

HELICASE PURIFICATION FOR DNA SEQUENCING

Leah Labib, B. Sc.

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE

in Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering
University of Ottawa
Ottawa, Canada

and

Medical Devices Innovation Institute
Ottawa Hospital Research Institute
University of Ottawa Heart Institute
Ottawa, Canada

April 2014

© Leah Labib, Ottawa, Canada, 2014

Abstract

BACKGROUND: A method to increase accuracy and ease-of-use, while decreasing time and cost in deoxyribonucleic acid (DNA) sequence identification, is sought after. Helicase, which unwinds DNA, and avidin, which strongly attracts biotin for potential attraction of biotinylated DNA segments, were investigated for use in a novel DNA sequencing method.

AIM: This study aimed to (1) purify bacteriophage T7 gene product 4 helicase and helicase-avidin fusion protein in a bacterial host and (2) characterize their functionality.

METHODS: Helicase and helicase-avidin were cloned for purification from bacteria. Helicase-avidin was solubilised *via* urea denaturation/renaturation. DNA and biotin binding were assessed using Electrophoretic Mobility Shift Assays and biotinylated resins, respectively.

RESULTS: (1) Helicase and helicase-avidin proteins were successfully purified. (2) Helicase protein was able to bind DNA and avidin protein strongly bound biotin.

CONCLUSION: Helicase and helicase-avidin can be purified in a functional form from a bacterial host, thus supporting further investigation for DNA sequencing purposes.

Acknowledgements

It is my honour to thank those, without whom, this research undertaking and incredible learning experience would not be possible.

First and foremost I would like to thank Dr. Tofy Mussivand and Dr. Jeff Dilworth for their roles as my co-supervisors. I feel very lucky and am immensely grateful that I have had the privilege of working with two great scientists who lead by example, and who are honest, kind, sensible, and upstanding. Without the guidance, mentoring, and encouragement of Dr. Mussivand, I would not have been introduced to the field of biomedical engineering or this innovative research topic; I would not have truly learned the paths of creativity, independence, team work, time management, and self-study, or the importance of looking at minute details as equally as the big picture. Dr. Mussivand also provided me with the privilege of exploring the world of Medical Devices along with many extracurricular opportunities that I would not have otherwise been exposed to. I am deeply grateful to Dr. Dilworth for his support, guidance, expertise and patience. Dr. Dilworth graciously welcomed me into his laboratory providing me with all the resources, encouragement, and advice I could hope for towards the successful completion of this project, and made me feel empowered and confident in my work. Dr. Dilworth's cool-headed nature, attention to detail, keen interest in his students' endeavours, and innate sense of being a teacher have been not only helpful to me in my pursuits, but also inspirational to me. I have truly felt welcomed and at home in his lab and at ease under his supervision.

I am grateful to Dr. Bashir Morshed, a former post-doctoral fellow in the lab of Dr. Mussivand, for his creativity and perseverance in innovating the novel concept that is the basis of this project, and for introducing me to the inner workings of an engineer's mind. Also in Dr. Mussivand's lab, I would like to thank Nazli Parast, RN, who has been immensely helpful in organizing the lab, mentoring, and showing me an example of strength and hard work. I would also like to acknowledge Peter Haddad, MA Sc, Massomeh Sheikh-Hassani, MSc, and John Szalas, MA Sc in the lab of Dr. Mussivand as their roles as consultants for my work.

I would like to thank the post-doctoral fellows in the lab of Dr. Dilworth for helping me with my many trials and tribulations in troubleshooting my lab experiments, above and beyond what was expected of them. Specifically, I would like to thank Dr. Kulwant Singh and his impressive knowledge of all things science, Dr. Hervé Faralli, who provided me with so much advice and moral support, and Drs. Arif Aziz and QiCai Liu whose advice and support were instrumental in the successful completion of my studies. I would also like to thank Dr. Soji Sebastian, Lifang Li, and Katherine Reilly for their help.

I would like to thank my parents, for all of the obvious reasons, as well as for providing the roof over my head, the food in my belly, and access to a car whenever I wanted. I'd also like to thank my sister Sarah and her husband Lionel for all of their help and support. I am very grateful to Sarah for helping in revisions of this document, and being so understanding of my stress.

Last but not least, I am indebted to all of my amazing friends for listening to, and helping me deal with, my many, and sometimes seemingly endless, frustrations and overworked schedule, and for providing me with socialization and more moral support than I could have ever dreamed of. I'd also like to thank my little baby nephew Noah Fishman for making me smile and making my stress melt away whenever I see him.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents	v
List of Figures and Tables	viii
Abbreviations.....	ix
1 - Introduction.....	1
1.1 Clinical Needs.....	1
1.2 Rationale	1
1.3 Hypothesis.....	2
1.4 Specific Aims.....	2
1.5 DNA Sequencing Overview	3
1.5.1 Chain Termination.....	4
1.5.2 Sequencing-by-Synthesis.....	5
1.5.3 Real Time Polymerase Reading Approaches	8
1.5.4 Non-polymerase Enzymatic Approaches.....	9
1.5.5 Non-enzymatic Sequencing Approaches.....	11
1.5.6 Areas Requiring Attention.....	12
1.5.7 Proposed Method and Device Complementing this Thesis	14
1.5.8 DNA Sequencing Overview Summary	15
1.6 Prior Related Research	16
1.7 Proposed Research	16
1.7.1 Objective and scope	16
1.7.2 Helicase Protein.....	17
1.7.3 Exploitation of Avidin-Biotin Relationship for Helicase-based Sequencing.....	17
1.7.4 Experimental Summary.....	19
1.7.5 Some Relevant Terms and Definitions Used in this Thesis	20
1.8 Expected results.....	21
1.9 Contribution and Significance	21
2 – Literature Review.....	22
2.1 Helicases.....	22
2.1.1 Significance of Helicase	22
2.1.2 Helicases in General.....	22
2.1.3 Descriptors for Helicases.....	23
2.1.4 Different Classes of Helicases	24
2.1.5 Hexameric Helicases.....	24
2.1.6 Bacteriophage T7 Gene Product 4 DNA Helicase.....	28
2.1.7 DNA Helicases Summary.....	33
2.2 Biotin and Avidin.....	33
2.2.1 Biotin.....	34
2.2.2 Avidin.....	34
2.3 Fusion Proteins	35
2.4 DNA	35
2.4.1 Ribose Sugar.....	35
2.4.2 Phosphate Group.....	36
2.4.3 Nitrogenous Bases.....	37
2.4.4 Replication.....	37
2.4.5 DNA Modification.....	38
3 - Aim 1: Purification of Helicase.....	39

3.1 Rationale	39
3.2 Purpose.....	39
3.3 Methods	39
3.3.1 PCR Amplification of Helicase Enzyme DNA Sequence.	40
3.3.2 Digestion and Ligation Reactions.....	40
3.3.3 DH5 α Ttransformation.....	41
3.3.4 BL21 Transformation.....	41
3.3.5 Induction of Protein Overexpression.....	41
3.3.6 Protein Purification and Dialysis.....	41
3.4 Results.....	42
3.5 Discussion.....	45
3.6 Conclusion	48
4- Aim 2: Purification of Helicase-Avidin.....	49
4.1 Rationale	49
4.2 Purpose.....	49
4.3 Methods	49
4.3.1 Fusion Protein Engineering	49
4.3.2 DH5 α Transformation.....	50
4.3.3 BL21 Transformation.....	50
4.3.4 Optimization of Growth Conditions	50
4.3.5 Protein Purification.....	50
4.3.6 Renaturation of Insoluble Protein	51
4.3.7 Protein Visualization.	51
4.4 Results.....	52
4.5 Discussion.....	55
4.6 Conclusion	57
5 – Aim 3: Testing Helicase Functionality	58
5.1 Rationale	58
5.2 Purpose.....	58
5.3 Methods	58
5.3.1 Electrophoretic Mobility Shift Assay	58
5.3.2 Crosslinker optimization	59
5.3.3 Helicase-DNA binding	59
5.3.4 Native Polyacrylamide Gel Electrophoresis.....	59
5.4 Results.....	60
5.5 Discussion.....	64
5.6 Conclusion	68
6 – Aim 4: Testing Avidin Functionality	69
6.1 Rationale	69
6.2 Purpose.....	69
6.3 Methods	69
6.4 Results.....	69
6.5 Discussion.....	70
6.6 Conclusion	71
7 – Key Results.....	72
7.1 Helicase Purification	72
7.2 Helicase-avidin Purification.....	72
7.3 Helicase Functionality	73
7.4 Avidin Functionality	73
8 – Conclusions and Recommendations.....	74

8.1 Conclusion	74
8.2 Recommendations	75
8.2.1 Helicase Characterization.....	75
8.2.2 Integration with Electronics for DNA Sequencing Device Prototype.....	76
9 – References	77
Appendix I – Ladders	89
Appendix II – Materials and Buffers	90
Appendix III – Methods	93
Appendix IV – Figures.....	126
Appendix V – Helicase Assays.....	133

List of Figures and Tables

Figures

Figure 1.1: Chain terminating DNA sequencing.

Figure 1.2: Pyrosequencing.

Figure 1.3: Single Molecule Real Time™ sequencing.

Figure 1.4: Cartoon representation of the Medical Devices Innovation Institute proposed novel DNA sequencing method.

Figure 1.5: Line representation of fusion proteins.

Figure 1.6: Schematic representation of helicase-avidin fusion protein.

Figure 1.7: Attracting specific sequences of DNA to helicase.

Figure 1.8: Artificial DNA fork.

Figure 2.1: DnaB family helicase conserved sequences.

Figure 2.2: T7gp4 DNA helicase structure.

Figure 2.3: T7gp4 DNA helicase subunit interface.

Figure 2.4: Deoxyribonucleic Acid Structure.

Figure 2.5: DNA Phosphodiester Bond.

Figure 3.1: Agarose gel of helicase PCR product, pET28b(+) vector, and final ligation product.

Figure 3.2: Verification of DH5α pET-helicase transformation.

Figure 3.3: Agarose gel electrophoresis verifying BL21 pET-helicase transformation by PCR screen.

Figure 3.4: Verification of helicase overexpression.

Figure 3.5: Purification of overexpressed helicase protein.

Figure 4.1: Coomassie stain of unsuccessful attempts at helicase-avidin purification

Figure 4.2: Western blot of sonicate supernatants of induced ER2566, RIL and BL21 cells at 37°C and 18 °C.

Figure 4.3: Helicase-avidin purification via various denaturing/renaturing conditions.

Figure 5.1: Effect of crosslinkers on MyoD binding.

Figure 5.2: Helicase binding to DNA.

Figure 5.3: Helicase-avidin binding to DNA.

Figure 6.1: Avidin-biotin binding assay.

Tables

Table 3.1: PCR primers for codon-optimized helicase Sequence.

Table 3.2: PCR Reaction Conditions.

Table 4.1: Components of the Helicase-avidin Fusion protein.

Table 5.1: Forked DNA substrate

Table 5.2: E-box DNA Sequence.

Table 5.3: Helicase reaction mixture.

Abbreviations

AAA+: ATPase Associated with various cellular Activities
A: adenine
AFM: atomic force microscopy
bp: base pairs
C: cytosine
dATP: deoxyadenosine triphosphate
ddH₂O: distilled deionized water
ddNTP: dideoxynucleotide Triphosphate
DNA: deoxyribonucleic acid
dNTP: deoxynucleotide Triphosphate
dsDNA: double stranded DNA
DSP: Dithiobis[succinimidyl propionate]
DSS: Disuccinimidyl suberate
DTT: dithiothreitol
dTTP: deoxythymidine triphosphate
E. coli: *Escherichia coli*
EtBr: Ethidium Bromide
EMSA: Electrophoretic Mobility Shift Assay
FRET: fluorescent resonance energy transfer
G: guanine
Gly: glycine
GT: guanidium thiocyanate
His-tag: polyhistidine tag
IMAC: immobilized metal-affinity chromatography
IPTG Isopropyl- β -D-thiogalactoside
LB Broth: Luria Bertani Broth
MCS: multiple cloning site
NTP: nucleotide Triphosphate
OD₆₀₀ and OD: Optical Density at wavelength of 600nm
PBS: phosphate buffered saline
PCR: polymerase chain reaction
Pi: inorganic phosphate
PPi: pyrophosphate
PNK: polynucleotide kinase
PVDF: Polyvinylidene fluoride
RIL: BL21-CodonPlus-RIL
Rpm: rotations per minute
SDS PAGE: Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis
Ser: serine
SET: single electron transistor
siNA: silicon nanowire
SMRT: single molecule real time TM
ssDNA: single stranded DNA
SWCNT: single walled carbon nanotube

T: thymine

T4gp41: bacteriophage T4 gene product 41

T7gp4: bacteriophage T7 gene product 4

UV: ultra violet

ZMW: zero mode waveguide

1 - Introduction

1.1 Clinical Needs

Genetic research has made the possibility of personalized medicine seem within reach. Despite the fact that great strides have been made towards identifying the genetic basis of many diseases, personalized medicine is still not effectively used in clinical settings due to cost, time, and the specialized training required to obtain genetic information by way of DNA sequence detection. Therefore, treating patients according to their genotype, and making use of this scientific knowledge is currently not easily attainable.

This issue is common to several fields of study in addition to personalized medicine, including field sciences, forensic sciences, pharmaceutical research, biochemistry, and molecular biology, among others, thus highlighting the need for a technique that addresses the aforementioned obstacles.

1.2 Rationale

A novel device and technique to automatically detect a DNA sequence from a sample input (e.g. a fingerprint or blood sample) to a genetic read out, which is relatively inexpensive, rapid, accurate, smart, portable, hand held, high-throughput, and integrated, is highly sought after.

The team of Dr. Mussivand at the Medical Devices Innovation Institute (MDI²) in Ottawa, Canada, is developing a novel device to this end, integrating various elements to obtain such a system. Helicase protein, which has a natural affinity for DNA, is being explored as a candidate protein in this method of DNA sequence detection which is being developed by Dr. Mussivand's team. This novel device will be capable of sequencing entire genomes, as well as targeted sequences. A mechanism to target the desired segments of a genome is very important for the fields of personalized medicine, forensics, security and many other applications. This project is a part of the system being developed at MDI² and focuses on a potential means of attracting specific fragments of genomic or mitochondrial DNA for nucleotide sequencing.

1.3 Hypothesis

DNA is stored in cells in the form of double stranded DNA (dsDNA). In order to determine the sequence of a strand of DNA nucleotides, there is a need to separate the strands of DNA into single stranded DNA (ssDNA) and to feed this ssDNA into the sequencing device. Current technologies lack the capacity to do this. Thus, the aim of this study was to investigate the use of helicase, a molecule which naturally attracts and unwinds dsDNA and translocates along ssDNA, towards a novel DNA sequencing device. Bacteriophage T7 gene product 4 helicase was identified as a candidate protein to this end because it is well studied, is highly processive on DNA, and is small in size.

Hypothesis 1: Bacteriophage T7 gene product 4 helicase can be purified from a bacterial host for potential use in a novel DNA sequencing method.

Hypothesis 2: A recombinant protein consisting of avidin, a protein with a naturally high affinity for the molecule biotin, fused with helicase protein can be solubly purified for potential use as a means of targeting the bacteriophage T7 helicase to the specific DNA sequences.

1.4 Specific Aims

In order to address the hypotheses, specific aims were outlined as a contribution towards the final aim. The specific aims of this research undertaking were as follows:

1. To purify bacteriophage T7 gene product 4 DNA helicase from *E. coli*
2. To purify the helicase-avidin recombinant protein from *E. coli*
3. To test the functionality of the purified helicase enzymes
4. To test the functionality of avidin in the recombinant protein

1.5 DNA Sequencing Overview

DNA sequencing is the process by which the order of nucleotides in a DNA molecule is identified. It is a vital tool used in various applications, from research science in genetics, genomics and biochemistry, to forensics or to point of care applications such as personalized medicine. Currently, there are dozens of proposed and commercialized DNA sequencing approaches and methods. While many are similar, primarily revolving around the use of the DNA polymerase enzyme, many are very different and innovative. These advanced methods, however, are slow, complex, require a large amount of space, multiple preparatory steps as well as specialized technicians. A rapid, portable, small, inexpensive, accurate, and fully automated all-in-one device has yet to be made commercially available. These factors collectively do not seem to be the focus of research of most groups who tend to focus on a few of the aforementioned issues. A novel method which aims to encompass all of these features will fill the gap in the current market of commercially available DNA sequencing devices.

The objective of this chapter is to present the various current and emerging DNA sequencing technologies, including those which are less well-known or publicized. It is important to present those that are well-known, because of the clout given to them by the greater scientific community, but it is equally as important to present the lesser-known proposed methods in order to have a complete literature review to justify the necessity of a proposed method.

While there are currently dozens of approaches for sequencing DNA, recently the development of novel methods has greatly increased. The first DNA sequencing method, known as “Sanger Sequencing” was published in 1977 [1]. It was only in 1996 that the “next-generation” method, commonly known as pyrosequencing was introduced, being commercialized only in 2005 [2].

DNA sequencing approaches may generally be placed under several categories: chain termination, sequencing by synthesis (cycle extension and polymerase approaches), non-polymerase enzymatic approaches (exonuclease, ligase and helicase-

based) and non-enzymatic approaches (including DNA nanopore and atomic force microscope-carbon nanotube).

1.5.1 Chain Termination

In 1977, the first two methods for sequencing DNA were introduced [1, 3]. These two methods both involved creating strands of DNA of various terminated chain lengths, with a knowledge of what the final nucleotide in the chain is, and using that information to determine the DNA sequence. For this reason, in this text, the two are classified under “Chain Termination”.

Maxam and Gilbert Sequencing

Maxam and Gilbert first developed a method of sequencing [3] in which chemicals break strands of DNA specifically at each of the four bases, Adenine (A), Guanine (G), Cytosine (C) or Thymine (T), or a combination of two of them. The target DNA is radioactively labeled at one end, and undergoes a partial reaction that chemically breaks off the DNA at different locations on the strand (Figure 1.1A). These fragmented reaction mixtures are then loaded into separate lanes on a polyacrylamide gel electrophoresis, separating the fragments by size, and are visualized by autoradiography from which the DNA sequence can be determined. This method was laborious and the sequence was not easy to decipher [1].

Sanger Sequencing

Later that year, Sanger proposed his method of DNA sequencing using DNA polymerase instead of chemical agents [1]. He proposed 4 reaction mixtures containing all four deoxyribonucleotides (dNTP), but with a small amount of radioactively labeled double deoxyribonucleotide (ddNTP) to inhibit the reaction at specific bases randomly. This would generate chains that terminated at different lengths, with the last nucleotide being a known value. Running the four reactions on a gel electrophoresis to separate them by size, the DNA sequence could be directly read from the visualized gel (Figure 1.1B).

Other variations of the radioactive Sanger sequencing involve radiolabeling the DNA primer rather than the ddNTPs. Current variations of the Sanger method use ddNTPs labeled with their own unique fluorescent label [4]. Capillary electrophoresis

then separates the DNA fragments by length [5]. A similar method was proposed in 2006 using an RNA polymerase instead of DNA polymerase [6].

1.5.2 Sequencing-by-Synthesis

Sequencing by synthesis approaches seek to determine the DNA sequence as it is being synthesized by a polymerase enzyme. This can be achieved by cycle extensions or with real time polymerase reading.

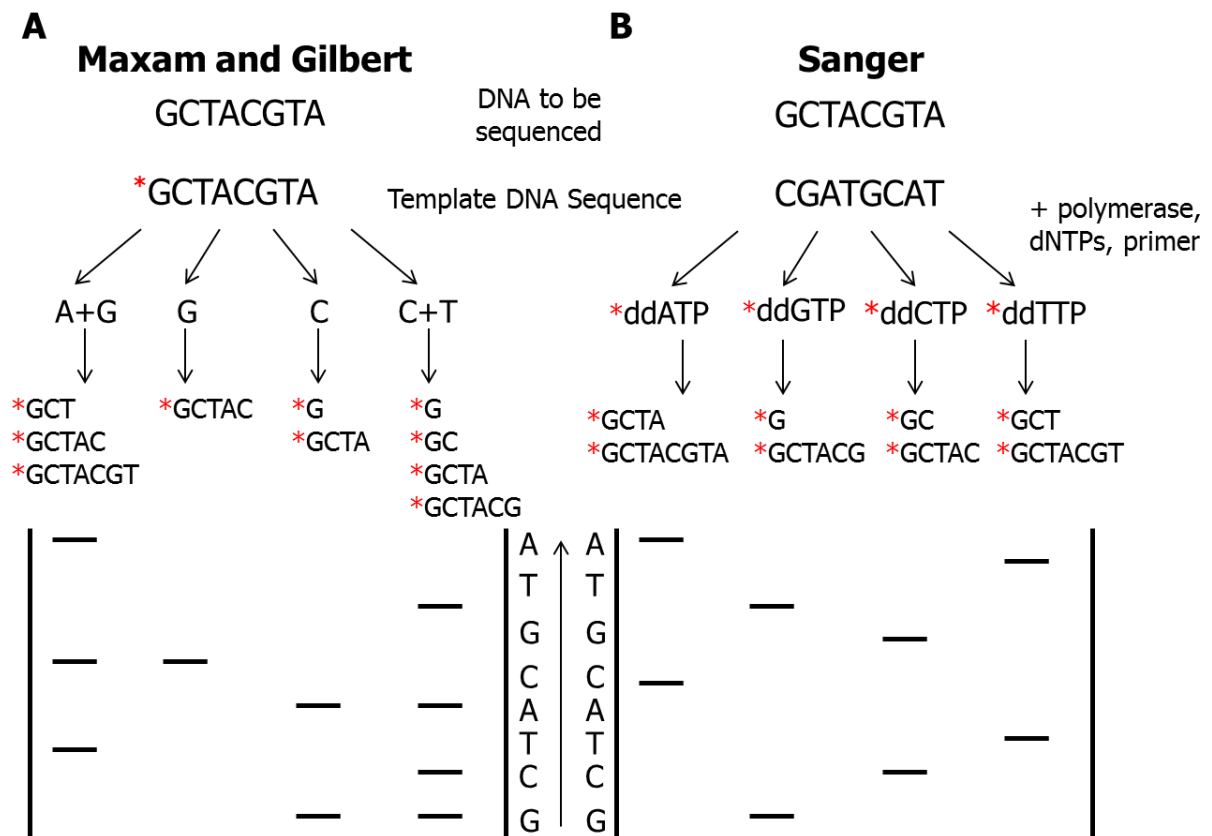


Figure 1.1: Chain terminating DNA sequencing. A) Maxam and Gilbert DNA sequencing method. DNA radioactively labeled at one end undergo partial reactions of chemical DNA cleavage specific nucleotides (A = Adenine, G = Guanine, C = Cytosine, T = Thymine) leaving fragments of different lengths. The reactions are separated by size on a polyacrylamide gel and the sequence can therefore be read. B) Sanger Sequencing Method. ddNTPs are used in small amounts in a polymerase reaction with radiolabeled primers (or ddNTPs, not shown) instead of chemical cleavage of the DNA strand, creating fragments of DNA that end in specific nucleotides. Four reactions, each with a different ddNTP, are run on a gel electrophoresis and the sequence can be directly read.

Cycle Extension Approaches

Instead of running four separate reaction mixtures to determine the DNA sequence, cycle extension approaches cycle each of the four bases within one reaction,

determining which has incorporated into the template strand sequentially [7]. For each base incorporation, 8 steps (4 base interrogation and 4 washings) are performed. This makes these methods inherently slow, however they have provided better accuracy and higher read-lengths compared to the chain termination approaches. The methods discussed in this review will be Pyrosequencing and similar methods, Solexa, HeliScope™, and ion sequencing.

Pyrosequencing

The first “next-generation” DNA sequencing method, known as “pyrosequencing” was commercialized in 2005 by 454 Life Sciences, as the Genome Sequencer FLX [2]. The method was first proposed in 1985 as a general tool for enzyme kinetic studies, and it is based on a series of enzymatic reactions that convert the pyrophosphate released by DNA polymerase, or other enzymes, upon incorporation of a nucleotide into a fluorescent signal (Figure 1.2) [8]. It was applied to the concept of DNA sequencing in 1996 [9] and commercialized only in 2005 [2]. Reaction mixtures carrying each of the four nucleotides are streamed through the reaction. When DNA polymerase incorporates a base, pyrophosphate (PP_i) is released, allowing the enzyme sulphurylase to create ATP, which fuels luciferase, thus generating a light signal. Once the signal is generated, the enzyme apyrase is streamed through to degrade any unincorporated dNTPs and leftover ATP, and the next cycle takes place. Later on, the method was altered to immobilize the DNA and repeat the dNTP extension cycle many times [9]. Since ATP fuels the luciferase enzyme in the cascade to produce the light signal, this modified version uses an analogue of ATP (dATP- α S) which can be incorporated by DNA polymerase but is not recognized by luciferase enzyme. The commercialized model by 454 Life Sciences immobilizes the luciferase and ATP sulphurylase enzymes, leaving the other reagents to flow through [4]. A patent also exists in which luciferase is replaced with peroxidase enzyme which produced hydrogen peroxide which oxidizes an electrode to generate an electrical signal upon base incorporation [10].

Illumina-Solexa Sequencing

The Illumina-Solexa method [2, 11], released in 2006, is somewhat a crossover between chain terminating methods and cycle extension. This method makes use of reversible terminator nucleotides for each of the four bases. Each reversible terminator nucleotide has a patented cleavable fluorescent cap on the 3'OH unique for each nucleotide [12]. In order for the fluorescent light to be intense enough to be detectable by the CCD camera, the target DNA is clonally amplified in a bead using PCR to create “colonies” of the fragment. The polymerase and reversible terminator nucleotides are then washed over the colonies, the CCD camera detects which terminator has been added to the DNA chain, a wash is streamed over the reaction to cleave the terminator group at the 3'OH of the nucleotide, and the cycle continues.

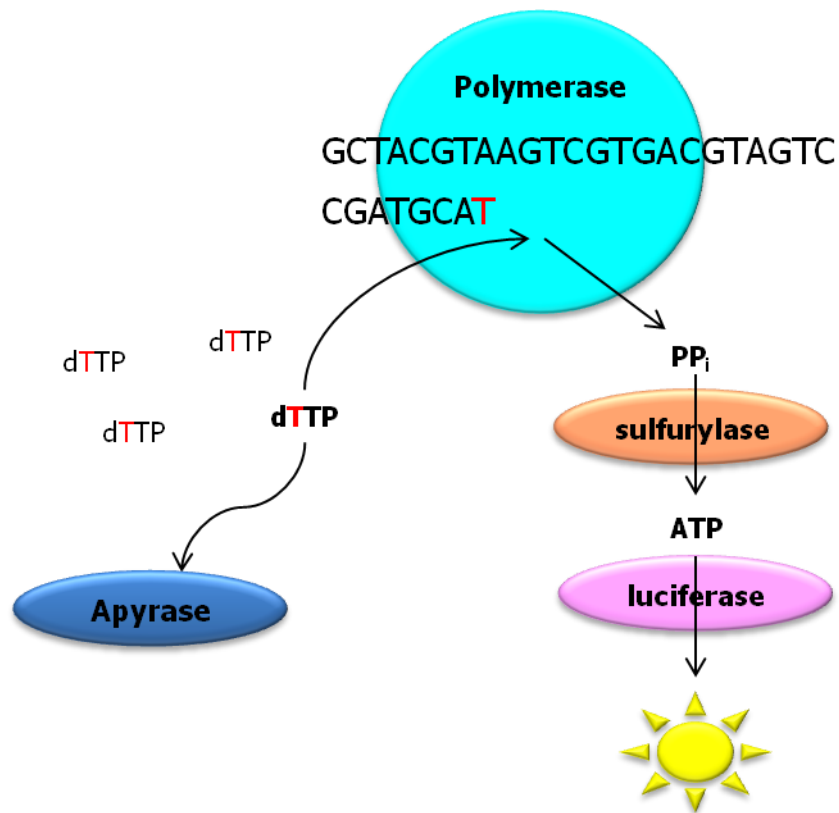


Figure 1.2: Pyrosequencing. The next generation sequencing technique, Pyrosequencing, involves the chain reaction detection of the pyrophosphate molecule released during the incorporation of a nucleotide to a DNA strand using a polymerase. The released pyrophosphate is converted by sulfurylase to ATP, which then fuels luciferase to produce light. Apyrase degrades the remaining unincorporated nucleotides, and the subsequent nucleotide is washed over.

Helicos HeliScope™ Sequencer

The HeliScope™ DNA sequencing device was released in 2007 [2]. In this method, the DNA is cut into fragments of 100-200 nucleotides. A polyA tail is added to the fragments in order to hybridize them to polyT templates on a flowcell. A charge-coupled device (CCD) camera records their locations. A reaction mixture containing DNA polymerase and a cleavable fluorescently labeled nucleotide is washed over, and the CCD camera records base incorporation. A wash is then streamed over to remove the polymerase and cleave the fluorescent label from the nucleotide, and the next cycle begins. This allows for simultaneous strand extension, which is asynchronous, unlike the other methods which run in parallel.

Ion Sequencing by Ion Torrent™ Systems

As with many new “third generation” methods, Ion Sequencing by Ion Torrent™ Systems avoids the use of optics that require specialized expensive reagents, such as modified nucleotides and enzymes [13]. As with other cycle extension approaches, each of the four nucleotides is cycled over the sample. However, when the DNA polymerase incorporates a nucleotide, one hydrogen atom (H^+) is released. This shifts the pH of the solution in the well containing the reaction, which in turn generates an electrical signal that is gathered by a semiconductor and processed [14]. This method does not require any modified nucleotides and is quick. A patent for this method describes two embodiments: one where pH changes are observed, and another where a heat sensor detects the heat of the polymerase reaction [15].

1.5.3 Real Time Polymerase Reading Approaches

Real time polymerase reading sequencing by synthesis approaches, as their title indicates, use DNA polymerase to monitor DNA sequence elongation in real time, at the sensitivity of single nucleotides [7]. Examples include SMRT™ and Visigen’s approach.

Single Molecule Real Time SMRT™ sequencing

The recently commercialized sequencer by Pacific Biosciences is based on the Zero-Mode Waveguide (ZMW) [2]. The concept involves nanometer scale apertures upon which a polymerase enzyme, which incorporates special gamma-labeled nucleotides, is immobilized. This chamber is placed on a metal film on a clear substrate.

Light shines on the opening, illuminating 20 zeptolitres, allowing for the measurement of fluorescence at the single molecule level as the labeled PP_i is released (Figure 1.3). The read lengths are high (1500+ bases), as polymerases are processive, and 3000 ZMW can be run in parallel.

Visigen Biotechnologies' Approach

Visigen biotechnologies' approach is very similar to the SMRT™ approach, however, this uses Fluorescent Resonance Energy Transfer (FRET) for nucleotide detection [2]. A DNA polymerase enzyme is engineered to contain a FRET donor close to the active site. The nucleotides are each modified with a unique FRET acceptor on the gamma phosphate. Upon nucleotide incorporation, the FRET acceptor and donor come in close proximity, creating a signal. Each nucleotide incorporated produces a different frequency, creating a unique DNA sequence signature. A patent exists for this method, also claiming to determine the sequence of DNA by measuring conformational changes in the polymerase enzyme as it binds different nucleotides [16].

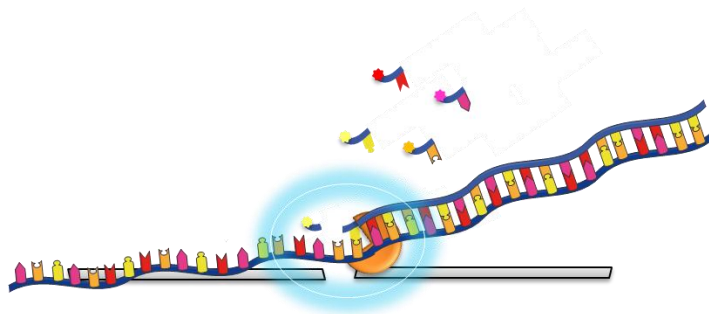


Figure 1.3: Single Molecule Real Time™ sequencing. This method involves a nanometer-scaled aperture with an immobilized polymerase protein attached (orange) that creates a zeptolitre sized detection range of fluorescence (blue halo). Specialized oligonucleotides, each with a different fluorophore on the gamma phosphate are incubated. As a nucleotide is incorporated by the polymerase, the fluorophore is released with the pyrophosphate and is sensed by the optical machinery. DNA images adapted from Servier Medical Art.

1.5.4 Non-polymerase Enzymatic Approaches

Since so many of the DNA sequencing approaches, as listed above and more, use polymerases as the basis of the sequencing reaction, many groups are looking towards other enzymes that naturally recognize and interact with nucleic acids. Exonucleases

are enzymes that cleave DNA, ligases are enzymes which ligate segments of DNA together in a specific manner, and helicases naturally unwind the DNA double helix.

Sequencing by Ligation

“Massively Parallel” Sequencing by ligation is a complex and slow means of sequencing by synthesis [17] commercialized by Applied Biosystems SOLiDTM. This method aims to increase the accuracy of DNA sequencing methods for the purpose of re-sequencing known regions of DNA to determine small variations among individuals [18]. The method involves a DNA template, sheared to 136 bp fragments which are clonally PCR amplified, attached to magnetic beads and set on an array for analysis [17, 18]. The targeted DNA sequences is then “mate paired” to this “shotgun library” [18] and ligase enzyme ligates correctly matches pairs. Optically labeled DNA octomers are interrogated by 4 channel imaging to determine matches, cleaving the last 2-3 base pairs at the end of each octamer each cycle, removing the fluorophore [17]. While providing an increase in accuracy, this method is time consuming, requires strong knowledge of the targeted DNA sequence, is reagent heavy, optically based and has very Exonuclease Sequencing

These methods use an exonuclease enzyme to remove fluorescently labeled nucleic acids from DNA in order of the sequence [7]. Exonucleases are enzymes that cleave nucleotides from the end of a strand of DNA, thus making the strand shorter by one nucleotide with each cleavage. They can do this, depending on the exonuclease, from the 5’ end or the 3’ end of the DNA strand. Once the labeled nucleic acid is released, it is detected in the flow stream by CCD cameras. A recent patent exists for this method [19].

Helicase-based Methods

Two patents were found using helicase for DNA sequencing. The first, an optical method, works similarly to exonuclease sequencing [20]. Briefly, fluorescently labeled nucleotides are hybridized to the target DNA. As the helicase unwinds the DNA, the fluorescently labeled nucleotide is detached from its complementary strand and is detected. The second patent describes the measurement of the radiation emitted by the interaction between helicase and the associated single stranded DNA (ssDNA) [21]. The

helicase is immobilized on a sensor chip. Modified ATP nucleotides with photocleavable blocking groups are used to fuel the helicase motor, making its behaviour controlled. Once one nucleotide is unwound and detected based on conformational changes detected by polarised light, light with wavelength of 260nm is pulsed to cleave the blocker and allow the reaction to detect the next base.

