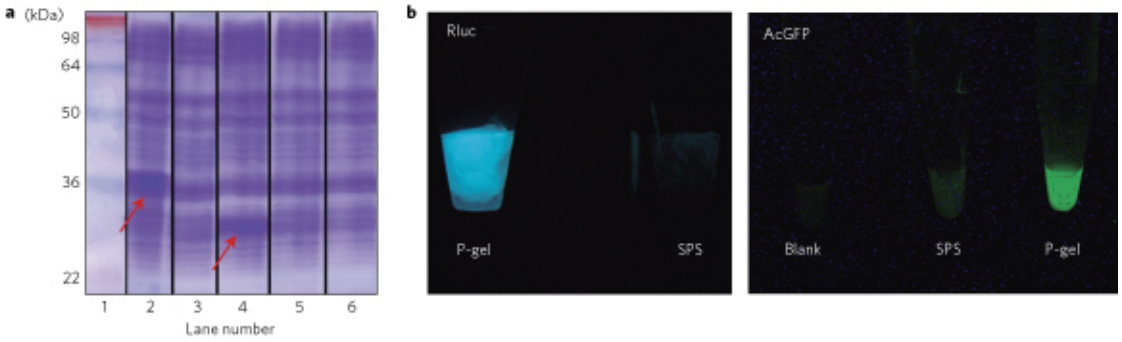


Figure 3: Comparison of protein production between P-gel and solution-phase (SPS) setups. (a.) SDS-polyacrylamide gel shows successful production of Rluc and AcGFP protein using P-gel in lanes 2 and 4, respectively. (b) P-gel demonstrated significantly higher protein yields when compared to SPS controls, as seen in fluorescent images of the reactions volumes[49,50].

Our laboratory has explored other methods of DNA gel formation that do not involve ligation of branched monomers. Rather than utilizing covalent linkages to maintain a gel format, these gels are synthesized based on non-covalent interactions of long strands of single- and double-stranded DNA interwoven at a very high density.

The mechanisms behind the synthesis of these materials are varied, where one uses thermostable, branched primers formed by psoralen crosslinking[51], and the other utilizes a modified rolling-circle amplification (RCA) procedure[52]. Interestingly, the material produced through the modified RCA process produces a network of high-density bird nest DNA structures that are held together weakly by relatively few DNA strands to form a mechanical metamaterial[52]. This structure makes these high-density regions amenable to separation into discrete microgels. Aspects of this process are explored further in Chapter 4 in regards to opening further areas of DNA materials to applications in cell-free protein expression.

Section 1.3 - Cell-free expression of proteins

The most significant advantage of *in vitro* protein production is the ability to produce and analyze a much larger mutation space than in a comparable cell-based process. In this thesis, I do not focus on the ability of cell-free systems to produce large quantities of protein, as cell-based methods have been optimized to such a large degree and incorporated into industrial processes so thoroughly in the last few decades. Rather, I focus on the concept of using DNA materials to reproduce the minimal level of cell function necessary to conduct directed protein evolution, as not only is transformation and cloning not required in such a non-living system, but proteins that would normally be harmful or toxic to a cell can now be produced outside of the cell using cellular machinary[53]. Certain molecules or proteins may not only affect the cell through the stress induced by the siphoning off of cell resources,

but directly through their molecular structure. This is especially true when utilizing the cell to produce exogenous proteins. Nonetheless, despite these limitations, the enormous potential of cell-based protein production can be seen in its widespread use across academia and industry. Traditional cloning work has been present for decades, and is taught as an essential lab technique for introducing plasmid DNA to cells[54].

Cells naturally and beautifully possess the ability to isolate gene sets and maintain a genotype-phenotype connection, a feat that is difficult to reproduce outside of living systems. The genotype of a protein is the actual DNA sequence responsible for its production, while the phenotype is the physical manifestation or particular trait of that protein. This section provides a background on a variety of methods used in the evolution and engineering of proteins, but it is important to keep in mind that every method must address three key requirements: isolation of gene sets, sufficient (assayable) production of protein, and maintenance of a genotype-phenotype connection.

The potential to directly use biological systems to mutate and engineer proteins is perhaps most clearly seen in the synthesis of a range of fluorescent proteins from the naturally occurring green fluorescent protein (GFP). Small mutations in the DNA sequence of GFP led to the creation of an array of proteins possessing different excitation and emission spectra[55,56]. This has allowed for multiplexed labeling capabilities, which has been an extremely important development in the exploration of biological systems as different cell components can be tracked simultaneously and protein emission profiles can be refined to provide more accurate imaging[57,58].

Since the initial works describing proteins as nature's molecular machinery, significant efforts have been made to discover the sequence-structure-function relationship based in the amino acid sequence that constitute a protein[59]. As a greater understanding of this connection has evolved through the field of proteomics, the goal of modifying proteins to increase enzymatic efficiency or to create entirely new protein functions has led to different approaches to altering and testing different amino acid sequences, and necessarily the underlying DNA sequence.

Though the desired goal of producing proteins geared to a specific process or function may be conceptually straightforward, the methods to achieve this are less so. There are two inherently different approaches towards the engineering of proteins. The first method requires in-depth study of individual proteins, whereby enough information about the sequence-structure-function relationship has been deciphered to successfully allow for rational and selective substitutions, additions, or deletions in the amino-acid sequence. Advances in protein crystallization and the ever-increasing power of computing to sift through complex interactions at the nanoscale has allowed for this process to become more streamlined. Importantly, it has become extremely useful in increasing our understanding of protein structure and elucidating the effects of specific amino acid changes on conformation. However, this approach towards creating protein variants can prove to be extremely time consuming and resource-intensive, and while contributing to basic knowledge of the proteins under investigation, it is often still very difficult to directly predict final enzymatic activity. This is especially true when mixtures of proteins act in concert with one another to produce synergistic interactions.

Another promising and widely used method of creating proteins with increased efficiency or novel functions is the process of directed evolution[60–65]. This method involves the introduction of mutations into a gene (or within a targeted region of a gene), producing proteins from this library of mutants, and then subjecting the system to selective pressure to determine which mutated genes produce a protein with the desired activity. Thus, the experimental design is focused more on the material platform in which proteins are tested and in the manner in which selective pressures are instituted. The examples of changes in protein structure or activity that one can search for covers a wide range of potential characteristics, including higher enzymatic efficiency, catalysis, or binding constants to target molecules.

Cell-based methods for evolution have been used for decades to tailor the activity of proteins. For plate growth assays, enzyme activity is connected in some manner to the survival or growth rate of the cell stock. In this manner, only those cells transformed with a gene encoding a protein with the desired activity will grow on a plate or in media. Perhaps the most straightforward application of directed evolution research involves the tying of cell survival to the activity of a target protein. In this manner, cells that possess higher activity mutants will grow faster and outcompete less efficient mutants. In industrial scale production, growth media and feed is a major cost, and thus much effort has focused on the growth of cells on different feed sources. For example, testing for the breakdown of complex carbohydrates has been a useful method for biofuel applications[65–76], whereby cells producing mutated cellulases are grown on a sugar-deficient media to select which cellulases most efficiently break down treated plant matter into simple sugars[68,69,75,77]. Such simple sugars then can be

reorganized into more complex molecules depending on the application, and further efforts are currently focused on industrial scale production of biofuels using cells evolved to act as biocatalysts for production[65–68,71,74]. These examples represent two of the main thrusts behind the evolution of proteins for industrial applications – to yield simple molecules from cheap, abundant feedstock for the repurposing of material flux to desired, and often more complex, molecules.

Methods such as microtiter selection, most often performed with fluorescent products, have allowed for the testing of catalytic activity outside of those processes connected with cell survival[78–81]. However, this process severely limits the number of gene variants that can be tested; a high-throughput, automated system can still only test $\approx 10^6$ variants, many magnitudes smaller than the potential library size[82]. Further advances have allowed for ultrahigh-throughput selection and analysis through methods such as affinity selection and fluorescence-activated cell-sorting (FACS). These methods allow for screening of binding interactions[83–86] and fluorescence products[82,87–93], respectively.

Though cell-based systems for directed evolution applications are widely accepted, there are certain disadvantages to selecting for protein mutations in a living system. Generally, when testing for protein phenotypes outside applications directly affecting cell processes, the use of living systems fails in the most basic regard when considering the requirements of testing a full mutation space. In order to produce mutations tailored to a specific function, the selection forces must only be specific to the activity of interest. In order to keep a gene in the selected set when using a living system, there is the additional requirement and selection pressure of maintaining or

increasing the growth of the cells. In most cases, the protein phenotype a researcher is seeking has an application outside cell growth, and thus this pressure proves to be extraneous and potentially disruptive to the experiment. When referencing potential sequence mutations, the term 'beneficial' mutation when testing in an *in vivo* system can often be describing an indirect measurement. As only genes that are maintained in live cells are selected for, the associated protein activity needs to be connected with cell-growth in some form (i.e. connecting products of an enzymatic reactions to an increase in metabolites), but certain protein mutations may spur cell growth without having a positive impact on the desired selection measure. On the other hand, beneficial mutations for the desired process can possibly decrease cell growth or prove to be toxic to the very cells that are serving as the platform for expression, testing, and evolution.

Based on these drawbacks, the ultimate goal would be to produce and test proteins without the needs of a living system. Before approaching such methods and my work, I first will introduce display systems that produce protein within a cell or virus, but yet allow for testing of the proteins on their associated surfaces. In this manner, the testing of proteins for activities other than catalysis and fluorescence has been expanded through the use of affinity selection, cell/phage display and yeast two-hybrid systems. These methods have allowed for the evolution and modification of protein binding affinities. Binding affinity testing requires that the proteins of interest be expressed in a manner that allows them to interface with other surfaces or molecules outside the volume in which they are produced. In order to accomplish this task, cell-display employs the use of membrane localization signals in the form of

peptides attached to the C- or N-terminus of a produced protein. These signaling peptides induce incorporation in the membrane of the cell while displaying the protein of interest to the extracellular environment. Such displays have been demonstrated and used in many varieties of cells, including gram-negative bacteria[94–96], yeasts[97–100], and even mammalian cells when higher level post-translational modifications may be required[101,102].

Phage display is a heavily-used technology in the study of protein-protein, protein-peptide, and protein-DNA interactions[86,103–118]. Bacteriophages are simple biological machines, containing a small DNA or RNA genome encapsulated by a protein coat. This protein coat is encoded within the phage's small genome, and thus the sequence of the protein of interest can be added manually to this sequence. In this manner, binding affinities can be tested while maintaining a genotype-phenotype connection without the direct use of live cells in the assay. Phage display has been heavily used for antibody selection[107,108,111,114,118], but has also led to improvements of protein stability[110,115] and even the organization of non-organic materials on viruses[119]. Phage and cell display technologies are shown in Figure 4, and demonstrate the presentation of protein to the external environment.

Phage has also been used outside of protein binding applications as an infection mechanism for a novel continuous evolution platform. Phage-assisted continuous evolution (PACE) uses a 'lagoon' that contains a replication population of infectious progeny phage and *E. coli*[120]. Desired protein activities can be selected for as long as there is a method that ties protein activity to production of protein III, which is required for the viral infection of cells and production of more phage. In the

particular case tested, T7 RNA polymerase was evolved to express specific binding with the T3 promoter, which it normally will not recognize as a binding sequence.

The drawback of all the methods listed thus far is that even if testing may take place using *in vitro* conditions, they still involve the use of living systems in one form or another. Viruses themselves are not considered 'living', but the process of gene library uptake as well as synthesis of new phage must occur through a cell intermediary; protein variants that are harmful to the cell may still be excluded from the final library by minimizing uptake and production of phage. In addition, there is still a needed transformation step, again limiting the size of the library significantly. Thus, even though the protein interactions are tested using *in vitro* binding platforms, many of the same fundamental problems surrounding undesired selection pressures are still found in phage display platform[121].

The methods discussed thus far have allowed those both in academia and industry to begin to explore what mutations lead to optimized protein activity. In order to truly explore the full potential of evolution platforms, fully *in vitro* systems have also been developed that perform both cell-free protein expression and create a genotype-phenotype link. Figure 4 shows examples of creating this link: ribosome, mRNA, and DNA display. Ribosome display utilizes the suspension of the ribosome activity at the end of the translation process, halting the complex's release of the RNA and associated protein, and thus leaving a physical connection between this translated protein and the mRNA template[122–124]. A similar method utilizes puromycin to covalently link an mRNA to the associated translated protein[125–130]. In this manner, binding assays can be performed on the protein while attached to the mRNA, and PCR

can be used to amplify the remaining nucleotide sequences in an assay after selection. The weakness of this approach is that there is only one protein associated with each mRNA. Phage display, on the other hand, can potentially display many proteins stemming from the same DNA sequence.

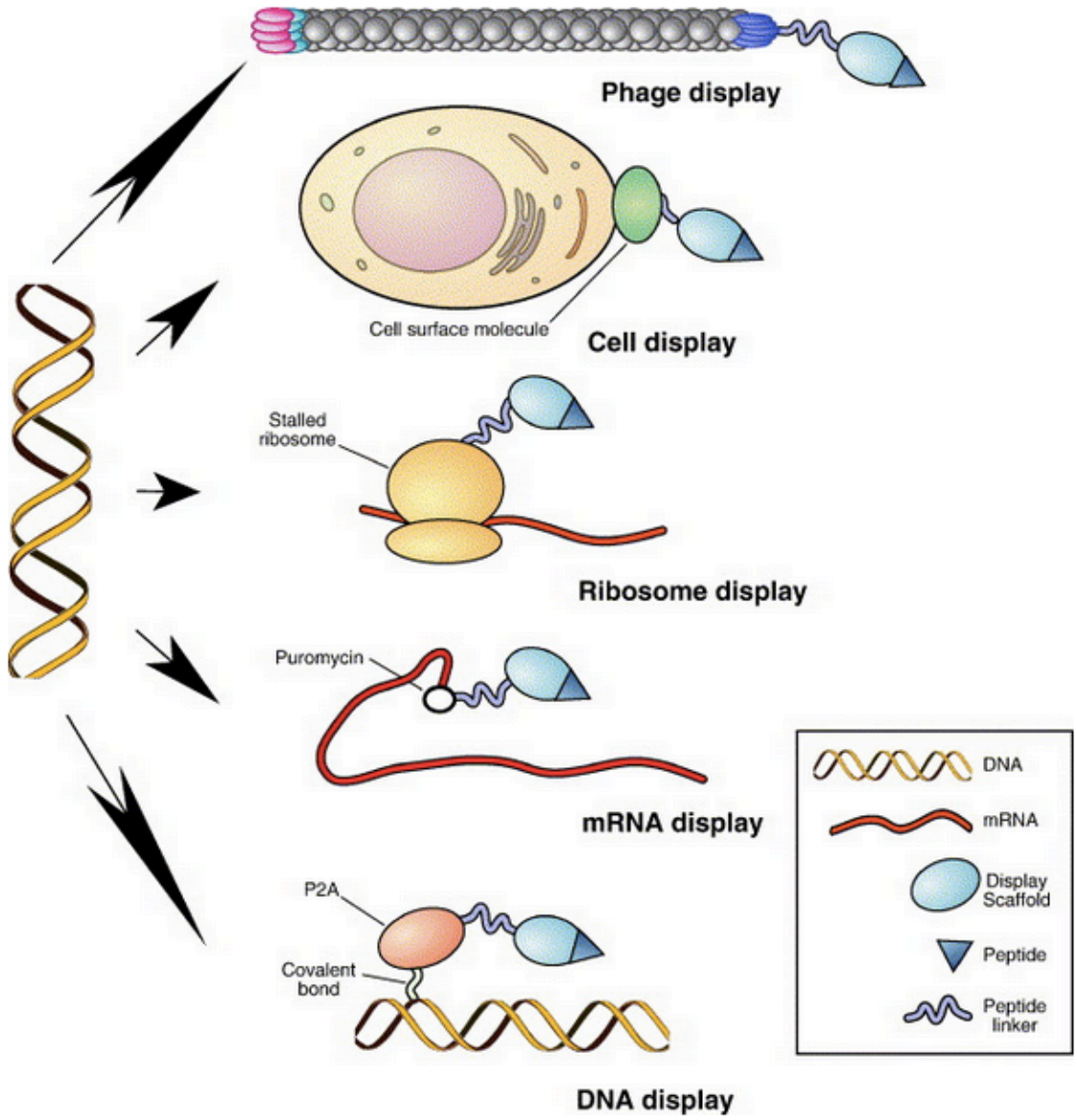


Figure 4: A comparison of display technologies for linking genotype and phenotype in selections involving binding target binding[131].

A different approach to creating a genotype-phenotype connection in a cell-free expression utilizes compartmentalization as the gene isolation mechanism. *In vitro* compartmentalization (IVC) has allowed for an artificial connection between genotype and phenotype; researchers have produced and assayed proteins that are not directly bound to either a surface or the associated DNA/RNA[90,92,132–138]. Use of this method has been mostly been limited to processes that have involved either fluorescence or direct DNA/RNA-binding products due to difficulties associated with assaying[132,136,138]. The fact that expression and assaying takes place in the same volume requires that neither of these processes significantly interfere with each other. This severely limits the variety of assays that can take place using a pure emulsion-based system. Another route to testing downstream binding interactions can be achieved through the previously mentioned DNA display, which is produced by binding protein to DNA within emulsion compartments. This emulsion-based gene isolation is the only current method by which DNA will associate *only* with the protein produced from its sequence.

Certain groups have attempted to address the drawbacks associated with having to both produce protein and assay activity within the same volume by separating these steps through the use of microbeads in droplets. Using a multistage process, Swartz and Stapleton evolved a FeFe hydrogenase to be more oxygen tolerant[92]. However, the overall procedure required three emulsification steps. Their results show that although this method is possible and offers a path to high-throughput analytics, it is indeed quite complex and is not as straightforward is it may seem at first – library losses occur at each step as does potential contamination and crosstalk

between gene and protein sets. In images of the emulsions, they tend to be mixtures of heterogeneous droplets that may be prone to aggregation[92]. The approach does yield a more active FeFe hydrogenase in the presence of oxygen, but the platform is limited and relatively complex for more general applications.

Such work demonstrates that bead-based expression, whereby DNA or RNA for a protein of interest is bound to a nano- or microparticle and placed in a cell-free expression solution, has been attempted. Furthermore, the immobilization of DNA and RNA onto surfaces has effects on the overall protein yield in cell-free expression - the ability to induce non-specific binding in a region around DNA or RNA has been shown to localize expression machinery, thus increasing transcription and translation[139]. The method of securing the gene of interest to the particle is a critical issue. Binding of DNA to surfaces shows lower expression efficiencies than a corresponding SPS expression[140–144], while attempts at localizing RNA shows drastically different results depending on the gene used and strength of binding interactions. Hamad-Schifferli specifically bound the RNA of eGFP and mCherry to nanoparticles through DNA-RNA interactions, and while certain binding conditions increased expression for mCherry, it actually reduced expression of eGFP[139]. The fact that these differences in expression are seen even with proteins as similar as these fluorescent reporters means potential differences among varying classes of proteins may be even more significant.

Recruitment of translational machinery through nonspecific interactions to regions of localized mRNA has been shown to provide an increase in translation efficiency[139,145,146]. Through using DNA instead of RNA as the template, the P-gel

produced by our laboratory crosslinks high local concentrations of gene to induce increased transcription and translation. The negative charge associated with using DNA as the scaffolding material can induce nonspecific interactions with the proteins involved in transcription and translation. Other materials have been used to induce gel formation in the presence of DNA, including agarose and alginate. However, our lab work has shown that these polymers actually reduced protein expression rather than increasing it.

Another material that has been shown to provide benefits for cell-free protein production is clay. Our lab has demonstrated that clay strongly binds DNA in solution, offers protection against DNAse and RNAse degradation, and increases the protein yield of luciferase protein[147]. The specific mechanism for this process was not analyzed heavily, but it demonstrates that the interaction of DNA, RNA, and proteins associated with cell-free protein production interact differently and specifically with different scaffolding materials. Yields are based on a complex combination of factors, where enzymatic recruitment, but also DNA and RNA protection, are perhaps the leading factors concerning the gene template.

Section 1.4 - Significance of this dissertation

The engineering of biological systems can take many forms and cover a wide variety of disciplines. They usually involve the use of natural, cell processes as a starting point for engineering higher molecular yields of particular metabolites, new cell pathways, or new proteins. In the case of protein production, cell-free synthesis of

proteins allows the focusing of energy and molecular flux towards the production of a protein of interest while removing the limitations on toxicity and cell stress induced by certain proteins. Protein labeling is more efficiently performed as direct control over amino acid makeup, including non-natural amino acids, can be exerted[148]. The range of proteins that can be produced, and thus, evolved from a cell-free system is dramatically greater than in cell-based systems.

Though cell-free protein synthesis sounds like the ideal platform for protein production, one loses the all-important genotype-phenotype connection that a cell so elegantly provides. In addition, though purification of the protein may be less time-consuming than cell-based systems, one still must go through common methods of purification, notably flow-through nickel columns. Lastly, although strides have been made in cell-free protein synthesis and the use of these systems for directed evolution, low yields and application-specific nature of these works have limited their use.

This dissertation provides insight into the potential of DNA materials to address fundamental needs in cell-free protein synthesis by creating a material that can both produce and capture protein. From this, there is the potential to create an artificial genotype-phenotype connection in a high-yield protein-producing platform. The use of engineered DNA gels offer the ability to naturally interface with DNA and RNA in a way that is not possible with other polymers. Furthermore, versatile functionalities can be introduced by either modifying the ends of engineered branched DNA structures or introducing molecules that actually intercalate and covalently bind the DNA double-helix.

In Chapter 2, I will discuss the synthesis of a microfluidic setup to produce cell-sized DNA gels and the method used to retrieve these gels. Chapter 3 will review the practical considerations associated with engineering X-DNA for gel production and protein synthesis. This includes the testing of different sticky ends to modify interactions between the X-DNA structures and how selection of a plasmid linearization site affects cell-free protein expression. Chapter 4 discusses different methods of modifying cell-free lysate and DNA gels to increase the platform's functionality. I first present the modification of the cell lysate and DNA to allow for protein capture, both in the P-gel format and in another form of DNA gel studied by our lab termed DNA bird nests, and then discuss a method of modifying the DNA to increase gel thermostability.

CHAPTER 2: SYNTHESIS OF DNA GEL MICRODROPLETS

Section 2.1 - Moving from bulk gel to microgels

The synthesis of bulk DNA gels through the crosslinking of branched DNA monomers has been studied previously in our laboratory[46]. In the process of exploring applications of this DNA gel format, genes were ligated into this bulk gel to form the material P-gel, or 'protein-producing gel'[49,50]. This gel was shown to increase gene expression significantly in cell-free lysate, many times over commercial solution-phase systems. The mechanism behind this process involves protection of the genes from enzymatic degradation and the significant enhancement of transcription, leading to an increased concentration of mRNA compared to a solution phase setup.

The protection of genes within this gel may stem from multiple causes, but at least one mechanism is clear. The capping of the gene ends through X-DNA reduces exonuclease degradation, and the simple addition of a large excess of DNA that is not associated with gene activity further reduces nuclease activity on the genes. This leads to longer reaction times, up to three times longer than solution phase, and the ability to reuse the gels with additional lysate. As mentioned, another extremely important effect of the gel was the approximately 73-fold increase in mRNA levels over the solution phase system. The mechanism for this significant increase was not intensively studied; however, the results speak to a promising leap forward for cell-free protein production.

The bulk P-gel format provided insight into using gel as a platform for cell-free protein production, but further work needed to be performed to scale this process up, make gel formation more consistent and reliable, and explore further functionality offered by a this networked DNA material. In this section, I will detail the microfluidic production of DNA microgels. This production provides a method to scale-up P-gel production while synthesizing homogenous gel structures that can be manipulated and processed as 'cell' structures based on their small size. This stands in contrast to earlier P-gel work, where gel molds of precursor solution were further broken up through pipetting of the gel solution. This leads to gel pieces that are heterogeneous in size and potentially gene constitution.

Section 2.2 - Microfluidic device fabrication and testing

In order to produce homogenous, cell-sized gel droplets, gelation must occur within a small confined volume. Water-in-oil emulsions as well as liposomes and vesicles provide an ideal microcompartment, and have previously been used as reaction vessels[129,131,141–144]. However, most bulk methods for emulsion formation produce size distributions that would lead to heterogeneous gels. Extrusion methods to reduce heterogeneity are taken after initial emulsification has occurred, and thus would disrupt gel formation as ligation begins immediately upon mixing with X-DNA subunits.

The use of microfluidics to produce large yields of highly monodisperse emulsions provides a platform in which to synthesize microscale DNA gel droplets. A

microfluidic device was designed that used a flow-focusing junction to allow for the production of water-in-oil emulsions[152,153] using three inputs: one input for the outer oil phase, and two inputs for the inner aqueous phase to allow for separation of X-DNA and ligase into separate precursor solutions[154]. The production of the device was undertaken through conventional photolithography processes, whereby a silicon master mold was created and from which PDMS replicas could be made. The first step in this process was a CAD design of the microfluidic layout. Two designs were eventually used in this project, but I will elaborate on the procedure for the first design as only the layout was changed for the second (the CAD layout and dimensions are different). The first design is shown in Figure 5A, demonstrating that two channels serving as aqueous inputs meet just before crossing over a third channel that will be carrying an oil flow. This double-junction design serves to pinch off the aqueous flow into water-in-oil droplets whose size depends on the ratio of flow between the aqueous and oil phases as well as the channel size where the flows meet. This relationship is shown in Figure 6 and Figure 7.

The CAD design was transferred to a chrome-glass mask using the Heidelberg Mask Writer DWL66 at the Cornell Nanoscale Science and Technology Facility (CNF). Following developing of the mask, a silicon wafer was coated with a 40 μm layer of the negative photoresist SU-8 (Microchem, Newton, MA). Prebaking was performed at 90°C for three minutes, and then contact-mode photolithography was undertaken using a Mask Aligner (ABM, Scotts Valley, CA) and UV shield to protect from the smaller wavelength tail of the 365 nm mask aligner output. A 40 second exposure was used, followed by a five minute post-bake at 95°C and three minute

developing step. Lastly, the dried wafer was heated at 110°C for five minutes to further harden the SU-8 pattern.

The SU-8 negative pattern master-mold was complete at this point, but in order to ensure the continued use of the mold after many repeats of PDMS casting, the hydrophobicity of the wafer was increased through molecular vapor deposition of (perfluorooctyl) trichlorosilane (FOTS). The wafer was first plasma oxidized, then coated with FOTS that serves as an anti-stiction layer, which reduces damage to the SU-8 pattern during removal of PDMS molds. After this deposition, degassed PDMS (placed under vacuum for one hour at room temperature) was poured on the wafer and heated at 70°C for four hours.

After the PDMS hardened, individual devices were cut off of the wafer. Vertical inlet and outlet holes were punched into the PDMS to allow for inflow and outflow pins to be inserted, and then the device and a PDMS-coated glass slide were plasma oxidized and bonded. Incubation in a dry chamber overnight ensured the secure bond between the two components. This entire production process is summarized in Figure 5.

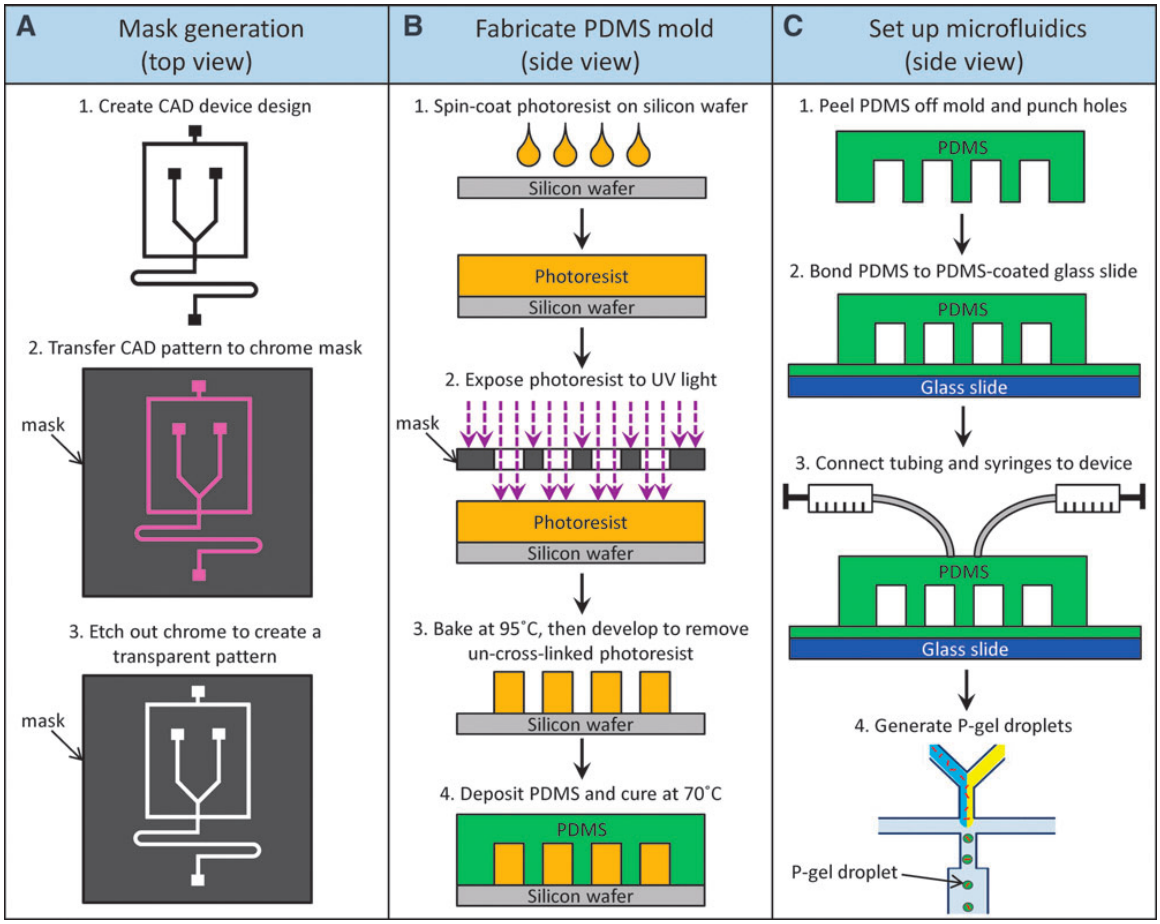


Figure 5: Overview of the microfluidic device construction[154].

The resulting emulsion and gel sizes based on input flow rates are shown in Figure 6 and Figure 7. The inputs for the device are also described, with one input containing X-DNA and gene and the other containing the ligation components, including ligase, ligation buffer, and excess ATP. However, after performing initial work with this microfluidic setup, it was noted that debris kept getting caught within the double junction and restricting flow; this required constant evaluation and device manipulation should flow be affected. We were able to dislodge most obstructions, but

this highlighted the need for filters within the device to ensure one could leave the setup running for long periods of time. Thus, we changed the microfluidic design to incorporate filters at each input, with the flow encountering parallel 50 μm channels followed by 10 μm channels before meeting other inputs. Being that the smallest channel width after this point is the 15 μm, found in the aqueous channels, there should not be any debris or gel aggregations large enough to clog the device at the flow junction. Figure 8 shows the in-device filters consisting of a two sets of a parallel series of size-exclusion channel. In addition, a redesigned double junction is shown with a smaller channel width than the previous device setup.

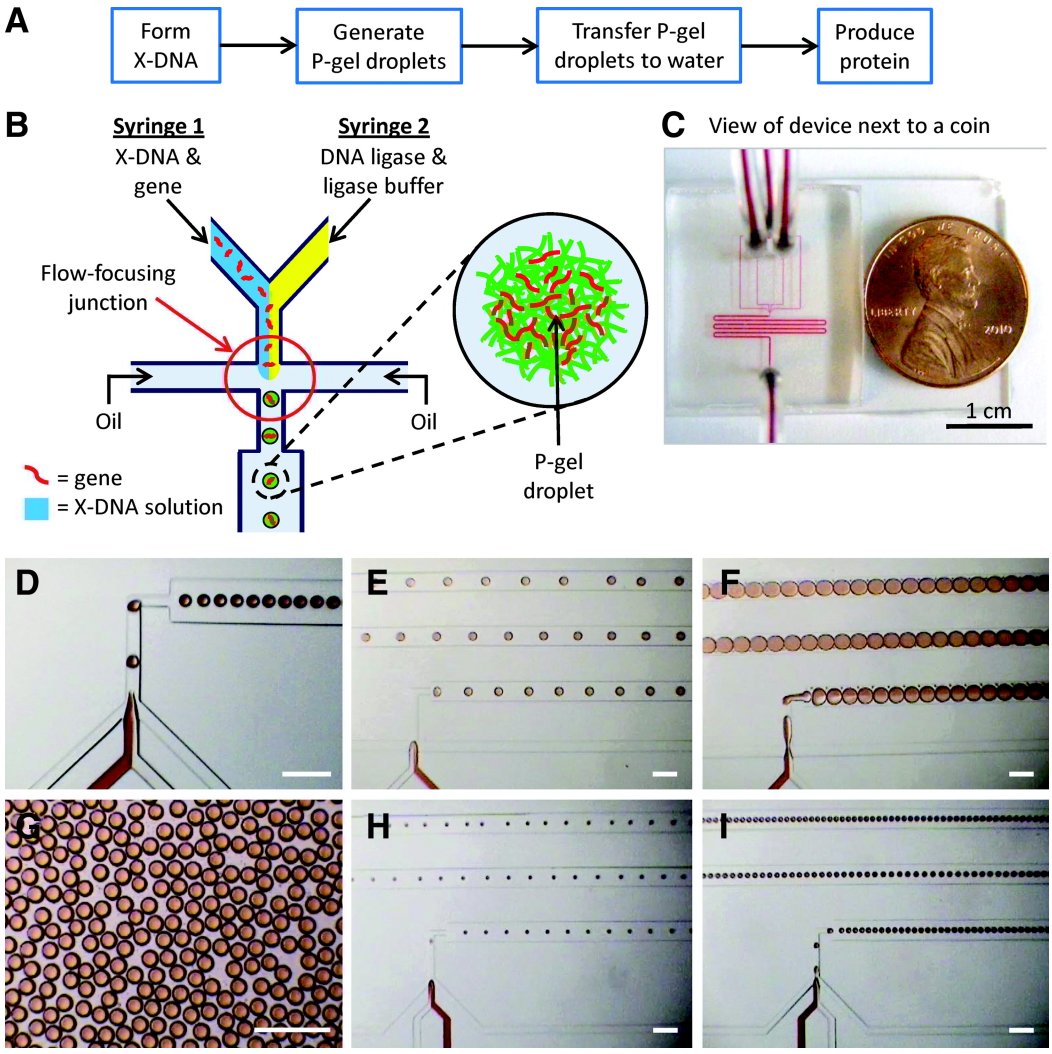


Figure 6: Overview of P-gel microdroplet expression[154]. (D) through (I) show flow under different ratios of aqueous flow versus oil flow. All scale bars are 100 µm.

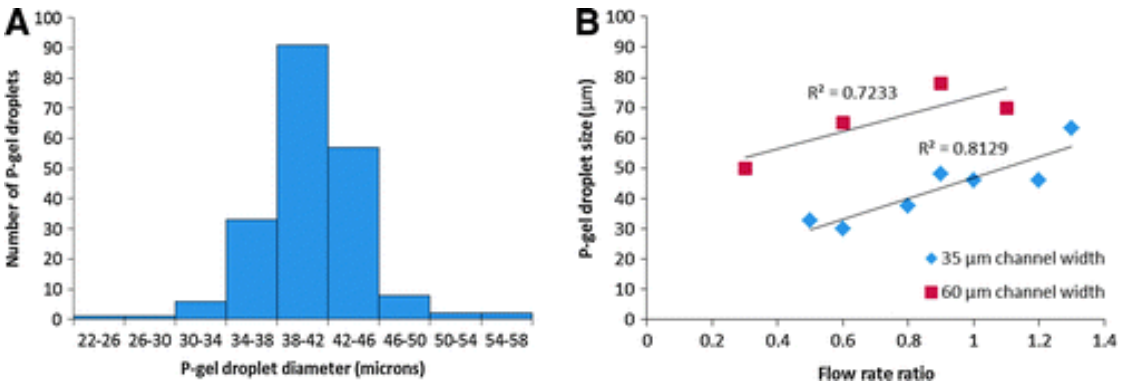


Figure 7: (A) P-gel droplet diameters for 35 µm channel width at 0.8 aqueous:oil flow rate ratio. (B) Relationship between P-gel droplet size and flow rate based on two channel widths[154].

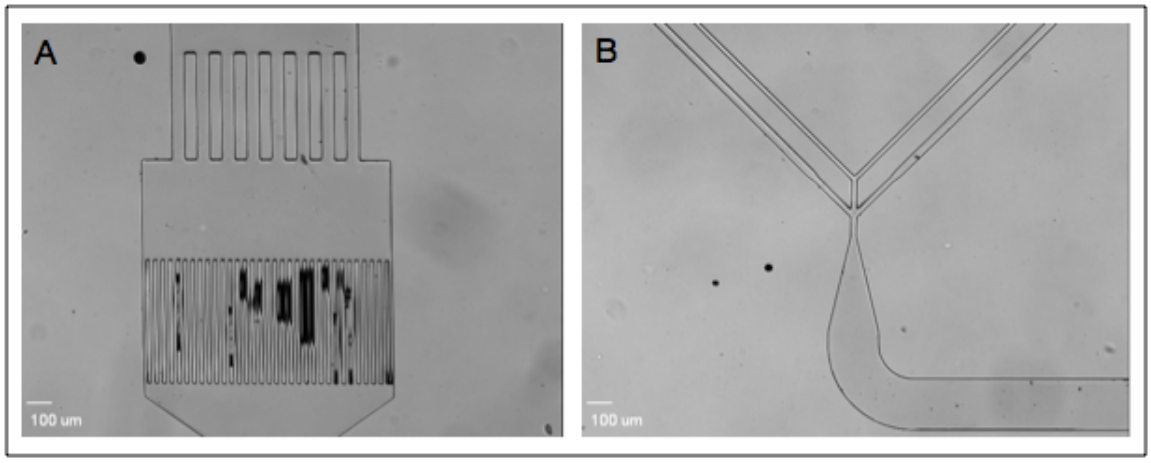


Figure 8: (A) Filters at the head of each input and (B) the double junction where aqueous inputs meet the oil phase.

For production of gel droplets in the redesigned device, the flow rates used were 1.0 µl/min for the aqueous inputs and 6.0 µl/min for the oil phase. This gave a flow rate ratio of 3:1 oil:aqueous phase, given that there are two aqueous inputs. The resulting droplets are approximately 40 µm in diameter. Table 1 shows more detailed

information regarding the resulting gel sizes based on changes in the flow rate ratio. In looking at the aqueous inputs (Table 3), the addition of ATP beyond the 1 mM present in typical ligation reactions is to account for both the instability, and thus breakdown, of ATP during its residency within the syringe and the prolonged ligation necessary to form the networked gel. Thus, a final concentration of 1.5 mM was used, representing a 50% excess over traditional ligation buffers.

Table 1: Mean diameter and standard deviation of DNA microdroplets using a 15 µm channel width and either a 1:1 (oil:aqueous phase) or 3:1 flow rate.

	1:1 Flow Ratio	3:1 Flow Ratio
Mean (µm)	38.03	24.77
SD (µm)	2.56	1.64

Furthermore, we did not use stock ligase buffer as provided by NEB or Promega, but rather CutSmart Buffer from NEB. Table 2 shows a comparison between the two NEB buffers. This change in protocol was enacted due to the presence of DTT in the buffer, which can interfere with downstream applications of the gels. More specifically, protein capture applications can be significantly hindered, as DTT can reduce divalent cations that will participate in NTA-metal ion chelation common in His-tag binding. As there is no ATP present in the CutSmart Buffer, this component is added to reach the 1.5 mM used in the supplemented ligase buffer reaction.

Table 2: Comparison of buffers used for gel ligation.

T4 DNA ligase Reaction Buffer (NEB)	CutSmart Buffer (NEB)
50 mM Tris-HCL	50 mM Potassium Acetate
10 mM $MgCl_2$	20 mM Tris-Acetate
1 mM ATP	10 mM Mg-Acetate
10 mM DTT	100 ug/mL BSA

Emulsions were collected in a 1.5 mL tube, prefilled with approximately 100 uL of mineral oil to reduce any droplet aggregation caused by collection in a dry tube. After collection, the droplets were incubated for 24 hours and then viewed under the optical microscope to check whether the gels had formed within the emulsions.

Droplet collection was performed by a rather rudimentary, but effective method. It has been found that the introduction of an electric field to a solution of water-in-oil emulsion droplets causes droplet instability by causing each emulsion to act as an individual dielectric as ions migrate to the their respective ends of the droplet based on the induced electric field[155–158]. Thus, to promote similar conditions, we transferred the emulsion solution from the 1.5 mL centrifuge tube to a smaller, 0.6 mL tubes and twisted the tube between our fingers with dry gloves for approximately one minute. We have made the conclusion that it is the static discharge from the twisting that causes the emulsions to break. The smaller diameter and larger surface of the smaller tube allows for more efficient breaking of the droplets. This conclusion is based on earlier experiments using both temperature changes and physical shaking to break the droplets; neither of these conditions effectively caused emulsion

aggregation. In addition, other methods such as hexane purification proved to be too harsh and often broke the DNA droplets apart.

To test the hypothesis that it is indeed the imparting of static charge to the tube that is inducing emulsion breakage, we plasma-treated a tube immediately before pipetting in a solution of water-in-oil emulsions. We compared this to a non-treated control tube and a tube on which the simple protocol previously described was performed. Figure 9 shows that compared to the non-treated control, the plasma treating and the 'static' charge breaking both separated the aqueous and oil phases significantly better. After breaking the emulsions and removing the oil now at the top of the solution, the microgels settle to the bottom of the tube, as seen in Figure 10. The gels were then collected, stained, and imaged (Figure 11) to confirm that they were synthesized from DNA and to view the porous structure of the resulting microgels. This porous structure should allow ample space for lysate diffusion into the gel.

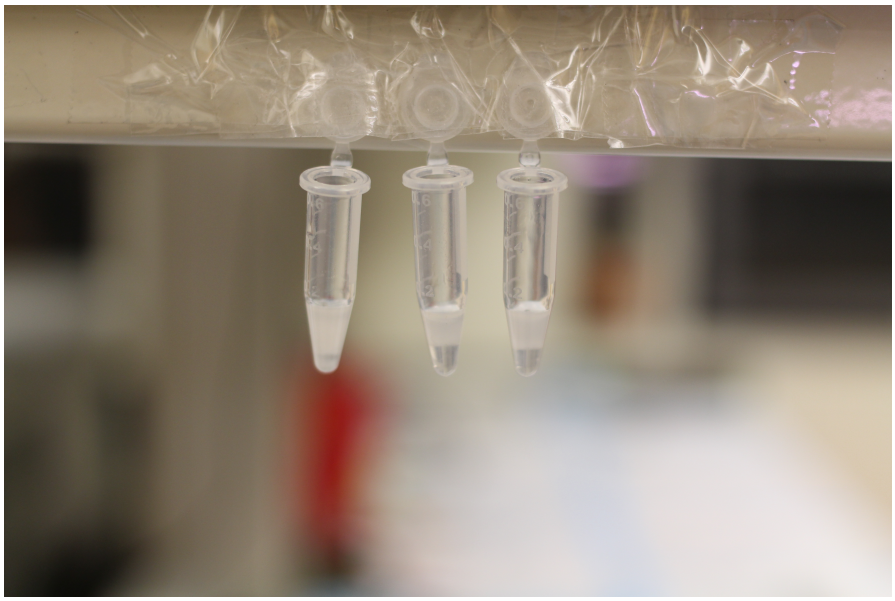


Figure 9: From left to right, breaking of emulsions in (A) non-treated tube, (B) plasma-treated tube, and (C) 'static' treated tube.

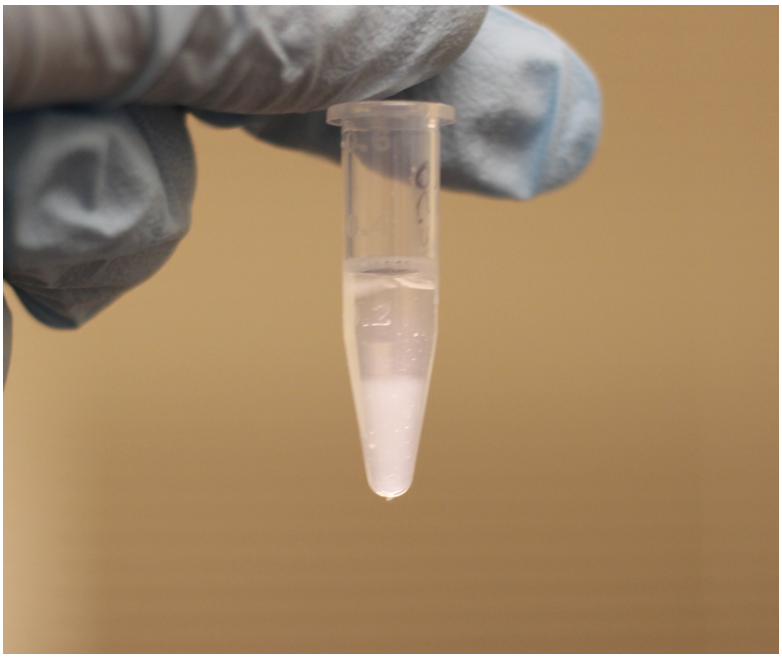


Figure 10: DNA microgels settled at the bottom of the extracted aqueous phase.

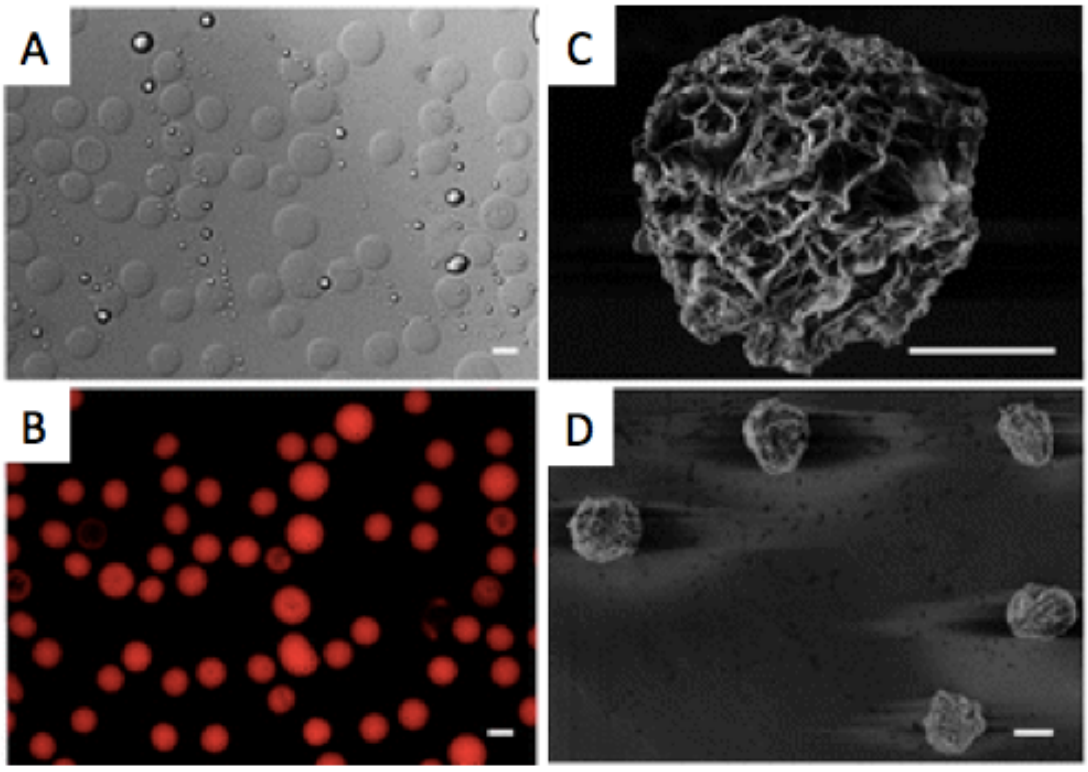


Figure 11: P-gel microdroplets viewed under (A) optical microscopy, (B) fluorescence microscopy, and (C-D) scanning electron microscopy. Scale bars are all 10 microns[154].

Section 2.3 – Summary

The P-gel format has previously been shown to increase protein yields in cell-free expression, but was relegated to either a bulk gel format or use of heterogeneous gel pieces after mechanical breaking. Here, we show the process by which we produce homogenous, cell-sized P-gel microdroplets through microfluidic emulsions. A straightforward method of gel collection from emulsions is also shown, which allows for the production and gathering of a large number of gels in a short period of time. The small size and homogeneity of the

microgels open up new applications in cell-free protein expression, especially greater control over gene isolation in mutant library testing.

Section 2.4 - Materials and methods

The oil used in these experiments was light mineral from Sigma-Aldrich, and was mixed with Abil-EM90 amphiphile (4% w/w) in order to allow for the stable compartmentalization of phases required for emulsion micelles. Before use, the oil-amphiphile mixture was syringe-filtered through 0.25 μm cellulose to reduce potential clogging of the microfluidic device. Aqueous inputs to the device are listed in Table 3.

Table 3: Aqueous inputs for the microfluidic production of P-gel microdroplets

Syringe 1		**Syringe 2**	
X-DNA (300 μM)	83.33 μL	NEB T4 Ligase (2,000 u/μL)	0.8 μL
Linear Plasmid (1500 ng/μL)	15.53 μL	Cutsmart Buffer (10X)	20 μL
Tris-NaCl Buffer (10 mM)	1.14 μL	ATP (10mM)	30 μL
		NF-H_2O	49.2 μL
Total	100 μL	Total	100 μL

CHAPTER 3: PROTEIN EXPRESSION – PLASMID AND X-DNA DESIGN

Section 3.1 - Plasmid selection and linearization

The ability of X-DNA to form a gel network is based on the covalent linking of monomers brought in proximity through attractive interactions induced by hybridization of single-stranded DNA sticky ends. Single-stranded DNA hybridizes specifically to opposing strands through Watson-Crick base-pairing, which is controlled through sequence design. In order to create conditions where each arm of the X-DNA can interact with any arm of an opposing X-DNA in a solution of identical X-DNA, the sticky end overhang must be palindromic. Additionally, self-ligation of X-DNA arms is minimized or eliminated completely by controlling the length of each arm. The double-stranded portion of each arm is 18 base pairs, well below the persistence length of dsDNA (≈50 nm, corresponding to approximately 150 bases), and thus the arms should remain stiff. The goal of crosslinking genes within this X-DNA gel network limits the types and number of sticky ends that can be used – the overhang must be complementary to that produced by the restriction enzyme used to linearize the gene/plasmid. The majority of restriction enzymes produce four base-pair palindromic overhangs.

In the original P-gel work performed in our laboratory, the plasmid pIVEX1.3 was used to express green fluorescent protein (GFP)[49]. This plasmid is optimized for expression in wheat germ lysate. However, my initial expression work was performed in *E. coli* lysate, as wheat germ lysate is significantly more expensive to purchase and

more time consuming/expensive to produce in the laboratory for initial work. At the time of this writing, *E. coli* lysate is cheaper when purchased commercially and is relatively straightforward to produce in the lab. Additionally, the decision to use *E. coli* became a matter of necessity rather than a cost consideration based on the need to make certain changes to the chemical makeup of the lysate, which will be discussed further in Chapter 4. For the purpose of this discussion, the takeaway is that a plasmid optimized for *E. coli* lysate was selected - pIVEX2.3d as supplied by 5' Prime, Inc.

Whereas pIVEX1.3 was linearized with ApaI (5' – GGGCCC – 3'), pIVEX2.3d does not contain this restriction site. Another site was needed that only presented at one location in the plasmid sequence and was not present in either of the cloned genes we would be expressing, GFP and mCherry. The AatII restriction site (5' – GACGTC – 3') was selected as it met these conditions, and additionally was not present in sequence regions near the promoter or terminator of T7 polymerase, which is responsible for the expression of the cloned gene. This reduces the chance that the digestion of the plasmid will affect expression. Figure 23 shows the location of the AatII restriction site.

There is a critical interplay between the restriction enzyme used for plasmid linearization and the resulting quality of P-gel, in regards to its physical appearance under bright field microscopy and its ability to withstand incubation under conditions associated with cell-free protein production. Though the details of these topics are discussed separately, the X-DNA design is dependent on the choice of restriction site because the linearized plasmid must find complementary sticky ends on the X-DNA. Restriction enzymes naturally recognize a broad range of sequences and produce a

variety of sticky ends, the majority of which are palindromic four base-pair overhangs. These overhangs vary in directionality (5' vs. 3' overhangs) and base makeup. In turn, sequence differences affect stability of hybridization, making them more or less likely to be ligated depending on the Gibbs free energy (ΔG) value of the particular sequence match. Thus, this description of hybridization stability affects the macroscale gel morphology, as higher ligation efficiencies yield a more highly networked structure.

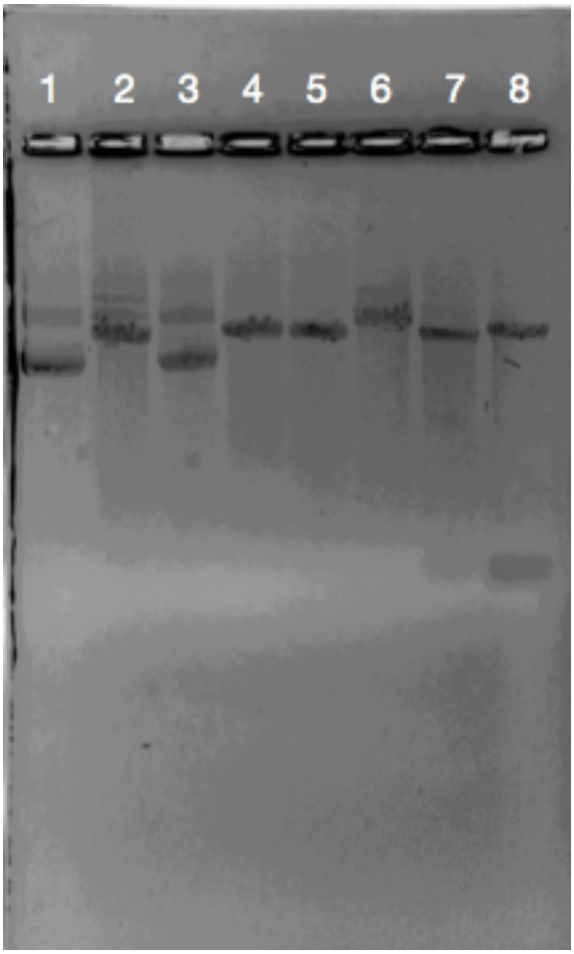
This interplay between gel strength and sticky end choice explains why AatII was not selected as the restriction enzyme for linearization. Though the linearized plasmid displays expression in cell-free lysate, as in seen in Figure 18 and Figure 19, the overhang sequence does not allow for a robust gel to form within the time conditions used for ligation. The negative gel formation results elucidated the importance of overhang sequence in DNA gel formation even when ligation, and not strictly hybridization, is the method of network formation. Thus, in searching for another restriction site, the sites were limited to overhangs with a lower ΔG than AatII and at least comparable to ApaI. The details of this process are detailed in Section 3.3. The sites BspEI and NgoMIV were selected, although their locations within the plasmid were a cause for concern due to their close proximity to the T7 terminator (BspEI) and T7 promoter (NgoMIV). Though the sites are not within the associated sequences themselves, digestion of the plasmid nearby may negatively affect expression through two mechanisms. Firstly, promoter and terminator sequences induce conformational changes within the DNA to have their desired effect on polymerase[159–163], and digestion of nearby DNA may interfere in this process. Secondly, and perhaps more importantly, linear DNA is prone to digestion by

exonucleases; the closer the gene itself is to the end of the linearized plasmid, the greater the likelihood that exonuclease will degrade the gene itself and lower overall expression. However, the details of this potential effect are not well understood, as the vast majority of gene expression is performed off of circular plasmids, and in those instances where linear genes are used, these design considerations are not deeply discussed.