1.5.5 Non-enzymatic Sequencing Approaches

Efforts to steer clear of polymerase-based sequencing have led other groups to explore the idea of entirely synthetic DNA sequencing methods. These include Nanowire, Nanopore, and Atomic Force Microscope-Carbon Nanotube based methods.

Nanowire Sequencing

The silicon nanowire (siNA) approach involves an array of siNA hybridized with peptide nucleic acid capture probes [22]. When these probes bind the complementary target DNA, altered resistance due to a change of charge density at the surface of the siNA are sensed. This is a label free method.

Nanopore Methods

The concept of the nanopore was first developed at Harvard University: the movement of DNA through a nanometer-scale pore by applying electric fields [2, 4]. The original method involved the protein α -hemolysin Protein embedded in a phospholipid bilayer membrane [23]. The patented concept involves detecting changes in ionic current through the nanopore as DNA flows through it, each base providing a different signal [24]. This concept with a protein in a membrane inherently had background noise and was unstable, therefore the solid-state synthetic pore was developed. Many variations of synthetic pores have been considered including holes drilled in silicon dioxide films, ion-beam sculpted, conical nanopores or pores etched in glass [2, 4], and several have also been patented [25]. However, while the α -hemolysin pore was neglected, due to issues with the creation of an appropriately sized synthetic nanopore, another protein pore, Mycobacterium smegmatis porin A (MspA) is being considered [5].

Another patented nanopore variation uses a graphene (single atom thick carbon sheet) nanogap as the semiconductor for nanopore sequencing [26]. When graphene is

rolled, a carbon nanotube is formed. Another patent exists for a nanogap between two carbon nanotube electrodes, passing the DNA through using optical tweezers [2, 27].

Nanopore methods are limited in that they cannot resolve single bases [2]. However, controlling and lowering the speed of movement of the DNA through the pore, by immobilizing it on a bead, can help resolve this issue. It would also allow for repeated measurements by moving the DNA back and forth through the pore.

Atomic Force Microscopy – Carbon Nanotube DNA sequencing

This method is currently under development and is in the proof of concept stage. Atomic Force Microscopy (AFM) uses a thin-tipped probe to directly record high resolution images of molecules [28]. A single-walled carbon nanotube (SWCNT) is grown at the tip of an AFM probe. The tip of the SWCNT is functionalized and a nucleotide is attached. This nucleotide is used to scan and bind complementary nucleotides on an immobilized target DNA strand. The location of the complementary nucleotide is measured. In the future, several probes with different nucleotides attached will be able to scan and sequence DNA.

1.5.6 Areas Requiring Attention

Polymerase Chain Reaction

Many of the systems discussed in the literature review, with the exception of a few of the “third-generation” sequencers such as SMRT™, require prior amplification of the target DNA to create a large amount of identical copies of the DNA molecule of interest. This process takes several hours. In addition, many of the methods require the emulsion PCR amplification step of DNA fragments within the sequencer, to make the light signal strong enough for reliable base detection by the CCD cameras. PCR amplification has revolutionised DNA analysis, but in some instances it may introduce base sequence errors into the copied DNA strands, or favour certain sequences over others, thus changing the relative frequency and abundance of various DNA fragments that existed before amplification

Time

Current DNA sequencing approaches require time consuming preprocessing steps, such as purification and PCR which add several hours to the process of obtaining DNA sequence information. Many of the current approaches to sequence DNA, such as the cycle extension approaches, are inherently time-consuming in their method due to the successive cycling of reagents. Many methods tether the DNA sample onto a bead in order to immobilize and be able to control their movement. This is another inefficiency in these methods, and increases time. A new method should seek to eliminate such inefficiencies, aiming for faster, yet still accurate, results.

Cost

While some of the new “third-generation” approaches have solved a few of the above mentioned issues, they still remain expensive. The requirement for modified nucleotides, specialized reagents, and the use of imaging equipment all add to the overall expense of the sequencing reaction. After the commercial mark-up for the company to make a profit, these methods are not nearly as inexpensive as a DNA sequencing reaction and machine could be.

Size of Device

Most methods require imaging detection systems. These systems inherently require space to set up, with the exception of the Zero-Mode Waveguide technology, which has been patented and thus cannot be used by other emerging technologies as a method of scaling down machine sizes that are currently very large (several meters in length and height). These technologies also require extensive computing abilities, which can take up much space.

Portability

Size and portability are very different things. Many of the current and emerging method require delicate imaging equipment. This imaging technology, while it has the potential to be scaled down as with the Zero-Mode Waveguide technology, it is less possible for it to be made portable. Non-optical methods solve this issue as they are sturdy like a computer, and can be transported with little risk of damage. This is very

important for the use of DNA sequencing in field sciences, and is appealing for point of care use as well.

Read Lengths

An issue with several of these systems is the ability to only read short stretches of DNA. Shorter read lengths are limiting because one cannot sequence most target DNA samples at once. For whole genome sequencing, shorter read lengths make the puzzle of overlapping sequences more difficult to put together. Several non-enzymatic methods based on hybridization to nucleotide tiling array microchips have read lengths that are so small (<50 bases) with little opportunity for improvement. As such, these techniques were not considered as part of the state of the art for inclusion in this review.

All-in-one Device

An automated all-in-one device that not only obtains DNA sequence information, but also prepares the sample for sequencing, is very appealing for all applications of DNA sequencing. Most of the methods simply focus on the DNA sequencing aspect ignoring prior steps that may be required. Some sources allude to this concept, however, developing an all-in-one device does not seem to be a focus for the major undertakings currently.

1.5.7 Proposed Method and Device Complementing this Thesis

Two helicase-based methods of DNA sequencing exist, and several make use of the SWCNT, however, none combine these two concepts. Helicase is a naturally occurring enzyme that attracts DNA, passes it through its core, and is capable of unwinding the double helix. Currently, the two methods which use helicase in DNA sequencing are optically based. SWCNT, as used in AFM, are highly conductive, and are therefore used as nanopores and nanogaps. They are also highly sensitive to small charge differences, like those differences between DNA nucleotides.

The properties of SWCNT used by the AFM approach, combined with helicase as a nano-biosensor can resolve many current issues such as cost, speed, and portability. The proposed method uses a helicase as a pore through which the DNA molecule is threaded. Unlike the Nanopore™ method, the base sequence will not be determined

based on the change in ionic current flowing through the nanopore due to the presence of nucleotides. Rather, as the helicase, which will be attached to the tip of a SWCNT, translocates along the DNA, the nucleotides will be sensed by a highly sensitive single-electron transistor (SET) sensing circuitry attached to the other end of the SWCNT (Figure 1.4). This method could take place as an array, and would not require synchronicity of parallel helicase detections, would not require any specially labeled nucleotides or reagents, could have the DNA and reagents flow in the flow-cell, and does not require costly imaging equipment.

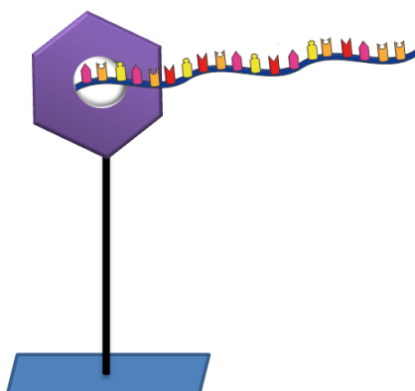


Figure 1.4: Cartoon representation of the Medical Devices Innovation Institute proposed novel DNA sequencing method. A single walled carbon nanotube (black line) is attached at one end to a helicase enzyme (purple hexagon), and at the other end to a single-electron transistor sensing circuitry (blue rectangle). As a strand of DNA passes through the central core of the helicase, the differential charge of each nucleotide will be passed through the nanotube to the electronic sensing circuitry. DNA images adapted from Servier Medical Art.

1.5.8 DNA Sequencing Overview Summary

DNA sequence information is required for scientific research (genetics, genomics, biochemistry, field sciences, ancient DNA, etc), forensic identification, and personalized medicine. Many researchers are inventing and developing new methods to determine the sequence of DNA either through alterations and modifications of existing polymerase-based approaches, or looking towards completely innovative, sometimes enzyme-free and synthetic, approaches. However, the market is lacking a device that is portable, small, rapid, inexpensive, accurate, and fully automated. The proposed helicase-carbon nanotube biosensor could resolve these issues and revolutionize the world of DNA sequencing.

1.6 Prior Related Research

The inception of this project related to DNA sequencing is a result and continuation of an ongoing DNA related venture at the Medical Devices Innovation Institute. The proposed novel DNA sequencing method is the final step towards an integrated all-in-one device in which the cell sample is input, and the DNA sequence data is provided.

The integrated device would consist of a portion that samples the cells, lyses and extracts DNA from the cells and finally provides the DNA sequence. A method and apparatus for collecting cells for macromolecular analysis [29] was developed in 2004 in which sufficient DNA was extracted for genetic analysis in less than 15 minutes from samples including fingerprints, nail clippings, hair, saliva and others.

The next set of results pertaining to the DNA related project involved a device for the electroporation and lysis of cells [30] in which cells could be lysed for DNA extraction in 7 ± 2 seconds.

The final stage towards attaining this novel integrated device is DNA sequencing. The present work is the first contribution towards this device involving a helicase enzyme.

1.7 Proposed Research

1.7.1 Objective and scope

The objective of this research project was to determine whether a functional bacteriophage T7 gene product 4 DNA helicase can be expressed in a bacterial host for potential use in a novel DNA sequencing device. A means of attracting specific sequences of DNA to the helicase molecule was also investigated by exploiting the strong natural affinity of the protein avidin to the small molecule biotin as outlined in this section. By engineering a recombinant protein consisting of helicase and avidin, and biotinylated oligonucleotides, specific sequences of DNA to be identified could be attractive to the helicase enzyme in this novel DNA sequencing method being investigated.

1.7.2 Helicase Protein

Helicases are enzymes that naturally bind and translocate along DNA. Bacteriophage T7 gene product 4 (T7gp4) DNA helicase is a hexameric helicase that threads single stranded DNA (ssDNA) through its central core unidirectionally in a highly processive manner. This property could be exploited in a novel method for DNA sequencing.



Figure 1.5: Line representation of fusion proteins. A) Bacteriophage T7 gene product 4 is naturally expressed as a helicase-primase fusion, with the primase sequence (orange) on the N-terminus, linked to helicase (purple) on the c-terminus via a linker region (green). B) The engineered helicase-avidin fusion protein replaces the N-terminal primase with the a monomeric avidin (cyan) sequence on the N-terminus.

T7gp4 is naturally expressed as a helicase-primase fusion protein [31] with primase, an enzyme that lays down the RNA primer for DNA replication, on the amino-terminus of the protein and helicase on the carboxyl terminus (Figure 1.5A). The two proteins are linked by a region of linker DNA, and are fully functional when truncated [32, 33]. The helicase portion is only functional when the linker region is included in the truncated form [33].

1.7.3 Exploitation of Avidin-Biotin Relationship for Helicase-based Sequencing

Avidin is a protein that can be fused with other proteins to create recombinant proteins [34]. In this instance, avidin was fused with a helicase monomer to create a helicase-avidin fusion protein (Figure 1.6). The avidin protein was placed on the N-terminal end of the truncated T7 helicase sequence (Figure 1.5A). This sequence includes helicase's amino acid linker region as well as an additional set of linker amino acids where the primase protein of bacteriophage T7gp4 DNA helicase has been excised (Figure 1.5B). The 36.5 kDa truncated helicase, along with the additional linker amino acids and added histidine tag for purification, yields a 51.7 kDa protein. Because the

natural form of T7pg4 is a fusion protein, swapping the primase portion at the N-terminus with avidin should yield a functional artificial recombinant protein.



Figure 1.6: Schematic representation of helicase-avidin fusion protein. Helicase (purple) is fused with avidin (blue) by means of a short sequence of amino acids as a linker region (green).

Avidin protein has a high affinity for biotin. Biotin is a very small molecule that can be annealed to one end of a DNA molecule in a process known as “biotinylation” (Figure 1.7A). Specific sequences complementary to the sequence of interest could be biotinylated in order to be drawn towards the Avidin protein.

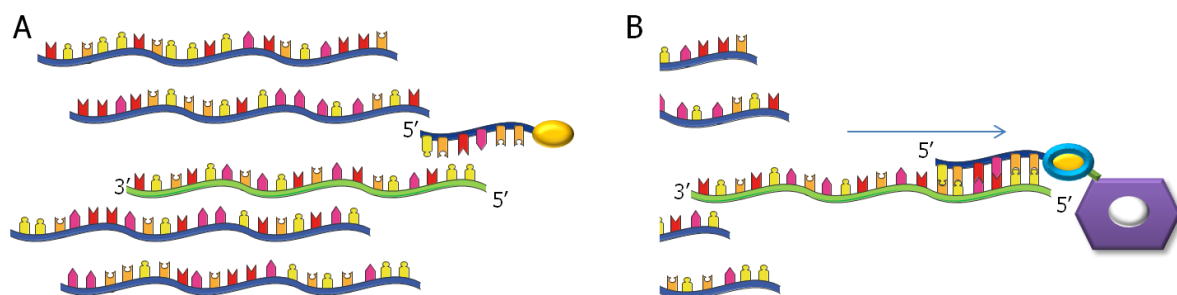


Figure 1.7: Attracting specific sequences of DNA to helicase. A) Natural action of biotinylation. A DNA sequence of interest is identified by its specific sequence of nucleic acids (green). A short strand of ssDNA complementary to the beginning of the desired stretch of ssDNA would be biotinylated (yellow). B) Exploitation of biotinylation. The short biotinylated ssDNA would anneal to the DNA strand of interest, and would be attracted to the avidin (cyan), bringing the DNA sequence of interest towards the helicase (purple) molecule for unwinding. DNA images adapted from Servier Medical Art.

The mechanism by which this would ultimately work in the novel device is described in Figure 1.6. Briefly, the DNA sequence of interest would be present within the whole genomic sequence pool of single stranded DNA. A short 3' biotinylated strand of DNA, complementary to the sequence of interest, would be introduced into the medium, and would anneal to the sequence of interest. After temperature-dependent

annealing occurs, a solution containing the preassembled helicase-avidin fusion proteins would be introduced to the area, thus attracting the biotin quickly. The helicase protein will thread the ssDNA through its central core in a 5'→3' directional manner, appropriate for DNA sequencing.

The short biotinylated strand of DNA would serve a similar function to primers used in PCR or polymerase based sequencing approaches, and could facilitate the creation of an artificial fork of DNA for recognition by helicase (Figure 1.8).

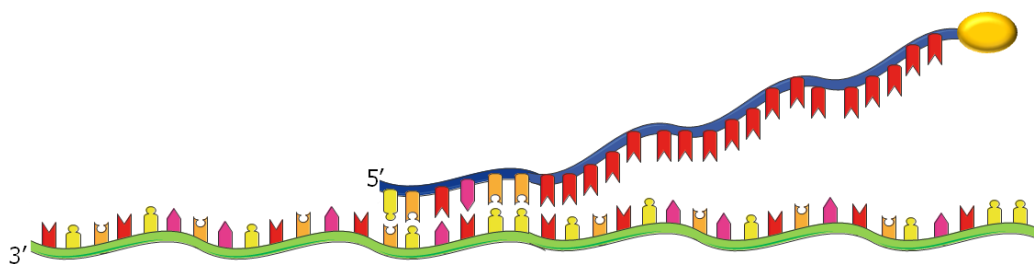


Figure 1.8: Artificial DNA fork. A short biotinylated strand of ssDNA, which is partly complementary to the DNA sequence of interest, could create an artificial DNA fork. An extension of the sequence containing only one nucleotide, for example, could help in the creation of the artificial fork as they would not interfere with appropriate annealing of the targeted ssDNA. DNA images adapted from Servier Medical Art.

Additionally, it has been shown that some hexameric helicases, similar in structure to T7 helicase, are capable of displacing streptavidin, a molecule homologous to avidin, from 3' biotinylated oligonucleotides [35]. If this were to be the case with avidin and T7 helicase, a potential issue of a loop forming when the DNA is attracted to the helicase would be avoided. Furthermore, the 3' tail in the fork required by T7 helicase for binding can be replaced with a biotin-streptavidin complex and the helicase can still bind and translocate along [36]. These two pieces of information further support the possibility of success in the proposed research undertaking.

1.7.4 Experimental Summary

To this end, T7 helicase and helicase-avidin fusion proteins were overexpressed for purification. Overexpression was induced in a bacterial host due to relative ease, cost-effectiveness, and large-scale industrial reproducibility, making it ideal for a potential commercialization venture. The enzymes were purified by means of a

histidine tag consisting of several histidine amino acid residues “tagged” at the end of the protein sequence, which is attracted to cobalt and nickel. This creates an easy and inexpensive means for purification, relative to the use of specific antibodies and other complex biochemical purification techniques. The purified proteins were characterized for function by electrophoretic mobility shift assays, to assess the DNA binding ability of the purified helicase, and biotin binding assays to assess the interaction between the avidin-helicase fusion protein and biotin.

1.7.5 Some Relevant Terms and Definitions Used in this Thesis

Helicase: truncated helicase portion of bacteriophage T7gp4 helicase-primase; possesses a natural affinity for DNA; unwinds dsDNA and translocates along ssDNA.

Biotin: small molecule with a strong affinity for the protein avidin; easily added as a modification of strands of DNA.

Avidin: protein with a strong affinity for the molecule biotin.

Oligonucleotide: a short single stranded polymer of DNA nucleotides.

Complementary DNA: DNA strand whose nucleotide sequence is complementary to the nucleotide sequence of another strand of DNA. In this thesis, complementary DNA is single stranded.

Fusion protein: the combination of two proteins linked together by peptide bonds and expressed from a single open-reading frame to generate a single multi-domain protein.

1.8 Expected results

Due to codon optimization of the DNA sequence, the helicase protein is expected to be expressed properly in *E. coli*. The engineered recombinant helicase-avidin fusion protein should also be found in the soluble fraction for this same reason, and due to helicase's natural state as a fusion protein, with the avidin being placed in the primase's location.

Functional activity will be verified in two key ways. Electrophoretic mobility shift assays should show a distinct band present representing helicase bound to a radioactive DNA probe above the band representing excess radioactively labeled DNA probe on a native polyacrylamide gel electrophoresis. The avidin assay should show that avidin is capable of binding to a biotinylated resin, whereas a control protein does not bind to the solid support and is washed out.

1.9 Contribution and Significance

In order to use helicase molecules in the proposed device and method for DNA sequencing, the enzyme must be synthesized and characterized for functionality; thus, the results from this thesis will be imperative for achieving this goal. The present research study aimed to explore a potential means of attracting specific segments of DNA to the helicase molecule by way of avidin-biotin attraction. A helicase-avidin fusion protein was engineered, synthesized, and isolated for study. These two engineered proteins, if purified as soluble proteins that are functional, could be integrated into a future helicase-based DNA sequencing method.

2 – Literature Review

This research undertaking requires the understanding of several components. Firstly, the proposed proof of concept method is centred on the action of the helicase enzyme, which naturally unwinds and threads DNA strands. Secondly, in order to be able to attract specific DNA sequences to the helicase enzyme, the helicase enzyme was covalently attached to the protein avidin, which has a naturally high affinity to the molecule biotin. Therefore, in this section, a literature on helicase enzymes (including T7gp4 DNA helicase), biotin and avidin, fusion proteins and DNA is reviewed.

2.1 Helicases

2.1.1 Significance of Helicase

The majority of DNA sequencing methods that make use of natural enzymes, as described previously (section 1.5), are DNA polymerase-based. The concept of using helicase is another way of similarly mimicking nature by use of an enzyme that naturally interacts with DNA. DNA helicase is capable of separating dsDNA into ssDNA as it translocates along a single strand of DNA. In theory, this property could be used to identify the nucleotides on a DNA strand as helicase translocates along it. This mimicry of nature, combined with electronic circuitry and other technology, could be used to obtain DNA sequence information.

The section below is designed to give an overview of helicases in general, as well as further detail on the chosen helicase for the present study. The choice of helicase protein was to fulfill a variety of criteria, including small size, desirable for an electronic helicase-based sequencing method, as is being explored by the Medical Devices Innovation Institute, to reduce noise, high processivity (translates to longer read lengths), ample functional characterization, and the ability to function on its own with as few as possible other proteins/reagents.

2.1.2 Helicases in General

Helicases are vital proteins to all organisms including viruses [37], bacteria and eukaryotes, including plants [38]. They are motor proteins that dissociate double stranded nucleic acid chains to single stranded nucleic acid chains. DNA helicases are motor proteins that unwind dsDNA to ssDNA through nucleoside triphosphate (NTP)

hydrolysis [39]. Other types of helicases also exist that unwind RNA (such as the Hepatitis C virus NS3 [40] or the human II/Gu [41]), both RNA and DNA (such as Nuclear DNA Helicase II [42] or SV40 large T antigen [43]), or RNA-DNA duplexes (such as Pif1p DNA helicase [44] or *E. coli* rho [39]).

Helicases remove the secondary structures in free ssDNA or RNA to allow for further DNA and RNA metabolic processing [39, 45]. They are involved in many cellular processes, including DNA replication [46], recombination [47], transcription termination [37], DNA repair [39], transcription, RNA splicing [48, 49], and dissociation of proteins from DNA [37].

In vivo, helicases work as part of a protein complex [48, 50], coordinating proteins on both strands that are being unwound [48]. Interestingly, they have been shown to function alone *in vitro* [39, 51]. Understanding the complex mechanism by which helicases work involves understanding the coupled coordination of the various catalytic subunits of the molecule working together, including NTP hydrolysis, DNA binding, and DNA unwinding. NTP binding and hydrolysis is a requirement for all DNA helicases [52].

2.1.3 Descriptors for Helicases

In order to understand the mechanisms of helicases, a familiarization with a few terms is advisable. Helicases use nucleic acids in two capacities: as a source of energy to fuel the NTPase motor, and as the substrate of DNA that the helicase translocates along. Throughout this document, the terms DNA or ssDNA refer to the strand of nucleotides threaded or unwound by the helicase, and NTP (such as ATP or dTTP) or nucleotide refer to the nucleotide factor that fuels the NTPase reaction.

The activities to describe helicases are as follows [53]:

- *Rate*: the number of unwound or translocated base pairs, or NTP hydrolyzed per unit of time. This can be from a few to thousands of base pairs per second and may be controlled by the NTPase rate.
- *Processivity*: The number of unwound or translocated base pairs before the enzyme dissociates from the DNA substrate.

- *Directionality*: Movement along the bound strand, from 3'-5' or 5'-3'.
- *Step size*: The mechanical distance travelled, or kinetic distance (bases unwound/translocated) per catalytic cycle.

2.1.4 Different Classes of Helicases

Helicases can be classified or grouped by several means [39, 53]:

- *Directionality*: translocation along ssDNA from 5'->3' or 3'->5'.
- *DNA/RNA*: interact with DNA, RNA, or both.
- *Superfamilies and families*: based on primary amino acid sequence, there are six superfamilies and families identified in total.
- *Structure*: based on quaternary structure/oligomerization, there are two structure based classifications in total.

The two structural classifications of helicases are (a) monomeric/dimeric helicases, which clamp onto a single strand of DNA and translocate along it, sometimes referred to as “crab-like” helicases [48], and (b) hexameric helicases, in which ssDNA passes through a central hole formed by the ring shaped structure of the hexamer. The hexameric conformation is more stable as it interacts with the entire DNA molecule, thus increasing processivity along the ssDNA. Increased processivity along ssDNA could translate to longer read-lengths in the application of DNA sequencing, therefore hexameric helicases are of further interest for this work.

2.1.5 Hexameric Helicases

Hexameric DNA helicases are found in viruses, bacteria, and eukaryotes [10;42], and are often involved in DNA replication, transcription termination, and recombination [42]. They are most often ring-shaped homohexamers (6 identical monomers). The ring-conformation of hexameric helicases, such as *E. coli*'s dnaB and bacteriophage T7 product of gene 4 (T7gp4), allows ssDNA to pass through its central hole, increasing processivity [41].

Hexameric helicases are capable of three catalytic activities: DNA binding, helicase (DNA unwinding), and NTPase. The NTP involved to fuel the motor is most often ATP, however for bacteriophage T4 gene product 41 (T4gp41) it is GTP, and for

T7gp4 it is dTTP [41]. The NTP may also be promiscuous, as with SV40 large T antigen helicase which binds ATP to unwind dsDNA, and binds to other NTPs to unwind RNA [405].

The monomeric helicase unit in helicases varies in size from around 30 kDa for RepA helicase to 160 kDa for human Bloom's syndrome helicase [41]. A smaller helicase is desirable for this project as there is potentially less interference with the electronic signal, and the strand of DNA will be closer to the electronic circuitry.

Hexameric helicases have five conserved motifs [10]: H1, H1a, H2, H3, and H4. The first 4 domains (H1, H1a, H2 and H3) are located at the C-terminus of the helicase where the monomers interact, which is also the site of NTP binding and hydrolysis (Figure 2.1) [10]. H1 is a sequence conserved in many NTP binding proteins and ATPases. It contains the conserved Walker A sequence, which is responsible in many

2.1.5 Hexameric Helicases

Hexameric DNA helicases are found in viruses, bacteria, and eukaryotes [37, 48], and are often involved in DNA replication, transcription termination, and recombination [37]. They are most often ring-shaped homohexamers (6 identical monomers). The ring-conformation of hexameric helicases, such as *E. coli*'s dnaB and bacteriophage T7 product of gene 4 (T7gp4), allows ssDNA to pass through its central hole, increasing processivity [39].

Hexameric helicases are capable of three catalytic activities: DNA binding, helicase (DNA unwinding), and NTPase. The NTP involved to fuel the motor is most often ATP, however for bacteriophage T4 gene product 41 (T4gp41) it is GTP, and for T7gp4 it is dTTP [39]. The NTP may also be promiscuous, as with SV40 large T antigen helicase which binds ATP to unwind dsDNA, and binds to other NTPs to unwind RNA [43].

The monomeric helicase unit in helicases varies in size from around 30 kDa for RepA helicase to 160 kDa for human Bloom's syndrome helicase [39]. A smaller

helicase is desirable for this project as there is potentially less interference with the electronic signal, and the strand of DNA will be closer to the electronic circuitry.

Hexameric helicases have five conserved motifs [48]: H1, H1a, H2, H3, and H4. The first 4 domains (H1, H1a, H2 and H3) are located at the C-terminus of the helicase where the monomers interact, which is also the site of NTP binding and hydrolysis (Figure 2.1) [48]. H1 is a sequence conserved in many NTP binding proteins and ATPases. It contains the conserved Walker A sequence, which is responsible in many proteins for coordinating and stabilizing the charge of the nucleoside phosphate group [54]. The walker A sequence forms a P loop that interacts with the β phosphate of the NTP and the Mg^{2+} cofactor [39, 45]. The H1a conserved sequence is responsible for activating H_2O molecules in a nucleophile attack [39], as well as playing a role in NTP binding and hydrolysis [45]. The H2 motif contains the conserved walker B sequence involved in coordinating Mg^{2+} with nucleotide binding [39] so that the Mg^{2+} binds to the β - γ phosphates of the NTP [45]. The role of the H3 motif is not very clear [55], however, it is known that these residues also interact with the γ phosphate of the bound nucleotide [39]. The H4 domain is near the centre of the ring shape and is involved in DNA binding [45, 48, 56]. It contains an arginine finger, which stabilizes the base on the bound NTP [39], and has a loop in the central hole “loop II” that interacts with H3 in the adjacent subunit to coordinate DNA binding [55]. This loop requires basic residues for functional DNA binding activity [57, 58].

The C-terminal ends of hexameric helicases are comprised of RecA-like folds [59]. This is a conserved activity in ATPase proteins where the nucleotide factor binds at the subunit interface of two adjacent RecA-like folds [60].

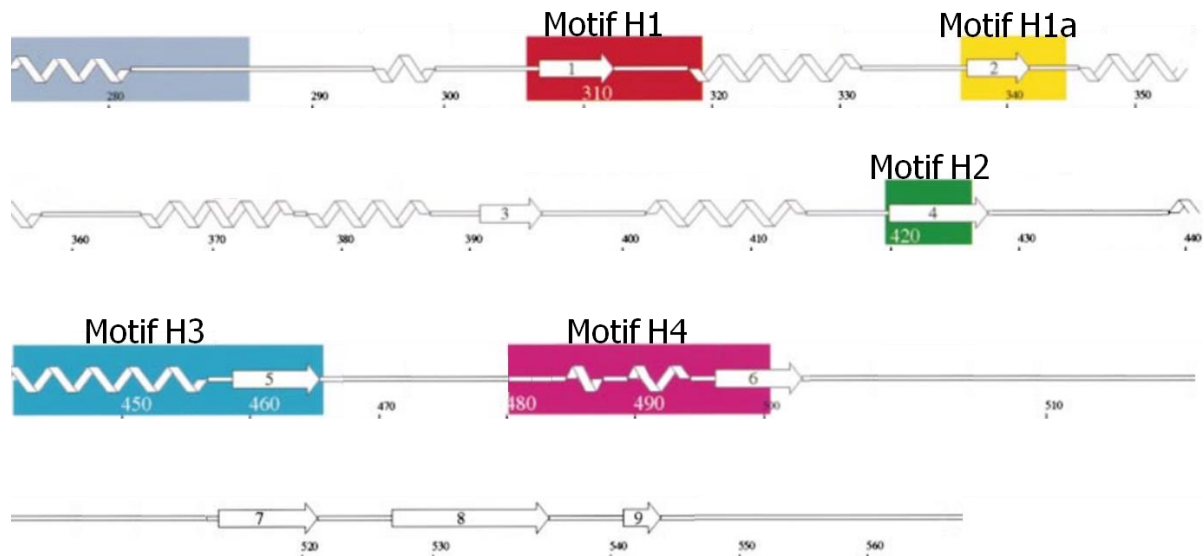


Figure 2.1: DnaB family helicase conserved sequences. T7gp4 DNA helicase amino acid structure with conserved regions highlighted: motif H1 in red, H1a in yellow, H2 in green, H3 in blue and H4 in violet. Figure adapted and modified from [48].

There are several hexameric helicases that have been characterized, for example:

- T4gp41: read lengths of over 1689 kb [61] at a rate of 400 nt/s [61] when stimulated by primase [39, 61].
- Plasmid encoded RSF1010 RepA: one of the smallest helicases [62], has 5'→3' directionality with optimal unwinding at pH 5.5. It does not interact with a primase protein [62], and requires a forked DNA substrate to load [39].
- *E. coli*'s DnaB: relatively well studied, however, it is not very stable [59], it cannot preload onto DNA [63], and requires DnaC protein to load onto DNA [64]. It is further stimulated by DnaG primase [59], which increases its NTPase and helicase activities to create Okazaki fragments during replication [59]. The complex of replisome proteins has an unwinding rate of 1000 bp/s [64, 65] but is significantly slower on its own, 35 bp/s [66] or 291 bp/s [67].
- Human hexameric helicases: Human Bloom's syndrome helicase [39], and mitochondrial "Twinkle" [68, 69], which is homologous to T7gp4 with a 46% sequence similarity [70]. T7gp4 will be discussed in detail below.

2.1.6 Bacteriophage T7 Gene Product 4 DNA Helicase

T7gp4 DNA helicase is a member of helicase superfamily 4. It is an ATPase Associated with various cellular Activities (AAA+) hexameric helicase. It is a relatively small, homohexameric [33] protein that has a two-tiered ring structure around ssDNA with six-fold symmetry [39, 71] (Figure 2.2). It has dTTPase activity and is highly processive on ssDNA [36]. T7gp4 translocates along ssDNA from 5'→3' [50, 72] and does not have any sequence specificity [73]. Electron microscopy studies have shown that T7 helicase has a ring diameter of 13 nm, with a centre at 2.5 nm [71]. In the presence of ssDNA, the ring tightens further [52].

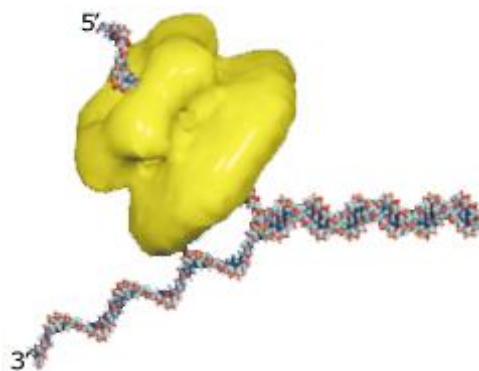


Figure 2.2: T7gp4 DNA helicase structure. The T7 DNA helicase hexamer (yellow) forms a two-tiered ring structure encircling the 5' ssDNA within its central hole at the replication fork, excluding, but interacting with (not shown) the 3' ssDNA, referred to as the strand exclusion model. Figure adapted from [39].

In the cell, T7gp4 is involved in replication [37, 45] and recombination [74]. It is naturally expressed as a helicase-primase fusion protein, as described below. T7gp4 has a complex mechanism of action in which oligomerization is a prerequisite for NTPase, DNA binding, and dsDNA unwinding activities [72].

In vivo, T7 helicase works as a part of a replisome complex [46, 75] in which the C-terminal last 17 residues interact with DNA polymerase and a processivity factor called thioredoxin [75, 76]. These interactions increase hexamer formation, and while T7gp4 is fully functional on its own *in vitro*, its activity is enhanced when incorporated into the complex [77]. Single stranded binding protein also helps to unwind dsDNA in this complex but has no effect on translocation along ssDNA [78].

In addition to its small size, lack of sequence specificity, and high processivity on ssDNA, T7gp4 is very well characterized. It is one of the most well studied helicases [55, 79], used as a model for studying ring shaped helicase mechanisms [37], and is often used as a model for studying replication [45, 46] due to the fact that only two proteins, T7 polymerase and T7 helicase-primase, are required to study replication *in vitro* [45]. There are also a significant number of mutagenesis studies that shed light on specific interactions and electron transfers that occur within the helicase structure [57, 80-89], important for deciphering an electronic sequence of nucleotides in a future helicase-based DNA sequencing method.

As mentioned above, T7gp4 is an AAA+ protein. AAA+ proteins are found in all kingdoms, and are often ring-shaped oligomers [90]. They are generally responsible for the disruption of molecular structures [90] by converting the chemical energy of NTPs to cellular events [91]. Commonly, AAA+ family proteins have a conserved ATPase region based in Walker A and Walker B motifs, which are present in T7gp4 helicase, as discussed earlier.

Bacteriophage T7 gene 4 encodes two protein products that both function as hexameric helicases [71, 89, 92]. Polypeptide A (4A) is 63 kDa and encodes the full length protein, which consists of an N-terminal primase portion and a C-terminal helicase portion linked by a linker region [31, 37, 89, 93-97] (Figure 1.5). Polypeptide B (4B) is a 56 kDa protein whose translation start site occurs at a second, internal initiation codon within gene 4 in the same reading frame as 4A, but at the methionine 56. Thus, T7gp4B is truncated for only the helicase functionality and lacks primase function. The functionality of the truncated helicase of T7gp4B is indistinguishable in terms of helicase-specific enzymatic activity from T7gp4A [92, 98]. T7gp4B includes a large portion of the primase amino acid sequence, however it is not necessary for helicase functioning since the helicase plus linker region sequence will form functional hexamers [33]. The truncated primase portion is also capable of functioning independently [32].

The linker region in T7gp4 spans from residues 246 to 271 of the T7gp4 primase-helicase sequence connecting the C-terminus helicase portion to the N-terminus primase portion [58]. It is essential to helicase oligomerization [50, 58, 86] and DNA loading [50, 58]. The linker region forms a helix-loop structural conformation. Helices in the linker region of one subunit interact with the loop structures of linker regions in an adjacent subunit in the formation of the hexameric ring [84]. In the truncated DNA helicase, within the linker region peptide sequence, the small size of the alanine residue 257 is essential for DNA loading, which occurs at the centre of the ring structure of the helicase [70, 86]. This DNA loading function which A257 is responsible is rescued when the primase portion of the helicase-primase protein remains [84], one reason why the T7gp4B protein may contain so much of the primase peptide.

T7 helicase can only form hexamers in the presence of a nucleotide factor [99, 100] and only if the NTP binding site is active/functional [50]. It can form a hexamer in the presence of dTTP, dTDP, dATP and ATP [31] at 0.2 μ M concentrations [101], however, it has a stronger affinity for dTTP ($k_m = 0.4$ mM) than dATP ($k_m = 1.2$ mM) or ATP ($k_m = 3.2$ mM) [31] and only hydrolyses and unwinds with dTTP [39]. The structure of the hexamer is more compact with dTTP than dTDP. The 5' methyl group on the Thymine base is postulated to explain the apparent preference for dTTP [72].

Mg²⁺ stabilizes the transition state of dTTP hydrolysis, and while Mg²⁺ is required for hydrolysis of dTTP, it is not required for dTTP binding or hexamer formation [31, 39, 73]. This property can be used to pre-assemble, or pre-load, helicase onto DNA. T7 helicase, as well as the truncated T7 helicase including the linker region, will form hexamers at high concentrations (≥ 100 μ M) without the presence of NTP or DNA [33, 101], however the presence of dTTP stabilizes the hexamer [101] and the DNA binding motif near the centre [48]. T7gp4 changes its conformation and interaction with DNA depending on the state of the bound NTP [72].

The nucleotide binding site of T7 helicase is at the interface of adjacent monomers' C-terminus, where the helices and loops of adjacent RecA-like domains

form a crevice [48, 59]. This is part of the conserved motor ATPase activity of AAA+ proteins [60].

Each monomer of the hexameric helicase has a nucleotide binding site located at the subunit interface [72] (Figure 2.3). The walker A and B motifs present in the NTP binding site in conserved motifs H1 and H2 interact with the phosphate groups of dTTP, relaying NTP hydrolysis charges to the DNA binding site [72]. This charge relay is coordinated by a conserved “glutamate switch” found in AAA+ proteins, in this case at position 343. Glu343 couples NTP binding activity to DNA binding [86]. If dTTP is bound at the subunit interface nucleotide binding site, Glu343 interacts with the γ -phosphate. Without dTTP present, Glu343 disrupts DNA binding [81].

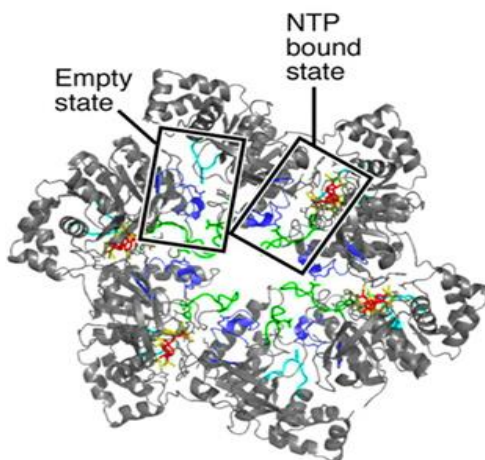


Figure 2.3: T7gp4 DNA helicase subunit interface. The six subunits of T7gp4 DNA helicase (gray) form a hexamer. The subunit interfaces (two examples in boxes) represent the NTP binding sites, forming different conformations when the NTP is bound or unbound. Figure adapted from [81].

Arginine fingers are also motifs found in GTPase activity proteins, and helicase's NTPase activity consists of conserved GTPase residues [48, 72, 82] at (a) Arg504, which interacts with the 4-carbonyl group of the Thymine [72], conferring preference for dTTP over dATP or ATP, and at (b) Arg522 which interacts with the gamma phosphate of dTTP [82].

The binding and hydrolysis of dTTP leads to unidirectional translocation along ssDNA [72] at 1 dTTP/3 nucleotides of ssDNA [49]. The bound nucleotide can have

three states: empty, loosely bound (as with dTDP), or tightly bound (as with dTTP) [58]. When bound to dTTP, the helicase takes on a more compact structure than when dTDP is bound [52]. EDTA stops the dTTP hydrolysis reaction [73]. Extensive studies on the dTTPase activity have been undertaken, showing that all of the subunits in the hexamer and all six dTTP sites are active for helicase translocation [102].

In order for helicase to bind to DNA, both dTTP binding [73] and hexamer formation activities must occur. DNA binding is a requirement and stimulus for NTPase activity and DNA unwinding [39], however helicase bound to DNA can still be deficient in ssDNA translocation [45].

T7 helicase does not have any sequence specificity for DNA binding [103], however it requires a forked substrate in order to bind an ssDNA tail on a dsDNA [73] within its central hole [55]. The helicase interacts with both strands of DNA, enclosing around the 5' tail in the forked substrate, and excluding the 3' strand [104, 105] (Figure 2.2). The central cavity can accommodate 25-30 nucleotides of ssDNA [71]. The forked substrate requires a 5' tail of 35 bases for maximum unwinding at any temperature [105] binding to a 3' tail of 15 bases at 18 °C [36, 105] and only 7 bases at 37 °C [95]. Such a fork can be created artificially for *in vitro* experiments by creating a dsDNA with two non-complementary tails [93, 94, 105]

It has been recently shown that the 3' tail of the forked substrate can be replaced by a biotin-streptavidin complex or a DNA hairpin [36]. This slows the binding reaction down by half, however, it is particularly interesting for this research undertaking due to the involvement of biotinylated DNA bound to (strept)avidin.

T7 helicase binds preferably to ssDNA [103], and travels unidirectionally from 5'→3' [94, 95, 106] highly processively along ssDNA, traveling more than 75 kb [49], but significantly less along dsDNA, dissociating after approximately 60 bp [36]. It is a relatively stable molecule, with a half-life of approximately 3 minutes at 37°C on circular DNA and 45s on linear ssDNA [73]. T7 helicase is also capable of transitioning from a bound ssDNA to a dsDNA during translocation [36]. Optimal unwinding has been

shown to be at 60 bp/s or higher at 18°C up to translocation of 132 nt/s at 18°C on ssDNA [49].

2.1.7 DNA Helicases Summary

Helicase molecules are of potential interest for DNA sequencing because of their natural affinity for DNA and capability of translocating along a strand of DNA. Helicases are motor proteins that couple NTPase activity to DNA binding and unwinding/translocating. Hexameric helicases translocate along DNA with high processivity. Bacteriophage T7 gene product 4 helicase is a good potential candidate for helicase based sequencing due to its small size, high processivity, extensive characterization, capacity to preload onto DNA in the absence of Mg^{2+} , capability of functioning *in vitro* alone, and ability to bind and translocate along any sequence of DNA.

2.2 Biotin and Avidin

The biotin-avidin relationship is unique within nature due to the unsurpassed strength of attraction between the two molecules. The two molecules have the strongest mutual attraction in all of nature, with a dissociation constant of $k_d = 10^{-16}$ nM [107] comparable to of $k_d < 100$ nM for helicase with ssDNA [49]. This relationship has been exploited for several different types of studies [108-112]. The only way to separate the two molecules from one another once bound is by denaturing the avidin protein for example by extreme changes in pH or with very high temperatures [112].

This strong attractive property could be harnessed as a part of the novel helicase based DNA sequencing technology in a number of capacities. First, if a recombinant protein consisting of a helicase monomer fused with an avidin protein were created, biotinylated DNA oligonucleotides could be used to attract specific DNA segments to the nano-biosensor (see sections 1.5.7 and 1.8 for more information). Additionally, the fusion protein could potentially be used as a means to attach the helicase to the carbon nanotube in the proposed nano-biosensor.

In order to understand how the properties of the two molecules can be exploited, a brief description of the two is necessary.

2.2.1 Biotin

Biotin, also known as vitamin H or vitamin B7 [34], is a small molecule that has been studied as early as 1939 [113]. The molecule consists of a double-ringed structure with a carboxyl group hanging off of a long carbohydrate tail. The carboxyl group that hangs from the molecule is transferred by carboxylating enzymes in biological processes [34], or is covalently bound to molecules, such as proteins or DNA [34, 114]. This process can occur either enzymatically or chemically [34].

When biotin is covalently bound to DNA, it is most often used to label one end of the DNA strand. This would allow for one molecule of DNA to be bound by streptavidin or avidin protein, a property that is exploited by several biochemical techniques, including the proposed attraction of the helicase-avidin recombinant protein in this work.

2.2.2 Avidin

Avidin is a protein with a very high affinity for the molecule biotin. It is a tetrameric glycoprotein [115] isolated from avian egg white oviducts [107]. In its non-glycosylated form, each monomer is about 14 kDa [34] and has been successfully cloned and expressed in *E. coli* [116].

Streptavidin, a homolog of avidin, was discovered from *Streptomyces avidinii* in 1964 [116] after avidin. It is so similar in its attraction to biotin, that the two proteins are often referred to as (strept)avidin [34]. Streptavidin is naturally a tetramer of about 15kDa [112] monomers, and can bind up to 4 biotin molecules with a dissociation constant of $K_d = 10^{-15}$ [107]. Unlike avidin, it is not a glycoprotein [115]. It is more widely used than avidin because of its lowered non-specific binding [34]. The two proteins are not evolutionarily related, sharing just 38% amino acid sequence similarity [107, 115], yet both consist of homotetramers that bind 4 biotin molecules strongly by β barrel structures [115].

Recently, an avidin monomer was developed [117, 118]. Non-glycosylated monomeric avidin was chosen as a candidate protein for this work over streptavidin primarily because of its slightly smaller size.

2.3 Fusion Proteins

When recombinant proteins are engineered, the design must ensure that each protein folds properly and that they do not interfere with each other. Inserting short peptides between the two combined proteins, known as linker peptides [119] helps to ensure proper folding of the two proteins. These linker regions are found in natural proteins, most commonly in eukaryotes, in between various catalytic domains [120] and are often serine and glycine rich [120, 121]. In engineered proteins, retaining biological activity for two fused proteins requires at least 11 inserted residues, and has maximal effect with 19 residues, especially when rich in glycine along with Alanine or Serine [121].

2.4 DNA

The genetic code is stored within the nucleus of each cell in the form of deoxyribonucleic acid (DNA). DNA is the blueprint for every living organism. The sequence of the four DNA nucleotides or nucleic acids (the genotype): adenine, guanine, cytosine, and thymine, determines every expressed appearance or mechanism that takes place in an organism (the phenotype). The four nucleic acids share a core structure composed of a ribose sugar, a phosphate group, and a nitrogenous base (Figure 2.4). Each one of these elements contributes an important role in the DNA structure.

2.4.1 Ribose Sugar

The ribose sugar may be present in various forms in a nucleotide. When bound to the phosphate group, at the 5' carbon, and nitrogenous base as a nucleotide at the 1' carbon, the ribose sugar has two free hydroxyl groups at the 3' and 2' carbons (Figure 2.4). In the case when both 3' and 2' hydroxyl groups are free, the nucleotide is called a ribonucleic acid, or RNA. This molecule is relatively unstable and is used by the cell to make transient copies of genes for expression. When the 2' carbon of the ribose sugar's hydroxyl group is replaced by a hydrogen, this is a deoxyribose sugar, and the molecule is a DNA nucleotide. If both free hydroxyl groups are replaced with hydrogens, the nucleotide cannot form a bond with another nucleotide to create a strand, and is

referred to as a double deoxyribonucleic acid. These double deoxyribonucleic acids are used in chain terminating DNA sequencing methods.

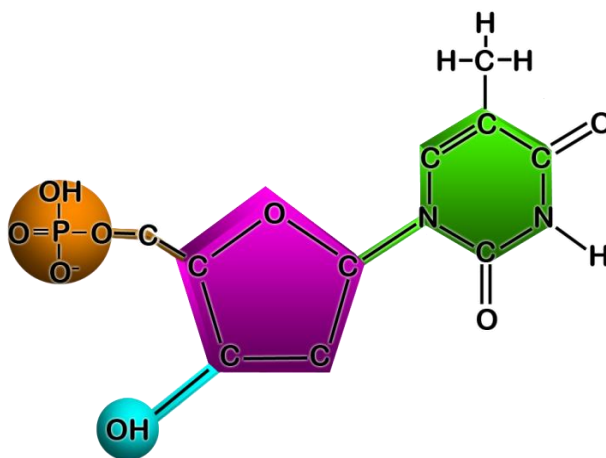


Figure 2.4: Deoxyribonucleic acid structure. DNA is composed of a ribose sugar (violet) with a nitrogenous base (green, in this case Thymine) bound to the C1 carbon of the ribose, a phosphate group on the C5 carbon (orange) and a single hydroxyl group (cyan) on the C3 carbon. The C2 carbon is not bound to a hydroxyl group.

2.4.2 Phosphate Group

The 3' carbon hydroxyl is involved in the reaction with an adjacent nucleotide's phosphate group in the creation of a strand of nucleic acids. The phosphate group of the nucleic acid, on the 5' carbon of the ribose sugar, forms a diester bond with the 3' carbon hydroxyl group of another nucleic acid to create the phosphate ribose backbone of DNA (Figure 2.5). This process releases a water molecule and a pyrophosphate molecule, which can be used in different DNA sequencing techniques such as pyrosequencing to track incorporations of bases to a strand of DNA. By convention, a DNA sequence is written from 5' to 3', indicating that the sequence of nucleic acids is read along the phosphodiester backbone from the 5' carbon phosphate to the 3' carbon.

A single nucleotide, such as adenosine triphosphate (ATP), can be used as energy storage for cellular processes. The energy stored in a nucleotide is stored in the triphosphate. When a pyrophosphate breaks off from the nucleic acid, energy is released to fuel the metabolic process.

2.4.3 Nitrogenous Bases

The four nitrogenous bases in DNA are the “letters” in the DNA alphabet. They align as a ssDNA as described above, and form dsDNA based on the complementarity of the nitrogen base pairing. Guanine and Cytosine pair together with a stronger bond of 3 hydrogen bonds, and Adenine and Thymine pair together with 2 hydrogen bonds. The order of nucleotide bases in a strand of DNA is known as its sequence.

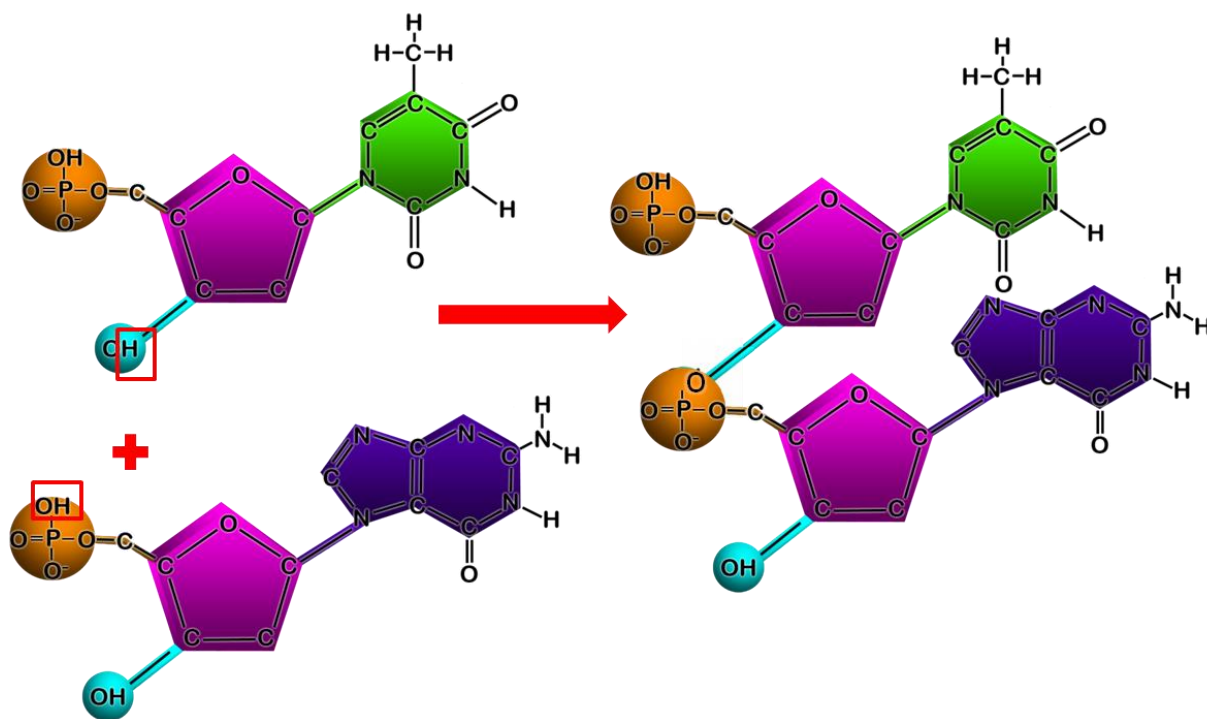


Figure 2.5: DNA phosphodiester bond. The hydrogen from the 3' hydroxyl group (cyan, red box) of one DNA nucleotide (green, thymidine in this case) undergoes a chemical bond with the hydroxyl group of the 5' phosphate group (orange, red box) of another nucleotide (purple, guanine) releasing a water molecule and creating a phosphodiester bond.

2.4.4 Replication

Since all cells in an organism contain the exact same DNA sequence, during cellular division, a cell must make an exact duplicate of the DNA in its nucleus. This process is known as replication. The main proteins involved in the replication process are: helicase, which unwinds double stranded DNA to two separate strands; primase, which creates a short complementary sequence of RNA; and polymerase, which utilizes the primer as a starting point to continue to sequentially create a duplicate of the strand of DNA being copied. Sequencing methods often make use of proteins that

naturally interact with DNA. DNA polymerase has been widely used in DNA analysis for amplification, as in PCR, and as the basis for many DNA sequencing techniques [1, 2, 4, 5, 7-16]. Due to the short unstable RNA strands created by primase, it is less frequently used in DNA analysis.

2.4.5 DNA Modification

DNA can be modified and manipulated in a variety of ways. Endonucleases, or restriction enzymes, recognize specific dsDNA sequences and cut them in a predictable fashion creating dsDNA fragments with blunt ends, or 2-3 nucleotide overhangs on one end. There are a vast variety of endonucleases commercially available, each that recognizes and cuts a specific sequence in a specific manner. Ligases are enzymes that can mend a phosphodiester backbone between two sets of dsDNA.

ssDNA can be modified at one end in a variety of ways. Radioactivity can be conferred to a strand of DNA by use of ^{32}P phosphate and the enzyme T4 polynucleotide kinase can add a phosphate group to the 5' end of a DNA strand. ssDNA can also have a biotin molecule added to one end.

3 - Aim 1: Purification of Helicase

3.1 Rationale

A means of rapidly, inexpensively, and easily purifying T7gp4 helicase with the potential for large scale production is desired for potential commercialization and device production. This would involve expressing and purifying the protein in a bacterial host, which is easy to work with, inexpensive, and could be easily performed on a large scale.

In the literature, T7 helicase is purified in a number of ways, most often by tandem purification. Scientific studies use tandem affinity purification, involving several separate purification steps in order to yield a purer protein product, which is time consuming, requires highly specialized training, and is costly. In these studies, a combination of 2 or more purification methods, including nickel resins, chromatography, complex fractionation, ATP-agarose affinity, or phosphocellulose column fractionation, are used [45, 50, 55, 72, 122, 123]. For example, many studies [45, 122, 123] by various research groups use modifications of a labourous 2 day long procedure to obtain purified helicase enzymes, which use 3 to 4 protein purification methods to obtain pure samples including “salting out” with ammonium sulfate, NaCl and PEI, phosphocellulose resins, DEAE sephacel columns HPLC and various concentrating membranes and cassettes. Studies using His-tags [50, 55, 72] also use 2 or more means of tandem purification with complex cellular lysis mechanisms combining chemical, freeze-thaw, and sonication methods which likely require 8 or more hours to complete. The presented method of purification requires approximately 5 hours of overall work, including waiting periods.

3.2 Purpose

The purpose of this portion of the study is to purify the helicase enzyme quickly, inexpensively, and easily in a bacterial host by means of a single purification step.

3.3 Methods

Detailed information of the methods performed for Aim 1 can be found in Appendix III under III.ii, III.iii, III.iv, III.v, III.vi, and III.vii.

3.3.1 PCR Amplification of Helicase Enzyme DNA Sequence.

For cloning into a vector for bacterial expression, the helicase enzyme DNA sequence was Polymerase Chain Reaction (PCR) amplified using PCR primers (Table 3.1) specific to the codon optimized helicase sequence with an annealing temperature of 64°C according to the New England Biosciences (NEB) Tm Calculator [124]. A helicase-avidin recombinant DNA sequence was obtained from GenScript™ for amplification. The primers were designed to contain a His-tag and EcoRI and NcoI restriction enzyme sites. PCR was performed for 35 cycles (Table 3.2) with the block temperature set to 98°C and the lid to 100°C as described in Appendix III, and verified by 1% agarose gel. PCR product purification was performed using a PCR purification kit purchased from Qiagen™. Further information on PCR amplification of DNA can be found in Appendix III.ii.

Table 3.1: PCR primers for codon-optimized helicase Sequence.

Forward	CAGTGCCATGGGACATCACCACCATCACCATCAGATCGCGAGATCATG
Reverse	CGTAGCTCGAGTCAGAAATCCGTGTC

Table 3.2: PCR Reaction Conditions.

Stage	Temperature	Time	
Initial Denaturation	98 °C	30 s	
Denaturation	98 °C	10 s	35 cycles
Annealing	64 °C	20 s	
Elongation	72 °C	30 s	
Final Extension	72 °C	5 minutes	
Hold	4 °C	Indefinite	

3.3.2 Digestion and Ligation Reactions.

The PCR product and empty pET28b(+) vector were digested with XhoI and NcoI restriction enzymes for 1 hour at 37°C. The digestion reaction products were run on a 1% agarose gel with ethidium bromide (EtBr) at 70 V for 1 hour. Visualization of the nucleic acids was performed by ultraviolet (UV) light and bands of interest were excised use a Qiagen™ gel extraction kit as described in Appendix III.ii.vi. The restriction digested vector and PCR product were ligated using T4 ligase enzyme at ratios of vector:insert of 1:0 (control), 1:3 and 1:7. Further information of DNA digestion, agarose gel DNA extraction and DNA ligation can be found in Appendix III.ii

3.3.3 DH5 α Transformation.

The ligation reactions were transformed into the DH5 α amplification bacterial cell line with 50 μ L/mL kanamycin antibiotic in LB broth for bacterial selection. Transformed bacterial cells were spread onto LB agar plates containing 50 μ L/mL kanamycin and colonies were picked and inoculated in LB broth with 50 μ L/mL kanamycin for screening. Colonies were screened for the presence of the insert by two methods: inoculating a colony into 10 μ L of PCR reaction mixture with T7 promoter primer and the helicase reverse primer as described in Appendix III.vi, and by digesting miniprep DNA with NcoI and XhoI restriction enzymes in a 20 μ L volume digestion reaction for 1 hour and 37°C. Both screening methods were run on 1% agarose gel electrophoresis. Further information on bacterial transformation, PCR and DNA digestion can be found in Appendix III under III.ii, III.iii. and III.vi.

3.3.4 BL21 Transformation.

Plasmid DNA was extracted from the successfully ligated colonies on a large scale using a midiprep kit purchased from QiagenTM as described in Appendix III.iv. The isolated plasmid was transformed into the BL21 *E. coli* expression cell line. Further information can be found in Appendix III.v.

3.3.5 Induction of Protein Overexpression

Positive clones were shaken at 37°C in Luria Bertani Broth (LB) medium containing 50 μ L/mL kanamycin until optical density of 0.4 at 600nm was obtained, and then induced with 1mM Isopropyl- β -D-thiogalactoside (IPTG) for 2-4 hours. Induced bacteria were lysed and overexpression of the target gene was verified running a 10% Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) at 120V for 1 hour and 20 minutes. The overexpressed proteins were visualized by Coomassie Brilliant Blue staining. Further information on bacterial protein overexpression and protein visualization can be found in Appendices III.v and III.vii, respectively.

3.3.6 Protein Purification and Dialysis.

Overexpressed helicase was purified by Talon cobalt resin. 50 mL aliquots of induced bacteria were centrifuged at 4000 rpm for 10 minutes at 4°C. The pellet was resuspended in 10 mL equilibration buffer with PMSF and sonicated on ice. The

sonicated solution was centrifuged at 11000 xg for 20 minutes at 4°C and the supernatant was applied to the resin, rotating for 2 hours at 4°C. The column was drained by gravity, washed once with 10 mL equilibration buffer, twice with 5 mL of wash buffer and eluted in 0.5 mL elution buffer 5-6 times. The elutions were verified by SDS PAGE run at 120V for 1 hour and 20 minutes and stained with Coomassie Brilliant Blue dye. The purified protein was dialyzed to replace elution buffer with a storage buffer consisting of 20 mM Hepes pH 7.6, 100 mM NaCl, 1mM MgCl₂, and 10% glycerol. Further information on bacterial protein purification and dialysis can be found in Appendix III.vi and III.vii, respectively.

3.4 Results

The helicase DNA sequence was successfully ligated into the multiple cloning site of the pET28b(+) expression vector. The helicase sequence including the linker region was PCR amplified from the helicase-avidin sequence purchased from GenScript™. The PCR yielded a single band at approximately 1 kb (Figure 3.1), consistent with the predicted size for the helicase sequence. The restriction digested empty pET28b(+) vector and PCR product were successfully ligated together. The 1 kb helicase PCR product, 5000 kb digested vector, and digested ligated vector with insert were visualized on a 1% agarose gel (Figure 3.1). The DNA sequence was verified by capillary based Sanger sequencing using the Applied Biosystems 3730 DNA Analyzer DNA sequencing device.

The ligation product was transformed into the DH5α strain of *E. coli*, and was selected for using kanamycin antibiotic resistance. The control ligation product was also transformed and did not yield any colonies. The 1:3 ligation had 91 and the 1:7 had 15 colonies (results not shown), thus the 1:3 ligation ratio was the most successful ratio. Four colonies from the 1:3 ligation and one from the 1:7 ligation transformation were screened for by PCR using the T7 promoter primer and the reverse helicase primer (Figure 3.2A). All five PCR screens yielded bands of approximately 1 kb, indicating that all five colonies contain the insert sequence next to the T7 promoter sequence. Therefore the insert was successfully ligated within the vector for all 5 colonies screened and the ligation reaction was successful.

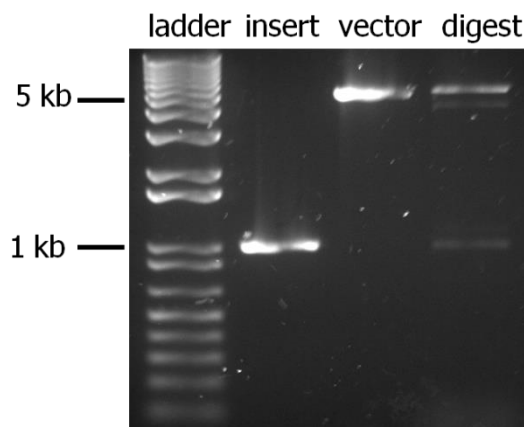


Figure 3.1: Agarose gel of helicase PCR product, pET28b(+) vector, and final ligation product. A 1% agarose gel containing ethidium bromide was loaded with 5 μ L of DNA ladder, 5 μ L of helicase PCR product (insert), 2 μ L of empty pET28b(+) vector, and 10 μ L of the final ligation product digested by NcoI and XhoI for 1 hour at 37°C.

The five colonies were also screened by DNA sequencing (Applied Biosystems 3730 DNA Analyzer DNA sequencing device) and by digesting miniprep DNA with NcoI and XhoI restriction enzymes (Figure 3.2B). The agarose gel visualized under UV light shows that all five screened colonies contain the vector with the insert. Colonies 1 and 4 did not undergo complete digestion as indicated by the two bands near 5000 kb. Colonies 2 and 4 contained the largest yield of extracted DNA from the minipreps, followed by colony 1 and colonies 3 and 5. Since all five colonies screened positively in both screening tests, colony 2, which yielded the largest concentration of DNA was selected for further study.

The helicase-pET28b(+) plasmid was transformed into an expression cell line for protein expression, and the transformation was verified by PCR. Plasmid DNA from colony 2 was purified on a large scale as a midiprep and transformed into the BL21 expression strain of *E. coli*. The agar plate plus kanamycin yielded 3 colonies (results not shown), all of which were screened by inoculating the colonies in a 10 μ L PCR reaction and loading on a 1% agarose gel. As with DH5 α , presence of the vector was successfully verified by PCR screen (Figure 3.3), as all three colonies yielded PCR products at approximately 1 kb.

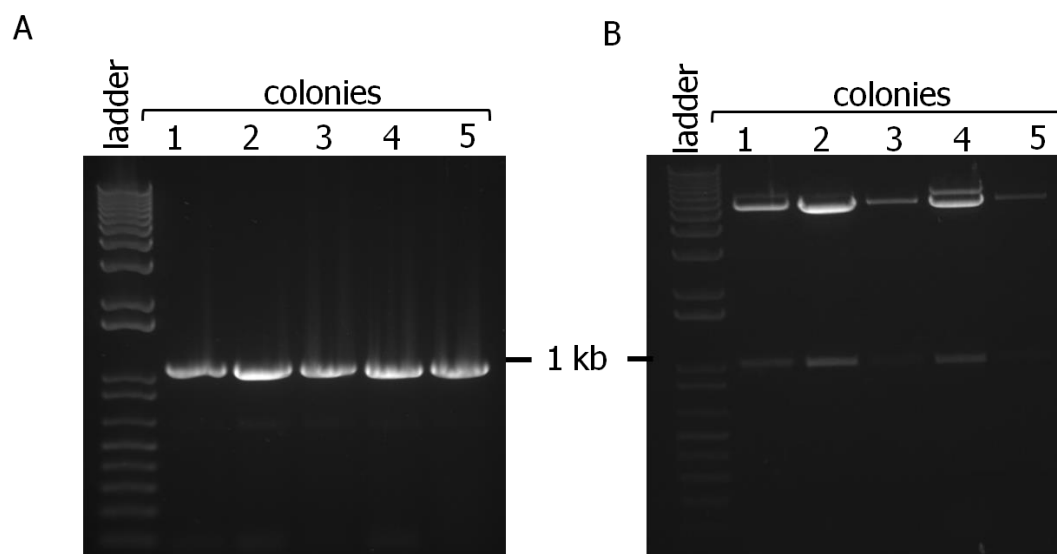


Figure 3.2: Verification of DH5α pET-helicase transformation. Five colonies of the pET-helicase DH5α transformation were picked for screening by two methods. A) PCR screen. Each colony was PCR amplified and run on a 1% agarose gel with EtBr for visualization. B) Digestion of isolated plasmid. Each of the same colonies were grown overnight in 3 mL of LB broth with kanamycin. Miniprep plasmid purification was performed, and isolated DNA was digested by NcoI and XhoI prior to agarose gel electrophoresis. All five colonies contained the helicase amplification product at 1 kb.

Overexpression was verified using Coomassie Blue staining of the protein from the transformed BL21 cells. Two of the three verified transformed BL21 colonies were induced with IPTG to overexpress the helicase protein (Figure 3.4). The Coomassie dye shows a strong band in the induced samples of both colonies at approximately 36.5 kDa, which is the size of the helicase protein. This large band is not present in uninduced samples and therefore represents the overexpressed protein. Both samples appeared to be equally viable, therefore sample 2 was chosen for further analysis.

After verification of overexpression of the correct sized protein, the helicase protein from a large scale induced batch of colony 2 was purified using the Talon cobalt based resin (Figure 3.5A). The purification process can be visualized in Figure 3.5A, showing the step wise purification towards a single protein product in elution lane 6. The induced sample appears to have a strong band at 36.5 kDa, which remains in the

supernatant of the sonication. As expected, the sonication pellet appears as a smear. The three washes loaded do not contain any protein, and the elutions all contain a strong band at 36.5 kDa, representing the purified protein. The elutions also contain trace amounts of non-specifically bound proteins that decrease in amount with each elution, no longer apparent by elution 5. All 6 elutions were pooled together and dialyzed for storage (Figure 3.5B). The resultant solution contained a strong band at 36.5 kDa with very little non-specific proteins. A faint second band just below the strong band is also apparent.

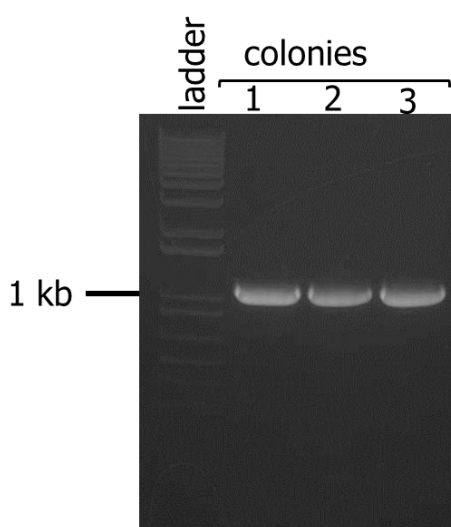


Figure 3.3: Agarose gel electrophoresis verifying BL21 pET-Helicase transformation by PCR screen. Three colonies were PCR amplified prior to agarose gel electrophoresis. The ~1 kb insert was present in all three colonies tested.