The solution-phase expression levels for both BspEI and NgoMIV digests were significantly lower than that produced by AatII, as shown in Figure 19. Though this may not directly correlate with the expression levels from P-gel microdroplets, we wanted a linearization site that did not reduce solution-phase expression so drastically. The relative increase in protein yield provided by P-gels is much greater when using linear plasmids with minimal solution-phase expression, as the P-gel reduces exonuclease degradation of plasmid by ligating the free ends into a DNA network. Thus, an expression comparison was needed that compared the P-gel sample to a linear plasmid control that could express to a significant degree. Based on the resulting gel strengths associated with different sticky end sequences that will be discussed in Section 3.2, we made the decision to introduce the MluI restriction site into the pIVEX2.3d plasmid.

Since this restriction site was being introduced manually, we could search for a possible mutation site as far from the gene of interest as possible to minimize the effect of linearization on cell-free expression. Figure 23 shows the selected region on the plasmid, which was selected based on its position opposite the gene, the minimal two base change needed for MluI site introduction, and its spacing in between the

antibiotic resistance gene and the origin of replication. As both the pIVEX2.d-GFP and pIVEX2.3d-mCherry plasmids had already been created, we needed to perform this process separately on each plasmid. The primers used for the site-directed mutagenesis are detailed in Table 6. Digestions of the mutated plasmids were undertaken to ensure that the plasmids isolated from different colonies did in fact contain the MluI mutation. These results are shown in Figure 12.



Lane 1	Circular pIVEX2.3d-GFP
Lane 2	AatII digestion of pIVEX2.3d-GFP
Lane 3	MluI digestion of MluIpIVEX2.3d-GFP – Colony 1
Lane 4	MluI digestion of MluIpIVEX2.3d-GFP – Colony 2
Lane 5	MluI digestion of MluIpIVEX2.3d-mCherry – Colony 1 – GC Buffer
Lane 6	MluI digestion of MluIpIVEX2.3d-mCherry – Colony 2 – GC Buffer
Lane 7	MluI digestion of MluIpIVEX2.3d-mCherry – Colony 1 – GC Buffer + DMSO
Lane 8	MluI digestion of MluIpIVEX2.3d-mCherry – Colony 2 – GC Buffer + DMSO

Figure 12: 1.0% agarose gel showing testing of different colonies to ensure the plasmids isolated contained the mutated MluI site. PCR buffer conditions are also mentioned for the red fluorescent proteins, as they required changes to the traditional PCR mix.

Based on the MluI digestion results, we selected colony 2 of GFP and colony 2 for mCherry. However, after running cell-free expression experiments to confirm whether the plasmids were functional in SPS, we gathered positive results for MluIpIVEX2.3d-GFP but not for the MluIpIVEX2.3d-mCherry. This mCherry sample was selected because it was produced in GC buffer, rather than GC+DMSO buffer, and thus presented a smaller possibility of extraneous base mutations[164]. In regards to proclivity for mutations during PCR, the buffer choice in increasing order of potential for base mutations is as follows: HF, GC, and GC+DMSO. MgCl2 and DMSO both serve to stabilize base pair hybridization, and can potentially maintain an unstable base pairing long enough for it to be incorporated into the growing DNA strand. The main difference between HF and GC buffer is the inclusion of a higher MgCl2 concentration in the GC buffer. As can be seen, HF buffer did not give positive results in the cases tested. Thus, only after colony 2 of GC mCherry was shown to have low expression were colonies 1 and 2 for GC+DMSO mCherry tested.

Section 3.2 - Selection of red fluorescent protein

Fluorescent proteins are often used as model or reporter proteins because their inherent fluorescence under UV excitation provides a direct and straightforward method of determining how much, and of which, protein has been produced. Green fluorescent protein is the most commonly used fluorescent reporter, and has been mutated many times over to create a large number of variants[55,56]. These variants have improved fluorescence properties in addition to having varied excitation and emission

spectra, producing different color signals to allow labeling and detection of multiple target simultaneously[57,58].

The company 5 Prime supplied the expression plasmid pIVEX2.3d with an accompanying control plasmid containing GFP with a C-terminal 6His-tag. We could then express wild-type GFP without having to perform the cloning manually. Being that one of the goals of this work is to demonstrate that, in lysate, microgels can express their own gene set and capture a representative protein sample produced from those genes, we required another fluorescent protein distinctive from GFP. This would allow us to produce two sets of P-gel microdroplets, one containing genes for green fluorescent protein and another for red fluorescent protein, and then optimize conditions whereby each gel fluoresced predominately green or red. Such a result would demonstrate that proteins from the gene set within the gel had a greater probably of securing themselves to the associated gel than do proteins produced by other gels. This is based off the assumption that even if translation is occurring in the solution and not just in the gel itself, which was the conclusion based off of our earlier work[50], the majority of mRNA in the immediate vicinity of each gel will be produced from the plasmids contained within. Thus, there should be a higher concentration of associated proteins in the volume surrounding the gel, leading to a representative display rather than simply an average of the larger solution volume.

There are many choices for red fluorescent protein, but in general red fluorescent proteins do not possess the same brightness as offered by green fluorescent proteins. At the beginning of this work, we would have selected Tdtomato based on its on its strong fluorescence signal compared to other red fluorescent proteins[56].

However, selection of a reporter protein is made more complicated by the fact that cell-free reactions often lack much functional cellular machinery beyond those for transcription and translation; thus, more complex proteins often need modifications made to the lysate to allow for their organization even if it appears to be a small structural change. Tdtomato, for example, is supposed to have the strongest fluorescence output but is a dimer rather than a monomer, and this extra level of organization may not take place efficiently in the cell-free reactions. Thus, it was felt to be prudent to clone and test three different red fluorescent reporter proteins: dsRed, mCherry, and tdTomato. We performed Gibson assemblies to create three pIVEX2.3d vectors containing each of these three proteins. The primers associated with the production of the plasmid backbone for each Gibson assembly are shown in Table 6. Colony growth demonstrated that the cloning was indeed successful, and Figure 13 shows the slightly different variations of red in the colonies

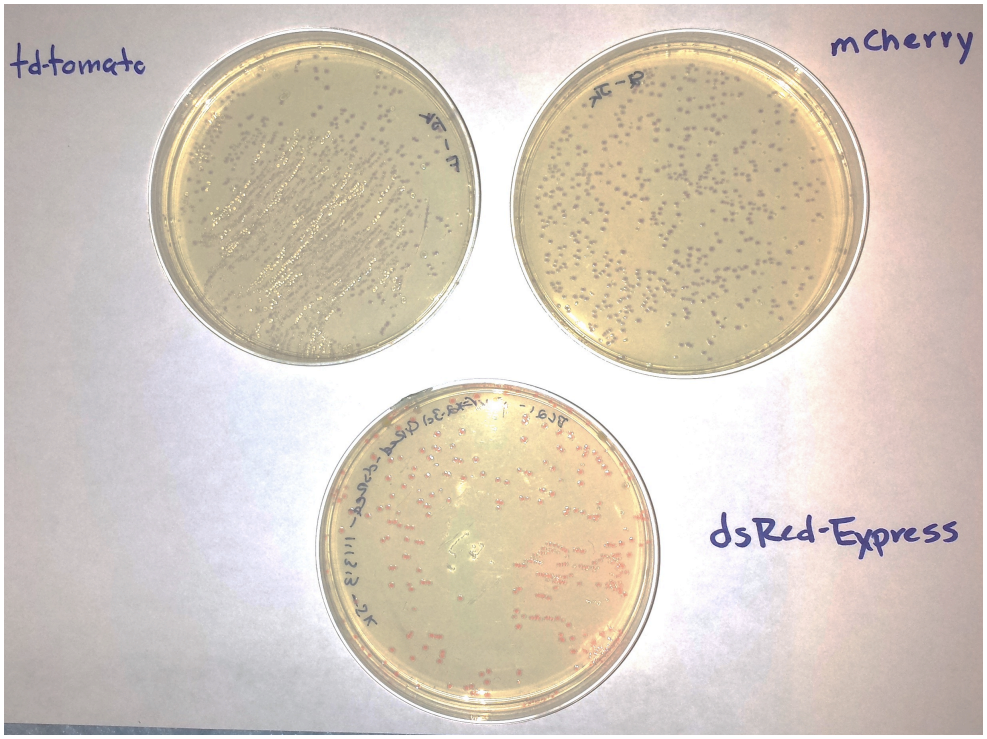


Figure 13: Colony growth showing the expression of various red fluorescent proteins in pIVEX2.3d

Figure 14 and Figure 15 demonstrate that mCherry provides the highest fluorescent signal after cell-free expression, using a 10 uL reaction with 100 ng of plasmid DNA. Selection of the mCherry as the red fluorescent reporter protein was based on this expression in cell-free lysate.

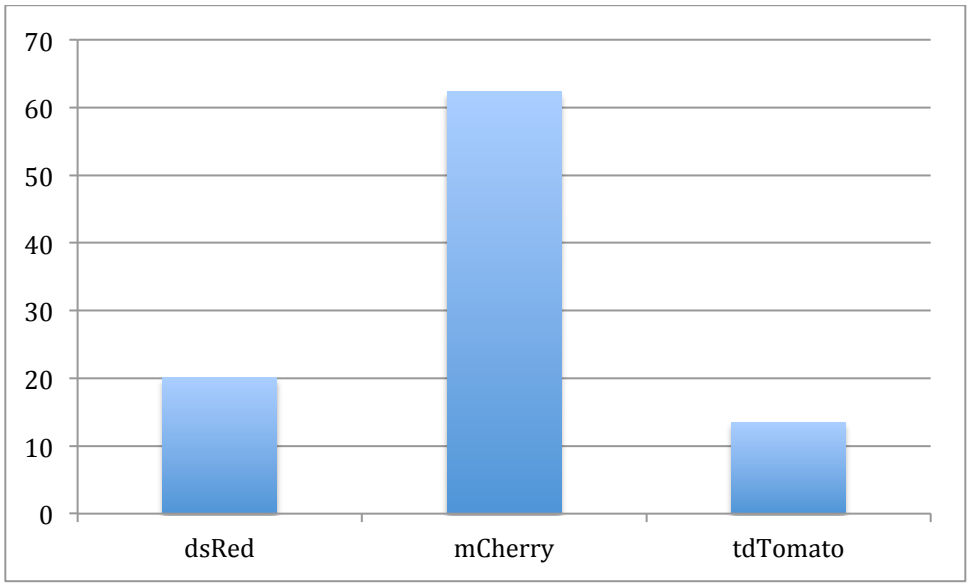


Figure 14: Fluorescence comparison between three red fluorescent proteins expressed in cell-free lysate using the pIVEX2.3d expression plasmid.

Figure 15: Solutions of (A) dsRed (B) mCherry (C) and tdTomato expressed in cell-free lysate and viewed through fluorescent microscope.

Section 3.3 - Gel formation based on sticky end sequence

Early P-gel experiments were performed using ApaI for the linearization digestion. This enzyme leaves a 5' – CCGG – 3' overhang sequence on the 3' ends of

the resulting linearized double strand, and thus this sequence also was included on the 3' ends of each strand making up the X-DNA. The majority of restriction enzymes leave four base-pair sticky ends, which alone are not sufficient to induce gel formation through simple hybridization. The melting temperature of this base-pair matching is too low (<10°C) to form stable, temporally significant interactions, but they do allow for ligation to occur. This is in contrast to gels that form solely through hybridization interactions; such gels utilize overhangs of at least eight base pairs[48].

As plasmid considerations came into play, discussed in Section 3.2, we began to use AatII linearized plasmid that presented overhangs of 5' – ACGT – 3' on the 3' ends of the double strand. Since we were not looking to create a gel through strong overhang hybridization but rather ligation, the base sequence of the overhangs was not initially considered critical to the design. However, microgel formation did not occur in emulsions using X-DNA with AatII overhangs, as shown in Figure 16. In order to maintain reproducibility in experiments, we held gel ligation times to 24 hours. Samples of AatII X-DNA left to ligate for significantly longer showed the ability to eventually form microgels, but it proved too long to use effectively and often varied between experiments. Within the 24 hour ligation time, gels were not seen in the solution extracted from emulsions or even in the emulsions themselves. After multiple attempts at attempting ligation at both 16°C, the optimal temperature for ligase, and 4°C, to ensure greater hybridization efficiency, we were not able to achieve successful DNA gel formation using AatII overhangs.

This stood in contrast to the work performed to synthesize P-gel microdroplets using ApaI-sticky end X-DNA. As compared to AatII X-DNA, gels produced using

ApaI sticky ends did form within emulsions, but had mixed results after extraction from the emulsions into pure aqueous phase. The gels tended to break up either from the extraction procedure itself or from pipetting and shaking, which is necessary for the cell-free expression.

These results suggested that the overhang sequence was critical to gel formation. As the melting temperatures of these four base-pair overhangs are quite low (less than 10°C), it is useful instead to compare their Gibbs free energy of formation. The 3' overhang of 5' – ACGT – 3' presented by AatII digestion possesses a ΔG of -6.5, while the 3' overhang of 5' – CCGG – 3' for ApaI digestion possesses a ΔG of -9.28. This matches with the experimental results showing that ApaI overhangs produce gels within emulsions while AatII does not. AatII may not be stable enough over a time period long enough to be amenable for ligation, and thus although it is theoretically able to undergo ligation, gelation requires a high ligation efficiency that is not being achieved with the smaller ΔG value.

As AatII did not provide usable microgels, another restriction site was needed. In analyzing suitable restriction enzymes, BspEI appeared to match the requirements needed. However, the restriction site (5' – TCCGGA – 3') is significantly closer to the cloned gene in the plasmid, and although it lies just outside the T7 terminator sequence, it is sufficiently close to possibly affect expression. The 5' overhang of 5' – CCGG – 3' provides a ΔG of -9.75 kcal/mole, and thus appeared to be a more suitable option for droplet formation than previous choices. Gel formation in emulsions was viewed, and in fact, the gels appeared to be more consistent in size, shape, and appearance than in the ApaI microgel sets. This can be attributed to the larger ΔG

value. Isolation from the emulsions also was successful, leaving behind full droplets and very little remnants and artifacts from broken droplets that could potentially interfere with protein capture further along in the experiment. The results of the gel droplet formation are shown in Figure 16, and corresponding bulk gel formation is shown in Figure 17.

This BspEI overhang gel seemed at first to be the solution to many of the problems plaguing gel formation and isolation, but another difficulty arose in the aforementioned location of the gene linearization. As seen in Figure 18 and Figure 19, cell free expression of pIVEX2.3d-GFP linearized with BspEI is lower than that provided plasmid cut with AatII. A greater amount of the linearized plasmid is needed to begin seeing signal significantly beyond background. It was considered beforehand that the close proximity of this restriction site to the T7 terminator may pose an issue, and indeed the proximity does negatively affect expression.

Another cutting site was selected based on its production of the same 5' – CCGG – 3' overhang as BspEI. NgoMIV is present just in front of the T7 promoter; again, this is closer than we would be comfortable with, but went ahead in testing the expression of the plasmid digested with this enzyme. Proximity to the cloned gene, specifically to the T7 promoter, interfered with expression, as seen in Figure 19.

It became clear that in order to take into consideration gel strength and expression efficiency, a restriction site would have to be added into the plasmid. Based on this manual sequence change, we had total control over what site we could use. Thus, we aimed to capitalize on the apparent trend seen in increasing gel robustness based on increasing ΔG value of overhang hybridization by analyzing the ΔG values

of all possible four base-pair overhang combinations consisting of 100% cytosine and guanine bases. For this analysis, we used the Oligoanalyzer Tool on the IDT® website[165]. As shown in Table 4, the 5' overhang 5' – CGCG – 3' presents the largest ΔG of -10.36 kcal/mole. With the physical data demonstrating significant morphology differences associated with small changes Gibbs free energy, this sequence was selected and introduced to the plasmid through site directed mutagenesis. The restriction enzyme MluI provides this selected overhang, and recognizes the sequence 5' - ACGCGT 3'. A site was chosen that required a two base mutation and was essentially directly opposite the cloned gene and not present in any of the other critical regions in the plasmid (origin of replication or ampicillin-resistance gene).

Table 4: Comparison of X-DNA sticky end overhangs, showing the enzyme recognition site, the resulting overhang, and the Gibbs free energy associated with the overhang hybridization.

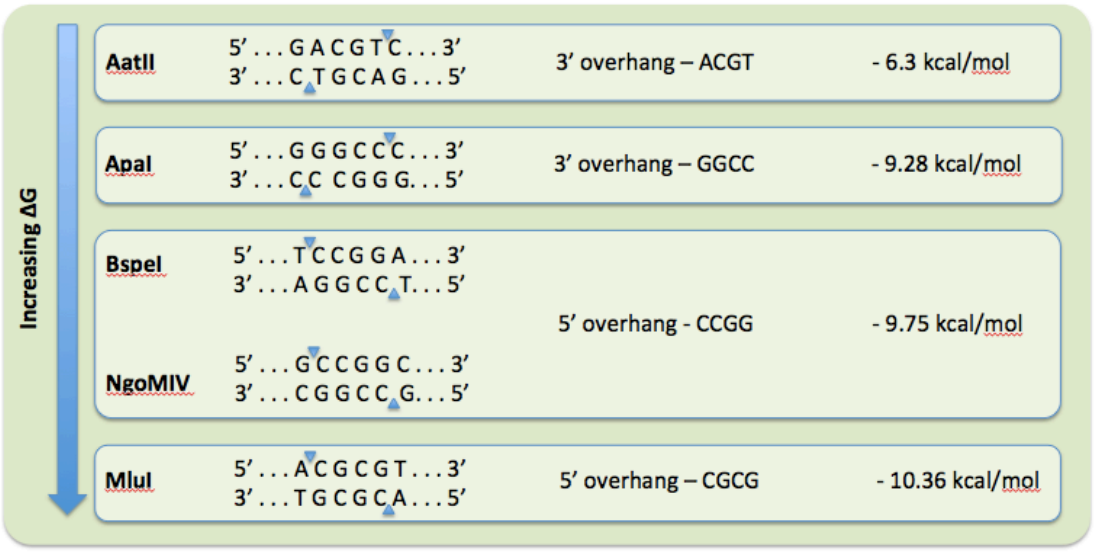


Table 5: Description of the different four base-pair overhangs used in the X-DNA design.

4 Base Pair Overhang (Overhang Direction)	Corresponding Restriction Enzyme Site	Gel Result
GGCC (3' overhang)	ApaI	No bulk Gels in emulsion Poor extraction
ACGT (3' overhang)	AatII	No bulk No gels in emulsion (Not applicable)
CCGG (5' overhang)	BspEI, NgoMIV	Bulk Gels in emulsion Good extraction
CGCG (5' overhang)	MluI	Bulk Gels in emulsion Good extraction

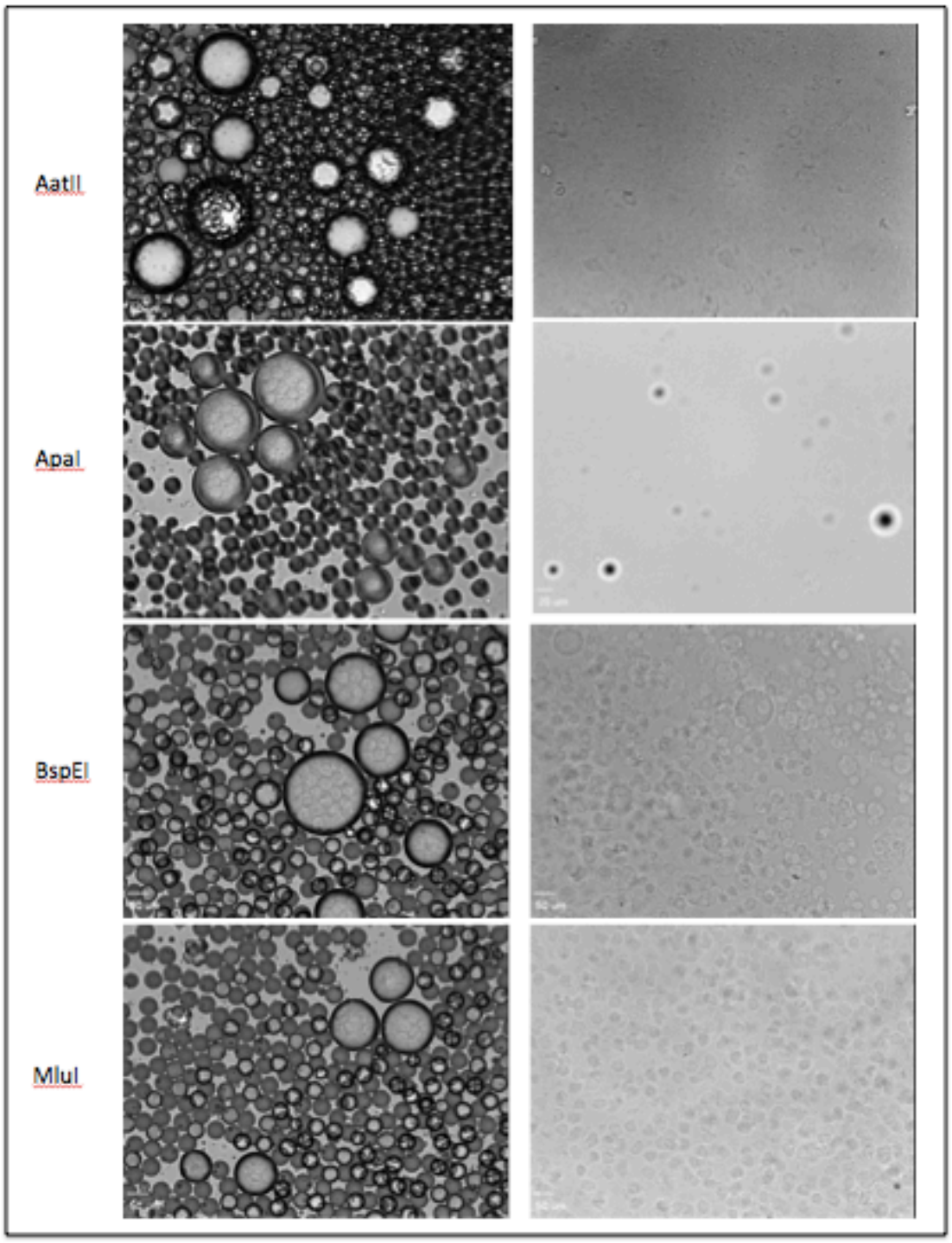


Figure 16: Views of X-DNA droplets in emulsions and after extraction into bulk aqueous phase. The restriction site names to the left define the overhangs used on the X-DNA.

After cloning, transformation, and plasmid isolation, the gene was digested to ensure that the site was introduced, and then the linear plasmid was tested in cell-free lysate to ensure that expression was higher than shown with BspeI and NgoMIV plasmids (Figures 18 and 19). Synthesis of X-DNA microgels was performed under the same conditions as the previous gel-emulsions experiments, which were described in Section 2, and the results of the synthesis are shown in Figure 16. One can see that in both the emulsion and extracted aqueous phase, the microgels produced from MluI-X-DNA appear to be robust and show little signs of break-up or fracturing. Based on the expression results and the strong gel formation using MluI, this X-DNA was used in the further experimental work on P-gel, including functionalization and expression work as discussed in Chapter 4. To ensure all X-DNA in these experiments was in fact being produced in high-yield and that results concerning gel formation were differences associated with X-DNA hybridization, we compared them on a 2.5% gel. Figure 20 shows that all X-DNA are being formed efficiently, and thus the overhang sequence is not affecting the efficiency of X-DNA hybridization.

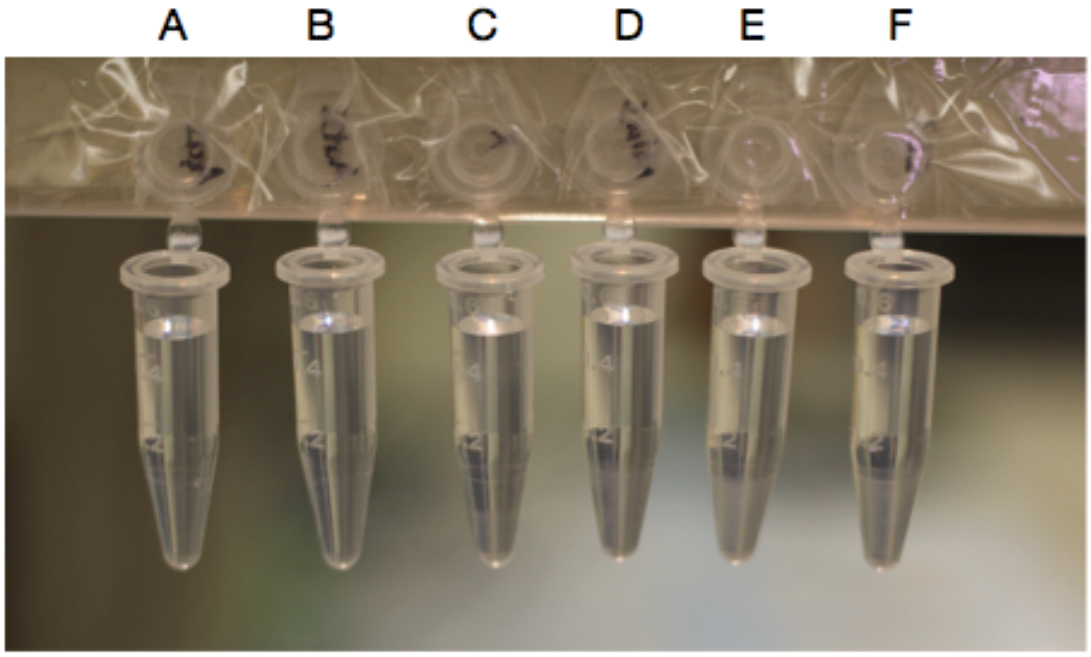


Figure 17: Comparison of bulk gel formation using X-DNA possessing four (A) AatII, (B) ApaI, (C) BspEI, and (D) MluI overhangs. Concentration of X-DNA (52.5 µm) was set at that used for P-gel bulk production in earlier work[49,50]. Tubes E and F are gels created by BspEI X-DNA that has been psoralen-crosslinked by 5:1 and 10:1 psoralen:X-DNA ratios, respectively.

Cell-free expression from the MluI P-gel microdroplets was found to be similar to that found in a system containing approximately 20 ng/uL of linearized plasmid. In looking at the solution used for P-gel synthesis in Table 3, the final P-gel sample contains approximately 17 ng/uL crosslinked into the gel used in the reaction. The expression was shown to be the approximately the same as 20 ng/uL MluI-linearized plasmid expression, which represents an approximately 18% increase in protein yield.