3.5 Discussion

The helicase sequence was successfully cloned into the DH5 α and BL21 bacterial cell lines. Furthermore, the helicase protein was successfully overexpressed in the BL21 cell line. These positive results were obtained despite low transformation efficiency in both DH5 α and BL21 transformations. The initial DH5 α transformation yielded 136 colonies, 5 of which were screened positive from two different screening methods. The control transformation did not yield any colonies, as expected. Possible reasons for the low transformation efficiency include cells that are not very competent, too little outgrowth of competent bacteria, high concentrations of kanamycin or low ligation

efficiency. Furthermore, the positive screening of all five colonies tested in the DH5 α and the lack of colonies on the control plate indicates that the vector used in the ligation reaction was most likely fully digested and did not self-ligate, which is expected because the digestion enzymes used did not have compatible sticky ends. The BL21 transformation yielded only 3 colonies on one plate compared to 91 on one plate at the optimal ligation. This cell line is known to have a lower transformation efficiency, and through experimental experience, requires more DNA (up to 100ng) that has been purified using a column in order to yield results. This transformation was screened only by PCR and DNA sequencing and not by digested plasmid DNA because BL21 is not as well suited to plasmid extraction as DH5 α . Only 2 of the 3 colonies were screened for small-scale induction because of a technical error, although in all likelihood the third colony would likely also overexpress helicase to the same degree upon IPTG induction.

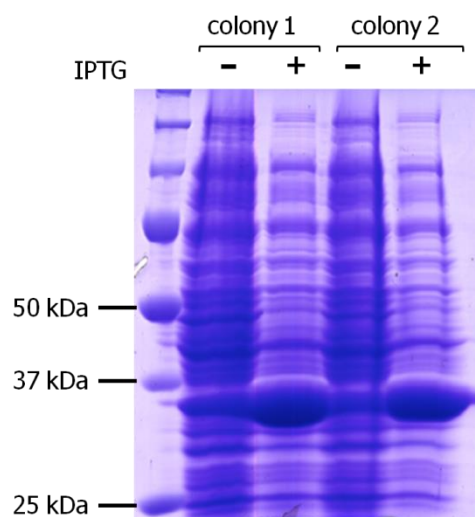


Figure 3.4: Verification of helicase overexpression. The BL21 colonies transformed with the pET-helicase vector were shaken in 2 batches of 3 mL LB broth with kanamycin each. One batch was induced with 1mM of IPTG, the other solution of each colony was left uninduced. The samples were run on a 10% SDS PAGE and stained with Coomassie Brilliant Blue Dye to visualize proteins. This was performed for two separate transformed colonies. The band at approximately 36 kDa represents the overexpressed helicase protein.

The BL21 cell line overexpresses the proteins under the control of the lac operon in the plasmid. Therefore, only successfully transformed cells containing a plasmid with a stably expressed protein will, upon induction by IPTG, create a strong band on an SDS PAGE indicating overexpression of the protein of interest. The IPTG

added to the bacteria creates a constant signal of RNA polymerases transcribing the gene on the vector, thus producing far more copies of the protein per cell than the other proteins naturally expressed by the cell.

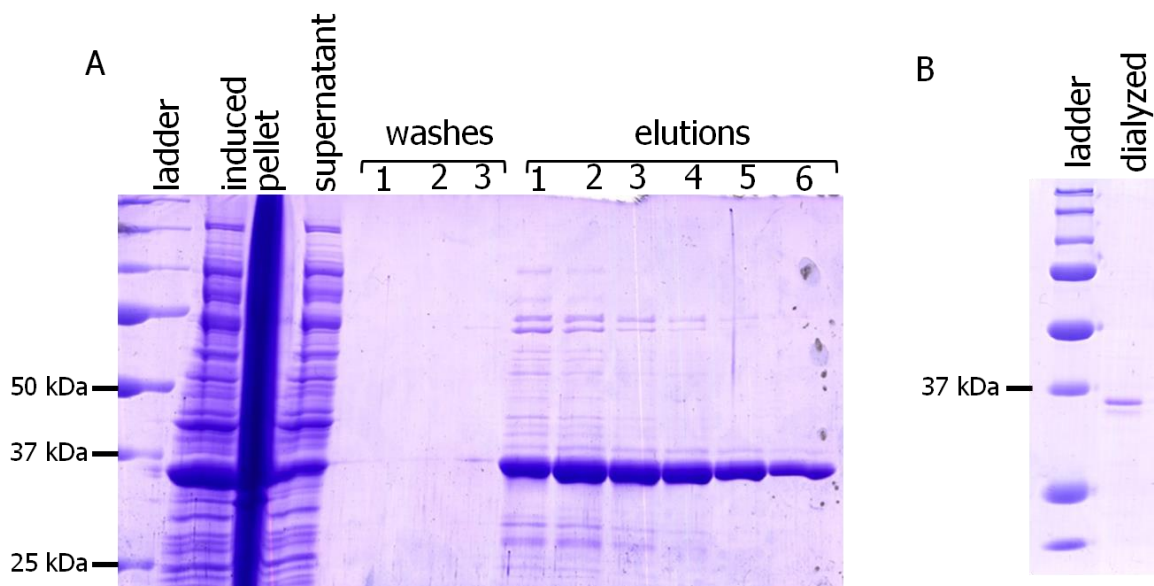


Figure 3.5: Purification of overexpressed helicase protein. A) The protein from the induced BL21 cells (induced), sonication products (pellet and supernatant), washes, and elutions were run on SDS PAGE and were stained with coomassie blue dye. The purification was performed using a talon cobalt His-affinity resin. B) The dialyzed protein was run on SDS PAGE and was stained with coomassie blue dye.

The coomassie stained SDS PAGE of the overexpressed helicase (Figure 3.5A) illustrates several steps in the purification process. The induced whole cell extract lane shows that the induction worked as it did on the small scale (Figure 3.4). The pellet of the sonication appears as a smear because it contains all other components of the cell, such as lipids and genomic DNA, which clump together and do not run well on an SDS PAGE. The supernatant of the sonication has a strong band at around 36.5 kDa which indicates that the induced protein is soluble. The lanes containing the washes do not get stained by the coomassie dye indicating that there is no protein present in those lanes, yet the elutions contain a strong band of the desired size. This indicates that the binding to the resin was selective and strong. The flow-through of the resin was not loaded onto the gel, but would likely have appeared as the same strong pattern of the

background proteins with a less strong band at 36.5 kDa representing the His-tagged proteins which did not bind to the column due to saturation. The elutions begin with a larger amount of helicase eluted which decreases slightly by elution 5 and even less for elution 6. This trend should be expected to be stronger, but as I will discuss in Chapter 4 with respect to the nickel based columns, where the helicase-avidin fusion protein did not elute from nickel based columns even at very high concentrations of imidazole, the helicase protein does not seem to elute easily from the resin due to the strength of the bond. The uninduced sample was not included on this gel due to lack of space, however, it is clear from Figure 3.4 that the pattern of protein expression between the induced and uninduced samples is different.

The purified helicase protein was dialyzed to exchange the elution buffer with one more suitable for storage. SDS PAGE and coomassie staining was used to verify the presence of the overexpressed helicase protein. This sample showed that the pooled elutions contained a strong amount of protein. A second, slightly smaller, band is apparent in this sample that is not in the elutions. This could be due to slight degradation during dialysis, although this is not likely since that would likely appear as a smear rather than a second distinct band because degradation would most likely create several protein fragments of many varied molecular weights rather than one single degradation product at a distinct molecular weight. An alternative, but less likely, explanation is that the second band represents a problem with the gel or loading mixture into the gel, preventing complete denaturation of the protein prior to loading, leading to two apparent sizes of the protein. It is also possible that the second band only appears due to the fact that less protein was loaded onto the gel; the apparent single large band in the elutions actually being composed of two distinct bands that blend together because of volume. This decreased amount of protein loaded also likely explains the lack of non-specifically bound proteins in the original elutions, which likely did not degrade during dialysis.

3.6 Conclusion

The helicase protein was successfully cloned into and purified (isolated) from *E. coli*, using the Talon cobalt based histidine affinity resin.

4- Aim 2: Purification of Helicase-Avidin

4.1 Rationale

T7 helicase does not recognize specific sequences for binding [73], which makes it a good candidate for sequencing any random portions or whole genomic DNA. However, to make this method more attractive for targeted sequencing for personalized medicine, a means of attracting those sequences is desired. As outlined in section 1.07, this could be accomplished by means of a helicase-avidin fusion protein. The avidin portion of this recombinant protein could attract specific strands of biotinylated DNA to the helicase.

4.2 Purpose

The purpose of this portion of the study is to purify a soluble helicase-avidin recombinant protein in *E. coli*.

4.3 Methods

Detailed information of the methods performed for Aim 2 can be found in Appendix III under III.i, III.ii, III.iii, III.iv, III.v, III.vi, and III.vii.

4.3.1 Fusion Protein Engineering

A specialized recombinant helicase-avidin fusion protein was designed (Table 4.1), including mutated residues in the avidin protein, linker DNA as well as a His-tag. The codon optimized DNA sequence for helicase and avidin was obtained from GenScript™ and inserted in the multiple cloning site of the pET28b(+) expression vector.

Table 4.1: Components of the Helicase-avidin Fusion protein.

Component	Purpose
His-Tag	Protein purification
Linker region	Prevention of secondary structure formation within functional domains
Avidin	Protein with high affinity for Biotin
N17I	Avidin Mutation: prevents glycosylation [118]
N54A	Avidin mutation: prevent oligomerization [117]
W110K	Avidin mutation: prevent oligomerization [117]
Helicase	Truncated bacteriophage T7gp4 protein, includes helicase and linker region. Unwinds double stranded DNA.

4.3.2 DH5 α Transformation

100 ng of vector was transformed into the Z-competent™ DH5 α amplification strain of *E. coli* and plated on LB agar plates with 50 μ L/mL kanamycin antibiotic pre-incubated at 37 °C. Colonies were picked and inoculated in LB broth with 50 μ L/mL kanamycin. In addition to DNA sequencing verification, two methods for the verification of plasmid transformation were performed as described in section 3.3, plasmid DNA was isolated by miniprep to verify the transformation and digested by NcoI and EcoRI restriction enzyme, and analyzed by PCR using a T7 promoter primer and the reverse primer for the helicase sequence. Further information on these methods can be found in Appendix III.ii, and III.iii.

4.3.3 BL21 Transformation

The extracted plasmid DNA from the DH5 α strain was transformed into an *E. coli* expression strain by heat shock transformation for 30 seconds at 42°C for protein expression. Further information can be found in Appendix III.iii.

4.3.4 Optimization of Growth Conditions

In efforts to optimize growth conditions to create a soluble protein product, several growth conditions were manipulated, as outlined and explained in Appendix III.v. The following is a list of the various growth conditions that were tested:

1. Expression cell lines: BL21, ER2566, or BL21-CodonPlus-RIL
2. IPTG induction concentration: 0.1 mM, 0.5 mM, or 1 mM
3. Induction OD₆₀₀: 0.2, 0.3, 0.4, 0.5, or 0.6
4. Glycerol: 0%, or 5%
5. Induction temperature: 18°C, or 37°C

4.3.5 Protein Purification

The bacteria containing overexpressed helicase-avidin protein were centrifuged down and then purified by various His-tag affinity resins:

1. Qiagen nickel-based resin
2. Biorad nickel-based resin
3. Talon cobalt-based resin

50 mL aliquots of the induced bacteria BL21 bacteria were centrifuged at 4000 rpm for 10 minutes at 4°C. The pellets were resuspended in resuspension buffer with PMSF, sonicated on ice to disrupt the membranes and the sonicate solution was centrifuged at 11000xg for 20 minutes at 4°C. The supernatant was applied to 0.5 mL of resin in a column and rotated for 2 hours at 4°C. The column was drained by gravity, washed with 10 mL equilibration buffer, twice with 5 mL wash buffer, and eluted in 0.5 mL elution buffer 5-6 times. Further information on His-tag affinity resins and protein purification can be found in Appendix III.vi.

4.3.6 Renaturation of Insoluble Protein

Renaturing of the insoluble protein was attempted by a number of approaches, as outlined and explained in Appendix III.vi:

1. Denaturing agent: Urea, or Guanidium Thiocyanate
2. Resuspension of the bacterial pellet: immediately into denaturing agent, or first without denaturing agent, centrifuged, and then with the denaturing agent
3. Renaturing by dialysis: large volume overnight, or several small volume 30 minutes each.
4. pH: 7.6, or 9.0
5. Renaturing on the column: single wash to remove denaturing agent, or successive small washes
6. With or without β -mercaptoethanol (reducing agent)
7. With or without NP40 (detergent)

A total of 49 permutations of the 11 growth and renaturing conditions were attempted, as highlighted in the results. Further information on these techniques can be found in Appendix III.vi.

4.3.7 Protein Visualization.

The purified helicase-avidin protein elutions were dialysed as described in section 3.3, loaded onto 10% SDS PAGE and run at 120V for 1 hour and 20 minutes, and either stained with Coomassie Brilliant Blue or visualized by Western Blot. Further information on the techniques can be found in Appendix III.vii.

4.4 Results

The first attempts at purification of the helicase-avidin protein were not successful. Three His-affinity resins were tested: 2 nickel-based resins from Qiagen and Biorad, as well as a cobalt-based resin from Talon (Figure 4.1A, B, and C, respectively). Very little to no protein seemed to be eluted from the resins. Concentrations of imidazole up to 1M were attempted with no improvement in amount eluted (data not shown). Western blotting was performed against the His-tag showing very faint banding of eluted protein (data not shown), however, it appeared that the protein was binding to the resin, but not eluting. Further, the majority of the protein was not in the soluble fraction and was therefore was not available to bind to the resin, explaining the lack of protein in the eluate. This could be indicative of improperly folded protein (Figure 4.1D).

Due to the presence of small amounts of protein in the soluble fraction, various means of altering growth conditions were attempted to increase the amount of protein. Coomassie stains were not sensitive enough to show the differences in soluble protein expression due to the large amount of background protein (results not shown), so western blotting was performed. Helicase-avidin overexpression was induced with IPTG for protein expression in all three cell lines tested (ER2566, RIL, and BL21) and at two temperatures tested (18°C and 37°C) (Figure 4.2A) with or without glycerol (results not shown) to try to enhance folding of the protein in the bacteria. The expression of soluble helicase-avidin in the BL21 cell line was the lowest of the three cell lines tested, followed by ER2566 and most with the RIL bacterial strain. Decreasing the temperature of induction to 18°C from 37°C also had a positive effect for all three bacterial strains (Figure 4.2B), ER2566 displayed a 6 fold enhancement and RIL was enhanced 10 fold by the reduction in temperature. BL21 cells did not appear to express any soluble protein at 18°C. Other means of optimizing growth conditions were explored, including changing the concentration of IPTG for induction, and changing the induction OD, however, they did not yield results sensitive enough to be seen on a coomassie stain. Therefore, renaturing protocols were explored, as discussed in further detail Appendix III.vi.

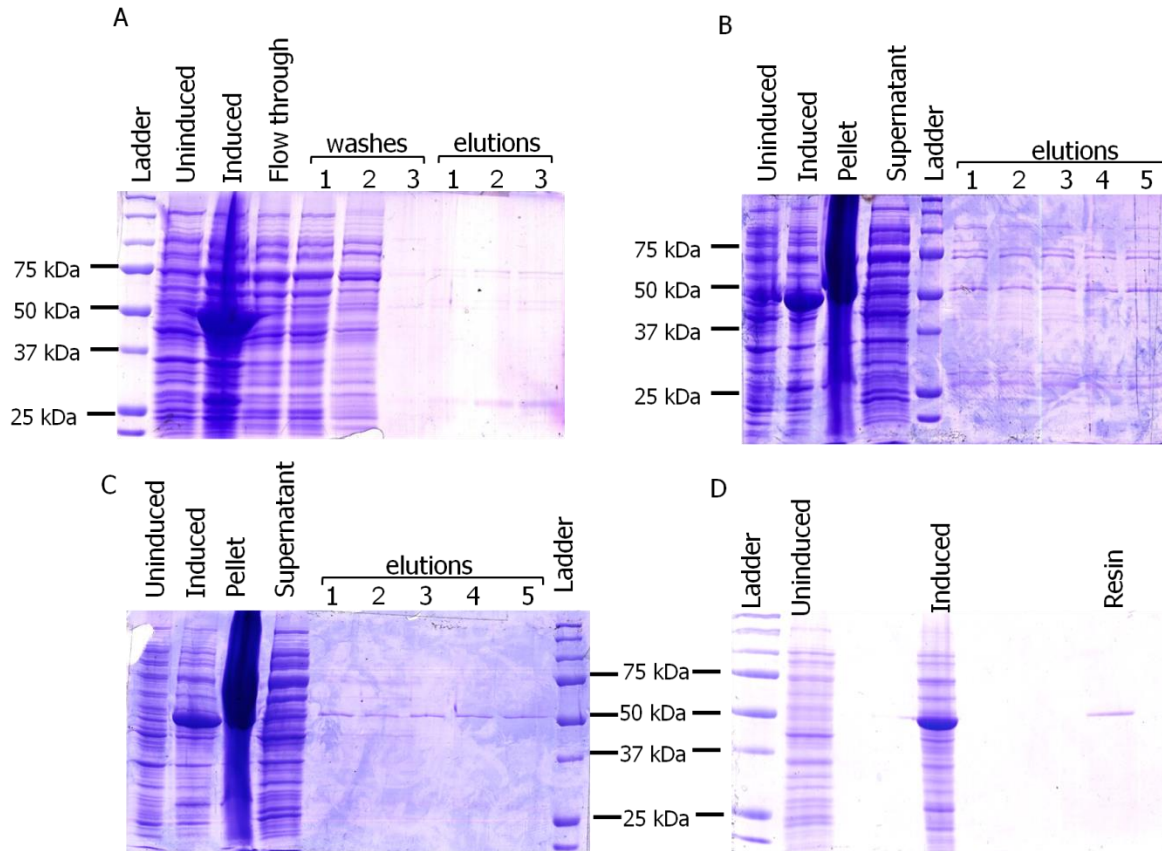


Figure 4.1: Coomassie stain of unsuccessful attempts at helicase-avidin purification with A) Qiagen nickel based, B) Biorad nickel based, and C) Talon cobalt based His-affinity resins. 5 μ L of ladder and 10 μ L each of uninduced (UI) and induced (I) bacterial samples, along with sonicate pellet (P) and sonicate supernatant (S) or resin flow-through (FT), and resin washes and elutions were loaded on 10% SDS PAGE run at 120V for 1 hour and 20 minutes, stained with coomassie brilliant blue. D) 5 μ L of ladder and 10 μ L each of uninduced (UI) induce (I) and resin were loaded, with several lanes separating samples, onto a 10% SDS PAGE run at 120V for 1 hour and 20 minutes and stained with coomassie brilliant blue dye.

Purification of the helicase-avidin fusion protein under denaturing conditions and renaturing the protein *in vitro* yielded a greater amount of soluble product. Several different renaturing conditions were tested. Figure 4.3 is a representative blot of the protein yielded from the seven techniques that were successfully able to yield soluble protein. Of these apparently successful renaturing conditions, the helicase-avidin renaturing condition (iv) (Figure 4.3), had a relatively strong band after dialysis and filtration. Briefly, the bacterial pellet was resuspended in non-denaturing conditions, the soluble fraction was discarded, the pellet was resuspended in denaturing conditions supplemented with urea, and the supernatant from this was retained. The soluble fraction from this resuspension was incubated with the His-affinity resin and the urea was subsequently washed off of in a series of washes with wash buffer supplemented

with β -mercaptoethanol, NP40, and 2M to 0M urea in 0.5M increments, which allowed the protein to refold. The protein was eluted with an elution buffers containing β -mercaptoethanol and NP40. Dialysis was performed at pH 7.5 to replace elution buffer with a buffer appropriate for storage. This sample was used for the enzyme assays in Chapters 5 and 6. Helicase-avidin renaturing conditions (i) (urea denaturation, resuspended twice, eluted with buffer containing β -mercaptoethanol and NP40, renatured by overnight dialysis at pH 7.5) and (vi), (guanidine thiocyanate denaturation, resuspended twice, renatured by overnight dialysis at pH 9 with small batches) both had faint bands indicating successful renaturing of the protein. Purified helicase from Chapter 3 was included to reference relative amounts of protein in the sample. Ovalbumin was loaded as a control to estimate helicase-avidin concentration.

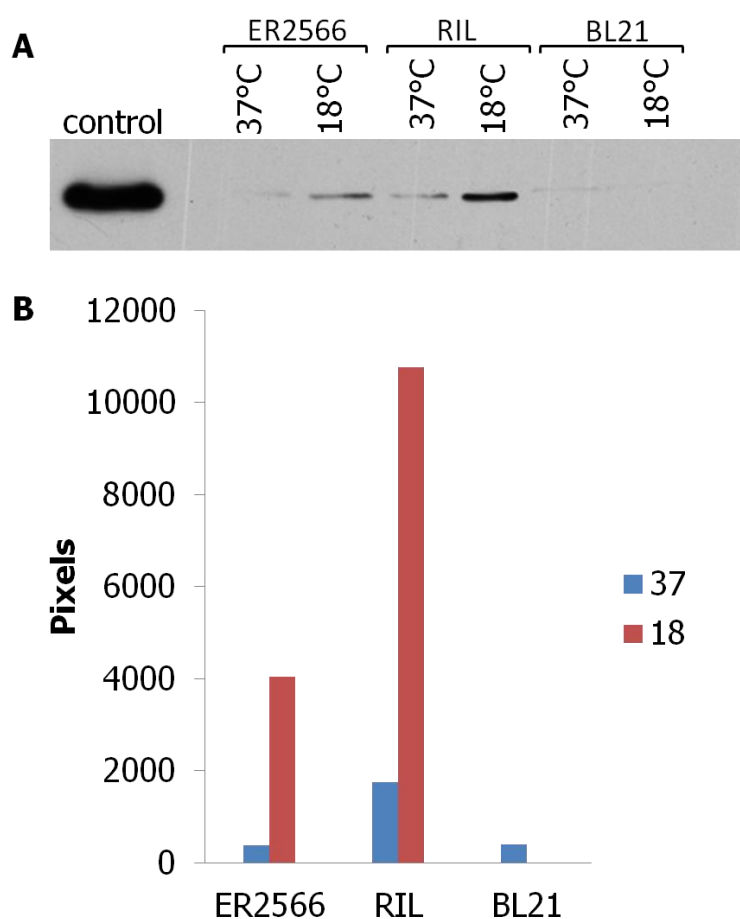


Figure 4.2: Western blot of sonicate supernatants of induced ER2566, BL21-CodonPlus-RIL and BL21 cells at 37°C and 18 °C. The three *E. coli* expression cell lines were transformed with the pET28b(+) vector containing the helicase-avidin fusion protein coding DNA sequence. A) Western blot was performed using an anti-His-tag antibody and visualized using chemiluminescence exposed to x-ray film. B) Densitometric analysis of the western blot exposure using ImageJ software.

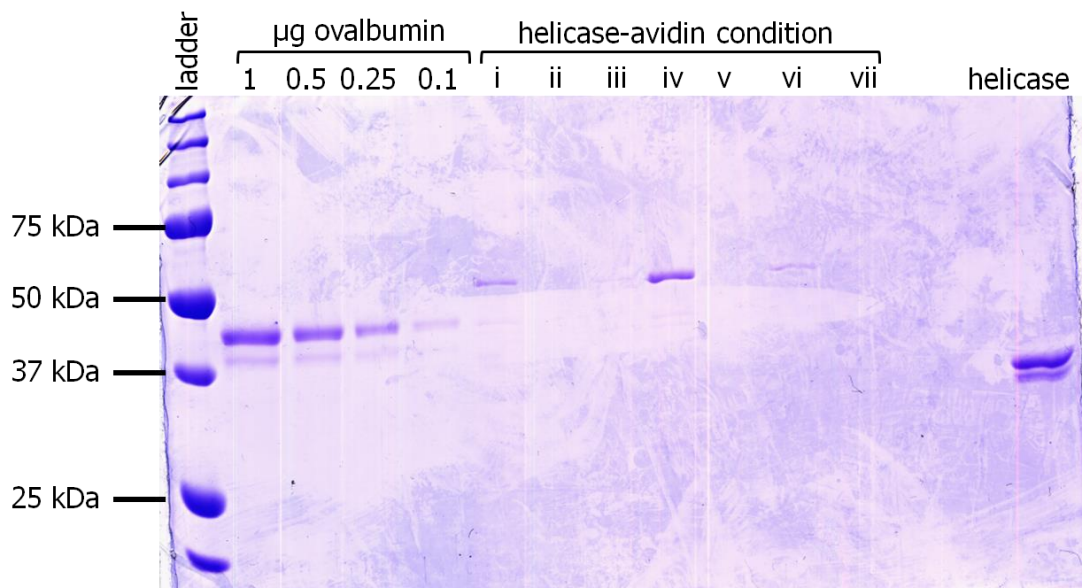


Figure 4.3: Helicase-avidin purification via various denaturing/renaturing conditions.

Protein from seven denaturing/renaturing conditions and a range of concentrations of ovalbumin were loaded on a 10% SDS PAGE along with a sample of purified helicase protein. The seven conditions tested were as follows: **(i)** denatured with urea with two resuspensions of bacterial pellet, eluted with β -mercaptoethanol and NP40, renatured by overnight dialysis at pH 7.5, **(ii)** denatured with urea with two resuspensions, eluted with β -mercaptoethanol and NP40, renatured by overnight dialysis at pH 9.0, **(iii)** denatured with urea, renatured on resin with washes of 2M, 1M, 0.5M and 0M urea, **(iv)** denatured with urea, renatured on resin with washes containing β -mercaptoethanol and NP40 at urea concentrations of 2M, 1.3M, 1M, 0.5M and 0M on the resin, **(v,vi,vii)** denatured with guanidium thiocyanate, renatured by dialysis at pH 7.5 with small batches (v), at pH 9 with small batches (vi) and overnight at pH 9 (vii).

4.5 Discussion

The purpose of this chapter was to purify the helicase-avidin fusion protein from a bacterial expression cell line. The protein was able to be induced in the bacteria, however, only a small amount of induced protein was eluting from the resin. This result indicated that the majority of the overexpressed protein was not being folded properly and remained in the insoluble fraction. This was further confirmed via Western blot analysis of the poorly purified protein samples (data not shown). Western blot analysis of the soluble fraction from various bacterial growth conditions (Figure 4.2) showed that these different growth conditions yielded varying relative amounts of soluble protein in the respective cell lines, namely BL21 being less efficient than ER2566 which is in turn less efficient than RIL for this protein, as well as enhanced expression when grown at lower temperatures. Thus, various means of altering and optimizing bacterial

growth conditions were attempted in order to enhance and shift more protein into the soluble fraction. Unfortunately, this did not yield results significant enough for a strong purification, therefore purification under denaturing conditions and renaturing the overexpressed protein in vitro was attempted.

The induced protein did not elute equally from the different resins tested. Rather, it had the most difficulty eluting from the nickel-based resins. This is likely due to the fact that the bond between nickel and the His-tag is much stronger than cobalt, due to their positioning on the periodic table, nickel being positioned to the right of cobalt within the transition metals, having just one more electron than cobalt. Copper, another atom with a strong affinity for the His-tag but which was not used, is positioned even further to the right on the periodic table, and possess an even stronger affinity for His-residues, possesses one electron more than nickel and two more than copper.

It is also interesting to note that in many of the purifications under denaturing conditions where the protein was renatured on the column, the protein did not elute from the column at all. It is possible that the poorly folded structure interacts more strongly with the resin, perhaps due to potentially exposed His residues or other exposed charges that would be folded within the molecule when properly folded. Subsequent attempts at reproducing the soluble purification of helicase-avidin elution were not efficient in the nickel-based resins, and were enhanced by using the Talon cobalt resin. Therefore, cobalt-based resins are preferable for this purification.

Ultimately, the purification of seemingly soluble protein often did not yield a soluble result after freezing and thawing for storage, despite proper storage conditions. This could be due to the fact that, for renatured proteins, if small amounts of misfolded proteins were present, they could cause properly folded proteins to misfold and aggregate because of the charges they possess [125]. Alternatively, the freeze-thaw process could cause the protein to unfold if the protein folding is not stable or due to pH changes in the solution due to temperature change [126]. In order to circumvent these problems, the protein was (a) filtered and centrifuged prior to freezing to increase the amount of properly folded protein in the sample and (b) centrifuged and

loaded onto an SDS PAGE and stained with Coomassie dye to verify the presence of a stable soluble protein, as shown in Figure 4.3.

The choice of denaturing conditions played a big role in the resulting protein stability and solubility. Experiments showed that urea was the better denaturant compared to guanidium thiocyanate (GT), possibly due to the fact that the two agents have different mechanisms by which they denature proteins [127]. It is possible that slowly removing the hydrogen-bonds formed with urea during denaturation allows for proper folding, which does not occur with GT which does not form hydrogen bonds during the process [127]. The GT denatured samples often did not elute at all when attempting renaturing on the resin.

Of the 49 ways of obtaining a soluble protein attempted, a total of 3 trials were somewhat successful after all stages, from bacterial growth, to freeze-thaw, were complete (Figure 4.3). Because two of those had a low product yield, only one sample (condition iv) was used for further analysis as it was deemed to be most likely to be active and functional. The other two conditions yielded results that were too dilute for further experimentation. These samples could have been concentrated, however, that would require further time consuming steps, added cost, and therefore was not desirable for our intended purposes.

An unexpectedly critical step to obtaining a soluble helicase-avidin yield was that of resuspending the pellet first under non-denaturing conditions followed by denaturing conditions, which were the only conditions that yielded any positive results. The reasoning behind this method was that removing the presence of other proteins would provide room for the misfolded protein to refold, without being obstructed.

4.6 Conclusion

The engineered helicase-avidin fusion protein can be stably expressed and rendered soluble when overexpressed in a bacterial host and isolated.

5 – Aim 3: Testing Helicase Functionality

5.1 Rationale

While it is very important to purify the proteins of interest in a bacterial host and obtain a soluble product, the purified proteins must be tested for functionality. To this end, the functionality of the helicase portion of the purified and solubilised helicase-avidin protein, and of helicase alone, was performed. Since helicase's NTPase, helicase, and translocation activities, as well as hexamer formation, are coupled to its ability to bind DNA [52], this was accomplished by assessing its capacity to bind to DNA via electrophoretic mobility shift assays (EMSA).

5.2 Purpose

The purpose of this chapter is to verify the functionality of the purified helicase proteins by assessing their ability to bind a DNA substrate.

5.3 Methods

Detailed information of the methods performed for Aim 3 can be found in Appendix III.viii.

5.3.1 Electrophoretic Mobility Shift Assay

EMSA was used to assess the DNA binding ability of the recombinant helicase and helicase-avidin proteins. Commercially available artificially forked DNA substrates that were radioactively labeled with ^{32}P at one end were used (Table 5.1). DNA binding reactions were conducted at room temperature in the helicase reaction mixture (Table 5.2) containing the purified helicase or helicase-avidin protein and the ^{32}P -labeled DNA. See appendix III.viii for information on helicase reaction mixture condition choices, DNA probe labeling methodology, and specific EMSA methodology

Table 5.1: Forked DNA substrate. ^{32}P labeled and complementary strands of forked DNA used for EMSA. Artificial fork regions are italicized and complementary sequences are in regular font.

^{32}P labeled	5' – <i>GCGCGCGCGCGCGCGCG</i> ATTGCGCGAATAGAGCTAGCTAGTCGAAGT – 3'
Complementary	5' – ACTTCGACTAGCTAGCTCTATTCGCGCAAT <i>ATATATA</i> – 3'

Table 5.2: Helicase reaction mixture. Reagents, concentrations and volumes of various components in the helicase reaction mixture used in this work, with role of the component provided.

Reagent	Final concentration	Volume (μL)	Role of Ingredient
1M DTT	10 mM	0.1	Stabilizes enzymes / proteins with free sulfhydryl groups [73]
DNA probe	varying	2 (varying)	Helicase's substrate
1M MgCl ₂	10 mM	0.1	Aids binding ssDNA [73] and dTTP hydrolysis and binding [31]
0.5M Hepes pH 7.0	20 mM	0.4	Maintains pH
100 mM dTTP	10 mM	1	DNA binding and protein oligomerization [31, 73, 101]
5M NaCl	50 mM	0.1	
1ng/ μL BSA	1 pg/mL	0.5	Prevents degradation of protein [73]
50% glycerol	2.5%	0.5	Stabilizes enzyme [128]
Protein	Varying	Varying	
ddH ₂ O		Varying	Nucleophile [31, 39]
Total		10	

5.3.2 Crosslinker optimization

MyoD protein was incubated with ³²P labeled “E-box” DNA sequence (Table 5.3) as a control for experimental functioning and cross-linker reaction optimization. The effect of various crosslinkers on MyoD binding was assessed by incubating the DNA with or without MyoD and the crosslinking agents: glutaraldehyde, formaldehyde or disuccinimidyl suberate (DSS), at different concentrations and incubation times.

Table 5.3: E-box DNA Sequence.

³² P labeled strand	5' – GATCCCCCAACACCTGCTGCCTGA – 3'
Complementary strand	5' – TCAGGCAGCAGGTGTTGGGGGGATC – 3'

5.3.3 Helicase-DNA binding

Helicase or helicase-avidin were incubated with the ³²P labeled artificially forked DNA in a Hepes-based helicase reaction mixture (Table 5.2) at room temperature for five minutes. The reactions were crosslinked with 0.02% formaldehyde for 30 seconds quenched with 5μL of 2.5 mM glycine. Various amounts of helicase (0-300 ng) and helicase-avidin (0-700 ng) were investigated over a range of incubation times (0, 2, 4, 6, 8, and 10 minutes).

5.3.4 Native Polyacrylamide Gel Electrophoresis

5% native acrylamide gels were pre-run at 100V for one hour. Reactions were mixed with loading dye and loaded onto the gels and run at 160V for 2 hours and 30

minutes. Gels were transferred to Whatman paper and dried for 1 hour at 80°C, and exposed to X-ray film at -80 °C overnight.

5.4 Results

Since helicase's binding to DNA is transient (binds to, unwinds, and then releases from DNA), the reaction must be stopped in time in order to be visualized. This can be effectuated using crosslinkers. Three crosslinking agents are known to crosslink DNA and helicase: glutaraldehyde [52], formaldehyde [129, 130], and dithiobis[succinimidyl propionate] (DSP) [35], which is very similar in structure to DSS. Glutaraldehyde, formaldehyde, and DSS were tested on MyoD's stable binding to the E-box sequence (Figure 5.1). Initial tests on crosslinkers were performed at times and concentrations prescribed by the original sources [35, 52, 129, 130] as outlined in Table III.6 in Appendix III (Figure 5.1A). The band at the bottom of the gel represents the radioactively labeled E-box DNA. When MyoD protein was added to the reaction, a shifted band appears (the middle band) indicating a DNA-protein interaction. The reactions with 0.02% glutaraldehyde for 20 or 60 seconds and 2.5 mM DSS for 4 minutes did not cause a band shift for protein-DNA interaction. Incubation with 0.365% formaldehyde, however, displayed a shift where the protein and DNA mixture are stuck in the wells of the native polyacrylamide gel (top band).