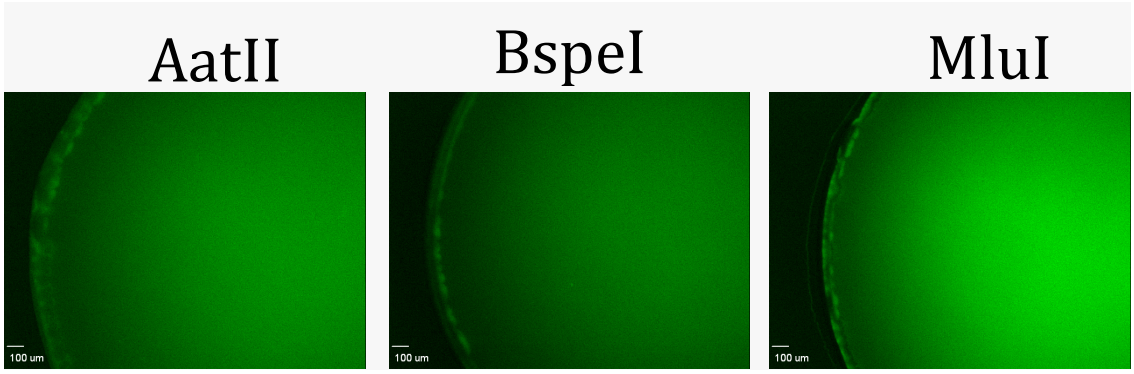


Figure 18: Cell-free expression comparison of GFP in droplets using different plasmid linearization sites. Solutions viewed were 1.0 uL and 2000 ms exposure with green UV filter was used. The expression vector is pIVEX2.3d linearized with (a.) AatII (b.) BspeI (c.) MluI. Each reaction was a 11.5 uL reaction containing 500 ng of vector.

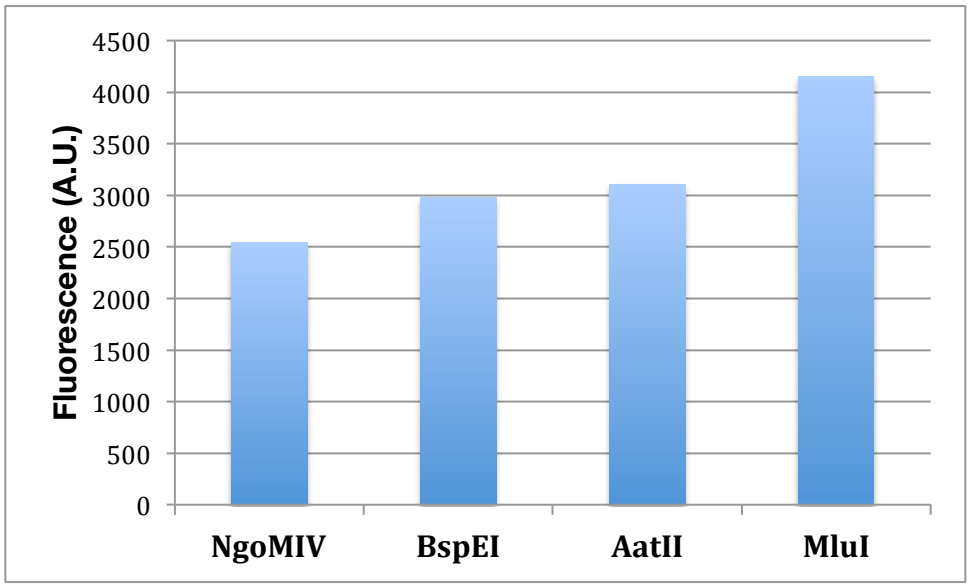
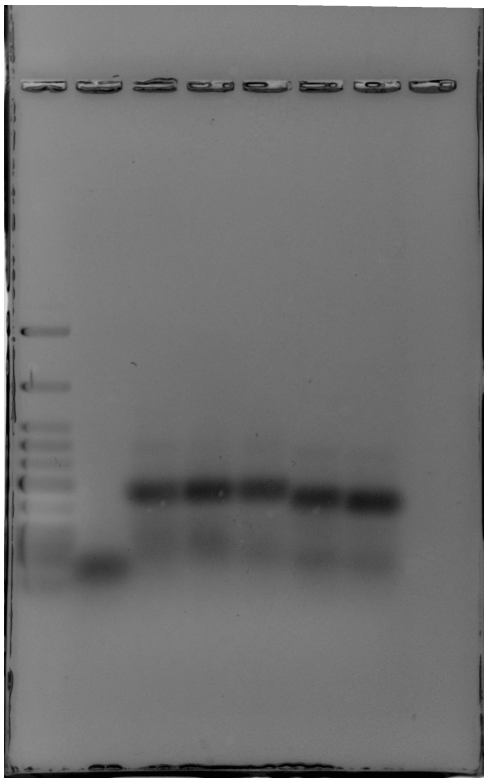


Figure 19: Cell-free expression levels of pIVEX2.3d-GFP linearized at different restriction enzyme sites.



Lane 1	NEB LMW Ladder
Lane 2	Oligo 4 of X-DNA
Lane 3	ApaI X-DNA (modified X-DNA sequence)
Lane 4	AatII X-DNA
Lane 5	ApaI X-DNA
Lane 6	BspeI X-DNA
Lane 7	MluI X-DNA

Figure 20: 2.5% agarose gel showing that all X-DNA with different sticky ends form in high yield. The modified X-DNA sequence for lane 3 is the sequence presented in Table 12 for psoralen-crosslinked X-DNA.

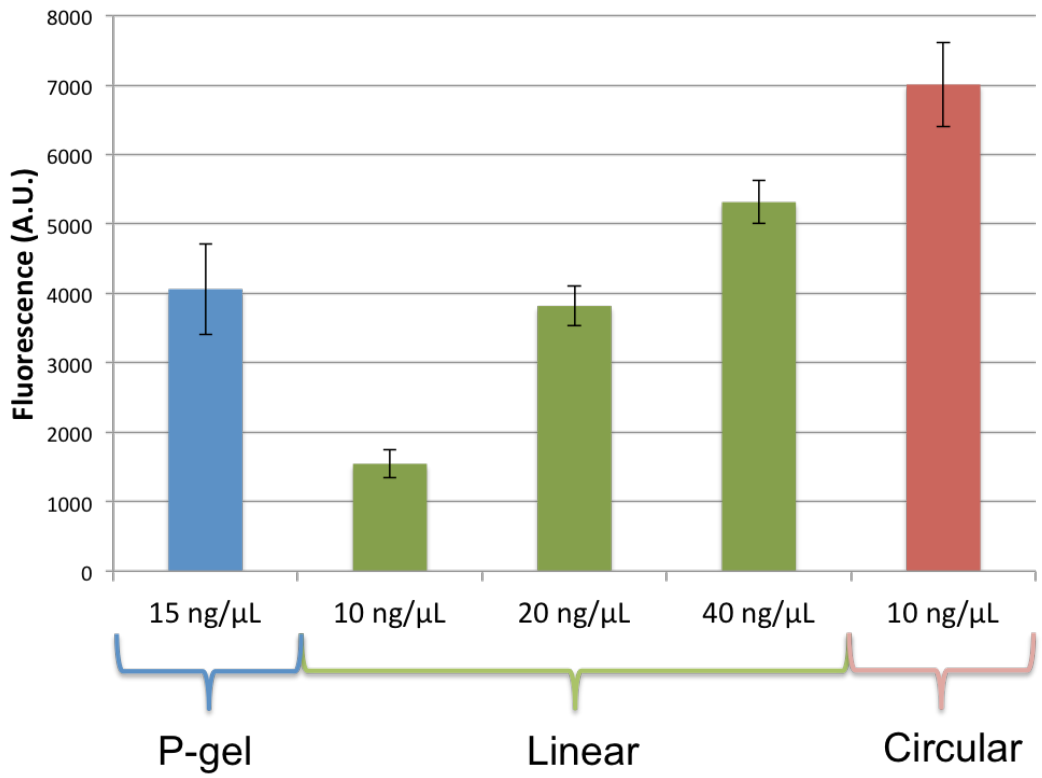


Figure 21: Cell-free expression of GFP in linear (MluI-digested) and circular plasmid formats.

Section 3.4 – Summary

The work discussed in this chapter serves to elucidate the design considerations that are important in P-gel microdroplet formation and gene expression. Though ligation is driving the crosslinking of X-DNA to form a networked structure, the sequence of the four base-pair overhangs is shown to be critical to final microgel strength. Furthermore, the restriction site used to linearize expression plasmid was shown to significantly affect the expression of protein in the solution-phase. This is an important aspect of system design

because P-gel expression yields are compared to solution-phase controls, which show drastically different expression based on the length of sequences flanking the gene of interest. Thus, the P-gel microdroplet format, which minimizes the need for long flanking sequences by crosslinking the gene into a larger DNA network, is extremely useful for linearized vectors with low expression and gene cassettes that often just contain the necessary promoter and terminator sequences flanking the gene.

Section 3.5 - Materials and methods

Protocol for preparing X-DNA

The annealing process for synthesis of X-DNA from four oligos is shown in Figure 22. It requires a slow ramp down in temperature to ensure proper hybridization, and is performed in 10 mM Tris buffer with 50 mM NaCl at a 100 μM concentration of each strand. After annealing, the solution is concentrated in a 30 kDa MWCO column and washed in the 10 mM Tris with 10 mM NaCl. The solution of X-DNA is aliquoted and stored at -20°C for long-term storage and at 4°C for a working solution.

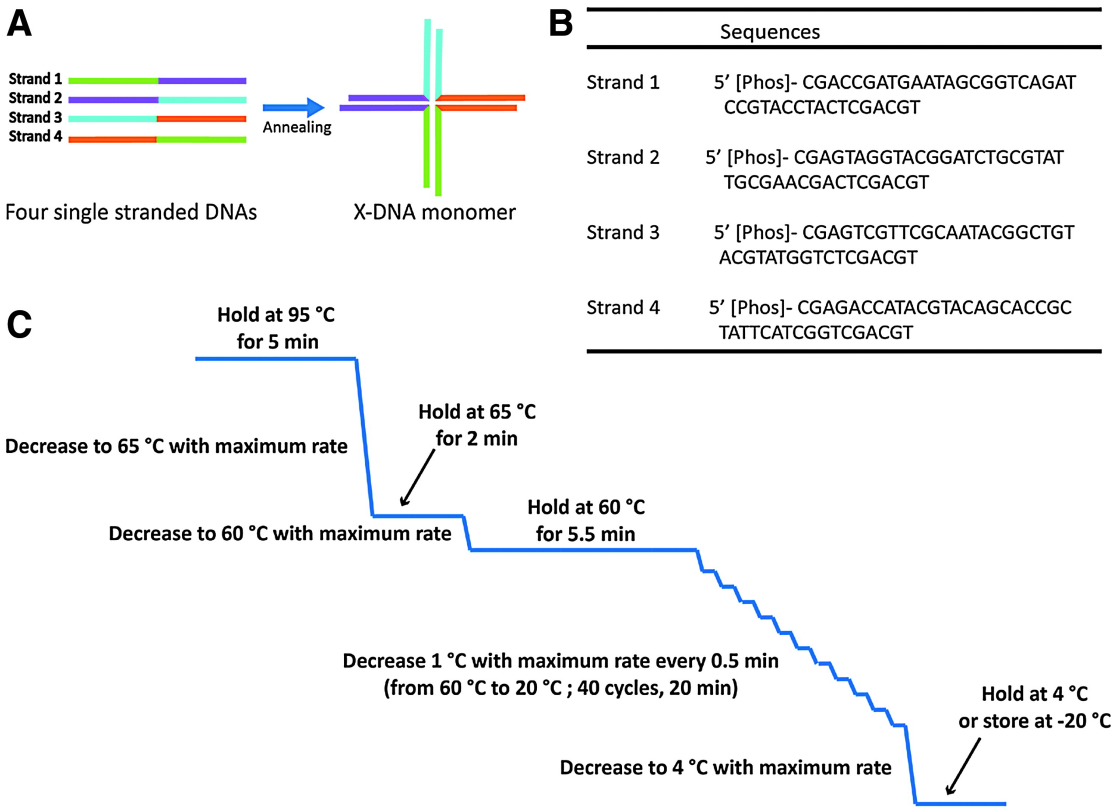


Figure 22: Four single-stranded oligos are annealed to form X-DNA. The colored segment in the (A) show sequence regions that are complementary[154].

Site-directed mutagenesis

Phosphorylated forward and reverse primers, shown in Table 6, were designed to incorporate a two base-pair change at the location of the new MluI site. PCR was performed using the Phusion® High-Fidelity DNA Polymerase from New England Biolabs® Inc., with protocol and buffer used according to kit instructions. Blunt-end ligation was performed following PCR to create circular plasmids that were then transformed into One Shot® Top10 Chemically Competent *E. coli* from Invitrogen™.

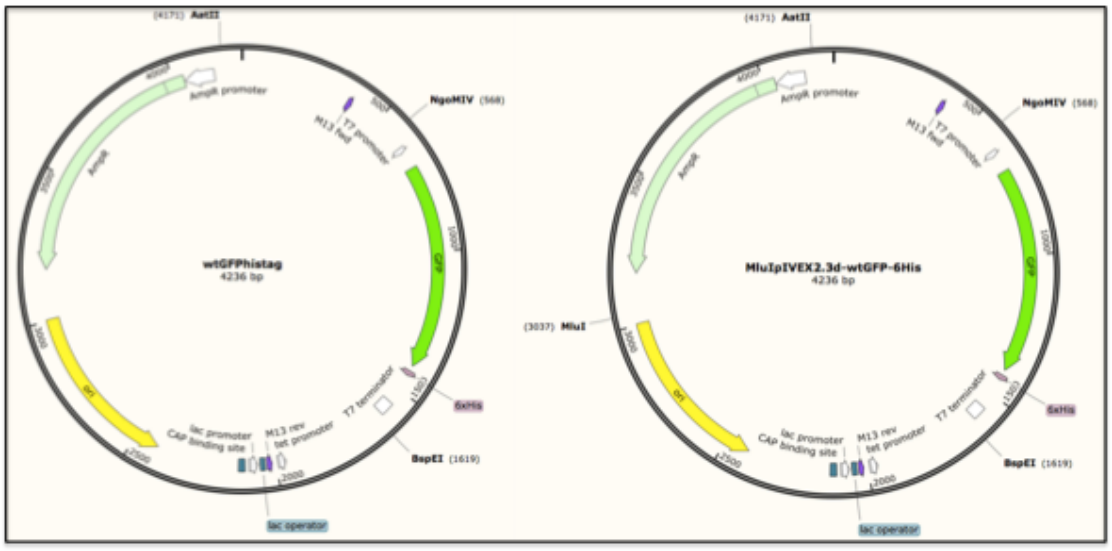


Figure 23: Plasmid map of pIVEX2.3d-GFP (left) and mutated MluIpIVEX2.3d-GFP (right). The addition of the MluI site was performed at a site opposite the gene of interest and between other coding regions.

Table 6: Primer sequences for different aspects of pIVEX2.3d manipulation.

Sequencing	
Forward	5' – TATAGGGAGACCACAACGGT – 3'
Reverse	5' – AGTGTGCTGGAATTCGC – 3'
Gibson Template	
Forward	5' - CAACCCGGGGGGGGTTC – 3'
Reverse	5' – GGTATATCTCCTTCTTAAAGTTAAACA – 3'
MluI mutation	
Forward	5' – /5Phos/GTAGTGGAACGAAAACTCACG – 3'
Reverse	5' – /5Phos/GCGTCAGACCCCGTAG – 3'

Cell-free expression reactions

Reactions in cell-free lysate were prepared under the same guidelines as reported in literature, specifically in our lab's P-gel expression work[50]. Briefly, 20 ng/uL of linear plasmid was used in cell-free reactions while 10 ng/uL circular plasmid was used. This is based on expression tests shown in Figure 21 that demonstrate these gene concentrations lead to similar expression levels. More information regarding the specific solution makeup of the lysates are found in Table 9.

CHAPTER 4: GEL FUNCTIONALIZATION

Section 4.1 - Production of homemade E.coli lysate

Due to the large number of expression experiments that were necessary to optimize expression and capture conditions, it was necessary to produce my own *E.coli* lysate. The original procedure for this lysate production has been documented previously, and is shown in Table 9. This procedure is an abbreviated lysate preparation outlined by Kim et al[166]. It is important to note that the final lysate solution is actually created from two solutions – the actual lysate gathered from cultured *E. coli*, and a buffered reaction mixture. The lysate itself mainly contributes the enzymes necessary for transcription and translation, while the reaction mixture contains the chemical energy, nucleotides, amino acids, and other molecules necessary for transcription and translation to continue over an extended time period. The combination of these two solutions to form the final mixture is termed 'lysate' unless otherwise noted.

Production of homemade lysate was not only a cost consideration, but was essential based on the goal of protein capture. The most widely used method of protein capture is His-tag binding, where a 6-histidine tag (though shorter or longer repeats can be used) on the end of protein specifically binds a chelated divalent metal cation. As we intended to functionalize DNA gels to achieve this capability, which will be discussed in Section 4.2, we needed to ensure the solution supported this method of binding. Commercial lysate kits, and homemade lysate recipes, include DTT as a

means to reduce protein-protein interactions by disrupting disulfide bond formation, thus increasing potential reaction time. As His-tag binding is dependent on Ni^{2+} chelation by NTA, DTT must be removed as it can reduce the nickel ions and render the binding method unusable. Figure 24 demonstrates that in a control experiment whereby GFP-6His is incubated with agarose-NTA-nickel beads from Qiagen, the DTT concentration in the lysate is sufficient to cause little or no protein capture. In the same lysate without DTT, the beads show a significant fluorescent signal.

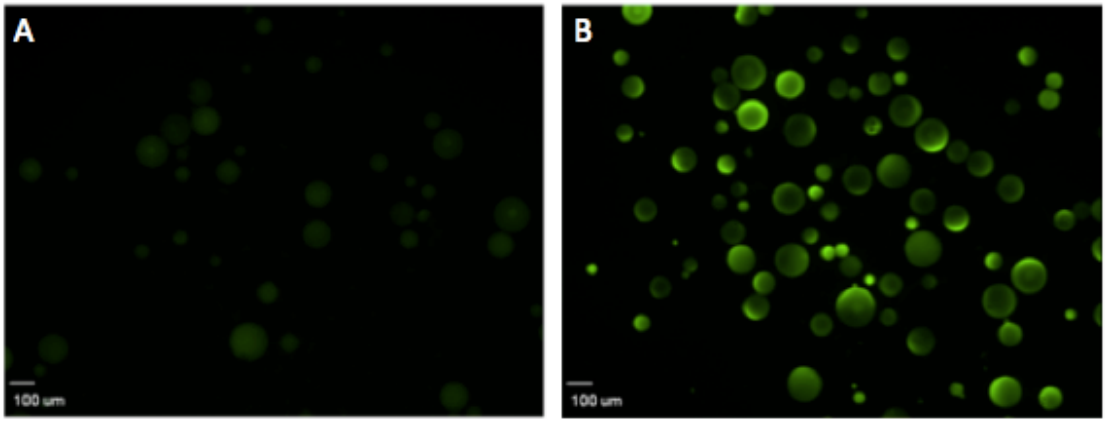


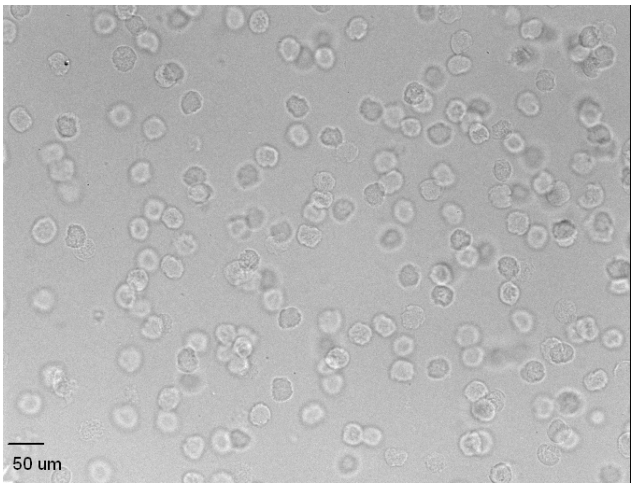
Figure 24: Agarose-NTA-Ni2+ gel beads incubated with GFP-6His in (A) traditional cell-free lysate and (B) cell-free lysate without DTT.

As a further consideration, in literature describing homemade lysate preparation the use of spermidine instead of PEG to increase solution viscosity has been shown to provide an increase in protein yield[167]. Viscosity-inducing reagents are necessary to create a more cell-like solution, aiding in the transcription/translation process[166–168]. Following the literature, we produced reaction mix with spermidine and saw a small increase in expression with circular vector (data not shown). However,

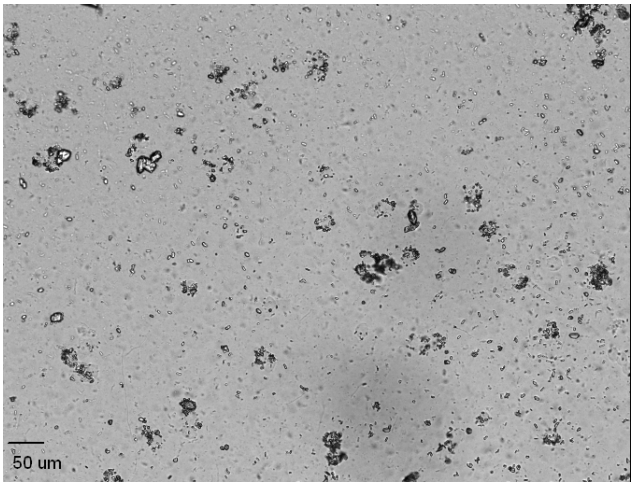
after attempting expression with P-gel microdroplets, gels were not present after a 6-hour expression. After multiple repeats of this setup, we concluded that the spermidine lysate was in fact disrupting the gel morphology. Though we did not perform further work to determine the cause of this result, it was noted that spermidine is a positively charged amine at the neutral pH that cell-free lysate reaction takes place (this reaction mixture becomes slightly acidic as the reaction progresses). Both spermine and spermidine, molecules that have the same +2 charge and similar chemical makeup, have been shown to interact strongly with DNA[169–171]. We concluded that the large amount of spermidine in the solution was disrupting the structure of the P-gel microdroplets.

Based on these results, we again produced a reaction mix without DTT, but replaced spermidine with PEG 8000 as the viscosity-inducing reagent[166]. Figure 25 shows that after incubation and shaking for six hours at 30°C, gel microdroplets are still visible and appear to be maintain their integrity as discrete, individual units with minimal aggregation. This is the solution that was used as the lysate solution for all lysate expression experiments in this thesis work.

(A)
Buffer



(B)
Lysate
(spermidine)



(C)
Lysate
(PEG)

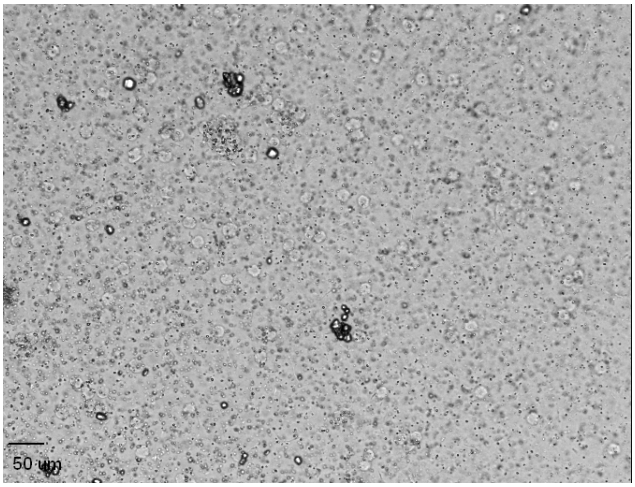


Figure 25: Optical microscopy showing the effects of lysate incubation on P-gel microdroplets compared to their original morphology in storage buffer (10 mM Tris-NaCl).

Section 4.2 - NTA-functionalization of X-DNA

As documented in previous work[49,50], DNA gel does not naturally possess the ability to capture, or aggregate, protein produced by the genes contained within. Protein isolation and purification from solution most commonly uses the specific binding of an N- or C-terminal His-tag to a chelated divalent metal cation such as nickel. This method of protein purification, whether that protein is produced through a cell-based or cell-free protein production method, is normally performed in a column format; columns provide access to immobilized nickel while allowing for constant flow. The nickel in these columns is chelated most often through a nitrilotriacetic acid group (NTA), though some commercial columns use a TALON^{TM} metal affinity resin[172–174]. Chelated nickel, copper, and cobalt have all been shown to specifically bind his-tagged proteins, though with different affinity and specificity[175,176].

In order to introduce a nitrilotriacetic acid group into the DNA, we needed to connect this molecule to a moiety that can interact with any of the chemical groups with which IDT could modify the 5' or 3' end of a DNA strand. Addition of a thiol to the 5' end of an oligo we use to construct our X-DNA monomers allows the DNA to interact with a maleimide group, and a literature search directed me to a chemical moiety containing an NTA group linked by three carbons to a maleimide moiety[177,178-179]. The structure of this molecule, maleimido-C3-NTA (MC3N), is shown in Figure 26.

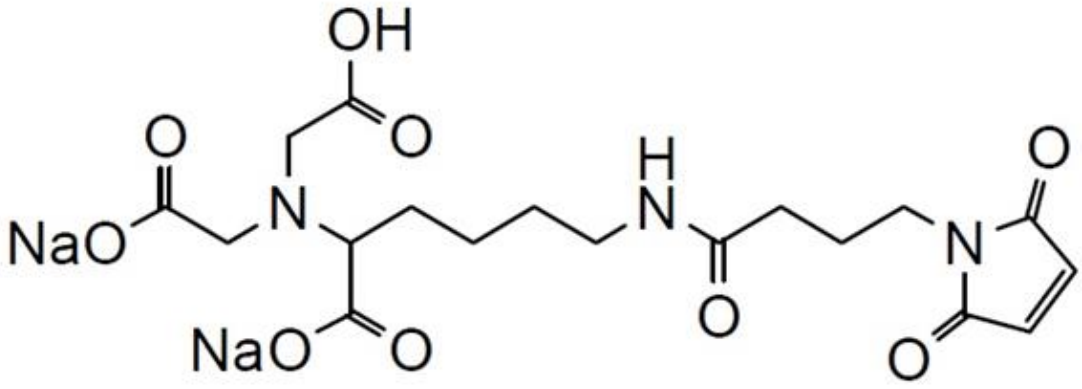


Figure 26: Chemical structure of maleimido-C3-NTA (MC3N). The maleimide group is included on the right side of the structure while the nickel-chelating moiety NTA is on the left.

MC3N functionalization required the deprotection of the C6-thiol modified DNA strand (Figure 47). This deprotection is performed with the reducing agent TCEP (tris(2-carboxyethyl)phosphine), which our lab has previously used in nanoparticle functionalization with DNA[24–28,180]. As opposed to DTT, which has also been commonly used for such applications, TCEP is non-toxic, does not emit a strong or foul odor, and is more stable towards oxidation[181,182]. TCEP also more specifically reduces disulfide bonds than DTT, which acts in a more general manner and is commonly removed before further functionalization steps are taken. The original protocol thus left TCEP in the reaction as we added the MC3N linker. However, upon further review it was noted that TCEP can in fact react with a maleimide group under certain conditions[183] and was potentially inactivating the MC3N linker. In order to ensure the removal of TCEP, we used 10,000 kDa MWCO spin filters from EMD Millipore, diluting TCEP concentration in the sample approximately 10,000 times. In

conjunction with a 25:1 molar excess of MC3N to DNA, TCEP interference through maleimide inactivation was made negligible.

The thiol deprotection process was undertaken in 10 mM Tris-NaCl, pH 8.0 buffer containing no EDTA, as this can interfere with His-tagging later in the experimental work. EDTA acts as a metal chelator for divalent ions such as Ni^{2+}, potentially stripping the nickel from the NTA groups linked to the DNA. The deprotection step is relatively short, with the sample being incubated for one hour at room temperature under 600 rpm shaking. The functionalization step was performed in TBS at pH 7.6 for 12 hours. An overview of this process is shown in Figure 27.

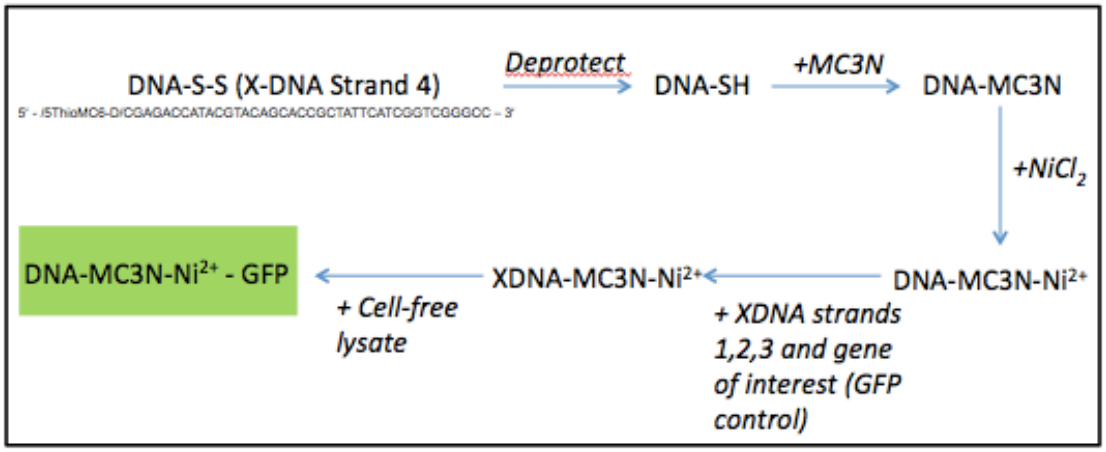


Figure 27: Overview of the MC3N functionalization procedure.

Section 4.3 - Testing of functional group presence and functionality

In order to test whether the MC3N was actually linked to the thiol-modified DNA, we performed both mass spectroscopy and a functional assay. The mass

spectroscopy was performed at the Cornell Core Laboratories Center. The two samples, the unmodified and MC3N-modified strand 4 of X-DNA, are shown in Figure 28. The mass change of approximately 1400 da corresponds to the combined molecular weights of the malemeido-C6-NTA group and the C6-thiol that are added onto the DNA.

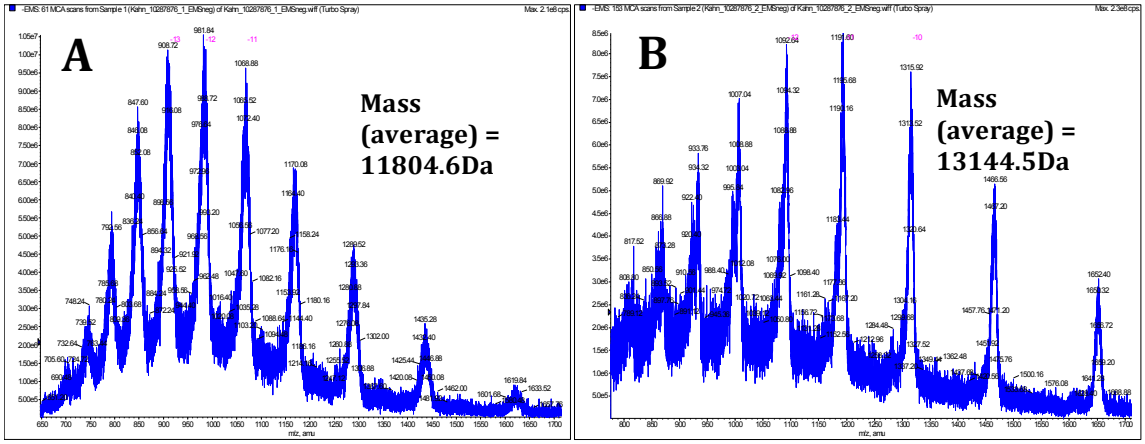


Figure 28: Mass spectroscopy results for A.) unmodified X-DNA strand 4 B.) Maleimido-C3-NTA modified strand 4.

The functional assay used a centrifugal filter with a mass cutoff above that of free GFP (~27 kDa) so that only GFP that was interacting strongly with MC3N modified X-DNA would be retained in the filter; the molecular cutoff used was 50 kDa. The modified X-DNA was synthesized using the same hybridization protocol shown in Figure 22, with the exception that the previously phosphorylated strand four is instead MC3N-functionalized strand four. Figure 29 shows that the MC3N-functionalized X-DNA forms with high efficiency, with a yield comparable to non-functionalized X-DNA. Functionalized and non-functionalized X-DNA were

incubated with commercial GFP in 1 mM and 10 mM $NiCl_2$ solutions at 25°C for one hour, shaking at 600 rpm. Additionally, incubation was performed in lysate supplemented with 10 mM $NiCl_2$.

After incubation, the solutions were washed with Tris-NaCl buffer, centrifuged through the columns, and the solution that was left in the filter was measured for fluorescence under the plate reader, as shown in Figure 30. One can see that although the amount of GFP captured on the gels does increase for both the non-functionalized and functionalized sets as nickel concentration increases and in the presence of lysate, the MC3N-functionalized X-DNA demonstrates a significantly higher retained yield of GFP.

Lane 1	ssDNA - Oligo 4 from X-DNA
Lane 2	X-DNA
Lane 3	MC3N-Functionalized X-DNA

Figure 29: 2.5% agarose gel showing X-DNA and MC3N-functionalized X-DNA. All lanes contain DNAs incubated in Tris-NaCl (10mM).

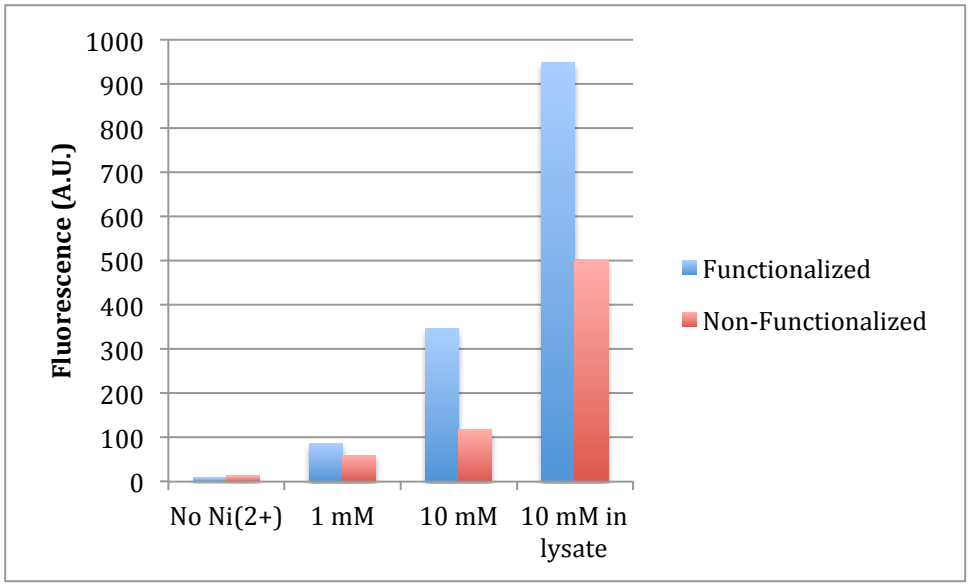


Figure 30: Plate reader values for GFP retained after washing and filtering through 50,000 MWCO columns. Values on the y-axis are arbitrary intensity units. The concentration of nickel refers to the concentration within the solution while incubating with GFP.

Section 4.4 - Incorporation of NTA into DNA gel

The MC3N-X-DNA must be incorporated into P-gel microdroplets without significantly affecting the strength of the resulting gel. Functionalization effectively removes one linking branch of the X-DNA monomer due to its inability to ligate another DNA strand. Thus, not all the X-DNA in the reaction solution can be functionalized, as earlier work performed by our lab demonstrated that Y- and T-DNA demonstrated a significantly lower tendency to form gel when compared to the X-DNA due to the lower crosslinking density inherent in the structure[46].

With this in mind, we aimed to have the greatest number of NTA groups in the gel as possible. Thus, we produced three different DNA gel samples containing 10%, 20%, or 50% functionalized X-DNA. We were seeking the percentage that could be included before the gel network was sufficiently compromised as to induce gel breakup upon not only extraction from the emulsions, but under the shaking and incubation that would be encountered during cell-free protein synthesis. Only the 10% functionalized gels survived both the extraction and shaking conditions. The 20% gels did form in the emulsions but did not survive extraction, while the 50% gels did not form gels within the emulsions to any visible extent.

The ability of the functionalized DNA gels to interact with His-tagged GFP was then tested. Functionalized and non-functionalized DNA gels were synthesized and incubated with 10mM $NiCl_2$ and GFP-His overnight at 30°C under 600 rpm shaking. After incubation, the gels were washed with Tris-NaCl by diluting 25X, shaking for one hour at 25°C, and then spinning down the gels at 0.2x1000 rcf for five min. The supernatant was removed, and the original reaction volume was viewed under the fluorescent microscope to observe if any difference was seen between the samples. Figure 31 shows the strong GFP signal on the modified gels, which is not viewed in the unmodified system.

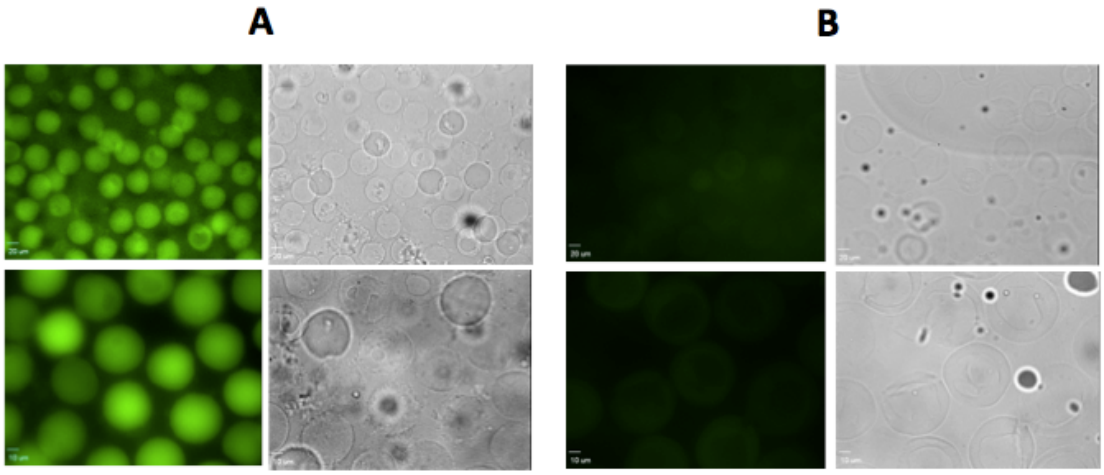


Figure 31: A.) MC3N-functionalized and B.) unmodified P-gel microdroplets incubated with GFP and washed to remove non-bound protein.

Section 4.5 - DNA bird nest functionalization

As mentioned in Chapter 1, the ligation of branched DNA structures is but one method of producing a DNA gel. Another method of gel synthesis performed in our laboratory is based on the novel combination of rolling-circle amplification (RCA) and multi-primed chained amplification (MCA). This process produces a gel that possesses unique metaproperties dependent on whether the gel is in a water or air environment, a phenomenon that our lab has studied and shown to be based on specific gravity of the surrounding environment[184]. I was interested in this method of gelation not based on these bulk material properties, but rather for another unique outcome of creating a gel through this procedure. The bulk gel structure is shown in Figure 32, where high-density bundles of DNA are linked weakly by small numbers of longer DNA strands.

These bundles are referred to as DNA bird nests, and after physical breaking through pipetting, the individual bird nests forming the bulk gel are broken apart into discrete units.

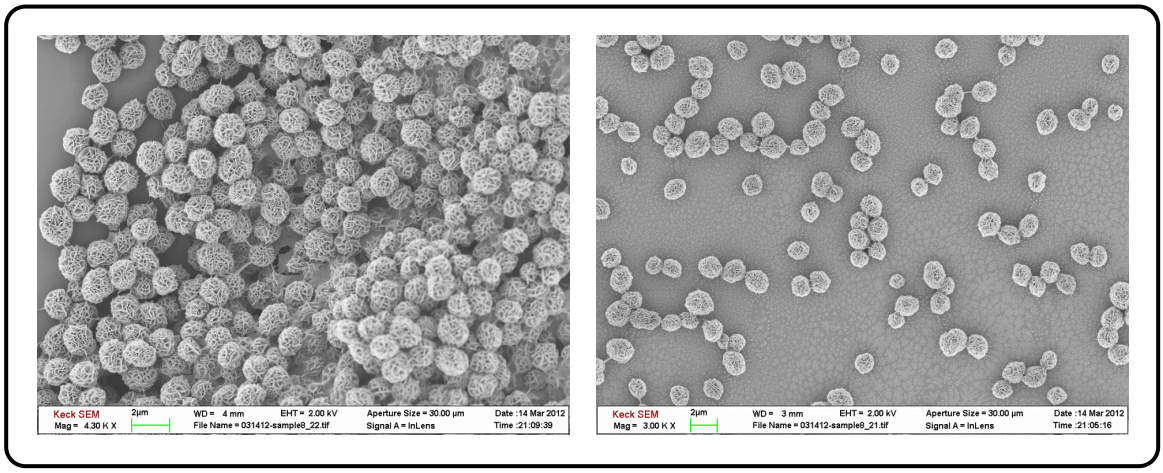


Figure 32: SEM images of Gel1 before (A) and after (B) mechanical breakup of gel linkages to form DNA bird nests.

In considering the mechanism of gel formation, each of these bird nests should be formed from one circular template (Figure 33). The elongation of this template during RCA provides many primer attachment points during the MCA process, thus yielding distinct regions of high DNA density. Each of these discrete bird nests should contain hundreds to potentially thousands of copies of the same circular template from which it was formed. This provides an extremely high copy number of the template covalently linked within a small, discrete gel volume. Such a phenomenon is indeed very difficult to replicate with other methods – the only other method capable of producing multiple, identical copies from a single template is droplet-based PCR amplification[185–188]. This unique aspect of the bird nests has been explored for its

potential in the delivery of large quantities of immunostimulatory CpG sequences to cells as well as the delivery of siRNA[189]. A precise measure of functional copy number is very difficult to ascertain as the gel structures contain a mixture of single- and double-stranded DNA; it is unclear how many of the template repeats are actual functional copies if a double-stranded DNA is required for transcription. Nonetheless, it is possible to determine the 'equivalent copy number' in each bird nest by comparing protein expression yields to solution-phase controls.

This platform holds great potential for directed evolution applications, as this method provides a cell-free platform for creating localized high-copy numbers of a single mutant from a gene library. When typical gene isolation is undertaken in cell-free expression, notably in *in vitro* compartmentalization platforms to ensure a genotype-phenotype connection, protein is normally produced from one molecule of gene template; this naturally leads to very low yields. Gene expression can now be performed off of a DNA gel possessing many copies of the same mutant. Combined with the ability to functionalize the gel with the same MC3N chemical group used in the functionalized P-gel microdroplets to link His-tagged protein, this format offers the possibility to address low expression yields present in current cell-free systems that *in vitro* compartmentalization technology while also providing a platform to link expressed protein.

An outline for the synthesis of bulk DNA gel from the RCA/MCA process is shown in Figure 33. The MC3N functionalization of primer 2 allows this group to be included in the gel as it is being formed. As protein expression and capture is the goal of this work, the single-stranded circular template is produced from pIVEX2.3d-GFP.

The sequences of the primers used for bird nest formation are listed in Table 11. Discrete bird nests are produced from the bulk gel by pipetting the solution, causing mechanical breaking of the weak DNA linkages between denser regions of DNA.

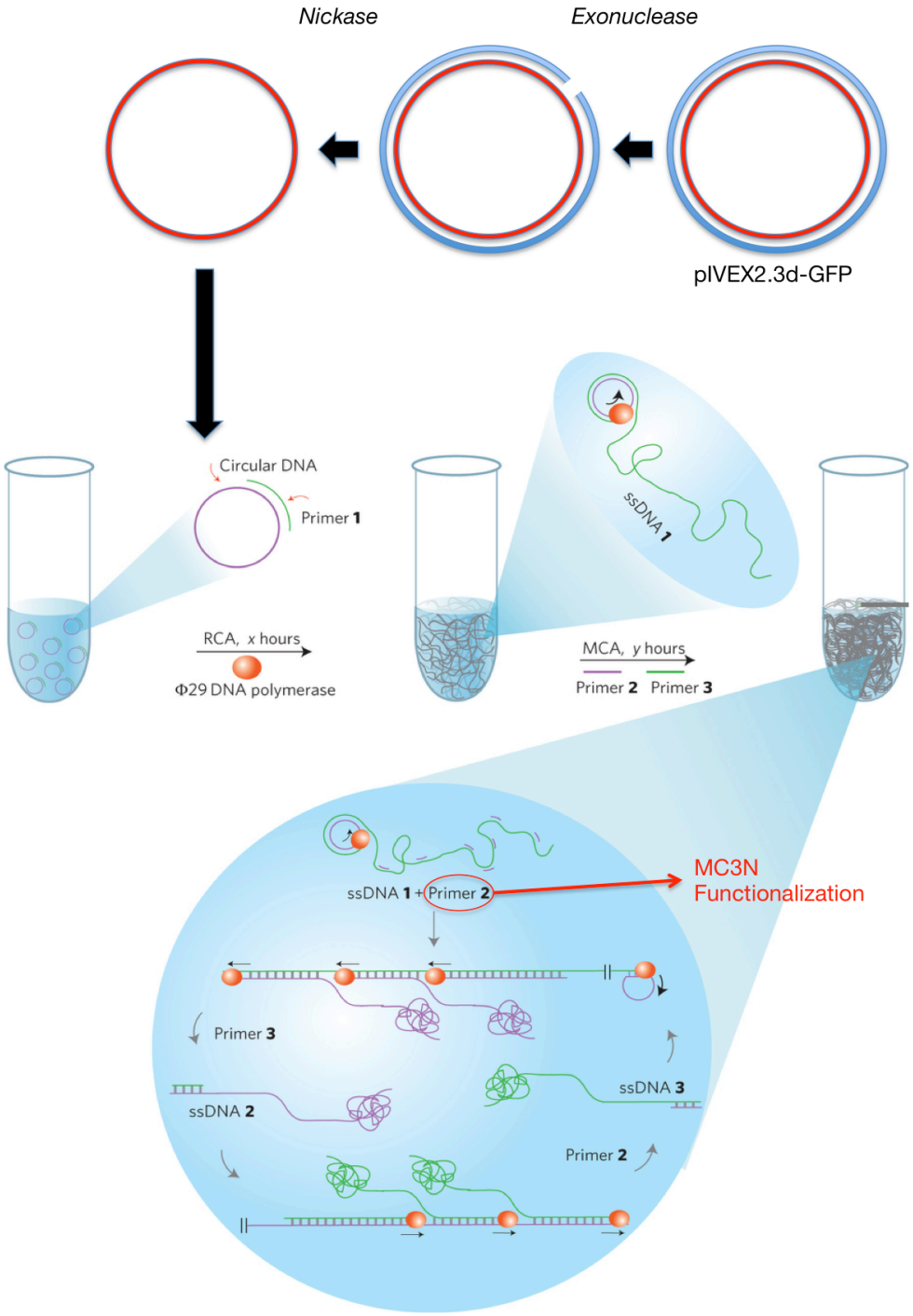


Figure 33: Synthesis of bird nest DNA gel. The MC3N functionalization is performed on Primer 2, and thus is incorporated in the multiprimed chained amplification[52].

The protein binding functionality of the bird nest gels containing the MC3N group was tested using both commercial GFP in buffered solution and in cell-free lysate. In buffered solution, both functionalized and non-functionalized gels were incubated first in 1 mM $NiCl_2$ for one hour of shaking at 600 rpm, and then introduced to commercial GFP. After one hour of incubation at 30°C (the temperature used in cell-free protein expression), the gels were washed with 10 mM Tris-NaCl solution and viewed under the microscope. Figure 34 shows a significant difference between the functionalized and non-functionalized bird nests. Functionalized bird nests possess a greater GFP signal than unfunctionalized bird nests.

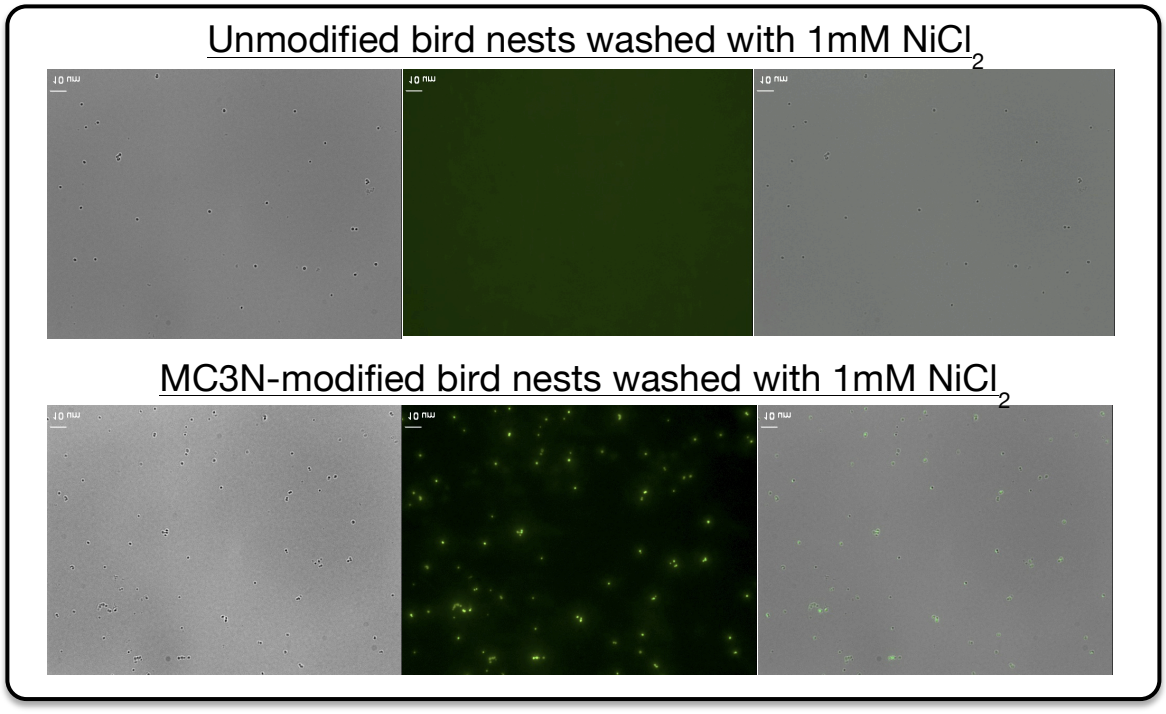


Figure 34: (A) MC3N-functionalized and (B) non-functionalized DNA bird nests incubated with GFP-6His and washed to remove non-bound GFP. The first, second, and third rows are bright field, fluorescent, and an overlay of the two, respectively.

A similar approach was taken to test whether the gels could capture protein in cell-free lysate. Gels were incubated with $NiCl_2$, and then were placed in cell-free lysate and incubated at 30°C for six hours with commercial GFP-His. Figure 35 shows that the modifications made in Section 4.1 to the cell-free lysate does allow for protein capture in the bird nest gels.

In order to test the 'equivalent copy number' of the bird nests, protein expression levels were measured and are shown in Figure 36. Though it may appear that expression levels are negatively affected by use of the bird nests when compared to the linear control, if one takes into account the actual number of bird nests introduced into the lysate then the expression levels prove to be much more impressive. The use of 9,500 bird nests/uL has approximately the same protein yield as 15 ng/uL of MluI linearized plasmid. Thus, each bird nest, in terms of its contribution to GFP output, represents the equivalent of nearly 500,000 functional copies of linear plasmid. In comparison to circular plasmid, this total is closer to 125,000 functional copies.

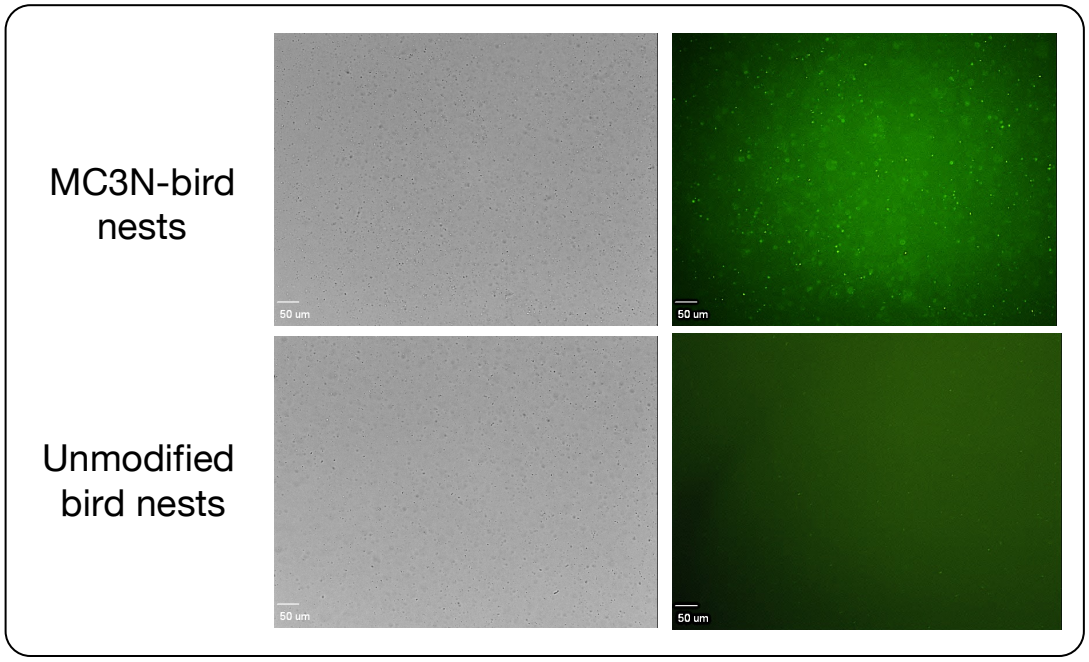


Figure 35: DNA bird nests after incubation in lysate containing commercial GFP-His. MC3N-functionalized bird nests are shown in (A) and unmodified bird nests are shown in (B).