The conditions of formaldehyde incubation were altered to optimize crosslinking by reducing the concentration (0.02% and 0.04%) and the time of treatment (20s, 60s, and 4 minutes) (Figure 5.1B). The lowest concentration of formaldehyde tested (0.02%) successfully caused a shift at all times tested. However, a shift is only visible at 20 seconds with the higher concentration (0.04%) and is stuck in the wells at the longer incubation time. Therefore, a concentration of 0.02% formaldehyde incubated for 30s was determined to be appropriate crosslinker conditions to use for the helicase reaction.

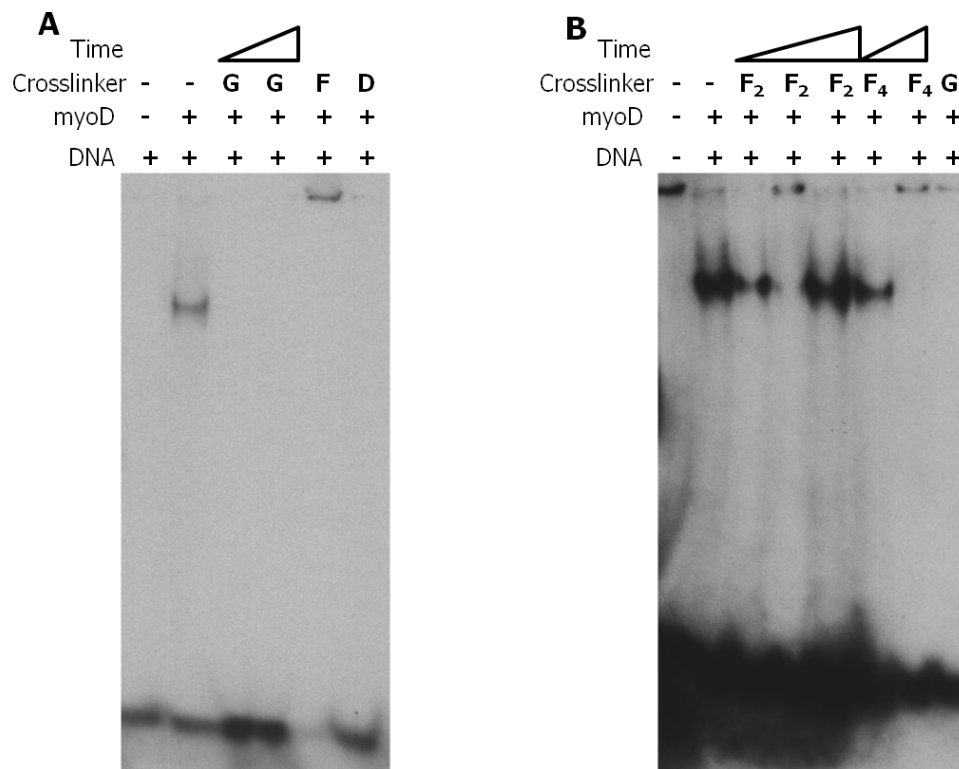


Figure 5.1: Effect of crosslinkers on MyoD binding. A) MyoD protein bound to E-box DNA sequence, with or without MyoD and crosslinking agent. G represents 0.02% glutaraldehyde at 20s (left) and 60s (right), F represents 0.365% formaldehyde for 4 minutes and D represents 2.5 mM DSS for 4 minutes. B) Further investigation on effects of formaldehyde. F₂ represents formaldehyde at 0.02% with times of 20s, 60s and 4 minutes (left to right), F₄ represents formaldehyde at 0.04% with times at 20s (left) and 60s (right). G represents 0.02% glutaraldehyde for 30 seconds. Alternate exposures can be found in Appendix IV, Figure A4.

Various amounts of helicase were incubated with artificially forked dsDNA (Figure 5.2A) and for a range of incubation times (Figure 5.2B). The dsDNA in the helicase reaction mixture buffer created a pattern of three distinct bands on the gel when loaded as a control, in addition to the dsDNA that remained in the wells of the gel at the top and a triangular shaped (inverted V) band that migrated with the loading dye at the bottom of the gel. When the dsDNA probe is loaded onto the gel without the helicase reaction buffer, the triangular shaped band is not present (Appendix IV-Figure A2). When helicase is added to the reaction mixture a fourth band appears within the gel above the three free probe dsDNA bands, indicated by a black arrow in Figure 5.2A. As the amount of helicase incubated with the dsDNA probe is increased, the intensity of the shifted band increases as well, and the amount of dsDNA probe stuck in the wells is

decreased. The effect of time was also investigated on the helicase reaction. 0.2 mg helicase was incubated with dsDNA for 0, 2, 4, 6, 8, and 10 minutes (Figure 5.2B). In the negative control reaction, only the three characteristic bands are present and the band of dsDNA probe stuck in the well of the gel is strong. The helicase reaction showed a shifted band for all time points, equally strongly, as well as displacing some of the signal of probe stuck within the wells to the gel creating bands of stronger intensity within the gel. Therefore, the purified helicase protein is capable of binding DNA at all concentrations and incubation times investigated.

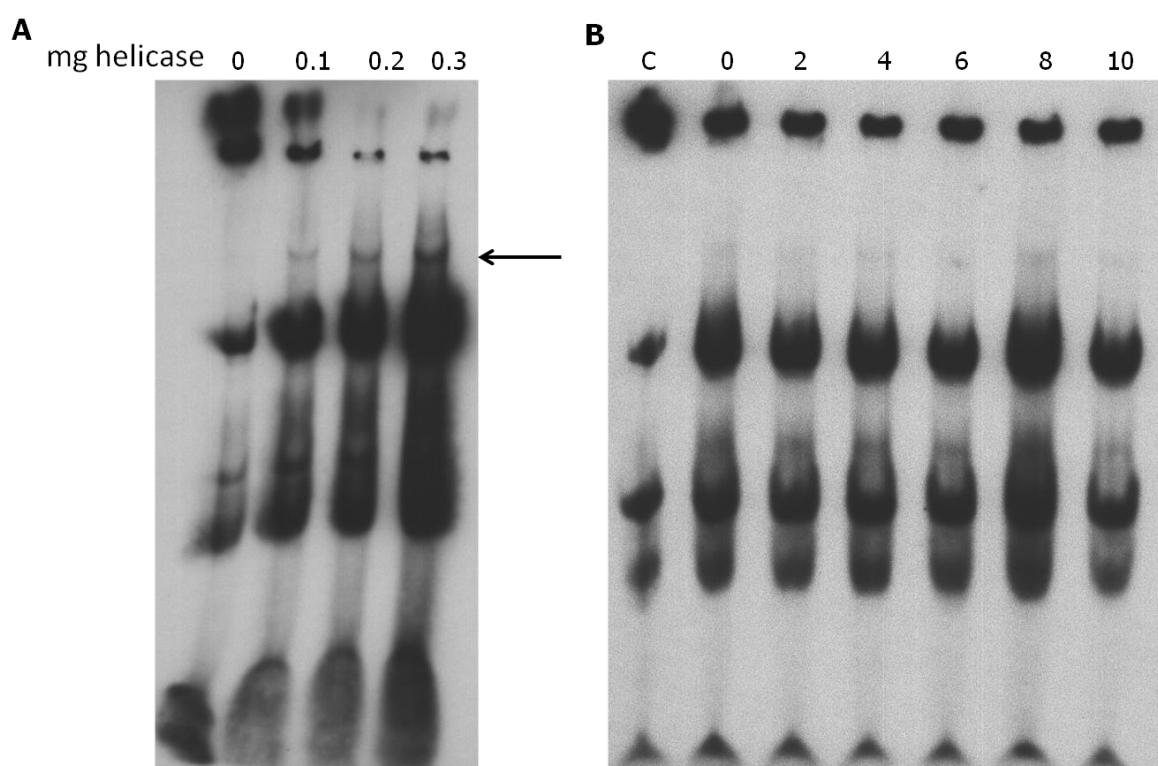


Figure 5.2: Helicase binding to DNA. A) 0.1, 0.2 and 0.3 mg helicase protein were incubated with artificially forked DNA. B) 0.2 mg of helicase was incubated with 0.75 μ L of radioactively labeled artificially forked dsDNA. C represents control which was dsDNA in the buffer without protein added. Numbered wells represent the time of incubation in minutes. The reactions were crosslinked with formaldehyde, loaded onto a 5% native acrylamide gel eletrophoresis and exposed to X-ray film for visualization. Alternate exposures can be found in Appendix IV, Figure A5.

Various amounts of purified helicase-avidin were incubated with the artificially forked dsDNA (Figure 5.3A) and for a range of incubation times (Figure 5.3B). As with the helicase reaction blots, the dsDNA created a pattern of three distinct bands on the

gel when loaded as a control. The triangular shaped bands which migrate with the loading dye forms a continuous line at the bottom of the gel because two adjacent lanes were loaded next to each other. When helicase-avidin was added to the reaction mixture, similar to what was observed with the helicase reactions, a fourth band appears within the gel, indicated by a black arrow in Figure 5.3B, however, the shift of the dsDNA probe stuck in the wells into the gel is less apparent, except at very high concentrations of fusion protein. Higher concentrations of helicase-avidin also change the migration of the triangular buffer band and the migration of the loading dye that appears further up the gel. The helicase-avidin band shift appears lower, ie migrates further, than the helicase band shift.

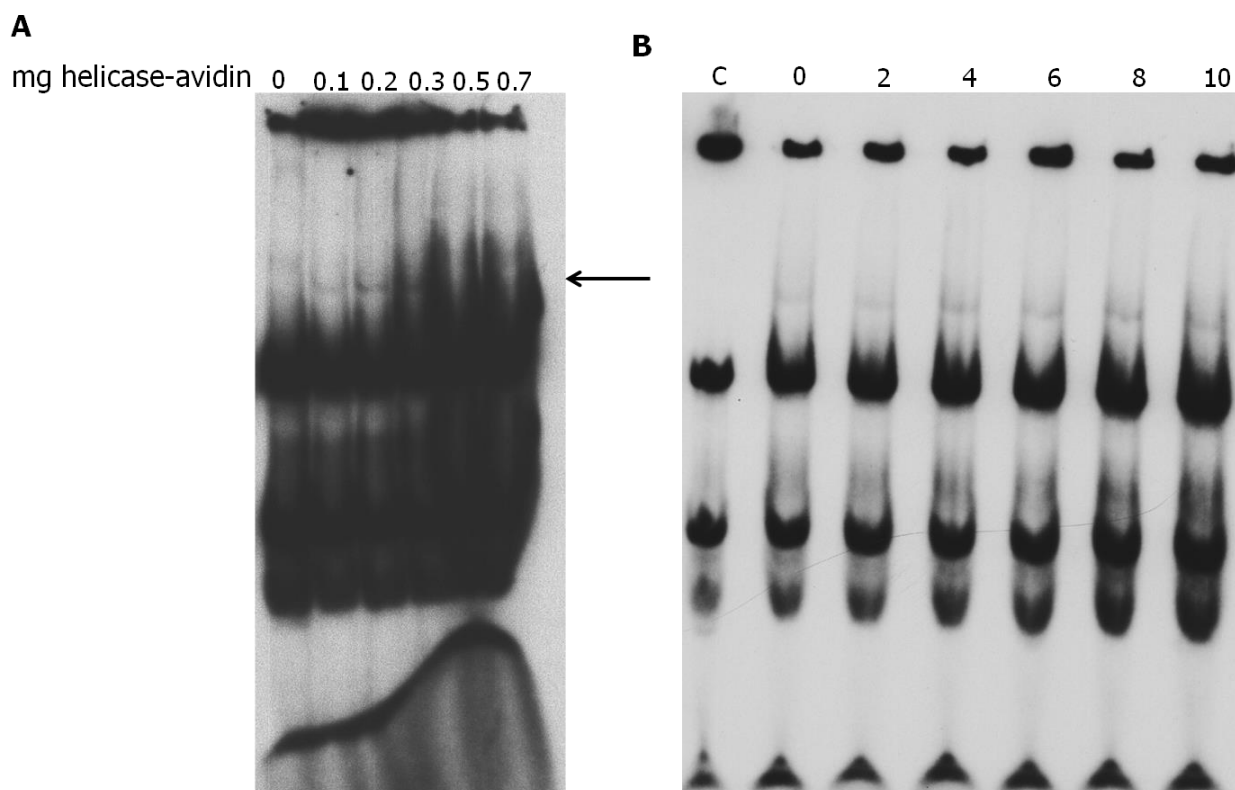


Figure 5.3: Helicase-avidin binding to DNA. A) 0.1, 0.2, 0.3, 0.5 and 0.7 mg helicase proteins were incubated with artificially forked DNA, B) 0.3 mg of helicase-avidin was incubated with 2 μ L of radioactively labeled artificially forked dsDNA for 5 minutes in a 25°C waterbath. C represents control which was dsDNA in the buffer without protein added. Numbered wells represent the time of incubation in minutes. The reactions were crosslinked with formaldehyde, loaded onto a 5% native acrylamide gel for eletrophoresis, and exposed to X-ray film for visualization. Alternate exposures can be found in Appendix IV, Figure A6.

In order to assess the effects of time on the helicase reaction, 0.3 mg of helicase-avidin was incubated with dsDNA for 0, 2, 4, 6, 8, and 10 minutes (Figure 5.3B). As above, the control reaction is missing the 4th shifted band above the free probe bands. Both helicase and helicase-avidin reactions showed a shifted band for all time points, equally strong, along with the displaced signal of probe stuck in the wells. Thus, the two purified proteins are capable of binding DNA at all concentrations and incubation times investigated.

5.5 Discussion

In this study, the ability of the purified helicase and helicase-avidin protein to bind DNA was assessed. Due to the transient nature of the reaction, the analysis of the three crosslinkers with MyoD and the E-box dsDNA probe was essential to the success of the experiment. If the experimental conditions cited from the literature were not capable of showing the interaction between MyoD and DNA, they would provide false negative results for helicase activity if applied to the helicase and helicase-avidin reactions. Optimization of these methodologies was necessary to find an optimal crosslinking reaction for the helicase tests, ultimately involving the concentrations used in the glutaraldehyde studies with formaldehyde incubation. The interaction between MyoD and E-box does not require crosslinkers to visualize by EMSA because the interaction is stable. Therefore, the optimization experiments only showed that the experiment could only show if crosslinking was efficient if concentrations or incubation time were too high, as with the initial formaldehyde assay (Figure 5.1A). MyoD binding to the E-box sequence was disrupted with the addition of glutaraldehyde and DSS, exemplified by the loss of the shifted band representing MyoD bound to the labeled E-box probe. This was unexpected due to the stability of the MyoD_E-box bond, and was essential to the success of the helicase experiments.

Formaldehyde at higher concentrations created a signal of probe that was stuck in the wells of the gel (Figure 5.1A). This was explored further as a good sign and indicator that too much formaldehyde had been used crosslinking not only MyoD to the DNA probe, but also to other MyoD molecules present creating aggregates that were too large to enter the gel. Thus formaldehyde concentrations were lowered and the

MyoD_E-box interaction was maintained (Figure 5.1B). Glutaraldehyde, a crosslinker similar to formaldehyde, but less strong, and DSS, an unrelated crosslinker, both caused a disruption in the MyoD_E-box interaction. This was not anticipated. It is possible that this occurred because the reagents denatured the proteins, or disrupted the binding reaction sterically or through charge interactions.

The banding pattern created by the artificially forked dsDNA may represent unannealed radioactively labeled ssDNA, dsDNA, and various interactions between DNA strands, such as the DNA fork sterically interacting with free ssDNA or dsDNA creating larger structures. Such larger structures likely get stuck in the wells, unable to migrate into the native gel. Loading dsDNA alone without buffer (but with loading dye) produces the same band pattern as dsDNA loaded with helicase buffer and increasing the salt content of the buffer to 150 mM of NaCl (Appendix IV – Figure A1).

The addition of buffer to the dsDNA creates an extra signal band that migrates with the loading dye in a triangular shaped band at the bottom of the gel. This band may contain the excess free ^{32}P that did not interact with the DNA and then separated from the mix due to the ionic charges in the buffer. It is also possible that the dsDNA migrates out of the wells and into the gel upon addition of the helicase protein for the same reason, however addition of high concentrations of NaCl does not yield the same effect (Appendix IV – Figure A1), indicating that it is not likely that the ionic presence is what causes the DNA aggregates to enter the gel, but that the protein interaction is real.

Another explanation for the DNA probe stuck in the well (topmost band in the gel) could be uneven or poor polymerization of the native polyacrylamide gel. However, filtration of the buffers, vortexing of the gel mixture, longer periods of waiting for the polymerization reaction to complete, freshly made reagents, different running buffer strengths (0.5x or 0.25x TBE), different gel percentages (5% and 6%), longer pre-running of gels and different pre-run conditions, among other changed aspects, did not affect any apparent experimental result including apparent band shifts, and migration of DNA aggregates in the wells into the gel (data not shown). Only the addition of larger

quantities of protein increased the signal of the shifted band, at specific locations for helicase and helicase-avidin.

Helicase and helicase-avidin, when bound to the dsDNA probe, each produced a unique band shift above the 3 free probe bands within the gel. This shift represents the interaction or binding of the protein with the dsDNA. The band shift for helicase alone was at a different relative migration to the band shift for helicase-avidin, which migrated further along the gel (closer to the top free probe band). This was seen consistently with all replicates of the experiment (data not shown) and was therefore not an artifact of running the experiments on two separate gels as displayed here. Helicase-avidin is larger in molecular weight than helicase alone, however, it migrated further along the gel than helicase did because native polyacrylamide gels do not denature the proteins and therefore maintain the secondary and tertiary structures. Thus, electrophoretic separation is not based on molecular weight as with SDS PAGE, rather it is based on molecular size and conformation. Therefore, whatever conformation the crosslinked helicase-avidin-DNA complex takes is perhaps more compact and/or possesses more external charge than helicase-DNA, allowing for its faster migration through the gel. This is unlikely to have any significant effect on a proposed sequencing technique.

The apparent shifted band for the two proteins likely represents the formation of the helicase hexamer at the forked dsDNA substrate. It is possible for helicase to interact with the DNA probe as monomers, dimers, or tetramers, or to be only partially crosslinked to the DNA. If this were the case within these experiments, it is possible that they would migrate at the same size as some of the free probe DNA bands. This could explain why the addition of helicase or helicase-avidin redistributes signal from the DNA probe aggregates stuck in the wells into the shifted 4th band, as well as to the “free probe” bands; these bands may contain DNA complexed with proteins. The addition of high concentrations of NaCl (150mM) to the reaction after the crosslinking has been quenched does not intensify the free probe bands nor add to the shifted band, corroborating this theory. It also helps to explain why the shifted band is so faint. If the helicase-DNA interaction includes monomeric, dimeric, or other interactions with the

dsDNA that are hidden in the free probe banding pattern, then not all of the helicase is forming hexamers, or not all of the hexamers are crosslinked properly. This could be visualized by a very strong shift in the banding pattern. However, as the amount of helicase or helicase-avidin incubated with the dsDNA probe is increased, the intensity of the shifted band increases as well. Thus, more protein is interacting with the dsDNA as a hexamer, intensifying the band shift.

The helicase-avidin protein was purified under denaturing conditions and was renatured on the resin in the presence of the detergent NP40 and the reducing agent β -mercaptoethanol. The purified helicase-avidin protein was subsequently dialyzed to exchange the elution buffer for the same storage buffer as the helicase protein, however, these two molecules are not permeable to the dialysis membrane and thus would remain in the stored solution. Small amounts of NP40 and β -mercaptoethanol, representing the same final concentration in the helicase reaction mixture as found in the helicase-avidin protein, were added to helicase trials to determine if they had any effect of the helicase-avidin binding to the DNA probe (data not shown). There was no significant effect on the binding of helicase to the DNA probe with the addition of these two molecules. Therefore small amounts of NP40 and β -mercaptoethanol do not have an effect on helicase binding to DNA. It is possible that other impurities in the solution will not affect helicase activity, which would be good for the potential use of these proteins in a DNA sequencing device as DNA samples containing impurities could still be sequenced.

The helicase-avidin protein, although a fusion protein and renatured *in vitro*, binds to the DNA probe, which indicates that it is likely catalytically active due to the fact that the various catalytic activities of helicase are coupled to DNA binding. The fact that it was renatured *in vitro* could raise concern of its catalytic activity because of potential misfolding in the process. One protein in a fusion protein may interfere with the activity of another due to proximity or interference with active sites, but this was not the case with helicase-avidin. This possibility was avoided by including extra linker amino acids in the sequences and by inserting avidin at the N-terminus of the helicase sequence, replacing the natural location of T7gp4 primase.

Helicase and helicase-avidin were incubated with the DNA probe at fixed concentrations from 0 to 10 minutes with the aim of visualizing helicase binding, unwinding and release from the DNA probe (Figure 5.3B). Theoretically, stopping the reaction with crosslinkers early on would show unbound probe, stopping it at an optimal timing would show bound probe, and stopping it after the helicase has translocated along the DNA would show unbound probe again. This, however, was not apparent in the results. The expected result would have been to see the band shift slowly appear, reaching a peak in intensity, and then disappear over the course of time. The observed result, however, is a band of steady intensity over the course of time. This may be due to malfunctioning NTPase activity or that a complete hexamer is not forming at the fork of DNA. Poor hexamer formation around the fork could be due to the steric interactions of the forks that are a possible cause of the multiple bands of free DNA probe. Kinetic assays for helicase functioning, as described in section 7.3, should be used to determine helicase NTPase activity.

In the literature, T7 helicase has been shown to be able to be pre-loaded onto DNA by omitting Mg^{2+} from the reaction mixture, adding it only to initiate unwinding [31, 39, 73]. This was attempted experimentally (data not shown) but the results were not clear. With the helicase protein, both the time course and pre-loading with varying concentrations of helicase, and various temperature conditions yielded results similar to Figure 5.2A. Experiments using Mg^{2+} to monitor helicase-avidin fusion protein (Appendix IV – Figure A4) yielded smears under pre-loading conditions, and bizarre overall increases in signal over time for single concentrations of helicase-avidin with Mg^{2+} added for different lengths of time despite the use of a master mix. These strange results were not isolated and occurred for several trials of Mg^{2+} pre-loading experiments. The lack of Mg^{2+} in the reaction mixture may hinder the ability of helicase-avidin to bind to the DNA probe, but further experiments using different means, as discussed in section 7.3, would better assess this.

5.6 Conclusion

The isolated and stable helicase and helicase-avidin fusion proteins bind to an artificially forked dsDNA probe as exemplified by EMSA.

6 – Aim 4: Testing Avidin Functionality

6.1 Rationale

One of the purified proteins being analyzed in this work is a helicase-avidin fusion protein. In order to verify that the recombinant protein is functioning, both subunits must be assayed: helicase and avidin. Helicase functioning was investigated in chapter 5. A functional avidin portion that is able to bind to the biotin molecule is also desired. If the protein has not folded properly, it will not bind to biotin, therefore the biotin binding ability of avidin had to be verified.

6.2 Purpose

The purpose of this chapter is to verify the functionality of the avidin portion of the helicase-avidin recombinant protein.

6.3 Methods

The biotin binding activity of avidin was assessed using a biotinylated resin. The experiment was conducted with minute amounts of protein and resin, therefore, many steps were taken to concentrate these small amounts for further analysis by SDS PAGE. Helicase-avidin was incubated with 10 μ l of the biotinylated resin for 1 hour. A disposable column was used to retain the small amount of resin and to facilitate washing and eluting. The resin was washed 10 times with 1x Phosphate Buffered Saline (PBS), and the final wash was concentrated using an Amicon™ centrifugal filter unit for further analysis. The resin was incubated with 0.1M Glycine-HCl at pH 2.8 to elute the protein bound from the resin three times for 5 minutes each at room temperature. The elution was concentrated using an Amicon™ centrifugal filter unit for further analysis. Detailed explanation of the methodology can be found in Appendix III.ix. The samples were loaded onto a 10% SDS PAGE run at 120V for 1 hour and 20 minutes and stained with Coomassie Brilliant Blue dye to visualize the proteins.

6.4 Results

A simple biotin binding assay was conducted to determine if the recombinant avidin is able to bind a biotin molecule attached to a solid support. Figure 6.1 shows a representative blot from a complete biotin binding assay. Ovalbumin was used as a negative control to account for false positive interactions. The input lane for helicase-

avidin and ovalbumin shows that protein was indeed loaded onto the column. No protein was found in the flow-through from the helicase-avidin mixture, however, for ovalbumin the flow-through contained a very strong protein band. For both sets of proteins the final wash did not contain any protein. The helicase-avidin elutions contained the protein that was bound to the resin during the flow-through and wash steps, whereas the ovalbumin samples did not. Neither resin contained any residual bound protein.

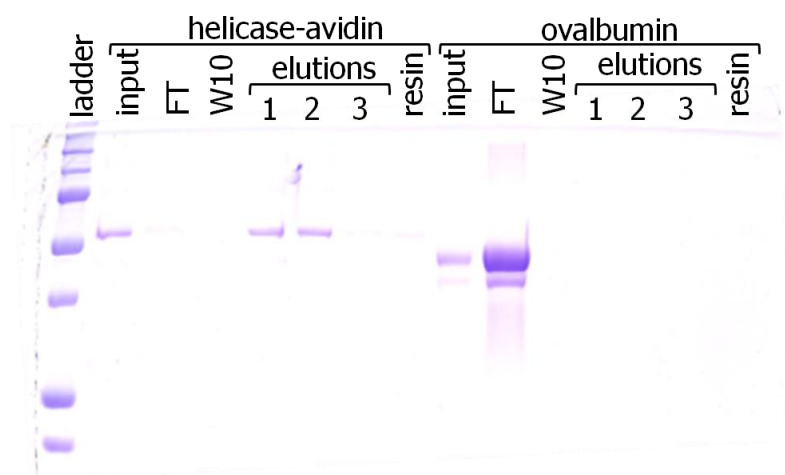


Figure 6.1: Avidin-biotin binding assay. Purified and renatured helicase-avidin protein was incubated with biotinylated resin for 1 hour. Input represents the input protein, FT the concentrated flow-through, W10 the 10th washing of the resin, elutions 1-3 the three elutions, and the remaining resin. Ovalbumin was used as a negative control of biotin binding using the same buffers. The samples were loaded on a 10% SDS PAGE and stained with coomassie brilliant blue dye.

6.5 Discussion

The helicase-avidin protein, when incubated with the biotinylated resin, was not washed out in the flow-through or washes of the resin, indicating a strong binding between the protein and the biotin molecules. Only when the pH of the solution was brought below pH 3, destabilizing the strong bond between avidin and biotin by denaturing the avidin, did the protein dissociate and elute. Two incubations with elution buffer at pH 2.8 were necessary to elute the protein, indicating a very strong bond between biotin and the solubilised avidin. To control for non-specific interactions between the proteins and the column, ovalbumin was used in parallel using the same buffers. Due to the similarity in molecular weight between helicase-avidin and

ovalbumin, this protein was selected as the control. After the same incubation period, ovalbumin did not bind to the biotinylated resin at all, completely washing out in the flow-through. Thus, the avidin portion of the helicase-avidin protein is functional in that it binds strongly to the biotin molecules on the resin.

6.6 Conclusion

The avidin portion of the solubilised helicase-avidin fusion protein has a strong affinity for biotin and is therefore functional.

7 – Key Results

7.1 Helicase Purification

Helicase protein was successfully purified from *E. coli*, using the Talon cobalt based histidine affinity resin.

- The helicase DNA sequence was ligated into the multiple cloning site of the pET28b(+) expression vector and successfully transformed in DH5 α bacterial cells.
- The subcloned plasmid was transformed into the expression cell line BL21.
- Helicase protein overexpression in the transformed BL21 cells was verified.
- The helicase protein was purified using the Talon cobalt based resin and dialyzed for storage.

Helicase purification from a bacterial host by a method which can easily be replicated on a large scale for commercial production was an important aim in this work for the integration of the helicase protein.

7.2 Helicase-avidin Purification

Helicase-avidin was successfully and solubly purified from *E. coli* by purification under denaturing conditions and renaturing *in vitro*.

- The Talon™ cobalt based resin yielded superior purification results compared with Qiagen™ and Biorad™ nickel based resins.
- Soluble over expressed helicase-avidin was optimized in the RIL bacterial strain, induced with 1 mM IPTG at 18°C overnight.
- Helicase-avidin was renatured and solubilized by urea purification. The successful means of obtaining soluble helicase-avidin involved the following:
 - resuspending the bacterial protein pellet in resuspension buffer without urea - centrifuging
 - resuspending the pellet of the first centrifugation in resuspension buffer with urea

- renaturing the protein bound to the resin by a series of washes of decreasing urea concentration containing β -mercaptoethanol and NP40 at pH 7.5.

Avidin protein was fused to helicase in order to create a recombinant protein that could potentially attract specific sequences of DNA to the helicase by means of biotinylated oligonucleotides. Obtaining a soluble and properly folded recombinant protein was an essential step towards this goal for the proposed DNA sequencing device.

7.3 Helicase Functionality

The purified helicase and helicase-avidin enzymes are capable of binding to a dsDNA probe as demonstrated by EMSA.

- The crosslinking reagent required to visualize helicase binding to dsDNA by EMSA is formaldehyde at 0.02% for 30s.
- Helicase and helicase-avidin bind to dsDNA, migrating at different rates within a native polyacrylamide gel electrophoresis.

It was essential to verify functioning of the isolated helicase and helicase-avidin proteins in order to move forward for further characterization. Helicase binding to DNA, as demonstrated by EMSA, is an indicator of proper hexamer formation and therefore can be further pursued for helicase unwinding and NTPase activities.

7.4 Avidin Functionality

The avidin portion of the purified solubilized helicase-avidin protein is functional.

- Avidin successfully and strongly bound to a biotin molecule attached to a solid support.

The helicase-avidin recombinant protein was engineered, purified and rendered soluble with the aim of attracting specific segments of DNA to the helicase enzyme for sequencing by means of biotinylated oligonucleotides. Therefore it was essential to confirm the avidin's ability to bind strongly to biotin, as demonstrated.

8 – Conclusions and Recommendations

8.1 Conclusion

DNA sequencing is an essential tool in various fields of study including personalized medicine, genetics, field sciences, forensic sciences, security, pharmaceutical research, biochemistry, molecular biology, and others. Many research groups seek to improve DNA sequencing techniques. A relatively inexpensive, rapid, accurate, smart, portable, hand-held, high-throughput, and integrated DNA sequencing device is sought out for development by the Medical Devices Innovation Institute.

In this thesis, several critical steps towards a novel helicase-based DNA sequencing technique being developed at the Medical Devices Innovation Institute have been achieved.

1. A functional bacteriophage T7 DNA helicase was isolated and purified from a bacterial host using means that are relatively easy, inexpensive, rapid and have a potential for large scale commercial production for use in a novel DNA sequencing method.
2. A recombinant protein containing helicase with DNA binding ability and functional avidin components that can be expressed in a bacterial host via cost effective and efficient means was engineered. This recombinant protein, capable of binding to DNA by helicase and to biotin by the avidin region, can be further explored as a potential means of targeting specific DNA sequences to the helicase molecule for sequencing *via* biotinylation.

8.2 Recommendations

This project must be explored further on several levels. 1) the helicase and helicase-avidin proteins that have been purified and characterized in this study must be further characterized, 2) the proteins must be integrated with a SWCNT for characterization within a functional prototype.

8.2.1 Helicase Characterization

Helicase and helicase-avidin require further characterization of catalytic activity and function prior to their integration in a prototype DNA sequencing device. This can be done by assays to detect kinetic activity. Kinetic activity refers to the movement and actions of an enzyme. Kinetic assays can be used to show helicase and helicase-avidin activity.

Helicase Assays

There are many assays available as options for helicase functionality testing, as described in Appendix V. Not all of these are suitable for the proof-of-concept requirements for this project, however the most suitable technique for this project is the reagentless biosensor for inorganic phosphate release [131, 132] which provides little interference with the system.

Heterohexamer Formation

Bacteriophage T7gp4 helicase forms homohexamers, as is studied in this work. However, it may be of interest to explore hexamer alterations such as mixtures of helicase and helicase-avidin to form heterohexamers, or exploring various means to obtain just one helicase-avidin monomer per hexamer.

DNA Substrates for Study

Changing the DNA probe being studied is another aspect to be considered in kinetic assays. Such DNA substrates could include larger/longer artificial forks, longer dsDNA sequences, ssDNA kinetics, and long ssDNAs with a partially complementary biotinylated ssDNA strand (as outlined in section 1.7).

Potential Concerns to Address

A potential concern is the formation of a loop as the biotinylated oligonucleotide remains bound to the avidin portion as the dsDNA is being unwound/threaded by the

helicase. This is unlikely to pose a real issue for two reasons. Firstly, the region of dsDNA being unwound would be relatively short and ssDNA would thread through for the majority of the time, leaving the biotinylated oligonucleotide simply hanging off of the helicase-avidin. Secondly, helicase may be capable of displacing avidin from biotinylated oligonucleotides [35] that would leave the short oligonucleotides to free float in solution while the ssDNA is read by the helicase.

8.2.2 Integration with Electronics for DNA Sequencing Device Prototype

Once helicase and helicase-avidin functionality has been further characterized, the protein can be integrated along with the SWCNT and characterized again for continued functionality. This attachment can be accomplished chemically [133] but the helicase-avidin protein provides another potential possibility for SWCNT attachment and the possibility of a reusable DNA sequencing chip. A SWCNT functionalized with biotin at the end instead of helicase directly could attract helicase-avidin monomers that are washed onto it. These can be the basis for hexamers formed to sequence DNA. Instead of inserting a new chip for the device for each read, the proteins can be denatured, by heat or pH, and washed off the biotin. Fresh enzyme can be washed over and attached to the biotinylated SWCNT for the next reaction.