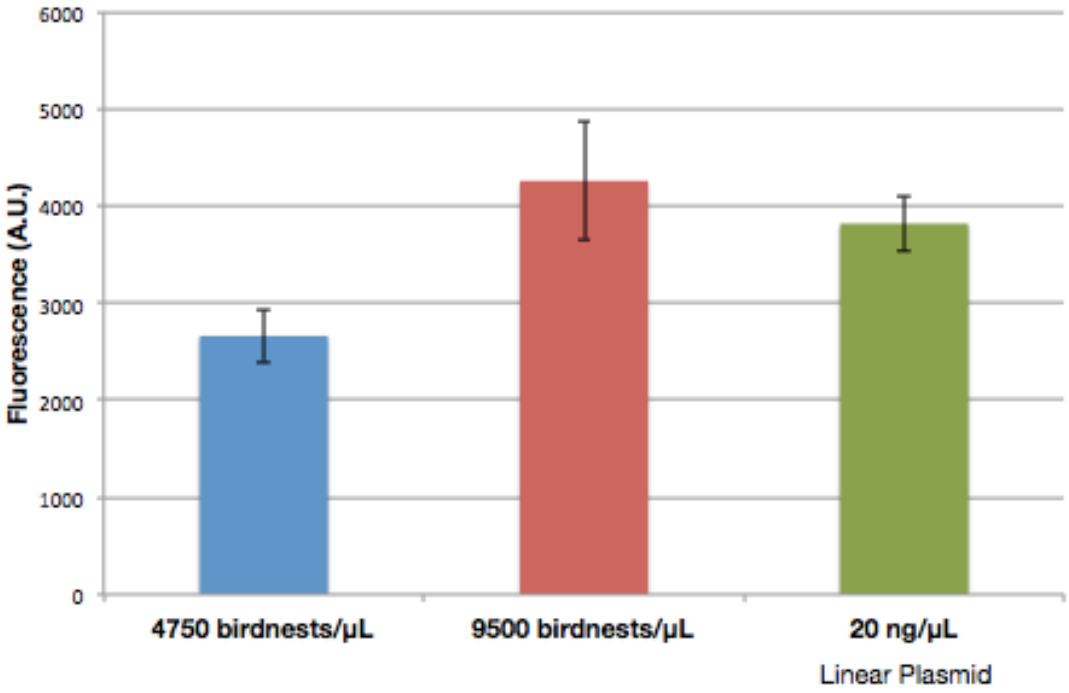


Figure 36: Comparison of GFP expression levels using bird nest DNA gels.

Section 4.6 - Increasing thermostability of DNA-based gels

After ligation into a gel network, complementary strands of the X-DNA network are held together through hybridization, and thus the gel is subject to the same temperature and salt variations as double-stranded DNA. In low salt conditions, the X-DNA can actually dehybridize into the four individual strands, as shown in Figure 43. The ligation of X-DNA into a network gel structure covalently links X-DNA monomers together, but still contains a double-stranded scaffold that is held together through hybridization. Due to the central crossover junction necessary to construct the X-DNA, the longest uninterrupted double-stranded binding created through ligation is 40 bases, regardless of how many X-DNAs are covalently linked. The melting temperature of these strand sequences is approximately 65°C, as seen in Figure 37. At this temperature, 50% of the sequences become dehybridized. As the temperature of the gel environment approaches and exceeds this temperature, the gel begins to fall apart. In order to interface with cell-free processes and proteins under extreme conditions of low salt, high temperature, and extreme pH, it is necessary to chemically stabilize the DNA gel. Interest in studying proteins in extreme environments has been growing as scientists have discovered and begun studying novel classes of organisms termed 'extremophiles' that thrive under conditions not tolerated by the majority of living systems. Such enzymatic systems offer the ability to conduct biological engineering under industrial conditions that would normally be destructive to more commonly encountered proteins[190,191].

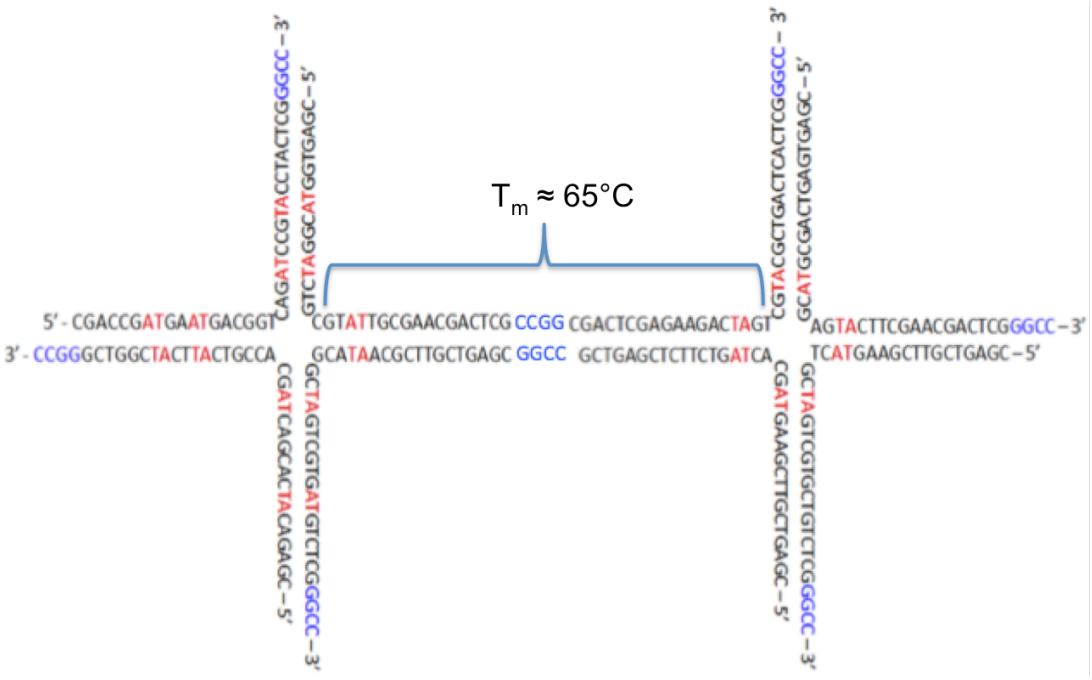


Figure 37: Depiction of longest continuous hybridized strands in ligated X-DNA structures.

One may consider that since X-DNAs are ligated together in the P-gel, the oligos should form continuous, covalently linked strands that span the whole gel; thus, the gel may be expected to expand at higher temperatures but not necessarily fall apart. However, this assumption would be based on 100% ligation efficiency, which the system certainly does not reach. This value has not been intensively studied, but as the ligation is accomplished through an enzymatic reaction the efficiency is assumed to be less than 100%. This efficiency is further compromised as gelation occurs and retards the diffusion of enzymes and molecules through the gel. Though multiple X-DNAs may be ligated and form long, covalently linked strands spanning multiple X-DNA structures, a non-ligated point will eventually be reached within the gel. Above

the melting temperature of the ligated arm-spans, single stranded DNA can detach from the scaffold and, in effect, dissociate and 'leave' the droplet.

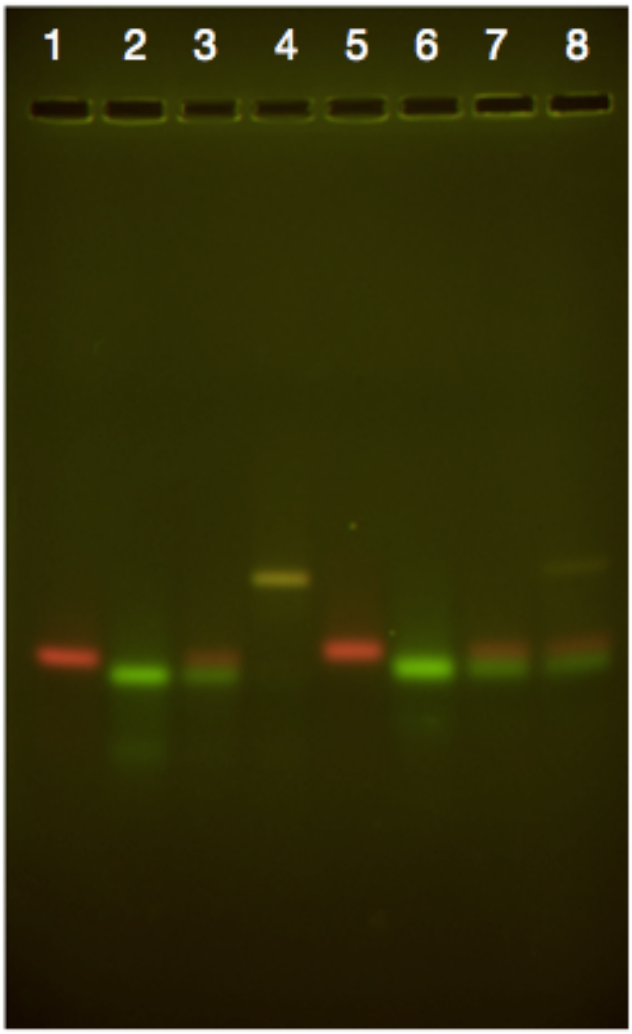
In addition to a lack of mechanical robustness at elevated temperatures, double-stranded DNA possesses a natural tendency to 'breathe', or display local denaturation and reclosing of the double strand[192–194]. This breathing process can potentially reduce the strength of the gel. This breathing tendency is elevated by the use of a crossover junction in the X-DNA structure[195], meaning that each node is potentially vulnerable to stresses on the gel structure. These stresses can come from multiple sources, including changes in temperature, pH, or mechanical stress induced through processes such as droplet purification and shaking during incubation.

In order to strengthen the structure of the gels without fundamentally changing the method of gelation and incorporation of gene, we developed a method to covalently link the double helix between complementary strands making up each X-DNA. Psoralen, a naturally occurring molecule that has been shown to specifically intercalate and crosslink DNA in the presence of ultraviolet light[51,196–198], was introduced to the X-DNA before ligation to produce covalently bonded X-DNAs. The introduction before ligation is necessary due to the end goal of this gel formation – gene expression. If psoralen is introduced after gene ligation into the gel network, the opposing strands of the gene itself will be crosslinked and thus rendered unusable for protein expression applications.

Psoralen is a molecule that displays an affinity to intercalate DNA at specific base pair combinations[199–202]. The 5' – TA – 3' sequence possesses the strongest tendency for psoralen intercalation, followed by 5' – AT – 3'. Other base

combinations show a low tendency for psoralen intercalation. Consideration of these intercalation properties of psoralen was, of course, not within the original design considerations of the X-DNA sequences. Thus, in analyzing the original X-DNA oligo sequences for the presence of 5' –TA – 3' as representative psoralen binding sites, three arms contain two bindings sites and one arm contains one. Furthermore, these binding sites are at different locations on the arms, which can introduce heterogeneity to what should be a symmetrical structure.

In order to gain greater control over psoralen intercalation, we altered the sequence of the X-DNA oligos to meet certain predetermined conditions. These conditions required that the same number of crosslinking sites be present on each arm, in the same location, and not in proximity to the sticky ends. The last condition is due to the fact that psoralen intercalation causes a deformity in the helical structure of the DNA and thus can interfere with ligation[196]. Figure 38 demonstrates that Y-DNA is able to ligate efficiently before psoralen treatment, but this efficiency is significantly reduced after psoralen crosslinking. Considering the original X-DNA sequences contained 5' – TA – 3' sites within five bases of the arms' sticky end overhangs, ligation efficiency would be significantly reduced had this sequence not been modified.



Lane 1	Untreated Y-DNA with red fluorophores
Lane 2	Untreated Y-DNA with green fluorophores
Lane 3	Untreated Y-DNA mixed together without ligase
Lane 4	Untreated Y-DNA mixed together with ligase
Lane 5	Crosslinked Y-DNA with red fluorophores
Lane 6	Crosslinked Y-DNA with green fluorophores
Lane 7	Crosslinked Y-DNA mixed together without ligase
Lane 8	Crosslinked Y-DNA mixed together with ligase

Figure 38: Effect of psoralen crosslinking on ligation efficiency. Note the yellow band in lane 4 is created by the ligation of two Y-DNAs containing either green or red fluorophores. The same amount of psoralen-crosslinked Y-DNA yields a significantly lower intensity yellow band, as seen in lane 8.

The four X-DNA oligo sequences were redesigned to have the same guanine and cytosine makeup as the original oligos as well as similar thermodynamic properties. So while the sequences were modified to incorporate the desired changes and introduce isotropicity to the psoralen crosslinking, the ability to form X-DNA should not be significantly affected. The specifics of this redesign are in Table 7.

Table 7: Considerations for sequence redesign of X-DNA to enable symmetrical psoralen intercalation and higher ligation efficiency after psoralen crosslinking[39].

Redesign of Thermostable X-DNA
The inclusion of one 5' – TA – 3' site at a distance of 2 bp away from the junction
The removal of all other 5' – TA – 3' as well as 5' – AT – 3' pairings
Sequence swapping to retain the same GC percentage in each of the arms as well as similar thermodynamic characteristics

The redesigned sequences are shown in Figure 39 and Table 12. One can see in Table 8 that this redesign was achieved while maintaining the thermodynamic characteristics of the original X-DNA sequences, which were selected based on design considerations discussed in our lab's previous dendrimer-DNA work[39]. Comparison of the X-DNA annealing yields are shown in Figure 40, and it is seen that they anneal in the same yields as the original X-DNA sequences.

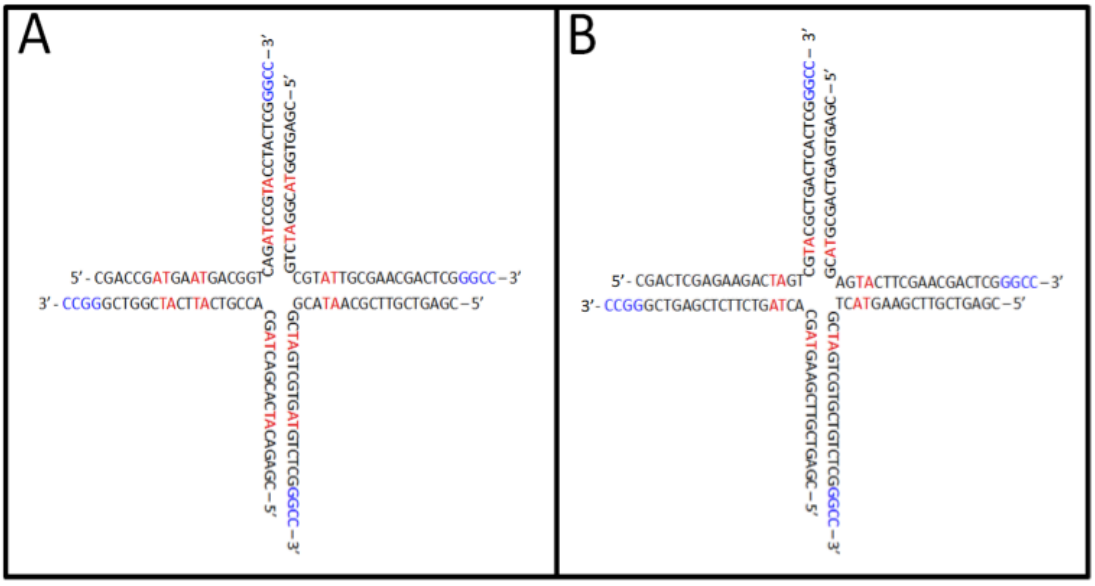
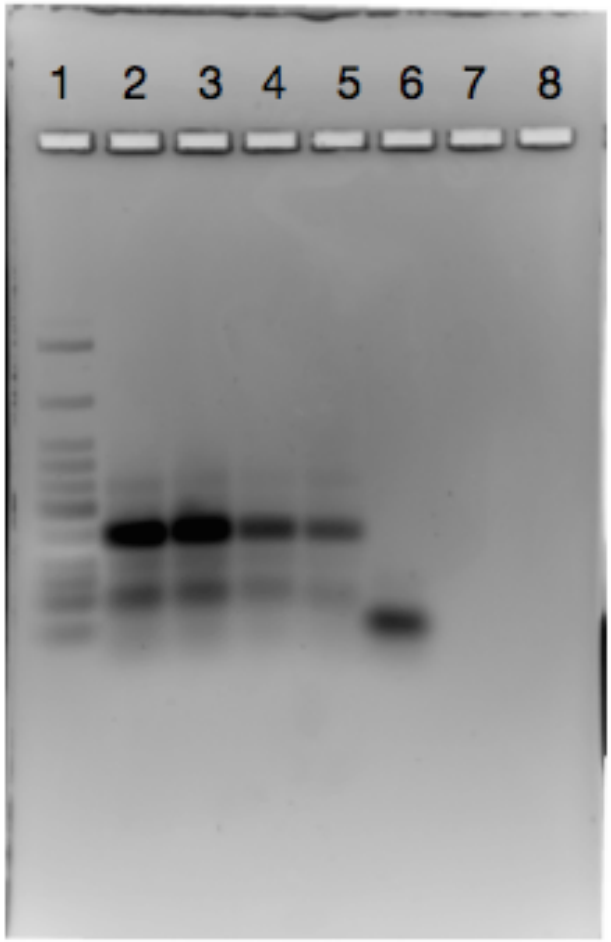


Figure 39: Redesigned X-DNA for greater control over psoralen intercalation. Bases highlighted in red depict a directed crosslinking site, while sequences in blue represent the ApaI restriction site overhang.

Table 8: Comparison of free energy of hybridization between sequences in the original X-DNA strands and redesigned X-DNA strands.

ΔG Values (kcal/mole)						
Oligo Strands	Self-Dimer	Hairpin	Arm Match		Wrong Arm Match	
XI_APAI	-9.6	-6.32	1-2	-36.93	1-3	-9.28
XII_APAI	-9.28	-3.54	2-3	-36.21	2-4	-9.28
XIII_APAI	-9.28	-3.58	3-4	-32.45		
XIV_APAI	-9.28	-2.05	1-4	-36.93		

ΔG Values (kcal/mole)						
Oligo Strands	Self-Dimer	Hairpin	Arm Match		Wrong Arm Match	
XI_APAI_TS	-10.87	-3.03	1-2	-35.31	1-3	-13.31
XII_APAI_TS	-10.87	-4.35	2-3	-30.02	2-4	-13.31
XIII_APAI_TS	-10.65	-2.76	3-4	-32.15		
XIV_APAI_TS	-9.96	-4.75	1-4	-35.47		



Lane 1	NEB LMW Ladder
Lane 2	Original X-DNA
Lane 3	Original X-DNA
Lane 4	Redesigned X-DNA
Lane 5	Redesigned X-DNA
Lane 6	Oligo 4 from X-DNA

Figure 40: Testing the ability of redesigned X-oligos to form branched X-DNA.

The formation of psoralen diadducts that covalently link complementary DNA strand requires UV exposure, but the process of UV-crosslinking hybridized X-DNA can be damaging to the DNA[203]. This DNA damage, in turn, can negatively affect ligation efficiency, thus negating any mechanical benefits of the psoralen crosslinking. This means that it is not only necessary to change the number and location of the directed crosslinking sites, but to optimize the reaction conditions to induce the least possible damage to the DNA.

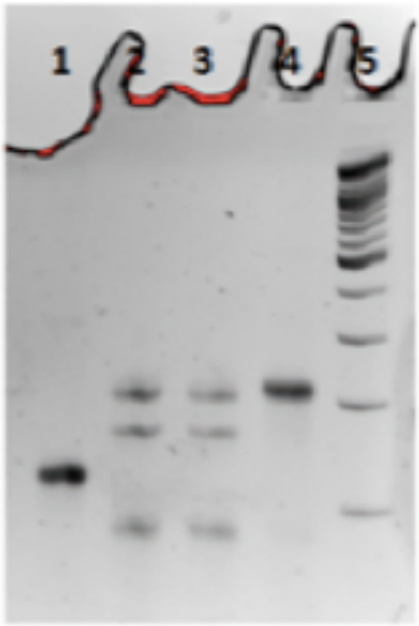
Whereas our lab's earlier work utilized a significant excess of psoralen (1000:1 psoralen:X-DNA excess) and a long UV exposure time (15 minutes), the DNA structures were acting as primers in a PCR reaction rather than branched units in a ligated network. Nucleotide damage to the ends of the DNA serving as a PCR primer may not make a significant difference in PCR efficiency, as the primer will still hybridize using the majority of the bases (15-20) even if multiple bases have been damaged. In using a double-stranded DNA-strand for ligation, the damaging of the terminal base yields an overhang that cannot be ligated. Thus, care must be taken to produce psoralen-crosslinked X-DNA that has not been compromised as to its ligation efficiency. To this end, it is necessary to minimize both the concentration of psoralen and UV exposure time. Psoralen was dissolved in DMF, an organic solvent which can be damaging to DNA if left in the solution, while UV exposure can lead to direct damage of the DNA through breaking of bonds as well as the production of free radicals which in turn can lead to DNA instability[203].

Initial experiments to test lower psoralen concentrations were performed at either 800:1 or 80:1 molar ratios of psoralen: X-DNA. Figure 41 shows that the lower

concentration actually does result in less efficient crosslinking. However, the need for every potential intercalation site to be crosslinked is not necessary. In reality, very few sites would need to be crosslinked in order to have a potentially significant effect on thermostability in a ligated X-DNA system. As stated previously, the ligated oligos form long single strands, and one psoralen-crosslinked site on that strand should be sufficient to keep the strand from escaping the gel under high temperature conditions above the melting temperature of ≈65°C. Experiments were thus performed with 5:1, 50:1, and 100:1 molar excess of psoralen:X-DNA. This yields a 1.25:1, 12.5, and 25 molar excess of psoralen to directed crosslinking sites, as each X-DNA contains four 5' – TA – 3' sequences. These particular excesses were selected as they represent 1X, 10X, and 20X molar ratios per intercalation site with an additional proportional excess to account for off-target crosslinking and molecular degradation.

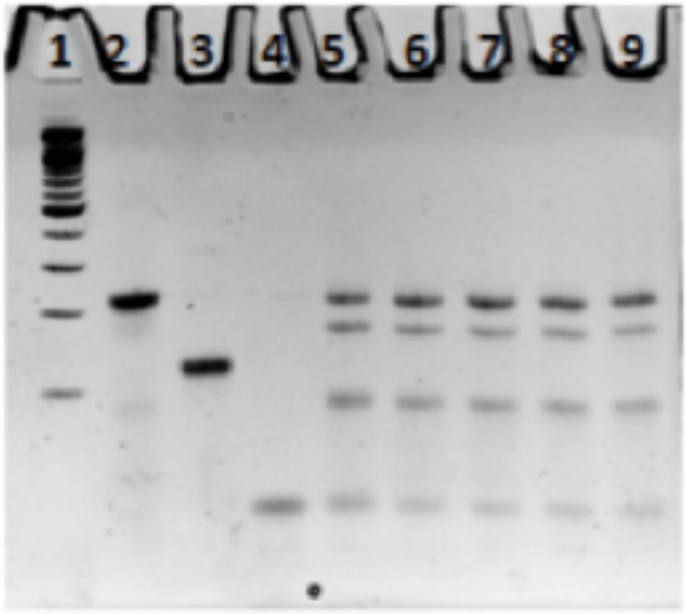
In addition to changing the reaction stoichiometry, we also modified the reactions times to determine whether the benefits of UV exposure (potentially higher yield of psoralen-crosslinked DNA) was outweighed by damage to the DNA. Figure 42 shows that reactions times 15 minutes and greater yield no difference in the amount of crosslinking occurring, and thus longer exposure times only serve to further damage the DNA. Exposure times were shortened based on this result to one minute and five minutes. Bulk DNA formation was used to provide an initial idea of whether the reactions were causing any physical change in the gel resulting from ligation of psoralen-crosslinked X-DNA. Based on Figure 17, which shows the results of gel using a one minute exposure time and a 5:1 psoralen:X-DNA ratio, we selected this condition; the resulting gel was more distinct and robust than other non psoralen-

crosslinked gels. In addition, when incubated under pure aqueous conditions (no salt) that would normally cause X-DNA to dehybridize into its constitutive four strands, the 5:1 psoralen-crosslinked X-DNA showed significantly less dehybridization (Figure 43). Pure water causes the dissociation of X-DNA due to a lack of charge screening, as DNA is negatively charged and requires salts to stabilize the helix and allow for hydrogen bonding between bases.



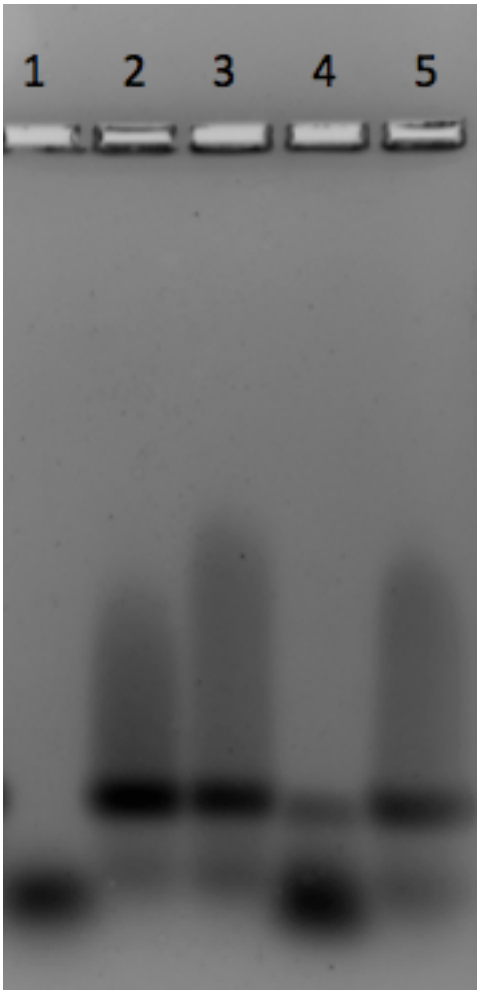
Lane 1	YTS (1000:1 psoralen:X-DNA)
Lane 2	XTS (20:1 psoralen:X-DNA)
Lane 3	XTS (20:1 psoralen:X-DNA)
Lane 4	XTS (1000:1 psoralen:X-DNA)
Lane 5	LMW-Plus Ladder (NEB)

Figure 41: X-DNA and psoralen-crosslinked X-DNA (XTS) stability in SDS-PAGE gel based on high and low ratios of psoralen: X-DNA in the crosslinking solution.



Lane 1	LMW-Plus Ladder (NEB)
Lane 2	XTS
Lane 3	YTS
Lane 4	X
Lane 5	XTS (15 min)
Lane 6	XTS (30 min)
Lane 7	XTS (45 min)
Lane 8	XTS (60min)
Lane 9	XTS (90 min)

Figure 42: SDS-PAGE gel comparing different UV exposure times for psoralen crosslinking (15-90 min exposures of 2.5 mWcm^{-2} UV light at 365 nm). There is no significant difference between crosslinked X-DNA (XTS) at the exposure times used.



Lane 1	ssDNA - Oligo 4 from X-DNA
Lane 2	X-DNA (Tris-NaCl incubation)
Lane 3	5:1 psoralen crosslinked X-DNA (Tris-NaCl incubation)
Lane 4	X-DNA (H_2O incubation)
Lane 5	5:1 psoralen crosslinked X-DNA (H_2O incubation)

Figure 43: 2.5% agarose gel comparing the dehybridization of X-DNA and psoralen-crosslinked X-DNA in nuclease-free water.

In order to test the whether the psoralen crosslinking procedure instituted above actually made a difference in the thermal stability properties of microgels, we produced a set of gels under the same ligation conditions as a non psoralen-crosslinked system. Briefly, we used a final concentration of 125 μm psoralen-crosslinked DNA and ligated at 16°C for 24 hours before collecting the microgels from emulsions. Both microgel sets were placed in the spectrophotometer, and the gels were allowed to settle at the bottom of the cuvettes before a temperature ramp to 95°C at 0.1°C/min. As seen in Figure 44, the different gels possess very different melting curve profiles. Absorbance levels in both solutions increase from approximately 25°C to 50°C, which we attribute to the release of non-ligated X-DNA from the gels (these may still be held in the gels through hybridization, which is disrupted even at these lower temperatures). At approximately 65°C, which is the melting temperature of the longest continuous stretch of DNA in the gel network, the non-crosslinked gels appear to begin rapidly falling apart, whereas the crosslinked gels appear to maintain the X-DNA in a gel format.