9 – References

- [1] F. Sanger, S. Nicklen and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 74, pp. 5463-5467, 1977.
- [2] W. J. Ansorge, "Next-generation DNA sequencing techniques," *N. Biotechnol.*, vol. 25, pp. 195-203, 2009.
- [3] A. M. Maxam and W. Gilbert, "A new method for sequencing DNA," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 74, pp. 560-564, 1977.
- [4] E. Pettersson, J. Lundberg and A. Ahmadian, "Generations of sequencing technologies," *Genomics*, vol. 93, pp. 105-111, 2009.
- [5] N. B. DeWalch, *High Speed DNA Sequencer and Method*, 2007.
- [6] W. J. Greenleaf and S. M. Block, "Single-molecule, motion-based DNA sequencing using RNA polymerase," *Science*, vol. 313, pp. 801, 2006.
- [7] E. Y. Chan, "Advances in sequencing technology," *Mutat. Res.*, vol. 573, pp. 13-40, 2005.
- [8] P. Nyrén and A. Lundin, "Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis," *Anal. Biochem.*, vol. 151, pp. 504-509, 1985.
- [9] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen and P. Nyren, "Real-time DNA sequencing using detection of pyrophosphate release," *Anal. Biochem.*, vol. 242, pp. 84-89, 1996.
- [10] N. Pourmand, L. Munishkina and M. Karhanek, *Rapid DNA Sequencing by Peroxidative Reaction*, 2010.
- [11] J. Ju, Z. Li, J. R. Edwards and Y. Itagaki, *Massive Parallel Method for Decoding DNA and RNA*, 2009.
- [12] J. Ju, D. H. Kim, L. Bi, Q. Meng and X. Li, *Four-Color DNA Sequencing by Synthesis using Cleavable Fluorescent Nucleotide Reversible Terminators*, 2011.
- [13] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth and J. Bustillo, "An integrated semiconductor device enabling non-optical genome sequencing," *Nature*, vol. 475, pp. 348-352, 2011.
- [14] M. Muers, "Technology: Getting Moore from DNA sequencing," *Nat. Rev. Genet.*, vol. 12, pp. 586, 2011.

- [15] H. Esfandyarpour and M. Ronaghi, *Heat and pH Measurement for Sequencing of DNA*, 2011.
- [16] D. H. Densham, *DNA Sequencing Method*, 2005.
- [17] J. S. Shendure Hanlee Ji, "Next-generation DNA sequencing," *Nat. Biotechnol.*, vol. 26, pp. 1135-1145, 2008.
- [18] Shendure, Jay Porreca, Gregory Reppas, Nikos Lin, Xiaoxia McCutcheon, John Rosenbaum, Abraham Wang, Michael Zhang, Kun Mitra, Robi Church, George, "Accurate multiplex polony sequencing of an evolved bacterial genome." *Science*, vol. 309, pp. 1728-1732, 2005.
- [19] J. Eid and A. Dewinter, *Nucleic Acid Sequencing Methods and Systems*, 2011.
- [20] A. Fehr, *Helicase-Assisted Sequencing with Molecular Beacons*, 2009.
- [21] D. H. Densham, *Polynucleotide Sequencing using a Helicase*, 2007.
- [22] Z. Gao, A. Agarwal, A. D. Trigg, N. Singh, C. Fang, C. H. Tung, Y. Fan, K. D. Buddharaju and J. Kong, "Silicon nanowire arrays for label-free detection of DNA," *Anal. Chem.*, vol. 79, pp. 3291-3297, 2007.
- [23] Rafael Mulero, Anmiv S Prabhu, Kevin J. Freedman, Min Jun Kim, "Nanopore-Based Devices for Bioanalytical Applications," *Jala*, pp. 1, 2010.
- [24] J. R. Sauer and B. J. Van Zeghbroeck, *Ultra-Fast Nucleic Acid Sequencing Device and a Method for Making and using the Same*, 2002.
- [25] D. K. Shenoy, E. Dalcanale, S. Willems and M. Busi, *Synthetic Nanopores for DNA Sequencing*, 2010.
- [26] H. W. C. Postma, "Rapid sequencing of individual DNA molecules in graphene nanogaps," *Nano Letters*, vol. 10, pp. 420-425, 2010.
- [27] H. Liu, J. He, J. Tang, H. Liu, P. Pang, D. Cao, P. Krstic, S. Joseph, S. Lindsay and C. Nuckolls, "Translocation of single-stranded DNA through single-walled carbon nanotubes," *Science*, vol. 327, pp. 64, 2010.
- [28] D. J. Burns and K. Toumi, "Measuring, shortening and functionalizing carbon nanotube tipped AFM probes for DNA sequencing," in *SMA Symposium*, 2007, .
- [29] T. Mussivand, *Method and Apparatus for Collecting Cells for Macromolecular Analysis*, 2004.
- [30] T. Mussivand, B. I. Morshed and M. Shams, *Device for Electroporation and Lysis*, 2011.

- [31] K. M. Picha and S. S. Patel, "Bacteriophage T7 DNA helicase binds dTTP, forms hexamers, and binds DNA in the absence of Mg²⁺. The presence of dTTP is sufficient for hexamer formation and DNA binding," *J. Biol. Chem.*, vol. 273, pp. 27315-27319, 1998.
- [32] N. K. Williams, E. Liepinsh, S. J. Watt, P. Prosselkov, J. M. Matthews, P. Attard, J. L. Beck, N. E. Dixon and G. Otting, "Stabilization of native protein fold by intein-mediated covalent cyclization," *J. Mol. Biol.*, vol. 346, pp. 1095-1108, 2005.
- [33] L. E. Bird, K. Hakansson, H. Pan and D. B. Wigley, "Characterization and crystallization of the helicase domain of bacteriophage T7 gene 4 protein," *Nucleic Acids Res.*, vol. 25, pp. 2620-2626, 1997.
- [34] E. P. Diamandis, "The biotin-(strept) avidin system: principles and applications in biotechnology," *Clin. Chem.*, vol. 37, pp. 625, 1991.
- [35] P. D. Morris and K. D. Raney, "DNA Helicases Displace Streptavidin from Biotin-Labeled Oligonucleotides," *Biochemistry (N. Y.)*, vol. 38, pp. 5164-5171, 1999.
- [36] Y. J. Jeong, V. Rajagopal and S. S. Patel, "Switching from single-stranded to double-stranded DNA limits the unwinding processivity of ring-shaped T7 DNA helicase," *Nucleic Acids Res.*, vol. 41, pp. 4219-4229, 2013.
- [37] Y. J. Jeong, M. K. Levin and S. S. Patel, "The DNA-unwinding mechanism of the ring helicase of bacteriophage T7," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 7264-7269, 2004.
- [38] R. R. Xu, S. D. Qi, L. T. Lu, C. T. Chen, C. A. Wu and C. C. Zheng, "A DExD/H box RNA helicase is important for K⁺ deprivation responses and tolerance in *Arabidopsis thaliana*," *Febs j.*, vol. 278, pp. 2296-2306, 2011.
- [39] S. S. Patel and K. M. Picha, "Structure and function of hexameric helicases," *Annu. Rev. Biochem.*, vol. 69, pp. 651-697, 2000.
- [40] D. W. Kim, Y. Gwack, J. H. Han and J. Choe, "Towards defining a minimal functional domain for NTPase and RNA helicase activities of the hepatitis C virus NS3 protein," *Virus Res.*, vol. 49, pp. 17-25, 1997.
- [41] B. C. Valdez, D. Henning, K. Perumal and H. Busch, "RNA-unwinding and RNA-folding Activities of RNA Helicase II/Gu — Two Activities in Separate Domains of the Same Protein," *European Journal of Biochemistry*, vol. 250, pp. 800-807, 1997.
- [42] S. Zhang, "Nuclear DNA helicase II unwinds both DNA and RNA," *Biochemistry (N. Y.)*, vol. 33, pp. 3906, 1994.
- [43] M. Scheffner, R. Knippers and H. Stahl, "Simian-virus-40 large-T-antigen-catalyzed DNA and RNA unwinding reactions," *Eur. J. Biochem.*, vol. 195, pp. 49-54, 1991.
- [44] *Nucleic Acids Symposium Series*, vol. 35, pp. 5809, 2007.

- [45] M. T. Washington, A. H. Rosenberg, K. Griffin, F. W. Studier and S. S. Patel, "Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding," *J. Biol. Chem.*, vol. 271, pp. 26825-26834, 1996.
- [46] S. J. Lee and C. C. Richardson, "Choreography of bacteriophage T7 DNA replication," *Curr. Opin. Chem. Biol.*, vol. 15, pp. 580-586, 2011.
- [47] K. Hanada, M. Iwasaki, S. Ihashi and H. Ikeda, "UvrA and UvrB Suppress Illegitimate Recombination: Synergistic Action with RecQ Helicase," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, pp. 5989-5994, 2000.
- [48] M. R. Sawaya, S. Guo, S. Tabor, C. C. Richardson and T. Ellenberger, "Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7," *Cell*, vol. 99, pp. 167-177, 1999.
- [49] D. E. Kim, M. Narayan and S. S. Patel, "T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA," *J. Mol. Biol.*, vol. 321, pp. 807-819, 2002.
- [50] S. Guo, S. Tabor and C. C. Richardson, "The linker region between the helicase and primase domains of the bacteriophage T7 gene 4 protein is critical for hexamer formation," *J. Biol. Chem.*, vol. 274, pp. 30303-30309, 1999.
- [51] S. Ghosh, S. M. Hamdan, T. E. Cook and C. C. Richardson, "Interactions of Escherichia coli thioredoxin, the processivity factor, with bacteriophage T7 DNA polymerase and helicase," *J. Biol. Chem.*, vol. 283, pp. 32077-32084, 2008.
- [52] Y. Yong and L. J. Romano, "Nucleotide and DNA-induced conformational changes in the bacteriophage T7 gene 4 protein," *J. Biol. Chem.*, vol. 270, pp. 24509-24517, 1995.
- [53] M. R. Singleton, M. S. Dillingham and D. B. Wigley, "Structure and Mechanism of Helicases and Nucleic Acid Translocases," *Annu. Rev. Biochem.*, vol. 76, pp. 23-50, 2007.
- [54] P. Linder, "Dead-box proteins: a family affair--active and passive players in RNP-remodeling," *Nucleic Acids Res.*, vol. 34, pp. 4168-4180, 2006.
- [55] S. J. Lee, U. Qimron and C. C. Richardson, "Communication between subunits critical to DNA binding by hexameric helicase of bacteriophage T7," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 8908-8913, 2008.
- [56] S. M. Notarnicola, "A domain of the gene 4 helicase/primase of bacteriophage T7 required for the formation of an active hexamer," *J. Biol. Chem.*, vol. 270, pp. 20215, 1995.
- [57] D. J. Crampton, S. Mukherjee and C. C. Richardson, "DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase," *Mol. Cell*, vol. 21, pp. 165-174, 2006.

- [58] M. R. Singleton, "Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides," *Cell*, vol. 101, pp. 589, 2000.
- [59] S. Bailey, W. K. Eliason and T. A. Steitz, "Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase," *Science*, vol. 318, pp. 459-463, 2007.
- [60] J. Ye, A. R. Osborne, M. Groll and T. A. Rapoport, "RecA-like motor ATPases--lessons from structures," *Biochim. Biophys. Acta*, vol. 1659, pp. 1-18, 2004.
- [61] R. D. Schrock and B. Alberts, "Processivity of the gene 41 DNA helicase at the bacteriophage T4 DNA replication fork," *J. Biol. Chem.*, vol. 271, pp. 16678-16682, 1996.
- [62] E. Scherzinger, G. Ziegelin, M. Barcena, J. M. Carazo, R. Lurz and E. Lanka, "The RepA protein of plasmid RSF1010 is a replicative DNA helicase," *J. Biol. Chem.*, vol. 272, pp. 30228-30236, 1997.
- [63] W. Bujalowski, M. M. Klonowska and M. J. Jezewska, "Oligomeric structure of Escherichia coli primary replicative helicase DnaB protein," *J. Biol. Chem.*, vol. 269, pp. 31350-31358, 1994.
- [64] N. Ribeck, D. L. Kaplan, I. Bruck and O. A. Saleh, "DnaB helicase activity is modulated by DNA geometry and force," *Biophys. J.*, vol. 99, pp. 2170-2179, 2010.
- [65] M. Chandler, R. E. Bird and L. Caro, "The replication time of the Escherichia coli K12 chromosome as a function of cell doubling time," *J. Mol. Biol.*, vol. 94, pp. 127-132, 1975.
- [66] S. Kim, H. G. Dallmann, C. S. McHenry and K. J. Marians, "Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement," *Cell*, vol. 84, pp. 643-650, 1996.
- [67] R. Galletto, M. J. Jezewska and W. Bujalowski, "Unzipping mechanism of the double-stranded DNA unwinding by a hexameric helicase: the effect of the 3' arm and the stability of the dsDNA on the unwinding activity of the Escherichia coli DnaB helicase," *J. Mol. Biol.*, vol. 343, pp. 101-114, 2004.
- [68] G. Farge, "The N-terminal domain of TWINKLE contributes to single-stranded DNA binding and DNA helicase activities," *Nucleic Acids Res.*, vol. 36, pp. 393, 2008.
- [69] E. Sarzi, "Twinkle helicase (PEO1) gene mutation causes mitochondrial DNA depletion," *Ann. Neurol.*, vol. 62, pp. 579, 2007.
- [70] J. N. Spelbrink, F. Y. Li, V. Tiranti, K. Nikali, Q. P. Yuan, M. Tariq, S. Wanrooij, N. Garrido, G. Comi, L. Morandi, L. Santoro, A. Toscano, G. M. Fabrizi, H. Somer, R. Croxen, D. Beeson, J. Poulton, A. Suomalainen, H. T. Jacobs, M. Zeviani and C. Larsson, "Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria," *Nat. Genet.*, vol. 28, pp. 223-231, 2001.

- [71] E. H. Egelman, X. Yu, R. Wild, M. M. Hingorani and S. S. Patel, "Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, pp. 3869-3873, 1995.
- [72] S. J. Lee and C. C. Richardson, "Molecular basis for recognition of nucleoside triphosphate by gene 4 helicase of bacteriophage T7," *J. Biol. Chem.*, vol. 285, pp. 31462-31471, 2010.
- [73] S. W. Matson and C. C. Richardson, "Nucleotide-dependent binding of the gene 4 protein of bacteriophage T7 to single-stranded DNA," *J. Biol. Chem.*, vol. 260, pp. 2281-2287, 1985.
- [74] J. C. Liao, Y. J. Jeong, D. E. Kim, S. S. Patel and G. Oster, "Mechanochemistry of t7 DNA helicase," *J. Mol. Biol.*, vol. 350, pp. 452-475, 2005.
- [75] A. W. Kulczyk, "Molecular interactions in the priming complex of bacteriophage T7," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, pp. 9408, 2012.
- [76] S. M. Notarnicola, "The acidic carboxyl terminus of the bacteriophage T7 gene 4 helicase/primase interacts with T7 DNA polymerase," *J. Biol. Chem.*, vol. 272, pp. 18425, 1997.
- [77] S. J. Lee, B. Marintcheva, S. M. Hamdan and C. C. Richardson, "The C-terminal residues of bacteriophage T7 gene 4 helicase-primase coordinate helicase and DNA polymerase activities," *J. Biol. Chem.*, vol. 281, pp. 25841-25849, 2006.
- [78] Z. G. He and C. C. Richardson, "Effect of single-stranded DNA-binding proteins on the helicase and primase activities of the bacteriophage T7 gene 4 protein," *J. Biol. Chem.*, vol. 279, pp. 22190-22197, 2004.
- [79] Y. J. Jeong, "Kinetic pathway of dTTP hydrolysis by hexameric T7 helicase-primase in the absence of DNA," *J. Biol. Chem.*, vol. 277, pp. 43778, 2002.
- [80] A. K. Satapathy, A. B. Kochaniak, S. Mukherjee, D. J. Crampton, A. van Oijen and C. C. Richardson, "Residues in the central beta-hairpin of the DNA helicase of bacteriophage T7 are important in DNA unwinding," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, pp. 6782-6787, 2010.
- [81] A. K. Satapathy and C. C. Richardson, "The Glutamate Switch of Bacteriophage T7 DNA Helicase: ROLE IN COUPLING NUCLEOTIDE TRIPHOSPHATE (NTP) AND DNA BINDING TO NTP HYDROLYSIS," *J. Biol. Chem.*, vol. 286, pp. 23113-23120, 2011.
- [82] D. J. Crampton, S. Guo, D. E. Johnson and C. C. Richardson, "The arginine finger of bacteriophage T7 gene 4 helicase: role in energy coupling," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 4373-4378, 2004.
- [83] S. Kunzelmann and M. R. Webb, "A biosensor for fluorescent determination of ADP with high time resolution," *J. Biol. Chem.*, vol. 284, pp. 33130-33138, 2009.

- [84] G. Patel, D. S. Johnson, B. Sun, M. Pandey, X. Yu, E. H. Egelman, M. D. Wang and S. S. Patel, "A257T linker region mutant of T7 helicase-primase protein is defective in DNA loading and rescued by T7 DNA polymerase," *J. Biol. Chem.*, vol. 286, pp. 20490-20499, 2011.
- [85] S. S. Patel, "The K318A mutant of bacteriophage T7 DNA primase-helicase protein is deficient in helicase but not primase activity and inhibits primase-helicase protein wild-type activities by heterooligomer formation," *Biochemistry (N. Y.)*, vol. 33, pp. 7857, 1994.
- [86] S. J. Lee and C. C. Richardson, "The linker region between the helicase and primase domains of the gene 4 protein of bacteriophage T7. Role in helicase conformation and activity," *J. Biol. Chem.*, vol. 279, pp. 23384-23393, 2004.
- [87] A. K. Satapathy, D. J. Crampton, B. B. Beauchamp and C. C. Richardson, "Promiscuous usage of nucleotides by the DNA helicase of bacteriophage T7: determinants of nucleotide specificity," *J. Biol. Chem.*, vol. 284, pp. 14286-14295, 2009.
- [88] S. M. Notarnicola and C. C. Richardson, "The nucleotide binding site of the helicase/primase of bacteriophage T7. Interaction of mutant and wild-type proteins," *J. Biol. Chem.*, vol. 268, pp. 27198-27207, 1993.
- [89] L. V. Mendelman, S. M. Notarnicola and C. C. Richardson, "Evidence for distinct primase and helicase domains in the 63-kDa gene 4 protein of bacteriophage T7. Characterization of nucleotide binding site mutant," *J. Biol. Chem.*, vol. 268, pp. 27208-27213, 1993.
- [90] A. F. Neuwald, L. Aravind, J. L. Spouge and E. V. Koonin, "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes," *Genome Res.*, vol. 9, pp. 27-43, 1999.
- [91] T. Ogura and A. J. Wilkinson, "AAA+ superfamily ATPases: common structure--diverse function," *Genes Cells*, vol. 6, pp. 575-597, 2001.
- [92] J. A. Bernstein and C. C. Richardson, "Purification of the 56-kDa component of the bacteriophage T7 primase/helicase and characterization of its nucleoside 5'-triphosphatase activity," *J. Biol. Chem.*, vol. 263, pp. 14891-14899, 1988.
- [93] R. L. Lechner and C. C. Richardson, "A preformed, topologically stable replication fork. Characterization of leading strand DNA synthesis catalyzed by T7 DNA polymerase and T7 gene 4 protein," *J. Biol. Chem.*, vol. 258, pp. 11185-11196, 1983.
- [94] S. W. Matson and C. C. Richardson, "DNA-dependent nucleoside 5'-triphosphatase activity of the gene 4 protein of bacteriophage T7," *J. Biol. Chem.*, vol. 258, pp. 14009-14016, 1983.
- [95] S. W. Matson, S. Tabor and C. C. Richardson, "The gene 4 protein of bacteriophage T7. Characterization of helicase activity," *J. Biol. Chem.*, vol. 258, pp. 14017-14024, 1983.

- [96] H. Nakai and C. C. Richardson, "Leading and lagging strand synthesis at the replication fork of bacteriophage T7. Distinct properties of T7 gene 4 protein as a helicase and primase," *J. Biol. Chem.*, vol. 263, pp. 9818-9830, 1988.
- [97] E. A. Toth, "The crystal structure of the bifunctional primase-helicase of bacteriophage T7," *Mol. Cell*, vol. 12, pp. 1113, 2003.
- [98] J. A. Bernstein, "Characterization of the helicase and primase activities of the 63-kDa component of the bacteriophage T7 gene 4 protein." *J. Biol. Chem.*, vol. 264, pp. 13066, 1989.
- [99] D. Kong, N. G. Nossal and C. C. Richardson, "Role of the bacteriophage T7 and T4 single-stranded DNA-binding proteins in the formation of joint molecules and DNA helicase-catalyzed polar branch migration," *J. Biol. Chem.*, vol. 272, pp. 8380-8387, 1997.
- [100] R. Kolodner, Y. Masamune, J. E. LeClerc and C. C. Richardson, "Gene 4 protein of bacteriophage T7. Purification physical properties, and stimulation of T7 DNA polymerase during the elongation of polynucleotide chains," *J. Biol. Chem.*, vol. 253, pp. 566-573, 1978.
- [101] S. S. Patel and M. M. Hingorani, "Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins," *J. Biol. Chem.*, vol. 268, pp. 10668-10675, 1993.
- [102] M. R. Webb, "A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, pp. 4884-4887, 1992.
- [103] M. M. Hingorani and S. S. Patel, "Interactions of bacteriophage T7 DNA primase/helicase protein with single-stranded and double-stranded DNAs," *Biochemistry*, vol. 32, pp. 12478-12487, 1993.
- [104] K. J. Hacker, "A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding," *Biochemistry (N. Y.)*, vol. 36, pp. 14080, 1997.
- [105] P. Ahnert and S. S. Patel, "Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate," *J. Biol. Chem.*, vol. 272, pp. 32267-32273, 1997.
- [106] S. Tabor and C. C. Richardson, "Template recognition sequence for RNA primer synthesis by gene 4 protein of bacteriophage T7," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 78, pp. 205-209, 1981.
- [107] O. Livnah, E. A. Bayer, M. Wilchek and J. L. Sussman, "Three-dimensional structures of avidin and the avidin-biotin complex," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, pp. 5076-5080, 1993.

- [108] C. Kendall, "Utilization of the biotin/avidin system to amplify the sensitivity of the enzyme-linked immunosorbent assay (ELISA)." *J. Immunol. Methods*, vol. 56, pp. 329, 1983.
- [109] G. S. Wood, "Suppression of endogenous avidin-binding activity in tissues and its relevance to biotin-avidin detection systems." *The Journal of Histochemistry and Cytochemistry*, vol. 29, pp. 1196, 1981.
- [110] S. M. Hsu, "Protein A, avidin, and biotin in immunohistochemistry." *The Journal of Histochemistry and Cytochemistry*, vol. 29, pp. 1349, 1981.
- [111] S. M. Hsu, "Use of avidin-biotin-peroxidase complex (ABC) in immunoperoxidase techniques: a comparison between ABC and unlabeled antibody (PAP) procedures." *The Journal of Histochemistry and Cytochemistry*, vol. 29, pp. 577, 1981.
- [112] M. González, "Extremely high thermal stability of streptavidin and avidin upon biotin binding," *Biomol. Eng.*, vol. 16, pp. 67, 1999.
- [113] P. M. West and P. W. Wilson, "The Relation of "Coenzyme R" to Biotin," *Science*, vol. 89, pp. 607-608, 1939.
- [114] A. Chapman Smith, "The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity," *Trends Biochem. Sci.*, vol. 24, pp. 359, 1999.
- [115] G. Gitlin, "Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine residues are involved in the binding site." *Biochem. J.*, vol. 269, pp. 527, 1990.
- [116] L. Chaiet, "The properties of streptavidin, a biotin-binding protein produced by Streptomyces," *Arch. Biochem. Biophys.*, vol. 106, pp. 1, 1964.
- [117] O. H. Laitinen, "Rational design of an active avidin monomer," *J. Biol. Chem.*, vol. 278, pp. 4010, 2003.
- [118] A. T. Marttila, "Recombinant NeutraLite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties," *FEBS Lett.*, vol. 467, pp. 31, 2000.
- [119] R. E. Bird, "Single-chain antigen-binding proteins," *Science*, vol. 242, pp. 423, 1988.
- [120] M. B. Howard, "Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans*," *Protein Science*, vol. 13, pp. 1422, 2004.
- [121] C. R. Robinson, "Optimizing the stability of single-chain proteins by linker length and composition mutagenesis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, pp. 5929, 1998.
- [122] S. J. Lee, "Essential lysine residues in the RNA polymerase domain of the gene 4 primase-helicase of bacteriophage T7," *J. Biol. Chem.*, vol. 276, pp. 49419, 2001.

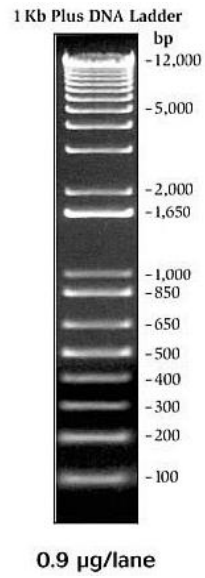
- [123] S. S. Patel, "Large scale purification and biochemical characterization of T7 primase/helicase proteins. Evidence for homodimer and heterodimer formation." *J. Biol. Chem.*, vol. 267, pp. 15013, 1992.
- [124] <https://www.neb.com/sitecore/content/nebshg/home/tools-and-resources/interactive-tools/tm-calculator>, .
- [125] V. Pelicic, J. Reyrat and B. Gicquel, "Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors," *Mol. Microbiol.*, vol. 20, pp. 919-925, 1996.
- [126] K. A. Pikal-Cleland, N. Rodríguez-Hornedo, G. L. Amidon and J. F. Carpenter, "Protein Denaturation during Freezing and Thawing in Phosphate Buffer Systems: Monomeric and Tetrameric β -Galactosidase," *Arch. Biochem. Biophys.*, vol. 384, pp. 398-406, 2000.
- [127] W. K. Lim, J. Rösger and S. W. Englander, "Urea, but Not Guanidinium, Destabilizes Proteins by Forming Hydrogen Bonds to the Peptide Group," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 2595-2600, 2009.
- [128] S. L. Bradbury and W. B. Jakoby, "Glycerol as an enzyme-stabilizing agent: effects on aldehyde dehydrogenase," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 69, pp. 2373-2376, 1972.
- [129] L. P. Blair, A. J. Tackett and K. D. Raney, "Development and Evaluation of a Structural Model for SF1B Helicase Dda $\dagger$," *Biochemistry*, vol. 48, pp. 2321-2329, 2009.
- [130] J. Toth and M. D. Biggin, "The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in Drosophila embryos," *Nucleic Acids Res.*, vol. 28, pp. e4-e4, 2000.
- [131] M. P. Okoh, J. L. Hunter, J. E. Corrie and M. R. Webb, "A biosensor for inorganic phosphate using a rhodamine-labeled phosphate binding protein," *Biochemistry*, vol. 45, pp. 14764-14771, 2006.
- [132] C. P. Toseland and M. R. Webb, "Fluorescence tools to measure helicase activity in real time," *Methods*, vol. 51, pp. 259-268, 2010.
- [133] J. Wang, "Carbon-Nanotube Based Electrochemical Biosensors: A Review," *Electroanalysis*, vol. 17, pp. 7-14, 2005.
- [134] <http://www.ncbi.nlm.nih.gov>.
- [135] sciencegateway.org/tools/proteinmw.
- [136] R. B. Tracy and S. C. Kowalczykowski, "In vitro selection of preferred DNA pairing sequences by the Escherichia coli RecA protein," *Genes Dev.*, vol. 10, pp. 1890-1903, 1996.
- [137] A. Greener, M. Callahan and B. Jerpseth, "An efficient random mutagenesis technique using an E. coli mutator strain," *Mol. Biotechnol.*, vol. 7, pp. 189-195, 1997.

- [138] B. J. Bornhorst and J. J. Falke, "Reprint of: Purification of Proteins Using Polyhistidine Affinity Tags," *Protein Expr. Purif.*, 2011.
- [139] S. K. Shukla and D. McCarthy, "An activity gel assay for the detection of DNA helicases and nucleases from cell-free extracts," *Nucleic Acids Res.*, vol. 22, pp. 1626-1631, 1994.
- [140] A. Lahaye, "PIF1: a DNA helicase in yeast mitochondria." *Embo j.*, vol. 10, pp. 997, 1991.
- [141] A. W. Dangel, L. Shen, A. R. Mendoza, L. Wu and C. Y. Yu, "Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family," *Nucleic Acids Res.*, vol. 23, pp. 2120-2126, 1995, 1995.
- [142] G. Salvi, P. De Los Rios and M. Vendruscolo, "Effective interactions between chaotropic agents and proteins," *Proteins*, vol. 61, pp. 492-499, 2005.
- [143] <http://rsb.info.nih.gov/ij/index.html>.
- [144] Z. Yang, K. L. MacQuarrie, E. Analau, A. E. Tyler, F. J. Dilworth, Y. Cao, S. J. Diede and S. J. Tapscott, "MyoD and E-protein heterodimers switch rhabdomyosarcoma cells from an arrested myoblast phase to a differentiated state," *Genes Dev.*, vol. 23, pp. 694-707, 2009.
- [145] N. W. Charter, L. Kauffman, R. Singh and R. M. Eglén, "A generic, homogenous method for measuring kinase and inhibitor activity via adenosine 5'-diphosphate accumulation," *J. Biomol. Screen.*, vol. 11, pp. 390-399, 2006.
- [146] A. L. Lucius, C. J. Wong and T. M. Lohman, "Fluorescence stopped-flow studies of single turnover kinetics of E.coli RecBCD helicase-catalyzed DNA unwinding," *J. Mol. Biol.*, vol. 339, pp. 731-750, 2004.
- [147] M. M. Martinez-Senac and M. R. Webb, "Mechanism of translocation and kinetics of DNA unwinding by the helicase RecG," *Biochemistry*, vol. 44, pp. 16967-16976, 2005.
- [148] M. M. Abdelhalim, *Helicases : Methods and Protocols*. New York, N.Y.: Humana Press, 2010.
- [149] A. F. Slatter, C. D. Thomas and M. R. Webb, "PcrA helicase tightly couples ATP hydrolysis to unwinding double-stranded DNA, modulated by the initiator protein for plasmid replication, RepD," *Biochemistry*, vol. 48, pp. 6326-6334, 2009.
- [150] M. S. Dillingham, D. B. Wigley and M. R. Webb, "Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine," *Biochemistry*, vol. 41, pp. 643-651, 2002.

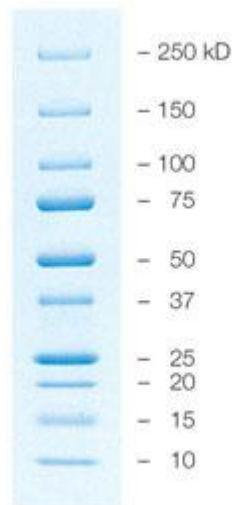
- [151] C. J. Fischer, N. K. Maluf and T. M. Lohman, "Mechanism of ATP-dependent translocation of E.coli UvrD monomers along single-stranded DNA," *J. Mol. Biol.*, vol. 344, pp. 1287-1309, 2004.
- [152] X. G. Xi and E. Deprez, "Monitoring helicase-catalyzed DNA unwinding by fluorescence anisotropy and fluorescence cross-correlation spectroscopy," *Methods*, vol. 51, pp. 289-294, 2010.
- [153] A. K. Eggleston, N. A. Rahim and S. C. Kowalczykowski, "A helicase assay based on the displacement of fluorescent, nucleic acid-binding ligands," *Nucleic Acids Res.*, vol. 24, pp. 1179-1186, 1996.
- [154] M. Brune, J. L. Hunter, S. A. Howell, S. R. Martin, T. L. Hazlett, J. E. T. Corrie and M. R. Webb, "Mechanism of Inorganic Phosphate Interaction with Phosphate Binding Protein from Escherichia coli," *Biochemistry (N. Y.)*, vol. 37, pp. 10370-10380, 1998.
- [155] M. Brune, "Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase," *Biochemistry (N. Y.)*, vol. 33, pp. 8262, 1994.
- [156] F. N. Biro, J. Zhai, C. W. Doucette and M. M. Hingorani, "Application of stopped-flow kinetics methods to investigate the mechanism of action of a DNA repair protein," *J. Vis. Exp.*, vol. (37). pii: 1874. doi, pp. 10.3791/1874, 2010.
- [157] M. Hirshberg, K. Henrick, L. L. Haire, N. Vasisht, M. Brune, J. E. Corrie and M. R. Webb, "Crystal structure of phosphate binding protein labeled with a coumarin fluorophore, a probe for inorganic phosphate," *Biochemistry*, vol. 37, pp. 10381-10385, 1998.
- [158] <http://products.invitrogen.com/ivgn/product/PV4406?ICID=search-product>.

Appendix I – Ladders

Invitrogen™ 1 Kb Plus DNA Ladder™



Biorad Precision Plus Protein™ Standards



Appendix II – Materials and Buffers

II.i Materials

1 kb plus DNA ladder, Invitrogen™

Difco™ Luria Bertani (LB) Miller Broth, BD

DNA primers, IDT technologies

GenElute™ HP Plasmid Midiprep Kit, Sigma

Glycine, Fisher Scientific

Helicase and avidin codon optimized sequences, Genscript™

Illustra™ Microspin™ G-25 Column, GE Healthcare

Immobilized Biotin Resin, G-Biosciences

Multicell Agarose, Wisent Inc.

Ni-NTA Agarose Resin, Qiagen

Phusion® High-Fidelity DNA Polymerase, New England Biosciences

Precision Plus Protein™ Standards, Biorad

Profinity IMAC Ni-Charged Resin, Biorad

QIAquick Gel Extraction Kit, Qiagen

QIAquick PCR Purification Kit, Qiagen

Spectra/Por molecularporous membrane tubing MWCO 6 – 8 000, Spectrum®

TALON® cobalt Metal Affinity Resin, Clontech

Ultra-0.5 Centrifugal Filter Unit with Ultracel-10 membrane, Amicon

Ultrafree-MC 0.45 µm PVDF centrifugal filter, Amicon

Urea, Fluka

II.ii Buffers

Agarose Gel 1%

1g agarose boiled in 100 mL 1X TAE. Add 1 μ L of 10 mg/mL ethidium bromide while cooling

Coomassie Blue Gel Colouring Solution

250 mL ethanol, 50 mL glacial acetic acid, 1.25g Coomassie Brilliant Blue R-250 in 500 mL water

Coomassie Destain Solution

100 mL ethanol, 100 mL glacial acetic acid, 800 mL water

Dialysis buffer

20 mM Hepes pH 7.6, 100 mM NaCl, 1mM MgCl₂, and 10% glycerol.

DNA digestion NEBuffer 4

1x Buffer contains: 50mM Potassium Acetate, 20mM Tris-acetate, 10mM Magnesium Acetate, 1mM DTT, at pH 7.9

DNA Loading Dye with Orange G

0.3% Orange G, 50% Glycerol, 5mM EDTA

Hepes pH 7.6 (500 mM K⁺ - Sigma #H0527)

138.2g Hepes (K⁺ salt) pH 7.6 with HCl in 1L water

Miller LB broth, Sigma™

10 g/L Tryptone, 10 g/L NaCl, 5 g/L Yeast Extract

PB Buffer, Qiagen

Composition is confidential. Guanidine hydrochloride (high concentrations) with isopropanol

PCR HF buffer, Phusion® NEB

25 mM TAPS-HCl (pH 9.3), 50mM KCl, 1.5 mM MgCl₂, 1 mM β -mercaptoethanol

PE buffer, Qiagen

Composition of this buffer is confidential and cannot be obtained

Phosphate Buffered Saline (PBS, 1x)

8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, 0.24 g KH₂PO₄ at pH 7.4

PNK buffer, NEB

70 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT, pH 7.6

Protein Loading Dye

200 mM Tris pH 6.8, 40% glycerol, 8% SDS, 1 mM EDTA, 0.1% bromphenol blue. At time of use add 20% β -mercaptoethanol

QG buffer, Qiagen

Composition of this buffer is confidential and cannot be obtained

SDS PAGE running buffer

20 g SDS, 288 g Glycine , 60.6 g of Tris Base in 2L of Water

SDS PAGE Drying Buffer

25 mL glycerol, 150 mL methanol in 500 mL water

Solution I (miniprep)

50 mM glucose, 10 mM EDTA, 25 mM Tris pH 8, 100 μ g/mL RNase

Solution II (miniprep)

0.2M NaOH, 1% SDS

Solution III (miniprep)

3M potassium acetate, 96% acetic acid

T4 DNA Ligase Reaction Buffer, NEB

50mM Tris-HCl, 10mM MgCl₂, 1mM ATP, 10mM DTT, pH7.5

TAE buffer (50x)

242 g Tris Base, 57.1 mL Glacial Acetic Acid, 18.6 g EDTA in 1L water

Talon Lysis/Equilibration Buffer

300mM NaCl, 50mM Hepes K⁺, 0.1% NP40

Talon Wash-2 Buffer

7.5mM Imidazole, 300mM NaCl, 50mM Hepes K⁺

Talon Elution Buffer

150mM Imidazole, 300mM NaCl, 50mM Hepes K⁺

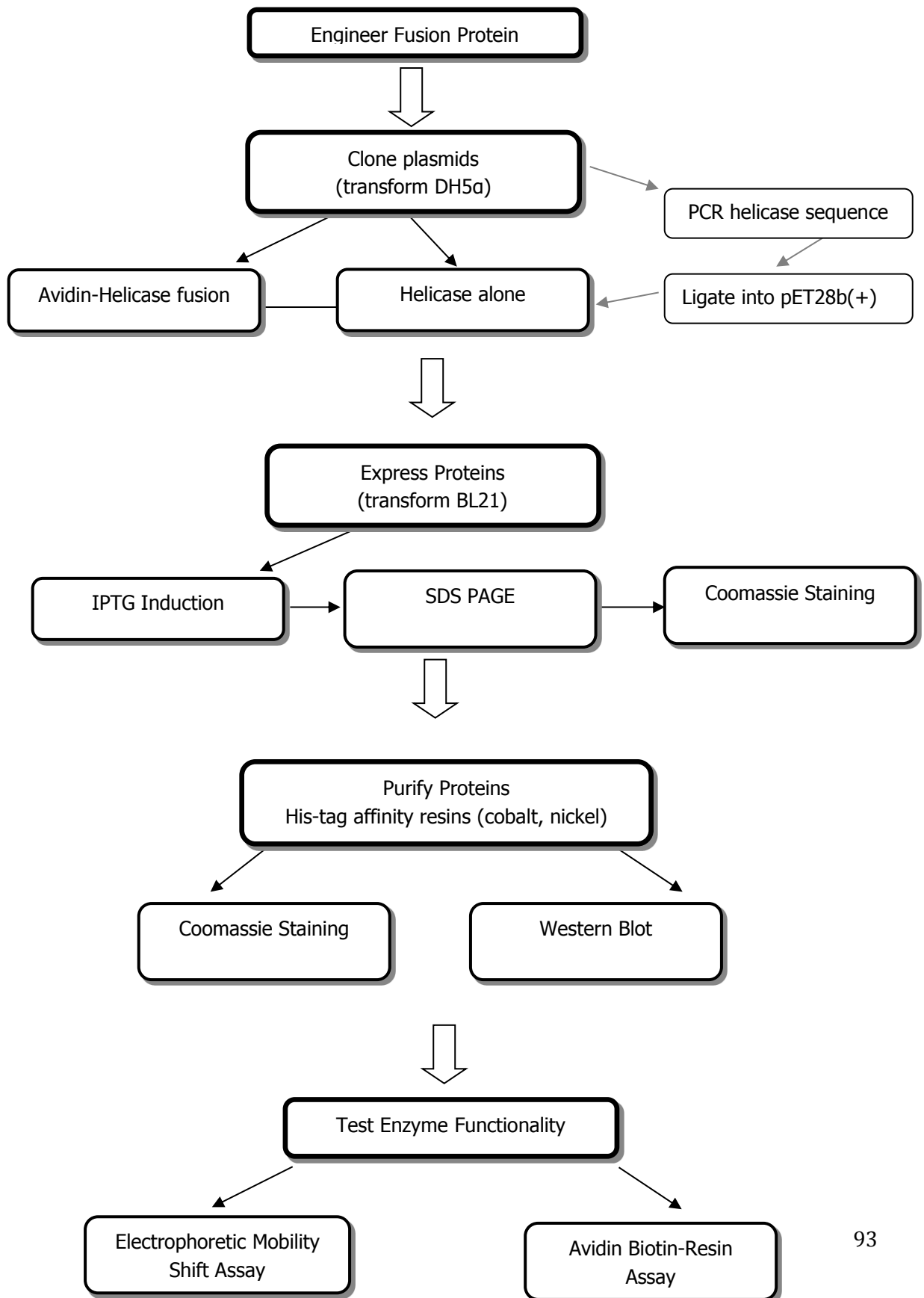
TBE buffer (5x)

108 g Tris Base, 55 g Boric Acid, 7.44 g EDTA in 4L water

Western Blot Transfer Buffer

288g glycine, 60.6g Tris base in 2L water.

Appendix III – Methods



III.i Engineering Fusion Protein

The focus of this project was to study whether or not specific DNA sequences can be attracted to DNA helicase for immediate unwinding using the natural strong affinity of the molecule biotin for the protein avidin. To this end, a specialized recombinant protein consisting of avidin and bacteriophage T7 DNA helicase was designed (Figure III.01A). This recombinant protein contains with several added features to ensure proper protein expression and purification (Table III.1) as outlined below. The nucleic acid sequences for avidin and helicase were found from the National Center for Biotechnology Information website [134].



Figure III.01: Helicase-avidin recombinant protein. A) Bar representation of the Helicase-avidin recombinant protein. Avidin (cyan) is attached to the His-tag (yellow) on the N terminus and to helicase (purple) on its C-terminus by two regions of linker amino acids (green). B) The amino

acid sequence, and C) nucleic acid sequences of the designed helicase-avidin recombinant protein. Three mutations are represented in the two sequences: N17I prevents glycosylation (red font), and N54A (yellow font) and W110K (dark green font) to create monomeric avidin.

III.i.i Fusion Protein Features

In order to purify the recombinant protein, six sequential histidine amino acid residues, known as a polyhistidine tag (His-tag), were added to the N-terminus of the helicase amino acid sequence. The His-tag has a strong affinity for nickel and cobalt, which can be exploited along with commercially available resins to bind and purify the protein (further explained in III.6). Using a His-tag to purify proteins is less complicated and less expensive than using avidin or helicase antibodies to purify the protein. Further, the commercially available metal-ion resins for His-tag purification are more stable than antibodies. Therefore, this method was deemed more suitable for a potential commercial venture.

Linker amino acid sequences were inserted between the His-tag and the beginning of the avidin protein sequence, as well as between the avidin and truncated helicase sequence, which includes the helicase-primase linker region. These inserted linker regions consist of 9 amino acids each and were inserted to prevent the possible formation of secondary structures within the functional domains of the avidin or helicase proteins. A Glycine Glycine Serine (GlyGlySer) pattern was used as this has been determined to produce neutral linker regions [119-121]. The choice of nucleic acid sequence to create the GlyGlySer amino acid pattern was chosen to minimize Guanine and Cytosine bonds, which are stronger than Adenine and Thymine bonds.

Bacteriophage T7 gp4 helicase is itself naturally expressed as a fusion protein with primase on its N-terminus [48]. On their own, both the helicase and primase sequences yield fully functional proteins [33, 89]. As a result, the avidin sequence was inserted at the N-terminus of the helicase sequence and not the C-terminus, to replace the natural site of Bacteriophage T7's primase protein. This should increase the likelihood of a functional helicase protein, as the avidin is less likely to interfere with the active sites of helicase since the space it consumes would normally be taken by a primase protein.

The complete sequence of the fusion protein has an expected molecular weight of 51.78 kDa as calculated by the molecular weights of each individual amino acid using Science Gateway's Protein molecular weight calculator [135]. The helicase sequence with a methionine and His-tag sequence inserted at its N-terminus has an expected molecular weight of 36.52 kDa.

Mutations were introduced within the avidin protein sequence in order to modify the molecule to suit its intended use in the recombinant protein. As avidin is naturally found as a dimer or tetramer, two mutations: W110K and N54A were introduced in order to create a monomeric mutant [117]. Additionally, avidin is naturally a glycoprotein. Although *E. coli* does not support glycosylation, in the event that in the future this protein sequence were to be produced in another host such as baculovirus, the mutation N17I was introduced to prevent glycosylation of the protein [118].

W	S	N	D	T	D	F
TCC	AAC	GAC	ACT	GAC	TTC	TGA
GAC	ACG	GAT	TTC	TGA	CTC	GAG

Figure III.02: Codon Optimization of DNA Sequence. A segment of the DNA sequence encoding bacteriophage T7gp4 protein is used as an example (middle row). Several DNA codons can be used in higher organisms to code for each amino acid, only some of which *E. coli* possesses. Therefore, single nucleotides must be changed (purple letters) in order to generate a codon optimized DNA sequence (bottom row) which yields the same amino acid sequence (top row).

III.i.ii Codon optimization

The bacteriophage T7 helicase and avidin recombinant DNA sequence was synthesized by GenScript. Along with the inserted mutations creating the appropriate mutant, as previously discussed, this DNA sequence was further codon optimized. *E. coli* are capable of producing all of the necessary amino acids for translation, however they do not possess all of the transfer RNA codon recognition sequences (Figure III.02). The codon optimized sequence, therefore, alters the DNA sequence but maintains the same final amino acid coding sequence, in order for the protein to be capable of expression in *E. coli*.

The DNA sequence was inserted in the multiple cloning site (MCS) of the pET28b(+) vector (Figure III.03). This vector was chosen due to several desirable properties. Firstly, it contains the gene for kanamycin resistance, which allows for the selection of only bacterial colonies that contain the plasmid in media containing kanamycin. Secondly, the vector also contains the T7 RNA polymerase promoter recognition site upstream of the inserted DNA sequence, which allows for inducible protein expression (explained in Appendix III.v) when transformed into the correct host.

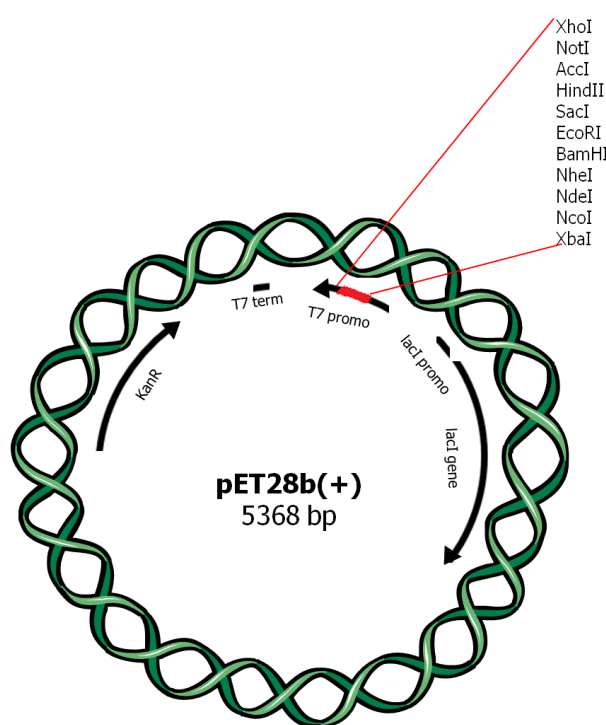


Figure III.03: The pET28b(+) expression vector. The 5368 bp expression vector contains a gene for kanamycin antibiotic resistance (KanR) for bacterial selection, a multiple cloning site (MCS) next to the T7 promoter (promo) sequence for the insertion of the gene of interest, the T7 terminator sequence to stop transcription, as well as the lacI promoter and lacI gene for expression of the lac repressor. Circular plasmid image from Servier Medical Art.

III.ii Plasmid Preparation

III.ii.i Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) DNA amplification is a means by which the enzyme polymerase creates copies of stretches of DNA. DNA polymerase requires a

nucleic acid primer consisting of a short complementary length of DNA in order to sequentially add nucleic acids. A heat resistant polymerase (Phusion™) is incubated with short DNA primer sequences complimentary to the ends of the desired amplicon. There are three stages to the PCR: denaturation, where the double stranded template is heated to separate the two strands; annealing, where the temperature of the medium is lowered to the ideal temperature for the primers to bind to the template (dictated by the number of G to C and A to T bonds); and elongation, where the temperature is set to the ideal temperature at which the polymerase protein ideally functions (Figure III.04). This process, when repeated, exponentially creates thousands of copies of DNA strands (Figure III.05).

III.ii.ii Primer Design

In order to isolate and amplify the Helicase DNA sequence from the vector containing the helicase-avidin recombinant sequence, forward and reverse DNA primers recognizing the start and finish of the helicase portion had to be created with the following criteria:

- Similar annealing temperature for both forward and reverse primers
- Restriction enzyme sites for sub-cloning into pET28b expression vector
 - o Both must be found in the multiple cloning site of the vector
 - o Both must function in the same buffer conditions
- Forward primer:
 - o NcoI restriction enzyme site for sub-cloning: C/CATGG
 - o Additional nucleotides for restriction enzyme binding
 - o Additional nucleotides to keep ATG in NcoI site in frame with His-tag (start codon)
 - o His-tag for purification
- Reverse Primer
 - o XhoI restriction enzyme site for sub-cloning: C/TCGAG
 - o Additional nucleotides for restriction enzyme binding

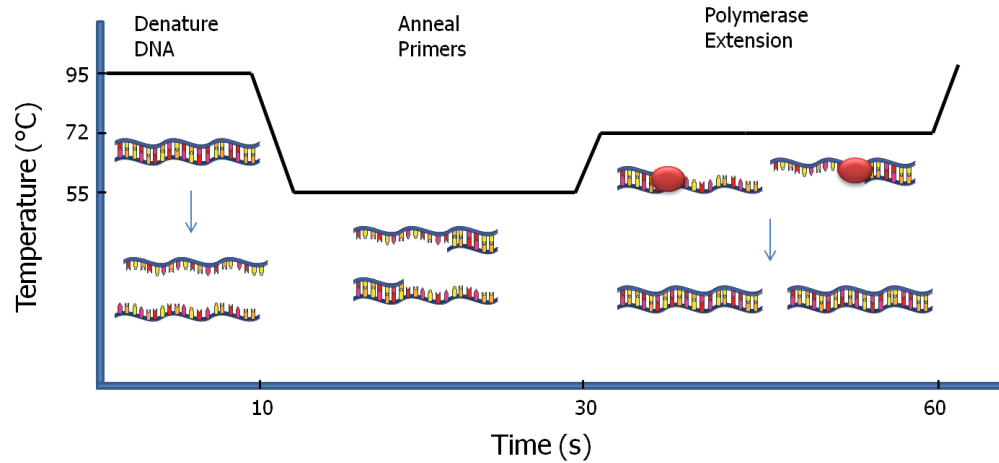


Figure III.04: Polymerase Chain Reaction. One cycle of the polymerase chain reaction involves increasing the temperature of the medium to denature dsDNA to ssDNA, lowering the temperature to the specific annealing temperature of the primers to anneal primers, and increasing the temperature to the active temperature of the heat-resistant polymerase being used for extension. The cycle is then repeated. DNA images modified from Servier Medical Art.

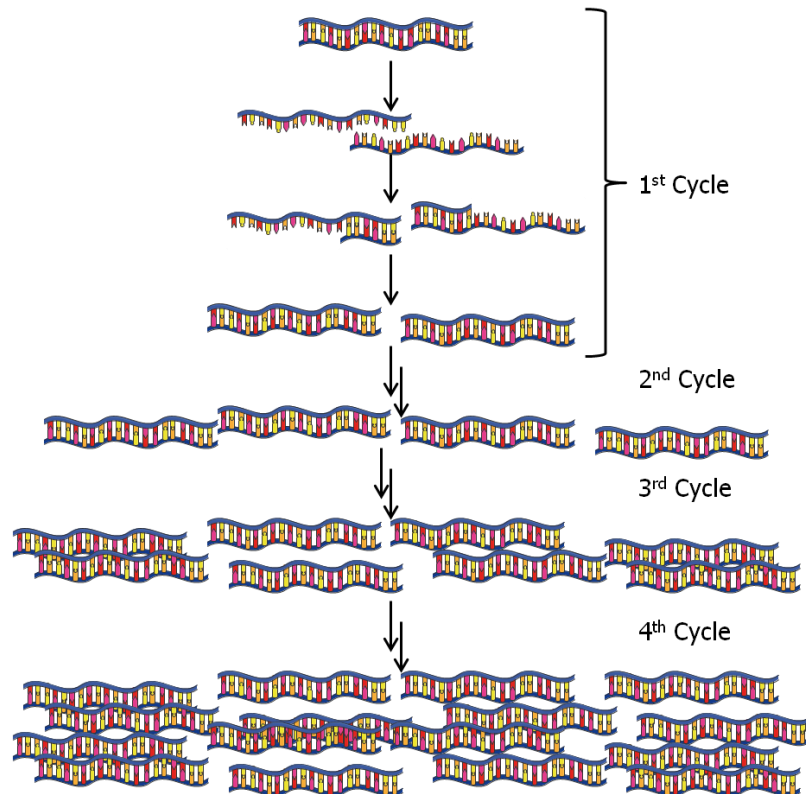


Figure III.05: Polymerase Chain Reaction Amplification. With each cycle, the number of DNA fragments is doubled. Double arrows indicate a complete cycle. The growth rate is exponential. DNA images modified from Servier Medical Art.

The codon optimized DNA sequence for helicase that was amplified by PCR is as follows. The primer annealing sites are underlined.

CACGATCGCGAGATCATGGAACAAGTGTGGAACGCCGGTCCGTGGATTCTTGACGGCGTGGT
CTCCGCTCTGTCTCTGCGTGAGCGCATCCGTGAACATCTGTCTTCAGAAGAGTCGGTCGGCCT
GCTGTTTAGCGGTTGTACGGGCATCAATGATAAAACCCTGGGTGCGCGTGGTGGCGAAGTCA
TTATGGTTACCTCTGGATCTGGTATGGGCAAGAGCACTTTCGTTTCGTCAGCAAGCACTGCAG
TGGGGAACCGCAATGGGAAAGAAAGTGGGTCTGGCCATGCTGGAAGAGTCCGTCTGAAGAGAC
CGCTGAGGATCTGATTGGACTGCACAACCGCGTGGTCTGCGTCAGAGTGAAGTGCCTGAAGC
GTGAGATCATTGAAAACGGCAAATTCGACCAATGGTTTGATGAACTGTTTCGGTAATGACACT
TTTCACCTGTACGATAGTTTTCGCCGAGGCTGAAACGGATCGTCTGCTGGCCAAACTGGCTTA
TATGCGCAGCGGACTGGGTTGCGACGTCATCATTCTGGATCATATCAGCATTGTTGTGAGCG
CATCCGGAGAGTCCGACGAGCGTAAGATGATCGATAACCTGATGACGAAGCTGAAAGGATTC
GCAAAATCTACCGGTCTCGTTCTGGTGGTCATCTGTACCTGAAGAATCCCACAAGGGCAA
AGCCCATGAAGAGGGACGTCCAGTGTCCATTACGGACCTGCGTGGATCTGGAGCACTGCGTC
AGCTGTCAGATACTATCATTGCACTGGAACGCAACCAGCAAGGTGACATGCCGAATCTGGTT
CTGGTGCATCTCTGAAATGCCGCTTTACTGGAGATACGGGTATTGCTGGCTACATGGAGTA
TAACAAGGAAACCGGCTGGCTGGAGCCGAGTTCGTATTCTGGTGAAGAGGAATCTCATTGAG
AAAGTACTGACTGGTCAAATGACACGGATTTCTGA

Methodology for PCR

For amplification and subsequent purification of the amplified sequences, 50 μ L reaction volumes were used (Table III.2). For PCRs used for verification, the reaction was scaled down to 20 μ L. The reaction conditions for the PCR is provided in Table III.3.

Table III.2: PCR Reaction Mixture.

Reagent	Volume
5x HF buffer	10 μ L
10 mM dNTP	1 μ L
10 mM forward and reverse primer mix	2.5 μ L
Template DNA (10 ng)	1 μ L
Phusion enzyme	1 μ L
ddH ₂ O	34.5 μ L
Total volume	50 μ L

III.ii.iii DNA Purification after PCR

DNA purification was performed after PCR using a commercially available kit from Qiagen, which makes use of a mini-column that retains the DNA under certain

conditions, washing out the other PCR constituents. The retained DNA is then eluted in a desired volume of either distilled deionized distilled water (ddH₂O) or TE buffer.

Methodology for DNA Purification after PCR

- Added 5 volumes of PB buffer to 1 volume of PCR Sample
- Applied the mixture to the column
- Centrifuged the sample at 13000 rpm for 1 minute
 - Discarded flow-through
- Washed the column with 0.75 mL of PE buffer and centrifuged again for 1 minute
 - Discarded flow-through
 - Centrifuged the column again to remove residual buffer
- Placed the column in a new 1.5 mL tube
- Added 30 µL of ddH₂O to the column
 - Let stand for 1 minute
 - Centrifuged at 13000 rpm for 1 minute.

III.ii.iv DNA Digestion

As described above, the PCR amplified DNA of the codon optimized helicase sequence was created with primers that inserted two restriction enzyme sites: NcoI and XhoI. These two sites are contained in the MCS of pET21b(+) vector and therefore could be digested in both the vector and PCR product by those specific endonucleases and subsequently ligated together to create a new vector. The reaction mixture for the DNA digestion is provided in Table III.4.

Table III.4: Digestion reaction conditions. The PCR product had a concentration of approximately 100 ng/µL and the vector had a concentration of approximately 888 ng/µL.

Reagent	PCR Product (µL)	Vector (µL)
XhoI	0.5	0.5
NcoI	0.5	0.5
DNA	30	4
Buffer 4	5	2
ddH ₂ O	9	13

The reaction was incubated overnight at 37 °C, then loaded onto a 1 % agarose gel.

III.ii.v Agarose Gel Electrophoresis

The product of a DNA reaction, such as digestion or PCR, can be verified using an agarose gel. Due to its negative charge, DNA can move through a polymerized agarose

gel by means of an electric current (electrophoresis) and visualized using the DNA intercalating substance ethidium bromide (EtBr). The size of the DNA fragment of interest can then be determined using a ladder consisting of DNA fragments of known sizes to determine whether or not the product of the reaction is of the correct size.

The ladder used for the DNA gel electrophoresis was the Invitrogen™ 1 kB marker (appendix I). 1 µL of DNA gel loading dye (appendix II) was used for each 5 µL of sample prior to loading on the gel in order to visualize the migration of the DNA and weigh the sample down in the wells. The gel was immersed in TAE buffer (appendix II) which provides a means for the electrophoresis current to run through.

Methodology for Agarose Gel

- Dissolved 0.7 g of agarose by boiling in 70 mL of TAE Buffer
- Added 0.7 µL of ethidium bromide and allowed mixture to cool in gel apparatus with a comb placed at the top.
- Once the gel solidified, removed the comb and immersed the gel in 1x TAE buffer
- Loaded one lane with 5 µL of DNA ladder
- Added loading dye to samples and loaded onto the gel
- Placed electrodes onto the apparatus and ran the gel at approximately 70 Volts for 1 hour
- Visualized bands under UV light and photographed.

III.ii.vi DNA Gel Extraction

DNA run on an agarose gel electrophoresis can be excised and purified. This allows for the choice of the correct sized band within a gel to be selected and isolated for further use. A commercial kit from Qiagen was used for this process.

Methodology for Agarose Gel DNA Extraction

- Under a UV light, and using a clean sharp blade, the gel slice containing the desired DNA fragment was excised
- The slice was weighed in a colourless 1.5 mL eppendorf tube
- 3 volumes of buffer QG was added for each for 1 volume of gel (where 100 mg = 100 µL)
- Incubated gel in a 50 °C water bath for 10 minutes, vortexing every 2-3 minutes, to dissolve gel
- Placed solution on ice
- Added 1 gel volume of isopropanol to the solution, and vortexed.

- Applied the sample to a QIAquick column
- Centrifuged 1 minute at 13000 rpm
 - Discarded flow-through
- Washed column with 0.5 mL buffer QG
 - Centrifuged 1 minute at 13000 rpm
 - Discarded flow-through
- Washed with 0.75 mL PE buffer
 - Centrifuged 1 minute at 13000 rpm
 - Discarded flow-through
 - Centrifuged again
- Placed column in a clean 1.5 mL microcentrifuge tube
- Added approximately 50 μ L of ddH₂O
 - Let stand for 1 minute
- Centrifuged 1 minute at 13000 rpm.

III.ii.vii DNA Ligation

The enzyme ligase can mend broken phosphodiester bonds along a DNA strand's backbone, thus ligating two separate strands of DNA together. The restriction enzymes used in this study created "sticky ends", meaning they will naturally anneal to the complementary ends (Figure III.06). Ligase enzyme will mend the backbone upon the annealing of the complementary sticky ends, creating a sealed bond. Therefore, when the digested vector and digested insert are incubated together with ligase, a circular plasmid will be created, which can then be verified and amplified by bacterial transformation. Only a circular plasmid will be accepted into a bacterial host, therefore only fully ligated samples will be transformed to create colonies on the LB agar plate.

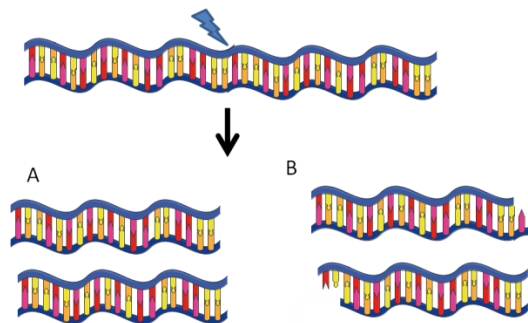


Figure III.06: Restriction Enzyme Ends. Restriction enzymes (blue zigzag) identify and cut at specific sequences of DNA. Depending on the endonuclease, this can lead to the creation of a) blunt ends, which will ligate to any other blunt ends without directionality or specificity, or B) sticky ends,

which have either a 5' or 3' overhang of 1-3 nucleotides and which will specifically only anneal and ligate to complementary ends. DNA images modified from Servier Medical Art.

Ligations are optimal under different conditions, including the ratio (as below) between insert and vector within the reaction mixture. Ideally, the reaction volume should be as small as possible to increase the likelihood of ligase encountering a ligatable end.

The equation used to determine the amount of DNA used in the ligation is as follows:

$$\text{Insert mass (ng)} = \text{ratio (3/1 or 6/1 etc)} * \frac{\text{insert length (bp)} * \text{vector mass (ng)}}{\text{vector length (bp)}}$$

Methodology for DNA Ligation

Table III.5: DNA ligation reaction mixture. Example assumes 100 ng of vector for ratios of vector to insert of 1:3, 1:5, 1:7 and 1:0 (control). Vector concentration for calculations was 888 ng/μL and insert concentration was 67.8 ng/μL

	1:3	1:5	1:7	1:0
ddH ₂ O	2.47 μL	1.93 μL	1.39 μL	3.28 μL
10x ligase buffer	0.5 μL	0.5 μL	0.5 μL	0.5 μL
Vector	0.22 μL	0.22 μL	0.22 μL	0.22 μL
Insert	0.81 μL	1.35 μL	1.89 μL	0 μL
Ligase	1 μL	1 μL	1 μL	1 μL
Total volume	5 μL	5 μL	5 μL	5 μL

- Incubated at room temperature for 1 hour
- Ligations can also be performed overnight at 15°C.

III.iii Bacterial Transformation

III.iii.i Competent Bacterial Cells

Competent bacterial cells are cells that have been treated in such a way so as to accept the insertion of a circular plasmid as a supplement to their native circular chromosome. The added plasmid not only contains the DNA sequence desired to be inserted in the bacteria, for either replicative or protein expressive purposes, but also confers antibiotic resistance to the bacteria (Figure III.03). This is essential as only successfully transformed bacteria will survive in growth media containing the particular antibiotic to which the plasmid confers resistance. The growth medium for

bacteria can either be liquid, in the form of Luria Bertani (LB) Broth (appendix II), or solid in an agar gel. The choice of liquid or solid growth medium depends on the application. The liquid form of growth medium as a broth will generally grow all bacteria which contain resistance as one large mass mixture. In order to create colonies of single bacterial lines, growth on a solid growth medium (agar gel plate) is optimal. Transformed competent cells are grown in broth, which would include a mixture of cells which have been transformed with the plasmid and which have not received the plasmid. This liquid mixture would be plated onto a solid gel where the bacteria will be separated, and each bacterium will replicate while immobilized creating a colony of the same strain. This colony can then be picked and grown again in broth for further expansion.

Various strains of bacteria, usually of *E. coli*, are made competent in order to serve different purposes based on the inherent nature of the bacterial strain. Two examples of such strains are DH5 α , which is used to amplify DNA, and BL21, which is used for protein expression.

Bacterial cells can be made chemically competent either using RbCl₂ for heat shock competence, in which the competent cells uptake the plasmid after shortly being heated up and then quickly placed on ice, or Z-competent™ in which they can uptake the plasmid at cold temperatures without heat shock. DH5 α strains were Z-competent, BL21 and other expression cell lines were heat-shock competent.

Methodology for Bacterial Transformation

- Petrie dish plates consisting of LB agar with kanamycin antibiotics (50 μ L/mL) were prepared and pre-incubated at 37 °C.
- 50 μ L of competent *E. coli* were thawed on ice in 1.5 mL tubes.
- 10-100 ng of pure plasmid DNA was added to the competent cells and mixed by finger tapping of the tube.
- The reaction mixture was incubated on ice for 30 minutes.
 - For heat shock competent cells only: placed solutions in 42°C water bath for 30 seconds and chilled on ice for 2 minutes.
- 500 μ L of LB broth (without kanamycin) was added to the mixture and shaken for 1 hour at 37 °C.

- The bacterial solution was then spread onto the pre-warmed agar plates and incubated overnight at 37 °C.
- Colonies formed on the agar plate were picked and grown in miller LB broth with 50µg/mL kanamycin.

III.iv DNA Purification

A method to amplify DNA other than PCR is by means of a bacterial host. As briefly described in section 1.9.3, competent cells can be transformed with DNA vectors containing the sequence of DNA desired to be amplified. The sequence of the vector itself may also be important, as in this study, where the vector containing the DNA coding sequence for helicase, was used to express the protein in the host. Thus, the choice of vector, as discussed in section 1.09.1, and the choice of bacterial host are important for different purposes.

III.iv.i DH5α competent bacteria

For the purpose of amplifying the DNA plasmid, and accessing the sub-cloned ligation product, the *E. coli* strain DH5α was used. DH5α's circular genomic chromosome contains several features that make it an appropriate host choice for DNA amplification:

- *recA* protein deficient: this protein is normally involved in DNA repair and aids in DNA recombination, which may lead to a rearrangement of the DNA in the plasmid. It also can cause slower growth, which is not desirable for rapid amplification of the plasmid [136].
- *EndA1* mutant: the gene for *EndA1*, an endonuclease, is mutated in order to prevent the degradation of the transformed plasmid to improve plasmid DNA quality [137]
- *hsdR17* mutant: another endonuclease which degrades DNA not methylated by the *EcoKI* protein (such as the transformed plasmid). This gene is mutated to prevent such an occurrence. [125]

Isolating the DNA from the transformed grown bacteria is known as DNA preparation. For small quantities of purified DNA a miniprep is performed. The minipreps performed for this research undertaking were done using the phenol chloroform extraction method and without a column. Midipreps were used to isolate

larger quantities of DNA at higher purity by use of commercial DNA purification columns as outlined below. Higher purity DNA was required for transformation of the BL21 cell line for protein expression and purification, as outlined in Appendix III.v.

Methodology for plasmid DNA Purification

Miniprep

- Colonies were picked and put into 3 mL of LB broth with kanamycin in a 14 mL bacterial transformation tube.
- Transferred 1.5 mL of the solution to a 1.5 mL eppendorf tube for the preparation of DNA, the rest was kept for a glycerol stock of the transformed bacteria.
- The transferred solutions were centrifuged at 13000 rpm for 1 minute.
 - The supernatant was aspirated.
- Resuspended the pellet in 250 µL of solution I with RNase
- Added 250 µL solution II, inverted tubes 4-6 times to gently mix
 - Incubated at room temperature for 5 minutes.
- Added 250 µL of pre-chilled solution III (potassium acetate),
 - mixed gently
 - incubated on ice for 5-10 minutes
- Centrifuged the solutions at 13000 rpm at 4 °C for 10 minutes.
- Transferred 750 µL of the supernatant to new tubes containing 500 µL of phenol/chloroform
 - ensured that no protein pellet was transferred.
- Vortexed for 30 seconds, and then centrifuged at 13000 rpm for 5 minutes
- Transferred 400 µL of the top (aqueous) phase to a new tube
- Added 45 µL of 3M sodium acetate pH 5.3
- Added 900 µL of 100% ethanol to samples
 - Incubated at room temperature for 2 minutes
- Centrifuged the solutions at 13000 rpm at 4 °C for 10 minutes
- Aspirated the supernatant and washed with 800 µL of 70% ethanol.
 - Centrifuged at 13000 rpm for 2 minutes, and aspirated the supernatant
- Left DNA pellet to dry on the bench for approximately 15 minutes.
- Resuspended the DNA in 50 µL of TE buffer or water

Midiprep (Qiagen)

- Centrifuged 50 mL of bacteria at 4000 rpm for 20 minutes at 4 °C
 - Discarded supernatant

- Resuspended in 2 mL of solution I + RNase
- Added 2 mL of solution II
 - inverted 4-6 times to gently mix
 - incubated at room temperature for 3 minutes
- Added 2 mL of solution III
 - inverted 4-6 times at room temperature
- Centrifuged at 20 000g for 30 minutes at 4 °C
 - transferred supernatant to a new tube
- Centrifuged at 20 000g for 30 minutes at 4 °C again
- Mixed supernatant with 2 mL of BB buffer
 - inverted 4-6 times
- Poured solution onto column and drained
 - discarded flow-through
- Washed with 750 µL of PE buffer
 - centrifuged 1 minute at 13000 rpm and discarded flow-through, twice
- Eluted in 100 µL of ddH₂O or TE buffer

III.v Bacterial Protein Expression

The intention of isolating a protein of interest grown in bacteria is to overexpress said protein within the cell beyond physiological proportions and to isolate it from the other proteins and molecules naturally produced by the bacteria to survive. In order to attain this goal, an appropriate bacterial host is required, as well as a vector containing the DNA coding sequence for the protein of interest.