After the temperature of the cuvettes returned to room temperature, we looked at the samples under the microscope to view whether there were in fact any microgel structures left in either sample. Figure 45 reveals that the crosslinked microgels gels did, in fact, remain in the sample while the control gels were not present. Thus, psoralen crosslinking did appear to significantly improve the thermostability of the microgels.

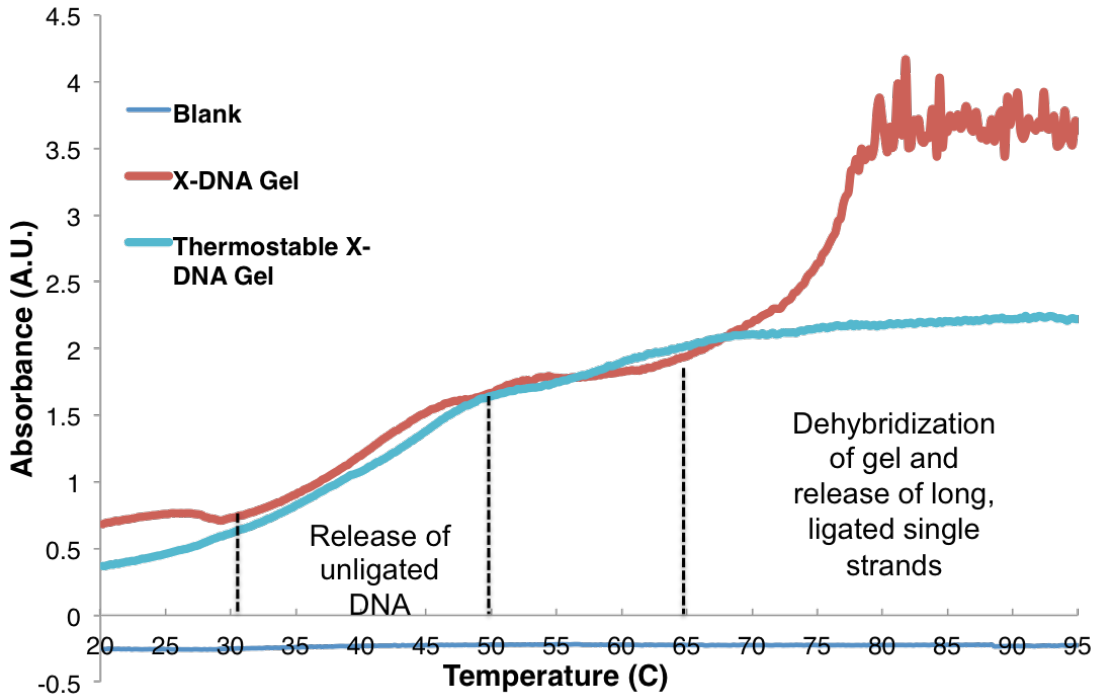


Figure 44: Melting curve for P-gel and thermostable P-gel microdroplets. Droplets were allowed to settle at bottom of cuvette before s temperature ramp at 0.1°C/min.

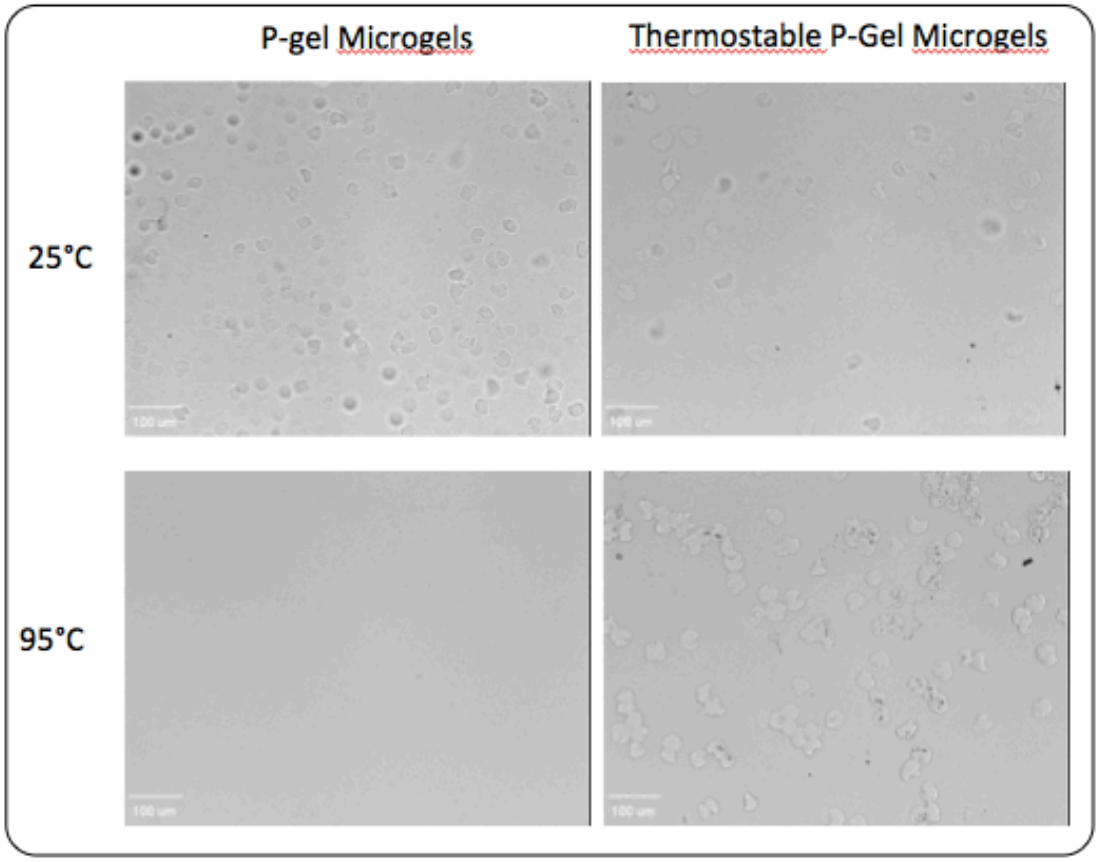


Figure 45: Demonstration of thermostability in psoralen-crosslinked microgels. The images labeled 95°C degree were taken after the solutions had cooled back down to 25°C.

Lastly, it was tested whether psoralen crosslinking influenced the ability of the P-gel microgels to engage in cell-free protein synthesis. The crosslinking does not affect the gene itself, as the gene is introduced after psoralen crosslinking, but may affect enzymes that recognize and bind DNA. Enzymes, such as polymerase, may use the DNA network as a scaffold, and thus interference with the natural DNA form may impede this process. Figure 46

demonstrates that while expression is significantly affected, the psoralen-crosslinked P-gel microdroplets still serve as a functional protein-producing platform.

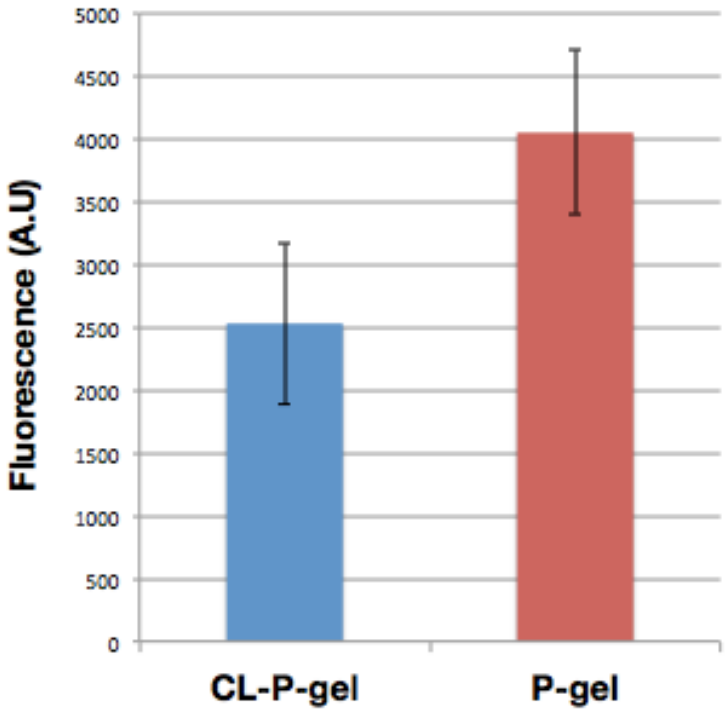


Figure 46: Cell-free protein expression of pIVEX2.3d-GFP using psoralen-crosslinked (CL-P-gel) and non-crosslinked P-gel microdroplets.

Section 4.7 – Summary

The use of DNA as the platform for hydrogel formation opens up different pathways of functionalization to expand the use and application of DNA microgels in cell-free protein synthesis. Here, we present methods to capture proteins on microgels, increase gene copy number in discrete gels, and enable gel

use under extreme conditions such as high pH and high temperature. Incorporation of the chemical group maleimido-C3-NTA allowed for effective binding of His-tagged proteins, which can be added to the gene sequence incorporated into the gel matrix. This same functional group was also incorporated in DNA bird nests, which consist of a different DNA microgel structure than the P-gel discussed in previous sections. These gels are significantly smaller and naturally are formed from one gene template, producing potentially thousands of copies of each gene within a discrete gel and expressing protein in levels equivalent to approximately 500,000 linear plasmid copies per bird nest. Lastly, the P-gel format is amenable to psoralen crosslinking of its double-stranded structure. This process enables gel survival at high temperatures while still allowing for gene expression.

Although the feasibility of these methods has been demonstrated here, future developments will have to focus on optimizing protein capture of the proteins produced from the gene sets contained within P-gels and bird nests and testing the validity of the genotype-phenotype connection. In order to tackle this task, a different red fluorescent reporter must be selected or optimization of gene concentration within the P-gel must be undertaken to ensure this protein is produced in assayable concentrations. This is further evidence that different genes can have varied responses to cell-free reaction conditions. The benefit of DNA-based hydrogels as the material interface means that conditions can be tuned to optimize expression of the gene-of-interest.

Section 4.8 - Materials and methods

Cell Lysate

The production of cell-free lysate was based on the production of an S12 lysate, which is faster and cheaper to produce while producing similar yields as the commonly used S30 lysate[166–168]. Briefly, BL21 (DE3) *E.coli* is grown in enriched media, and synthesis of T7 polymerase under the *lacU5* promoter is induced by introduction of IPTG. The cells are centrifuged, pelleted, and resuspended in buffer. The internal cell components are extracted by use of a French press and further centrifugation to remove cell debris and the large amount of lipid from the membranes.

This BL21 lysate is supplemented with further components necessary for energy production, specifically creatine phosphate, as well as additional monomers for RNA and protein production. The complete reaction mixture is shown in Table 9, with PEG and spermidine recipes included. DTT concentration was also modified as discussed in Section 4.2.

Table 9: Components included in the cell-free protein expression reaction mixture.

Component	Standard Reaction	Spermidine Reaction	Final Concentration
1 M Hepes-KOH Buffer (pH 8.2)	5.7 µl	5.7 µl	57 mM
50 mM ATP	2.4 µl	2.4 µl	1.2 mM
50 mM CTP	1.7 µl	1.7 µl	0.85 mM
50 mM GTP	1.7 µl	1.7 µl	0.85 mM
50 mM UTP	1.7 µl	1.7 µl	0.85 mM
500 mM DTT	0.4 µl	0.4 µl	2 mM
10 mg/ml *E. coli* total tRNA mixture (from strain MRE600)	1.7 µl	1.7 µl	0.17 mg/ml
15 mM cAMP	4.27 µl	4.27 µl	0.64 mM
2M Potassium Glutamate	4.5 µl	4.5 µl	90 mM
2 M Ammonium Acetate	4 µl	4 µl	80 mM
1 M Magnesium Acetate	1.2 µl	1.2 µl	12 mM
1 mg/ml Folinic Acid ($_L$-5-formyl-5,6,7,8-tetrahydrofolic acid)	3.4 µl	3.4 µl	34 µg/ml
Each of 20 Amino Acids	17.98 µl	17.98 µl	1.5 mM/AA
20% PEG (8000)	10 µl	N/A	2%
1 M Creatine Phosphate (CP)	6.7 µl	6.7 µl	67 mM
500 µg/ml Creatine Kinase (CK)	0.64 µl	0.64 µl	3.2 µg/ml
pIVEX 2.3d-Gene of Interest (500 µg/ml)	1.34 µl	1.34 µl	6.7 µg/ml
S12 *E. coli* Lysate – 2x YT	27 µl	27 µl	27% (v/v)
100 mM M Spermidine	N/A	1.5 µl	1.5 mM
Nuclease-Free Water	3.67 µl	12.17 µl	To 100 µl

Maleimido-C3-NTA Functionalization of DNA

Maleimido-C3-NTA was ordered from Dojindo Industries (Japan). We brought the stock to 20 mM in TBS solution and left in 4°C in dark conditions for long-term storage, as MC3N is light sensitive.

Strand four of the X-DNA sequence was ordered with 5' Thiol Modifier C6 S-S from Integrated DNA Technologies (IDT®). The DNA was deprotected by addition of 100:1 molar TCEP excess in a 100 μM solution of DNA, and incubated and shaken for one hour at 25°C and 600 rpm. The TCEP was removed using a 10 kDa MWCO filter column from EMD Millipore, and the DNA was eluted in TBS buffer. MC3N was immediately added in 25:1 molar excess, and incubated and shaken for 12 hours at 16°C and 600 rpm under dark conditions. MC3N was removed using a 10 kDa MWCO filter column and incorporated into X-DNA through the hybridization protocol discussed in Section 3.5.

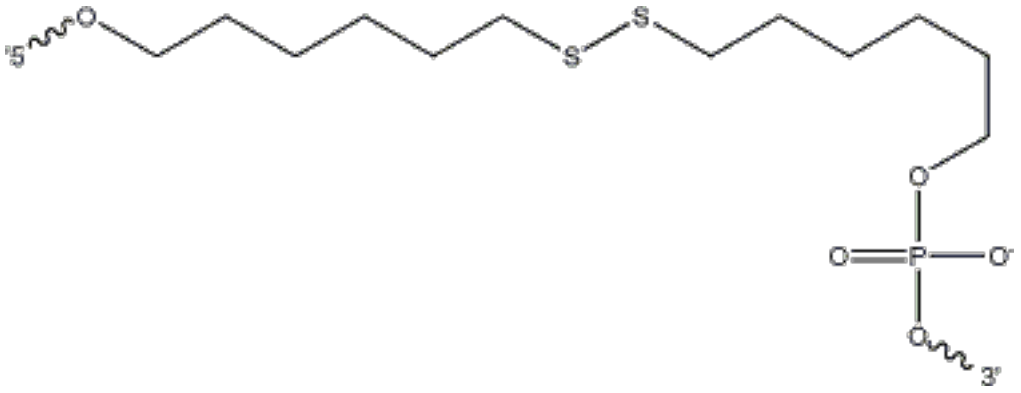


Figure 47: The functional group 5' Thiol Modifier C6 S-S as provided by IDT on the ordered oligo.

Mass Spectrometry

Data was acquired by an AB/Sciex (Foster City, CA USA) 4000 Q Trap outfitted with the Turbo Ion Spray source. The instrument was operated in negative ion, enhanced (EMS) mode scanning at 4000 amu/sec from m/z 400-1800Da. The samples were diluted to 250ul in 50% acetonitrile/water and infused directly into the ion source at 5 µl/minute. The Total Ion Chromatogram (TIC) was acquired using MCA (Multiple Chromatogram Addition) until the base peak signal reached at least 8.0e6 counts/sec. Further parameters for operation are shown in Table 10. Three adjacent, highest abundance ions in the charge envelope were selected and deconvoluted using Analyst 1.4.2 software (AB/Sciex) to obtain the molecular weight of the dominant species.

Table 10: Operating conditions for mass spectrometer.

Curtain Gas (CUR)	30.0 (arbitrary units)
Ion Spray Voltage (IS)	-4200V
Collision Gas (CAD)	Low
Interface Temperature (TEM)	200°C
Gas 1 (GS1)	15.0 (arbitrary units)
Gas 2 (GS2)	20.0 (arbitrary units)
Declustering Potential (DP)	-150V
Collision Energy (CE)	-5.0V

SEM Imagery

SEM images were obtained using the LEO 1550 FESEM at a 2.0 kV setting. Samples were freeze-dried on silicon wafer and held under vacuum until imaging.

Table 11: Primers used for DNA bird nest synthesis with pIVEX2.3d-GFP as the template. Primer 1 is used during RCA, and primers 2 and 3 are used during MCA. Primer 2 is thiolated to allow for further functionalization with MC3N.

Strand	Modification	Sequence (5' – 3')
Primer 1	None	CCAGCGTTTCTGGGTGAGCAAAAACAGGAA
Primer 2	5'-Thiol	TATTACCGCCTTTGAGTGAGCTGATACCGC
Primer 3	None	GCGGTGTGAAATACCGCACAGATGCGTAAG

Table 12: Sequences for the original X-DNA and redesigned X-DNA for psoralen crosslinking. All sequences listed here contain a 3' ApaI overhang sequence.

Strand	Original X sequence (5' to 3'), 5'-phoshorylated
1	CGACCGATGAATGACGGTCAGATCCGTACCTACTCGGGC
2	CGAGTGGTACGGATCTGCCGTATTGCGAACGACTCGGGCC
3	CGAGTCGTTCGCAATACGGCTAGTCGTGATGTCTCGGGCC
4	CGAGACATCACGACTAGCACCGTCATTATCGGTCGGGCC
Strand	**Redesigned X sequence (5' to 3'), 5'-phoshorylated**
1	CGACTCGAGAAGACTAGTCGTACGCTGACTCACTCGGGCC
2	CGAGTGAGTCAGCGTACGAGTACTTCGAACGACTCGGGCC
3	CGAGTCGTTCGAAGTACTGCTAGTCGTGCTGTCTCGGGCC
4	CGAGTCGTTCGAAGTAGCACTAGTCTTCTCGAGTCGGGCC

CHAPTER 5: CONCLUSION AND FUTURE PERSPECTIVE

One of the fundamental challenges in bioengineering is the consideration of how materials interface with biological systems. When the materials themselves are constructed out of biological molecules, the ability to control the type and degree of interaction is greatly enhanced. DNA materials engineering provides a path towards designing and synthesizing materials that not only specifically interact with other DNA and RNA, but with an enormous range of enzymes and cell processes. Though the 'generic' properties of DNA are touted – such as the precise control over length, monodispersity – the real benefits of DNA as a material emerge precisely because of the genetic nature of the polymer. The ability to precisely control the interactions between single strands of DNA and RNA allows an unmatched ability to form different structures at the nanoscale beyond the common double-stranded format. These more complex structures can be engineered to form larger, networked materials at the microscale, providing opportunities for incorporation of different DNA and RNA structures or targets as well as further functionalization.

The synthesis of X-DNA as a crosslinkable monomer for gel synthesis, and the process as described in this work to optimize and modify the system, demonstrates the potential and versatility of DNA nanotechnology. Sticky end sequences were changed not only to accommodate plasmids cut with different restriction enzymes, but also to improve the robustness of the microgels. The flexibility to make targeted changes in the sequence of the plasmid itself allowed

us to optimize expression of linear plasmids and still interface these plasmid with the X-DNA showing the best gel characteristics. The benefits to the optimization of linear gene expression are important for cell-free expression, but this issue is also alleviated when the gene itself is ligated into the DNA gel network. The significant decrease in protein yield from linearized plasmids as opposed to circular plasmids (approximately a 5x decrease in yield) most likely stems from elevated nuclease degradation of the linear template. The results showed that the further the cutting site was from the gene of interest, the higher the protein yield. The P-gel has been shown to protect DNA from nuclease degradation, and thus would aid in this regard. In cases where either a PCR gene construct is used, or only small DNA sequences flank the ends of a gene within a linearized plasmid, gene incorporation into P-gel can provide a significant increase in protein yield.

Though protein yields were not greatly increased as in our original work using P-gel, one must keep in mind the important difference in the lysate systems used. Whereas that work used wheat germ lysate, a eukaryotic lysate, the use of *E. coli* lysate was more apt in this case. The large number of reactions needed and the changing of the lysate itself to accommodate the functionalization required for protein capture made *E. coli* lysate a better choice for testing. Eukaryotic and prokaryotic systems differ in enzyme makeup and functionality, and thus where the DNA gel may have provided an ideal scaffold for eukaryotic transcription and translation, it appears to have less of an effect in the prokaryotic system. Regardless of the specific increase in protein yields, the initial bulk gel format has

been modified into a controllable and scalable microgel platform that successfully produces protein from incorporated gene sets.

The functionalization of DNA microgels in this thesis offers a proof-of-concept platform for expanding the use of DNA materials in cell-free protein expression applications. A critical issue surrounding cell-free directed evolution systems is the isolation of gene sets and the maintenance of a genotype-phenotype connection. The microfluidic synthesis of microgels provides a platform for the isolation of a mutant gene library, with genes being incorporated into ligated DNA network surrounded by a large, shared solution volume without utilizing the confines of an emulsion.

Protein capture to maintain a representative genotype-phenotype connection of a microgel's gene and associated protein set is a complex task. Although further work is needed to optimize protein capture functionality of the gels and to show that captured proteins are indeed produced from the associated gene set, the work presented here represents a first step in marrying the concepts of protein expression and capture with DNA materials. Protein capture onto DNA scaffolds has been achieved previously, but most often on origami platforms and not in a high-throughput material system.

Though I did not discuss in detail the methods behind DNA bird nest synthesis as this was not the focus of my work, the refocusing of this process using plasmid as the material template offers great potential for directed evolution applications. In regard to gene isolation, each bird nest is formed from a single gene template and thus the mechanism of formation itself ensures library

separation. The gelation process provides the only method, outside emulsion PCR, that will amplify a gene and keep the copies of that gene within the same physical volume. Furthermore, the physical linking of the gene copies in bird nests makes this material more convenient to interface with other processes when compared to gene copies in emulsions. As shown in this thesis work, cell-free protein expression per bird nest is extremely high. Protein capture did function on this material, but again, further work is needed to determine whether each gel is capturing mainly proteins from its own gene makeup or simply a sampling of the larger solution volume. Bird nests using other fluorescent reporter proteins should be synthesized and tested to confirm this genotype-phenotype connection. However, even should this not be case, the DNA bird nest format can interface with existing emulsion formats to provide an unmatched method of increasing the copy number of gene mutants during protein expression. Current gene library isolation and emulsion expression processes use infinite dilutions to express protein from one gene molecule; the method presented here provides a straightforward path to increasing protein yield in an emulsion-based system.

Lastly, I explored the concept of thermostable gels within the context of cell-free protein expression and directed evolution. Interest in proteins that act in a variety of different environments, under potentially extreme temperature and/or pH conditions, is growing as extremophiles have been shown to not only survive, but thrive, in such conditions. In areas such as biofuels, proteins that maintain enzymatic activity under high temperatures can more efficiently act

under conditions that aid in cellulose degradation. Tailoring proteins to better suit industrial processes can be performed in cell-free systems, but materials used in these systems also need to be able to withstand the selection pressures. By utilizing the great sequence flexibility of DNA, I was able to redesign the X-DNA structures previously used in the lab to accommodate a spatially controlled covalent crosslinking of the double helix. It is interesting that although effectively the same monomer is being used for creating the gel network, the covalent linking near the X-DNA junction significantly increases the thermostability of the resulting gels.

The achievements presented here open up avenues for further work in the research space of protein engineering. The *in vitro* DNA microgel platform can be applied to the production of any gene that does not require significant post-transcriptional modifications, yet each gene may require different gel conditions for optimal expression. The original P-gel work demonstrated the production of various functional proteins in wheat germ lysate, although I did not achieve sufficient levels of mCherry expression to continue work using fluorescence microscopy. A possible reason behind this difference between expression from GFP and mCherry is that the genes may operate differently under differing X-DNA:gene ratios. Similar to the original work that demonstrated an optimal ratio for GFP in the plasmid pIVEX1.3, it appears that expression of different genes may require further optimization regarding the amount of X-DNA used in gel formation; this can be further tested with an array of different proteins. Further insight into these mechanisms will potentially allow for better control over

genotype-phenotype connection, as control over expression levels can aid in ensuring proteins have sufficient nickel binding sites on their associated microgel and thus reducing their free diffusion to gels containing other gene sets.

The major thrust of this thesis is to demonstrate that in the field of cell-free protein synthesis, DNA materials offer multiple avenues for tailoring a platform for protein expression and capture that is not achievable through other materials. There is much left to explore and optimize within the realm of possibilities presented in this work. Application of these methods to gene selection and ensuring genotype-phenotype connection within DNA gels is the next logical step in this process, presenting an entirely *in vitro* method of exploring full mutation spaces in the engineering and discovery of new protein activity.

BIBLIOGRAPHY

1. De Jong, H. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* **9,** 67–103 (2002).

2. Kitano, H. Systems biology: A brief overview. *Science* **295,** 1662–1664 (2002).

3. Kitano, H. Computational systems biology. *Nature* **420,** 206–210 (2002).

4. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402,** C47–C52 (1999).

5. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 425–430 (2006).

6. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Sci.* **329 ,** 52–56 (2010).

7. Benner, S. A. & Sismour, A. M. Synthetic biology. *Nat. Rev. Genet.* **6,** 533–543 (2005).

8. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. CELL Biol.* **10,** 410–422 (2009).

9. Haseltine, E. L. & Arnold, F. H. Synthetic Gene Circuits: Design with Directed Evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 1–19 (2007).

10. Billerbeck, S., Härle, J. & Panke, S. The good of two worlds: increasing complexity in cell-free systems. *Curr. Opin. Biotechnol.* **24,** 1037–1043 (2013).

11. Trucksis, M., Michalski, J., Deng, Y. K. & Kaper, J. B. The Vibrio cholerae genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci.* **95 ,** 14464–14469 (1998).

12. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem. Soc. Rev.* **39,** 1388–1405 (2010).

13. Moses, J. E. & Moorhouse, A. D. The growing applications of click chemistry. *Chem. Soc. Rev.* **36,** 1249–1262 (2007).

14. Roh, Y. H. *et al.* DNAsomes: Multifunctional DNA-Based Nanocarriers. *Small* **7,** 74–78 (2011).

15. Owczarzy, R. *et al.* Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **43,** 3537–3554 (2004).

16. You, Y., Tataurov, A. V & Owczarzy, R. Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **95,** 472–486 (2011).

17. Cunliffe, D., Pennadam, S. & Alexander, C. Synthetic and biological polymers––merging the interface. *Eur. Polym. J.* **40,** 5–25 (2004).

18. Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17,** 265–286 (1988).

19. Jones, M. R. *et al.* DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nat Mater* **9,** 913–917 (2010).

20. Macfarlane, R. J. *et al.* Nanoparticle Superlattice Engineering with DNA. *Sci.* **334 ,** 204–208 (2011).

21. Park, S. Y. *et al.* DNA-programmable nanoparticle crystallization. *Nature* **451,** 553–556 (2008).

22. Senesi, A. J. *et al.* Stepwise Evolution of DNA-Programmable Nanoparticle Superlattices. *Angew. Chemie Int. Ed.* **52,** 6624–6628 (2013).

23. Zhang, C. *et al.* A general approach to DNA-programmable atom equivalents. *Nat Mater* **12,** 741–746 (2013).

24. Cheng, W. *et al.* Probing in Real Time the Soft Crystallization of DNA-Capped Nanoparticles. *Angew. Chemie* **122,** 390–394 (2010).

25. Cheng, W., Park, N., Walter, M. T., Hartman, M. R. & Luo, D. Nanopatterning self-assembled nanoparticle superlattices by moulding microdroplets. *Nat Nano* **3,** 682–690 (2008).

26. Cheng, W., Campolongo, M. J., Tan, S. J. & Luo, D. Freestanding ultrathin nano-membranes via self-assembly. *Nano Today* **4,** 482–493 (2009).

27. Campolongo, M. J. *et al.* Crystalline Gibbs Monolayers of DNA-Capped Nanoparticles at the Air–Liquid Interface. *ACS Nano* **5,** 7978–7985 (2011).

28. Tan, S. J. *et al.* Crystallization of DNA-Capped Gold Nanoparticles in High-Concentration, Divalent Salt Environments. *Angew. Chemie Int. Ed.* **53,** 1316–1319 (2014).

29. Luo, D. The road from biology to materials. *Mater. Today* **6,** 38–43 (2003).

30. Seeman, N. C. DNA in a material world. *Nature* **421,** 427–431 (2003).

31. LaBean, T. H. *et al.* Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes. *J. Am. Chem. Soc.* **122,** 1848–1860 (2000).

32. Shen, Z., Yan, H., Wang, T. & Seeman, N. C. Paranemic Crossover DNA: A Generalized Holliday Structure with Applications in Nanotechnology. *J. Am. Chem. Soc.* **126,** 1666–1674 (2004).

33. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350,** 631–633 (1991).

34. Zhang, Y. & Seeman, N. C. Construction of a DNA-Truncated Octahedron. *J. Am. Chem. Soc.* **116,** 1661–1669 (1994).

35. Aldaye, F. A. & Sleiman, H. F. Sequential Self-Assembly of a DNA Hexagon as a Template for the Organization of Gold Nanoparticles. *Angew. Chemie Int. Ed.* **45,** 2204–2209 (2006).

36. Aldaye, F. A. & Sleiman, H. F. Dynamic DNA Templates for Discrete Gold Nanoparticle Assemblies: Control of Geometry, Modularity, Write/Erase and Structural Switching. *J. Am. Chem. Soc.* **129,** 4130–4131 (2007).

37. Seeman, N. C. DNA Nanotechnology: Novel DNA Constructions. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 225–248 (1998).

38. Elbaz, J., Wang, Z.-G., Wang, F. & Willner, I. Programmed Dynamic Topologies in DNA Catenanes. *Angew. Chemie Int. Ed.* **51,** 2349–2353 (2012).

39. Li, Y. *et al.* Controlled assembly of dendrimer-like DNA. *Nat. Mater.* **3,** 38–42 (2004).

40. Li, Y., Cu, Y. T. H. & Luo, D. Multiplexed detection of pathogen DNA with DNA-based fluorescence nanobarcodes. *Nat Biotech* **23,** 885–889 (2005).

41. Lee, J. B. *et al.* Multifunctional nanoarchitectures from DNA-based ABC monomers. *Nat. Nanotechnol.* **4,** 430–436 (2009).

42. Yang, D. *et al.* Novel DNA materials and their applications. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **2,** 648–669 (2010).

43. Tan, S. J., Kiatwuthinon, P., Roh, Y. H., Kahn, J. S. & Luo, D. Engineering Nanocarriers for siRNA Delivery. *Small* **7,** 841–856 (2011).

44. Lee, J. B. *et al.* DNA-based nanostructures for molecular sensing. *Nanoscale* **2,** 188–197 (2010).

45. Campolongo, M. J. *et al.* Adaptive DNA-based materials for switching, sensing, and logic devices. *J. Mater. Chem.* **21,** 6113–6121 (2011).

46. Um, S. H. *et al.* Enzyme-catalysed assembly of DNA hydrogel. *Nat. Mater.* **5,** 797–801 (2006).

47. Um, S. H., Lee, J. B., Kwon, S. Y., Li, Y. & Luo, D. Dendrimer-like DNA-based fluorescence nanobarcodes. *Nat. Protoc.* **1,** 995–1000 (2006).

48. Xing, Y. *et al.* Self-Assembled DNA Hydrogels with Designable Thermal and Enzymatic Responsiveness. *Adv. Mater.* **23,** 1117–1121 (2011).

49. Park, N. *et al.* High-yield cell-free protein production from P-gel. *Nat. Protoc.* **4,** 1759–1770 (2009).

50. Park, N., Um, S. H., Funabashi, H., Xu, J. & Luo, D. A cell-free protein-producing gel. *Nat. Mater.* **8,** 432–437 (2009).

51. Hartman, M. R. *et al.* Thermostable Branched DNA Nanostructures as Modular Primers for Polymerase Chain Reaction. *Angew. Chem. Int. Ed. Engl.* **52,** 8699–8702 (2013).

52. Lee, J. B. *et al.* A mechanical metamaterial made from a DNA hydrogel. *Nat Nano* **7,** 816–820 (2012).

53. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23,** 150–156 (2005).

54. Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **10,** 411–421 (1999).

55. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotech* **22,** 1567–1572 (2004).

56. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Meth* **2,** 905–909 (2005).

57. Shcherbo, D. *et al.* Bright far-red fluorescent protein for whole-body imaging. *Nat Meth* **4,** 741–746 (2007).

58. Subach, F. V *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat Meth* **6,** 153–159 (2009).

59. *Protein Structure-Function Relationship*. (Springer US, 1996).

60. Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **21,** 473–480 (2011).

61. Dougherty, M. J. & Arnold, F. H. Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **20,** 486–491 (2009).

62. Johannes, T. W. & Zhao, H. Directed evolution of enzymes and biosynthetic pathways. *Curr. Opin. Microbiol.* **9,** 261–267 (2006).

63. Lutz, S. Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21,** 734–743 (2010).

64. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci.* **99 ,** 14652–14657 (2002).

65. Zhao, H., Chockalingam, K. & Chen, Z. Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr. Opin. Biotechnol.* **13,** 104–110 (2002).

66. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).

67. Bornscheuer, U. T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5,** 137–143 (2001).

68. Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14,** 438–443 (2003).

69. Himmel, M. E., Ruth, M. F. & Wyman, C. E. Cellulase for commodity products from cellulosic biomass. *Curr. Opin. Biotechnol.* **10,** 358–364 (1999).

70. Lin, H., Li, W., Guo, C., Qu, S. & Ren, N. Advances in the study of directed evolution for cellulases. *Front. Environ. Sci. Eng. China* **5,** 519–525 (2011).

71. Nestl, B. M., Nebel, B. A. & Hauer, B. Recent progress in industrial biocatalysis. *Curr. Opin. Chem. Biol.* **15,** 187–193 (2011).

72. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* **10,** 866–876 (2009).

73. Turner, N. J. Directed evolution of enzymes for applied biocatalysis. *Trends Biotechnol.* **21,** 474–478 (2003).

74. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5,** 568–574 (2009).

75. Wilson, D. B. Cellulases and biofuels. *Curr. Opin. Biotechnol.* **20,** 295–299 (2009).

76. Zhang, Y.-H. P., Himmel, M. E. & Mielenz, J. R. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol. Adv.* **24,** 452–481 (2006).

77. Percival Zhang, Y.-H., Himmel, M. E. & Mielenz, J. R. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol. Adv.* **24,** 452–481 (2006).

78. Lueking, A. *et al.* Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270,** 103–111 (1999).

79. Galarneau, A., Primeau, M., Trudeau, L. E. & Michnick, S. W. beta-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat. Biotechnol.* **20,** 619–622 (2002).

80. Xiao, Z. Z., Storms, R. & Tsang, A. Microplate-based filter paper assay to measure total cellulase activity. *Biotechnol. Bioeng.* **88,** 832–837 (2004).

81. Cianchetta, S., Galletti, S., Burzi, P. L. & Cerato, C. A Novel Microplate-Based Screening Strategy to Assess the Cellulolytic Potential of Trichoderma Strains. *Biotechnol. Bioeng.* **107,** 461–468 (2010).

82. Yang, G. & Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **10,** 2704–2715 (2009).

83. Flynn, C. E., Lee, S.-W., Peelle, B. R. & Belcher, A. M. Viruses as vehicles for growth, organization and assembly of materials. *Acta Mater.* **51,** 5867–5880 (2003).

84. Flynn, C. E. *et al.* Synthesis and organization of nanoscale II-VI semiconductor materials using evolved peptide specificity and viral capsid assembly. *J. Mater. Chem.* **13,** 2414–2421 (2003).

85. Lee, S.-W., Mao, C., Flynn, C. E. & Belcher, A. M. Ordering of Quantum Dots Using Genetically Engineered Viruses. *Sci.* **296 ,** 892–895 (2002).

86. Sarikaya, M., Tamerler, C., Jen, A. K. Y., Schulten, K. & Baneyx, F. Molecular biomimetics: nanotechnology through biology. *Nat. Mater.* **2,** 577–585 (2003).

87. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **12,** 1281–1289 (2005).

88. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173,** 33–38 (1996).

89. Fu, A. Y., Spence, C., Scherer, A., Arnold, F. H. & Quake, S. R. A microfabricated fluorescence-activated cell sorter. *Nat. Biotechnol.* **17,** 1109–1111 (1999).

90. Ma, F., Xie, Y., Huang, C., Feng, Y. & Yang, G. An Improved Single Cell Ultrahigh Throughput Screening Method Based on In Vitro Compartmentalization. *PLoS One* **9,** (2014).

91. Nolan, G. P., Fiering, S., Nicolas, J. F. & Herzenberg, L. A. Fluorescence-Activated Cell Analysis And Sorting Of Viable Mammalian-Cells Based On Beta-D-Galactosidase Activity After Transduction Of Escherichia-Coli LacZ. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2603–2607 (1988).

92. Stapleton, J. A. & Swartz, J. R. Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PLoS One* **5,** e15275 (2010).

93. Valdivia, R. H. & Falkow, S. Fluorescence-based isolation of bacterial genes expressed within host cells. *Science (80-.).* **277,** 2007–2011 (1997).

94. Lång, H. Outer membrane proteins as surface display systems. *Int. J. Med. Microbiol.* **290,** 579–585 (2000).

95. Lee, S. Y., Choi, J. H. & Xu, Z. Microbial cell-surface display. *Trends Biotechnol.* **21,** 45–52 (2003).

96. Wittrup, K. D. Protein engineering by cell-surface display. *Curr. Opin. Biotechnol.* **12,** 395–399 (2001).

97. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *METHODS* **60,** 15–26 (2013).

98. Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17,** 467–473 (2007).

99. Wang, K. C. *et al.* Yeast surface display of antibodies via the heterodimeric interaction of two coiled-coil adapters. *J. Immunol. Methods* **354,** 11–19 (2010).

100. Kondo, A. & Ueda, M. Yeast cell-surface display—applications of molecular display. *Appl. Microbiol. Biotechnol.* **64,** 28–40 (2004).

101. Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc. Natl. Acad. Sci.* **103 ,** 9637–9642 (2006).

102. Beerli, R. R. *et al.* Isolation of human monoclonal antibodies by mammalian cell display. *Proc. Natl. Acad. Sci.* **105 ,** 14336–14341 (2008).

103. Azzazy, H. M. E. & Highsmith, W. E. Phage display technology: clinical applications and recent innovations. *Clin. Biochem.* **35,** 425–445 (2002).

104. Barbas, C. F., Kang, A. S., Lerner, R. A. & Benkovic, S. J. Assembly Of Combinatorial Antibody Libraries On Phage Surfaces - The Gene-III Site. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7978–7982 (1991).

105. Beaber, J. W., Tam, E. M., Lao, L. S. & Rondon, I. J. A new helper phage for improved monovalent display of Fab molecules. *J. Immunol. Methods* **376,** 46–54 (2012).

106. Benhar, I. Biotechnological applications of phage and cell display. *Biotechnol. Adv.* **19,** 1–33 (2001).

107. Bratkovic, T. Progress in phage display: evolution of the technique and its applications. *Cell. Mol. LIFE Sci.* **67,** 749–767 (2010).

108. Fujii, I. Directed evolution of antibody molecules in phage-displayed combinatorial libraries. *YAKUGAKU ZASSHI-JOURNAL Pharm. Soc. JAPAN* **127,** 91–99 (2007).

109. Gao, J., Wang, Y., Liu, Z. & Wang, Z. Phage display and its application in vaccine design. *Ann. Microbiol.* **60,** 13–19 (2010).

110. Jung, S., Honegger, A. & Pluckthun, A. Selection for improved protein stability by phage display. *J. Mol. Biol.* **294,** 163–180 (1999).

111. Koscielska, K., Kiczak, L., Kasztura, M., Wesolowska, O. & Otlewski, J. Phage display of proteins. *ACTA Biochim. Pol.* **45,** 705–720 (1998).

112. Li, Y. *et al.* Directed evolution of human T-cell receptors with picomolar affinities by phage display. *Nat. Biotechnol.* **23,** 349–354 (2005).

113. Petrenko, V. A. Evolution of phage display: from bioactive peptides to bioselective nanomaterials. *Expert Opin. Drug Deliv.* **5,** 825–836 (2008).

114. Rader, C. & Barbas, C. F. Phage display of combinatorial antibody libraries. *Curr. Opin. Biotechnol.* **8,** 503–508 (1997).

115. Sieber, V., Pluckthun, A. & Schmid, F. X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16,** 955–960 (1998).

116. Smith, G. P. & Petrenko, V. A. Phage display. *Chem. Rev.* **97,** 391–410 (1997).

117. Wang, K. C., Wang, X., Zhong, P. & Luo, P. P. Adapter-Directed Display: A Modular Design for Shuttling Display on Phage Surfaces. *J. Mol. Biol.* **395,** 1088–1101 (2010).

118. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making Antibodies By Phage Display Technology. *Annu. Rev. Immunol.* **12,** 433–455 (1994).

119. Nam, K. T. *et al.* Virus-Enabled Synthesis and Assembly of Nanowires for Lithium Ion Battery Electrodes. *Sci.* **312 ,** 885–888 (2006).

120. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472,** 499–503 (2011).

121. Nishikawa, T., Sunami, T., Matsuura, T. & Yomo, T. Directed Evolution of Proteins through In Vitro Protein Synthesis in Liposomes. *J. Nucleic Acids* **2012,** 11 pages (2012).

122. Hanes, J., Schaffitzel, C., Knappik, A. & Pluckthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18,** 1287–1292 (2000).

123. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Pluckthun, A. Tailoring in vitro evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 75–80 (2001).

124. Xu, L. H. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9,** 933–942 (2002).

125. Lipovsek, D. & Pluckthun, A. In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290,** 51–67 (2004).

126. Nemoto, N., MiyamotoSato, E., Husimi, Y. & Yanagawa, H. In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* **414,** 405–408 (1997).

127. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12297–12302 (1997).

128. Traut, R. R. & Monro, R. E. The puromycin reaction and its relation to protein synthesis. *J. Mol. Biol.* **10,** 63–72 (1964).

129. Ueno, S., Kimura, S., Ichiki, T. & Nemoto, N. Improvement of a puromycin-linker to extend the selection target varieties in cDNA display method. *J. Biotechnol.* **162,** 299–302 (2012).

130. Yanagawa, H. Exploration of the Origin and Evolution of Globular Proteins by mRNA Display. *Biochemistry* **52,** 3841–3851 (2013).

131. Sergeeva, A., Kolonin, M. G., Molldrem, J. J., Pasqualini, R. & Arap, W. Display technologies: Application for the discovery of drug and gene delivery agents. *Adv. Drug Deliv. Rev.* **58,** 1622–1654 (2006).

132. Griffiths, A. D. & Tawfik, D. S. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* **22,** 24–35 (2003).

133. Lu, W.-C., Levy, M., Kincaid, R. & Ellington, A. D. Directed Evolution of the Substrate Specificity of Biotin Ligase. *Biotechnol. Bioeng.* **111,** 1071–1081 (2014).

134. Ostafe, R., Prodanovic, R., Nazor, J. & Fischer, R. Ultra-High-Throughput Screening Method for the Directed Evolution of Glucose Oxidase. *Chem. Biol.* **21,** 414–421 (2014).

135. Takeuchi, R., Choi, M. & Stoddard, B. L. Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4061–4066 (2014).

136. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16,** 652–656 (1998).

137. Zinchenko, A. *et al.* One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution. *Anal. Chem.* **86,** 2526–2533 (2014).

138. Miller, O. J. *et al.* Directed evolution by in vitro compartmentalization. *Nat Meth* **3,** 561–570 (2006).

139. Park, S. & Hamad-Schifferli, K. Enhancement of In Vitro Translation by Gold Nanoparticle–DNA Conjugates. *ACS Nano* **4,** 2555–2560 (2010).

140. He, M. & Taussig, M. J. Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Res.* **29 ,** e73–e73 (2001).

141. Angenendt, P., Kreutzberger, J., Glökler, J. & Hoheisel, J. D. Generation of High Density Protein Microarrays by Cell-free in Situ Expression of Unpurified PCR Products . *Mol. Cell. Proteomics* **5 ,** 1658–1666 (2006).

142. Chandra, H. & Srivastava, S. Cell-free synthesis-based protein microarrays and their applications. *Proteomics* **10,** 717–730 (2010).

143. He, M. Cell-free protein synthesis: applications in proteomics and biotechnology. *N. Biotechnol.* **25,** 126–132 (2008).

144. Berrade, L., Garcia, A. & Camarero, J. Protein Microarrays: Novel Developments and Applications. *Pharm. Res.* **28,** 1480–1499 (2011).

145. Fraser, C. S., Hershey, J. W. B. & Doudna, J. A. The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat Struct Mol Biol* **16,** 397–404 (2009).

146. Gingras, A.-C., Raught, B. & Sonenberg, N. eIF4 INITIATION FACTORS: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annu. Rev. Biochem.* **68,** 913–963 (1999).

147. Yang, D. *et al.* Enhanced transcription and translation in clay hydrogel and implications for early life evolution. *Sci. Rep.* **3,** (2013).

148. Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat Rev Mol Cell Biol* **13,** 168–182 (2012).

149. Noireaux, V., Maeda, Y. T. & Libchaber, A. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci.* **108 ,** 3473–3480 (2011).

150. Libchaber, V. N. and R. B.-Z. and J. G. and H. S. and A. Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* **2,** P1 (2005).

151. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. United States Am.* **101 ,** 17669–17674 (2004).

152. Qin, D., Xia, Y. & Whitesides, G. M. Soft lithography for micro- and nanoscale patterning. *Nat. Protoc.* **5,** 491–502 (2010).

153. McDonald, J. C. & Whitesides, G. M. Poly(dimethylsiloxane) as a Material for Fabricating Microfluidic Devices. *Acc. Chem. Res.* **35,** 491–499 (2002).

154. Ruiz, R., Kiathwuthinon, P., Kahn, J. S., Roh, Y. H. & Luo, D. Cell-Free Protein Expression from DNA-Based Hydrogel (P-Gel) Droplets for Scale-Up Production. *Ind. Biotechnol.* **8,** 372–377 (2012).

155. Ha, J. & Yang, S. Breakup of a Multiple Emulsion Drop in a Uniform Electric Field. *J. Colloid Interface Sci.* **213,** 92–100 (1999).

156. Raut, J. S., Akella, S., Singh, A. & Naik, V. M. Catastrophic Drop Breakup in Electric Field. *Langmuir* **25,** 4829–4834 (2009).

157. Ha, J.-W. & Yang, S.-M. Deformation and breakup of a second-order fluid droplet in an electric field. *Korean J. Chem. Eng.* **16,** 585–594 (1999).

158. Zagnoni, M., Le Lain, G. & Cooper, J. M. Electrocoalescence Mechanisms of Microdroplets Using Localized Electric Fields in Microfluidic Channels. *Langmuir* **26,** 14443–14449 (2010).

159. Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M. & Record, M. T. One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc. Natl. Acad. Sci.* **107 ,** 10418–10423 (2010).

160. Hartvig, L. & Christiansen, J. Intrinsic termination of T7 RNA polymerase mediated by either RNA or DNA. *EMBO J.* **15,** 4767–4774 (1996).

161. Sousa, R. & Mukherjee, S. in (Biology, B. T.-P. in N. A. R. and M.) **73,** 1–41 (Academic Press, 2003).

162. Tang, G. Q. & Patel, S. S. T7 RNA polymerase-induced bending of promoter DNA is coupled to DNA opening. *Biochemistry* **45,** 4936–4946 (2006).

163. Újvári, A. & Martin, C. T. Evidence for DNA bending at the T7 RNA polymerase promoter. *J. Mol. Biol.* **295,** 1173–1184 (2000).

164. Cirino, P., Mayer, K. & Umeno, D. in *Dir. Evol. Libr. Creat. SE - 1* (Arnold, F. & Georgiou, G.) **231,** 3–9 (Humana Press, 2003).

165. Integrated DNA Technologies, I. OligoAnalyzer 3.1. (2004). at <https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>

166. Kim, T.-W. *et al.* Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.* **126,** 554–561 (2006).

167. Calhoun, K. A. & Swartz, J. R. Total amino acid stabilization during cell-free protein synthesis reactions. *J. Biotechnol.* **123,** 193–203 (2006).

168. Schwarz, D. *et al.* Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat. Protoc.* **2,** 2945–2957 (2007).

169. Thomas, T. & Thomas, T. J. Polyamines in cell growth and cell death: molecular mechanisms and therapeutic applications. *Cell. Mol. Life Sci. C.* **58,** 244–258 (2001).

170. Feuerstein, B. G., Pattabiraman, N. & Marton, L. J. Molecular dynamics of spermine-DNA 1nteractioas sequence specificity and DNA bending for a simple ligand. *Nucleic Acids Res.* **17 ,** 6883–6892 (1989).

171. Feuerstein, B. G., Williams, L. D., Basu, H. S. & Marton, L. J. Implications and concepts of polyamine-nucleic acid interactions. *J. Cell. Biochem.* **46,** 37–47 (1991).

172. Porath, J. Immobilized metal ion affinity chromatography. *Protein Expr. Purif.* **3,** 263–281 (1992).

173. Chaga, G., Hopp, J. & Nelson, P. Immobilized metal ion affinity chromatography on Co2+-carboxymethylaspartate–agarose Superflow, as demonstrated by one-step purification of lactate dehydrogenase from chicken breast muscle. *Biotechnol. Appl. Biochem.* **29,** 19–24 (1999).

174. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S. Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **41,** 98–105 (2005).

175. Block, H. *et al.* in *Guid. to Protein Purification, 2nd Ed.* (Enzymology, R. R. B. and M. P. D. B. T.-M. in) **463,** 439–473 (Academic Press, 2009).

176. Sidenius, U., Farver, O., Jøns, O. & Gammelgaard, B. Comparison of different transition metal ions for immobilized metal affinity chromatography of selenoprotein P from human plasma. *J. Chromatogr. B Biomed. Sci. Appl.* **735,** 85–91 (1999).

177. Shen, W., Zhong, H., Neff, D. & Norton, M. L. NTA Directed Protein Nanopatterning on DNA Origami Nanoconstructs. *J. Am. Chem. Soc.* **131,** 6660–6661 (2009).

178. Sacca, B. & Niemeyer, C. M. Functionalization of DNA nanostructures with proteins. *Chem. Soc. Rev.* **40,** 5910–5921 (2011).

179. Shimada, J. *et al.* Conjugation of DNA with protein using His-tag chemistry and its application to the aptamer-based detection system. *Biotechnol. Lett.* **30,** 2001–2006 (2008).

180. Cheng, W. *et al.* Free-standing nanoparticle superlattice sheets controlled by DNA. *Nat Mater* **8,** 519–525 (2009).

181. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry. *Anal. Biochem.* **273,** 73–80 (1999).

182. Rhee, S. S. & Burke, D. H. Tris(2-carboxyethyl)phosphine stabilization of RNA: comparison with dithiothreitol for use with nucleic acid and thiophosphoryl chemistry. *Anal. Biochem.* **325,** 137–143 (2004).

183. Shafer, D. E., Inman, J. K. & Lees, A. Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with Maleimide and α-Haloacyl Groups: Anomalous Elution of TCEP by Gel Filtration. *Anal. Biochem.* **282,** 161–164 (2000).

184. Xu, X. *et al.* Gravity and Surface Tension Effects on the Shape Change of Soft Materials. *Langmuir* **29,** 8665–8674 (2013).

185. Zhu, Z. *et al.* Highly sensitive and quantitative detection of rare pathogens through agarose droplet microfluidic emulsion PCR at the single-cell level. *Lab Chip* **12,** 3907–3913 (2012).

186. Zhu, Z. *et al.* Single-molecule emulsion PCR in microfluidic droplets. *Anal. Bioanal. Chem.* **403,** 2127–2143 (2012).

187. Tewhey, R. *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotech* **27,** 1025–1031 (2009).

188. Leng, X., Zhang, W., Wang, C., Cui, L. & Yang, C. J. Agarose droplet microfluidics for highly parallel and efficient single molecule emulsion PCR. *Lab Chip* **10,** 2841–2843 (2010).

189. Lee, J. B., Hong, J., Bonner, D. K., Poon, Z. & Hammond, P. T. Self-assembled RNA interference microsponges for efficient siRNA delivery. *Nat Mater* **11,** 316–322 (2012).

190. Demirjian, D. C., Morís-Varas, F. & Cassidy, C. S. Enzymes from extremophiles. *Curr. Opin. Chem. Biol.* **5,** 144–151 (2001).

191. Van den Burg, B. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* **6,** 213–218 (2003).

192. Altan-Bonnet, G., Libchaber, A. & Krichevsky, O. Bubble Dynamics in Double-Stranded DNA. *Phys. Rev. Lett.* **90,** 138101 (2003).

193. Campa, A. Bubble propagation in a helicoidal molecular chain. *Phys. Rev. E* **63,** 21901 (2001).

194. Peyrard, M. Using DNA to probe nonlinear localized excitations? *EPL (Europhysics Lett.* **44,** 271 (1998).

195. Wei, X., Nangreave, J. & Liu, Y. Uncovering the Self-Assembly of DNA Nanostructures by Thermodynamics and Kinetics. *Acc. Chem. Res.* (2014). doi:10.1021/ar5000665

196. Shi, Y.-B., Griffith, J., Gamper, H. & Hearst, J. E. Evidence for structural deformation of the DNA helix by a psoralen diadduct but not by a monoadduct. *Nucleic Acids Res.* **16 ,** 8945–8952 (1988).

197. Gaboriau, F., Vigny, P. & Moron, J. Secondary Structure Of Dna Modified By Monofunctional Psoralen Derivatives. *Biochemistry* **28,** 5801–5807 (1989).

198. Yoakum, G. H. & Cole, R. S. Cross-Linking And Relaxation Of Supercoiled Dna By Psoralen And Light. *Biochim. Biophys. Acta* **521,** 529–546 (1978).

199. Marciani, S., Dall'acqua, F., Vedaldi, D. & Rodighiero, G. Receptor sites of DNA for the photoreaction with psoralen. *Farmaco. Sci.* **31,** 140—151 (1976).

200. Takasugi, M. *Et Al.* Sequence-Specific Photoinduced Cross-Linking Of The 2 Strands Of Double-Helical Dna By A Psoralen Covalently Linked To A Triple Helix-Forming Oligonucleotide. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 5602–5606 (1991).

201. Kittler, L. & Lober, G. Sequence Specificity Of Dna-Psoralen Photoproduct Formation In Supercoiled Plasmid DNA (pUC19). *J. Photochem. Photobiol. B-BIOLOGY* **27,** 161–166 (1995).

202. Dall'Acqua, F., Vedaldi, D., Bordin, F. & Rodighiero, G. New Studies on the Interaction Between 8-Methoxypsoralen and DNA in Vitro. *J Investig Dermatol* **73,** 191–197 (1979).

203. Roh, Y. H., Park, J. H., Ye, J. J., Lee, J. E. & Luo, D. Systematic Studies of UV Stability and Photopolymerization Efficiency of DNA-Based Nanomaterials. *ChemPhysChem* **13,** 2517–2521 (2012).