The vector to be transformed in the expression cell line must contain specific features for the overexpression of the protein of interest. The vector used for the present study was the pET28b(+) vector . This vector contains a gene for kanamycin antibiotic resistance for selection of transformed bacteria, as well as a modified lac operon.

III.v.i Lac operon induced Protein Expression

The lac operon DNA sequence consists of a promoter region, which binds RNA polymerase, followed by a repressor sequence, followed by the lacZ, lacY and lacA genes, which are under control of the promoter/repressor operon genes (Figure III.07). The lac repressor protein lacI binds to the repressor sequence, effectively blocking RNA

polymerase from moving forward to transcribe the lacZ, lacY, and lacA genes. When lactose is present, it binds to the repressor protein, releasing it from the DNA, allowing for RNA transcription of the genes to occur.

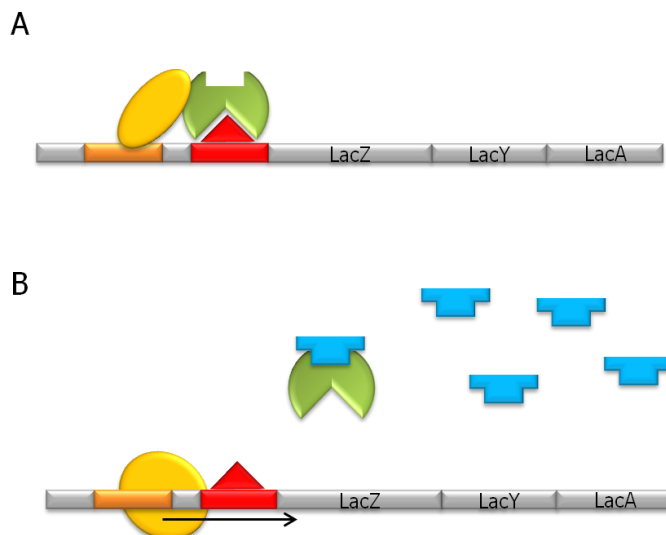


Figure III.07: The Lac Operon. Represented in A) repressed state and B) Active state. The bar represents the DNA sequence of the lac operon, orange being the promoter sequence, which is bound by RNA polymerase (yellow circle), red is the repressor sequence bound by the lac repressor (green), followed by the genes under the lac operon's control (lacZ, lacY, lacA). When bound by a lactose molecule (blue), the lac repressor releases from the repressor DNA sequence, and the RNA transcription can begin.

The lac operon is exploited in bacterial expression vectors which work in concert with modified specialized bacterial hosts to specifically target the expression of a protein of interest that is inserted in a vector containing the operon, directly downstream of it. The vector's lac operon's promoter sequence is specific to T7 RNA polymerase so that only bacteriophage T7 RNA polymerase can transcribe the gene of interest upon induction and not *E. coli*'s own RNA polymerase. The vector also contains the lacI gene encoding the lac repressor. This repressor binds to the operator and promoter sequences of the lac operon, preventing the binding of RNA polymerase and repressing transcription of the protein of interest.

In order to induce the transcription of the inserted gene of interest, two things must occur: 1) the lac repressor must be removed from the operator, and 2) T7 RNA polymerase must bind to the T7 promoter sequence on the vector (Figure III.08).

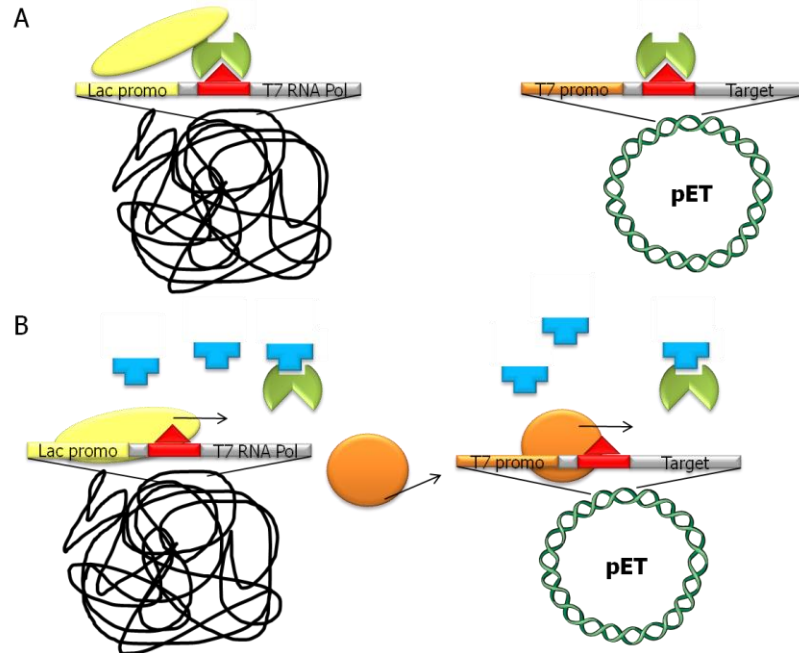


Figure III.08: Induced Bacterial Protein Expression Under the Control of the Lac Operon.

In order for induction of overexpression of the gene of interest to occur in an expression host transformed with an expression plasmid containing the gene of interest, the lac operons in both the bacterial genome (black) and expression vector (green circle) must be activated. A) In its repressed state, the lac repressor sequences (red) are bound by the lac repressor (green) which prevents RNA polymerase (yellow oval) from binding the lac promoter sequence (yellow) and moving forward to transcribe T7 RNA polymerase on the bacterial genome. B) In its activated state, the IPTG or lactose (blue) release the lac repressor from the lac repressor sequence, allowing RNA polymerase to transcribe the T7 RNA polymerase gene on the bacterial chromosome yielding T7 RNA polymerase (orange circle). T7 RNA polymerase then binds the free T7 promoter sequence in the expression vector and transcribes the target gene of interest. DNA plasmid image from Servier Medical Art.

The choice of a strain of *E. coli* engineered for protein expression is important as the host works in complement to the vector. The T7 RNA polymerase gene is present in the host chromosome, downstream from the lac operon inserted in the host chromosome. Therefore, when the lac repressor is removed from the lac operator on the vector freeing the transcription start site for the gene of interest, it is also removed from the lac operator on the host chromosome transcribing T7 RNA polymerase which can in turn go and occupy the vector's T7 promoter site. The host chromosome also contains the *lacI* gene to ensure repression of the lac operon in the absence of an inducer.

Naturally, the molecule allolactose binds to the lac repressor, which changes the conformation of the repressor, releasing it from the operon thus allowing for the start of gene transcription of genes downstream from the lac operon. IPTG (Isopropyl- β -D-thiogalactoside), an analog of allolactose, is used instead of allolactose to induce the transcription of the gene of interest in the host. IPTG contains a sulfur group which renders it non-hydrolysable, thus preventing the molecule from being metabolized/degraded by the cells. This ensures that the concentration of IPTG to remain constant with each cellular division, allowing for bacterial induction to occur from one generation to the next.

Therefore, when IPTG is added to the bacterial medium, the lac repressor is removed from the lac operators on both the host chromosome and the plasmid, freeing the transcription start site on the host chromosome to *E. coli*'s RNA polymerase that transcribes T7 RNA polymerase, which can in turn bind the T7 promoter site, transcribing the gene of interest. After several cycles of cellular division, the protein of interest becomes the most abundant protein in the cells. It has been empirically determined that the best time to induce the cells is when the optical density of the solution, measured at 600nm (OD₆₀₀) is between 0.4 and 0.6.

III.v.ii Expressed Protein Stability

There are several factors that mediate the stability of the expressed protein. If the gene of interest's DNA sequence inserted in the vector is out of phase with the start codon, an unstable protein will be transcribed which will be easily degraded by the cell and induction will not be successful. Furthermore, in the case of over expressed protein, the host may be overwhelmed by the volume of protein being expressed, resulting in improperly folded protein. Under normal circumstances, the host provides chaperone proteins to help guide the tertiary folding of the protein, but under conditions of inducible overexpression of the protein of interest, the cell may not be able to keep up (Figure III.09). This could lead to misfolded proteins and protein aggregates that are insoluble in aqueous solution. A properly folded protein will have charged residues hidden inward, but misfolded proteins do not have such a conformation and may be attracted to other misfolded proteins creating aggregates.

These aggregates can also encourage the further misfolding of properly folded proteins. Thus, it is imperative that conditions are made perfect for enhancing protein stability.

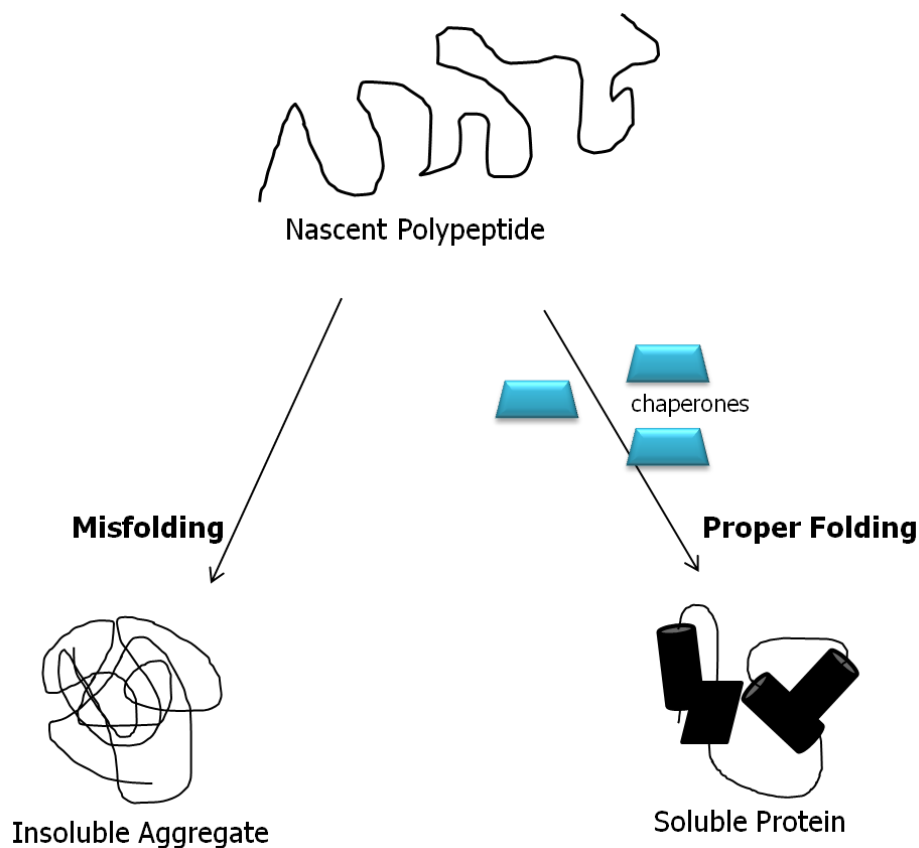


Figure III.09: Protein Folding. A nascent polypeptide produced by a ribosome is guided by chaperones (blue pentagons) to yield a properly folded soluble protein in which hydrophobic residues are hidden in pockets and hydrophilic amino acid residues are exposed. If there are not enough chaperones present, the protein will not fold properly and will form an insoluble pellet.

If the cell is not producing a properly folded protein, several steps can be taken during expression to enhance this. Reducing the concentration of IPTG added for induction would lead to less aggressive stimulation of the lac operon expression system, which may allow time for the chaperones provided by the host to aid in the proper folding of the protein. Lowering the temperature and/or adding glycerol inhibit the rate of growth of the bacteria, which would also slow down the rate of protein expression. Inducing the cells at a lower optical density (OD_{600}) may also have a similar effect. Additionally, selection of the expression hosts system is important as it is the

host that supplies the T7 RNA polymerase. Bacterial expression hosts proficient at protein expression include the BL21, BL21-CodonPlus-RIL (RIL), and ER2566 *E. coli*.

III.vi Protein Purification

Protein purification is required in order to separate a protein of interest from the other constituents of a cell including lipids and DNA. Although the protein of interest in this case is being overexpressed to the point that it is present in significantly larger quantities than the other proteins in the cell, separation of this protein from the others will help to ensure specificity of action in the enzymatic activity assays. In order to isolate a particular protein, 1) the cell must be lysed, 2) the protein must somehow be attracted and bound to a matrix, 3) the other constituents of the cell must be washed off, and 4) the protein must be eluted. Each of these steps may be undertaken in a variety of ways in this thesis. Following protein purification, the eluted protein solution must be dialyzed to replace the elution buffer with a storage buffer, as described below.

The cells may be lysed either chemically (e.g. detergents) or mechanically (e.g. sonication). In the present study both chemical lysis with detergents such as NP40 and sonication were used in succession to disrupt membranes during purification were used. Further details can be found in the methodology outlined below.

Specific binding of the protein to a matrix requires the use of a matrix or resin affixed with molecules that recognize the protein, such as antibodies. An economical way to bind proteins to a resin is by use of tags, which are amino acid sequences artificially added onto the end of a protein of interest. These tag sequences possess particular properties that make them attractive to other molecules for the purpose of attraction to a resin and isolation.

Therefore, the histidine tag of the protein binds to the metal bound to an immobilized metal-affinity chromatography (IMAC) resin [138], and the other proteins are washed off with buffer. The His-tag is then eluted out by competition using high concentrations of imidazole, a molecule similar in structure to the histidine side chain. Copper binds to histidine most tightly, then nickel and most weakly by cobalt. For this

reason, cobalt based resins provide purer, but lower concentrations, of eluted protein, as the more weakly bound protein allows for less non-specific binding to the resin. Copper is least frequently used as the binding is too strong and the protein cannot always be eluted from the resin. In the present study two sets of nickel resins, produced by Qiagen and Biorad, as well as a cobalt resin produced by Talon, were used.

Methodology for His-tag protein purification

Buffers are different for different resins

- Centrifuged induced bacteria in 50 mL aliquots at 4000 rpm for 10 minutes at 4°C.
- Discarded supernatants and resuspended pellets in 10 mL of equilibration buffer + 0.5 µL of 0.5M PMSF
- Shook the solution gently at 4°C for 10 minutes
- Sonicated the solution on ice, at 25% 10seconds on 10 seconds off for 5-15 minutes
- Centrifuged the sonicated solution at 11000 xg for 20 minutes at 4°C
 - For denaturing conditions, equilibration buffer in step 2 was either replaced with the same buffer containing urea, or a second centrifugation step was added after sonication and the 2nd pellet was resuspended in a buffer containing urea by sonication.
- Applied the supernatant was applied to 0.5 mL of resin in the column and rotated for 2 hours at 4°C.
- At room temperature, drained the column by gravity, and washed with 10 mL of equilibration buffer.
 - In denaturing conditions, equilibration buffer contained the denaturing agent.
- Washed the column was washed twice with 5 mL of wash buffer.
 - For renaturing of the protein on the column, several washes of decreasing concentration of the denaturing agent were used.
- Incubated the protein resin room temperature with 0.5 mL of elution buffer for 10 minutes to elute the protein, 5 times.
 - In denaturing conditions, the elution buffer contained the denaturing agent

III.vi.i Dialysis

The elution buffer used to elute the protein from the resin can interfere with experimental reactions. Further, for storage purposes, the protein must be resuspended in glycerol. Therefore, dialysis is required to exchange the solution containing the

eluted protein. This process makes use of a semi-permeable membrane in which small molecules contained in the buffer can pass through, but the protein cannot (Figure III.10). The molecular weight cut off can vary depending on the semi-permeable membrane. In this work, the Spectrum™ Spectra/Por dialysis bag, that has a molecular weight cut off of 6 – 8 000 Daltons, was used. The eluted protein mix is sealed in a dialysis bag which is then placed into a beaker containing a large volume (1-2 litres) of the buffer desired to replace the elution buffer. Due to differences in concentration, the molecules move along the concentration gradient thus replacing the elution buffer with the desired buffer.

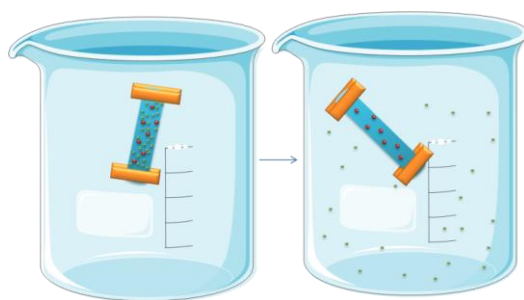


Figure III.10: Dialysis. The solution to be dialyzed is placed in a dialysis bag tied off at the ends by two clamps (orange rectangles) and placed into a beaker with a large volume of storage solution. Small molecules (green dots) pass through the dialysis bag along the concentration gradient, whereas large molecules (red dots) cannot. Images modified from Servier Medical Art.

Methodology for Dialysis

- Soaked a dialysis bag in dialysis buffer (appendix II) and clamped at one end to seal.
- Pipetted the protein elution solutions in to the sealed bag.
- Folded the dialysis bag and clamped at the other end to seal.
- Placed the dialysis bag into 1-2 litres of dialysis buffer.
- Placed the beaker containing the dialysis buffer on a magnetic stirrer overnight at 4°C
 - or in smaller volumes (500 mL) for 30 minutes, replacing the buffer solution several times.

III.vi.ii Denaturing and Renaturing Proteins

As discussed in Appendix III.v, an expressed protein may remain stable but still be improperly folded by the host cell. If the cell is not producing a properly folded

protein, several steps can be taken during expression to enhance this, yet if these measures are not successful, the stable polypeptide that is misfolded may be renatured during or after purification by a variety of means.

In the case where the bacteria expresses a misfolded protein of interest, the misfolded protein will be found in the bacterial pellet, while the little properly folded and soluble protein will be in the supernatant. The bacterial pellet can be lysed with a lysis buffer in order to drive the protein to solubilise in water. This can be accomplished using a denaturing agent, such as a detergent [139], or as in the present case, a chaotropic agent [140, 141]. Urea and Guanidium thiocyanate are two chaotropic agents, the former being less aggressive than the latter. Chaotropic agents disrupt clathrate structures of water, that are tight orderly network of low energy water molecules which surround hydrophobic amino acids [142]. This allows room for the high energy, non-ordered, water molecules in the rest of the solution to come near the protein and thus disrupt the protein's structure.

After a protein has been denatured by a specific denaturing agent, the removal of the agent gives the protein the chance to refold itself properly. This can be effectuated *via* dialysis after the protein has been eluted from the column, or by washing the protein bound to the affinity resin with buffers consisting of decreasing concentrations of the denaturing agent. Either of these processes can be enhanced by reducing the concentration of denaturant slowly. The presence of low concentrations of reducing agents such as dithiothreitol (DTT) or β -mercaptoethanol, or detergents such as NP-40 [141], can help coordinate ionic charges aiding in the refolding of the protein. A protein may also fold preferentially under different pH conditions [140].

III.vii Protein Visualization

III.vii.i SDS PAGE

Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis (SDS PAGE) is a method by which proteins are separated based on their molecular weight. SDS is a detergent that denatures proteins, ridding them of their secondary and tertiary structures, allowing the separation of the protein to occur only based on the primary

amino acid sequence. Acrylamide is polymerized as a slab gel, creating a matrix through which the proteins must migrate, allowing smaller molecules to pass through more easily than larger molecules, thus separating them by size. A protein ladder consisting of polypeptides of known molecular weight is run alongside samples as a reference for the size of the proteins being analyzed.

The percentage of acrylamide affects the migratory separation of the proteins. A higher percentage of acrylamide creates a more dense mesh network through which the proteins must travel, thus would increase resolution for larger molecules. In the present study, SDS PAGE were all prepared at 10% acrylamide (appendix II). The SDS PAGE is polymerized with wells along the top in which protein samples, in loading buffer, are loaded. The loading buffer (appendix II) contains a small anionic dye, such as bromphenol blue, to track the migration of the sample, glycerol to increase the density of the solution to sink to the bottom of the well, and a denaturing agent such as β -mercaptoethanol to reduce disulphide bonds.

SDS PAGE can be visualized either by staining of the proteins to visualize all proteins present in the gel, or by use of specific antibodies in what is known as a western blot.

III.vii.ii Coomassie Staining

Coomassie Brilliant Blue dye, dissolved in a solution containing methanol and acetic acid, stains the SDS PAGE gel and the proteins in it. The acetic acid fixes the proteins to the gel during this process. A solution containing ethanol and acetic acid removes the blue dye from the gel, revealing blue bands indicating the presence of protein. This can be used to demonstrate the purity of a protein sample, its size and/or approximate concentration. However, this method cannot be used to identify specific proteins.

III.vii.iii Western Blotting

A western blot makes use of antibodies raised against particular proteins to visualize their presence on an SDS PAGE. The gel must first be transferred to a membrane (e.g. nitrocellulose or polyvinylidene fluoride, PVDF) by electric current. The

membrane is blocked with milk proteins to reduce nonspecific binding. The membrane is washed over by the primary antibody, which recognizes the protein of interest. The primary antibody is then bound by a general appropriate secondary antibody, which then reacts with chemiluminescent molecules creating a signal that can be exposed onto film and visualized.

Methodology for Western Blot

- Placed SDS PAGE in a sandwich consisting of the following constituents, all pre-soaked in transfer buffer containing methanol, in the following order:
 - a sponge
 - 2 sheets of Whatman paper
 - a PVDF membrane (activated in 100% methanol) or nitrocellulose membrane
 - the gel
 - 2 sheets of Whatman paper
 - a sponge
- Placed the sandwich in an electrophoresis apparatus to transfer the membrane
 - 100V for 2 hours with an icebox or at 20V overnight.
- Blocked the membrane with 5% milk in 1x phosphate buffered saline (PBS) rocking for 30 minutes at room temperature.
- Applied the primary antibody (4 µg in 2 mL milk in PBS) and rotated at 4°C overnight.
- Washed the membrane three times for 5 minutes with 1x PBS,
- Incubated at room temperature with 15mL milk containing 5 µg of secondary antibody for 1 hour.
- Washed the membrane three times for 5 minutes with 1x PBS
- Covered with 2mL of chemiluminescent detection reagent
- Exposed to x-ray film.

III.vii.iv Densitometric Analysis

Coomassie stains and western blots are analyzed visually. This can be a very subjective process. Densitometric analysis of the bands that appear allows quantification of the bands on a gel. This can also be used for DNA gels. Densitometry involves the digital analysis of an image to quantify the number of pixels in a particular band. The values generated can be graphed or manipulated for analysis. In the present study the free NIH software ImageJ [143] was used to quantify western blots.

III.viii Helicase Activity Assay

In order to assess helicase's capacity to bind to DNA, a technique known as Electrophoretic Mobility Shift Assay (EMSA), also known as a gel shift or gel retardation assay, was employed. This technique displays the interaction of proteins with DNA by visualizing the position of the DNA run through a native poly-acrylamide gel (Figure III.11). The method uses DNA that is radioactively labeled on one end. In one lane on the gel, pure DNA is loaded. In another lane, DNA that has been incubated with the protein of interest is loaded. If the protein of interest binds to DNA, the DNA will be sterically larger and will move more slowly through the acrylamide gel, thus shifting the location of the radioactive signal on the DNA. In the present study, the positive control for the functioning experiment was MyoD protein bound to a radioactively labeled E-box sequence (Table 5.3), which has been showed in the literature and in the laboratory of Dr. Jeffrey Dilworth to have a positive signal in an EMSA [144].

The DNA binding ability of helicase has been assessed by EMSA by other researchers [35, 52, 56, 76, 77, 81, 129]. The experimental conditions used in these publications are highlighted in Table III.6. β,γ -methylene dTTP is a non-hydrolyzable form of dTTP and is not commercially available. Therefore, the helicase reaction performed in the present study was stopped using a cross-linking reagent, such as glutaraldehyde and formaldehyde.

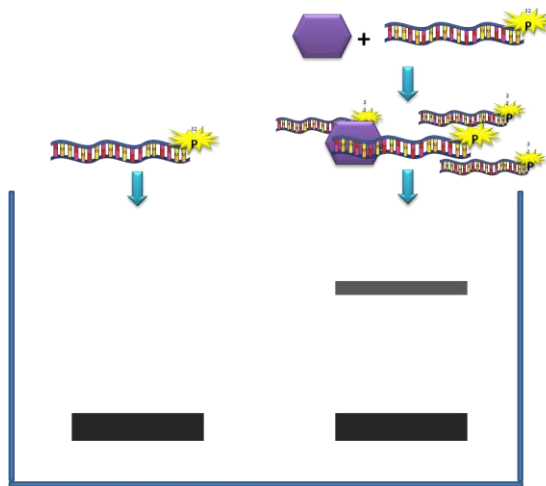


Figure III.11: Electrophoretic Mobility Shift Assay. Two reactions are loaded onto a native polyacrylamide gel (blue outline). The first (left) lane is loaded with radioactively labeled DNA, the second (right) is loaded with radioactively labeled DNA incubated with a DNA binding protein. When some of the radioactively labeled DNA is bound by a protein, this complex migrates more slowly through the gel, and creates a shift in the signal. The gel is exposed to film, and the band pattern appears (black and gray bands). DNA images modified from Servier Medical Art.

Table III.6: Helicase reaction conditions for electrophoretic mobility shift assays. B, γ -me represents B, γ -methylene. TRX represents thioredoxin, BB represents bromphenol blue, XC represents xylene cyanol and pol represents polymerase.

Enzyme studied	DNA	DTT	MgCl ₂	NaCl	Tris-HCl	T7 gp4	β , γ -me dTTP	other	Incubation time/temp	Crosslinker	Stop	Year	Ref
T7gp4A	0.1 μ M	10 mM	10 mM	50 mM	40 mM pH 7.0	vary	1 mM		Room temp 5 minutes	No	loaded onto native PAGE	1995	[56]
T7gp4	0.24 μ M	1 mM	10 mM		20 mM, pH 7.5	0.24 μ M	1 mM	(up to) 5% glycerol	30 degrees 10 minutes	Gluteraldehyde (glutaric dialdehyde) 0.02% 20 seconds	100mM EDTA 40% glycerol 0.1% BB 0.1% XC	1995	[52]
T7gp4 (with T7 pol)	1.8 μ M	10 mM	10 mM	150 mM	40 mM pH 7.5	3 μ M	1 mM	0.8 μ M pol	30 degrees 10 minutes	no	loaded onto native PAGE	1997	[76]
T7gp4, T7pol TRX	100 nM					2.4 μ M	0.5 mM	1.8 μ M DNA pol	37 degrees 10 minutes			2006	[77]
T7gp4	1 mM	10 mM	10 mM		40 mM pH 7.5	vary	1 mM	50 mM potassium glutamate	37 degrees 30 minutes	No	nitrocellulose membrane on ζ probe membrane	2011	[81]
Gp41	25 mM				25 mM Hepes, pH 8.2	5.8 μ M		100 μ M EDTA 10 mM ATP	37 degrees 10 minutes	DSP crosslinker 4 mins at 37 degrees	Quenched by 0.1M glycine	1999	[35]
SF1B helicase Dda T4	4 μ M	1 mM	4 mM MgOAc		20 mM Hepes, pH 7.5	2 μ M			2 min at room temperature	Formaldehyde	quenched at 0.5, 1, 5, or 10 min 200 mM glycine.	2009	[129]

Methodology for Radioactively Labeling Probes from EMSA

Table III.7: Reaction mixture for Labeling DNA probe. Radioactive phosphate (P32) is incubated with polynucleotide kinase (PNK) enzyme isolated from bacteriophage T4 along with the ssDNA oligonucleotide to be labeled in the reaction mixture.

Reagent	Volume (μL)
ddH ₂ O	5.8
1 M DTT	0.2
25 pmol Oligonucleotide	1
³² P	1
10x PNK buffer	1
T4 PNK enzyme	1
Total	10

- Incubated the reaction mixture at 37°C for 30 min
- Heated the mixture at 95°C for 3 min to deactivate the kinase
- Added 3 μL of 25 pmol complementary ssDNA
- Heated the solution at 95°C for 3 min
 - Transferred the tubes to boiling water
- Allowed the tubes to cool to room temperature in the water bath slowly for annealing of the dsDNA
- Added 50 μL of nuclease free or autoclaved water to the reaction
 - Centrifuged at 3000rpm for 1 minute.
 - Broke the bottom of a G-50 filter column was broken, and the column was placed in a disposable eppendorf tube and centrifuged for 1 minute at 3000 rpm.
- Switched the filter to a fresh, labeled eppendorf tube
 - Transferred the reaction solution to the filter
- Centrifuged at 3000 rpm for 1 minute
- Stored the dsDNA at -20 °C.

Table III.8: Helicase reaction mixture. Reagents, concentrations and volumes of various components in the helicase reaction mixture used in this work, with role of the component provided.

Reagent	Final concentration	Volume (μ L)	Role of Ingredient
1M DTT	10 mM	0.1	Stabilizes enzymes / proteins with free sulfhydryl groups [73]
DNA probe	varying	2 (varying)	Helicase's substrate
1M $MgCl_2$	10 mM	0.1	Aids binding ssDNA [73] and dTTP hydrolysis and binding [31]
0.5M Hepes pH 7.0	20 mM	0.4	Maintains pH
100 mM dTTP	10 mM	1	DNA binding and protein oligomerization [31, 73, 101]
5M NaCl	50 mM	0.1	
1ng/ μ L BSA	1 pg/mL	0.5	Prevents degradation of protein [73]
50% glycerol	2.5%	0.5	Stabilizes enzyme [128]
Protein	Varying	Varying	
ddH ₂ O		Varying	Nucleophile [31, 39]
Total		10	

Methodology for EMSA

The experimental parameters varied with each experiment, therefore only a vague description is provided.

- Incubated the protein with DNA in a 10 μ L reaction volume of the helicase reaction mixture at 37°C or 25°C for the indicated period of time
- Stopped the reaction using EDTA and/or a crosslinking reagent.
- Quenched the crosslinker.
- Added native gel loading dye to the reaction mixture and loaded onto a 5% native polyacrylamide gel at 160V for 2.5 hours.
 - The gel had been pre-run at 100V for 1 hour.
- The gel was dried onto Whatman paper using a gel dryer at 98°C for 1 hour, and exposed to x-ray film at -80°C overnight or longer, as indicated. Or room temperature for 3 hours, or as necessary depending on the intensity of the radioactive signal.

III.ix Avidin Activity Assay

In order to assess whether or not the avidin portion of the helicase-avidin fusion protein is enzymatically active and functional, a simple binding assay was devised. The assay utilised to assess the activity of the avidin portion of the helicase-avidin recombinant protein was similar to the protein purification protocols presented in section 2.4. In this

case, instead of cobalt or nickel resin as with the histidine purification, an immobilized biotin resin consisting of agarose beads covalently bound to biotin molecules was used (G biosciences). The helicase-avidin fusion protein binds strongly to the biotin on the resin, withstanding several washes, and eluting at a very low pH by disrupting the high affinity bond between avidin and biotin. The collected supernatant, washes, and elutions are then run on an SDS PAGE that is subsequently stained by Coomassie Brilliant Blue dye. A successful result for this assay would be indicated if the washes do not contain any protein, but the elutions do (Figure III.12).

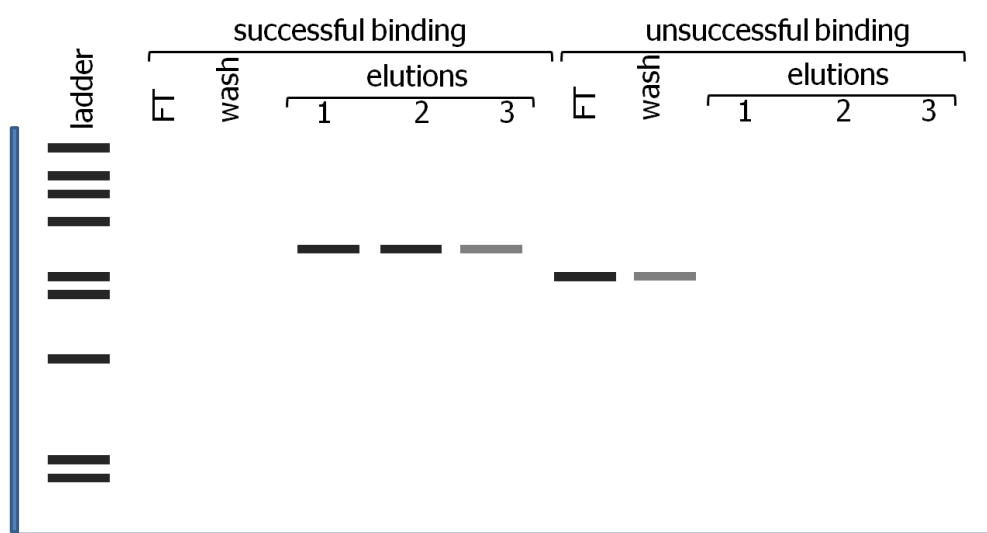


Figure III.12: Expected results of an SDS PAGE run with the results of an avidin-biotin binding experiment. A representative PAGE gel is presented, with visualized protein bands of strong intensity in black and weaker intensity in gray. A successful experiment where avidin binds to a biotinylated resin would yield no protein in the flow-through (FT) or wash, and only in the elutions. An unsuccessful experiment would involve the majority of the protein to wash out in the flow-through and washes.

Methodology for Avidin Biotin Binding Assay

- 10 μ L of biotin resin was incubated with approximately 8 μ g (200 μ L) of protein rotating at 4°C for 1 hour.
- The resin was then transferred to an Ultrafree-MC 0.45 μ m PVDF centrifugal filter tube and centrifuged at 1,000 xg for 2 minutes.
- The supernatant was saved and concentrated using an Amicon Ultra-0.5 mL Centrifugal Filter for Protein Concentration.
- The resin was then washed with 200 μ L of 1x PBS buffer and centrifuged for 1,000 xg for 2 minutes, 10 times

- Concentrated the last wash using an Amicon Ultra-0.5 mL Centrifugal Filter for Protein Concentration in order to ensure binding of the protein.
- The protein was then incubated for 20 minutes with 10 μ L of 0.1M Glycine-HCl, pH 2.8 for elution, 4 times.
- The elutions were neutralized with 1 μ L of 1M Tris, pH 8.5.
- Samples of the input, concentrated supernatant, concentrated final wash, 4 elutions and the resin were loaded onto a 10% SDS PAGE
 - Run at 120V for 1 hour and 20 minutes
- SDS PAGE was stained with Coomassie Brilliant Blue dye.

Appendix IV – Figures

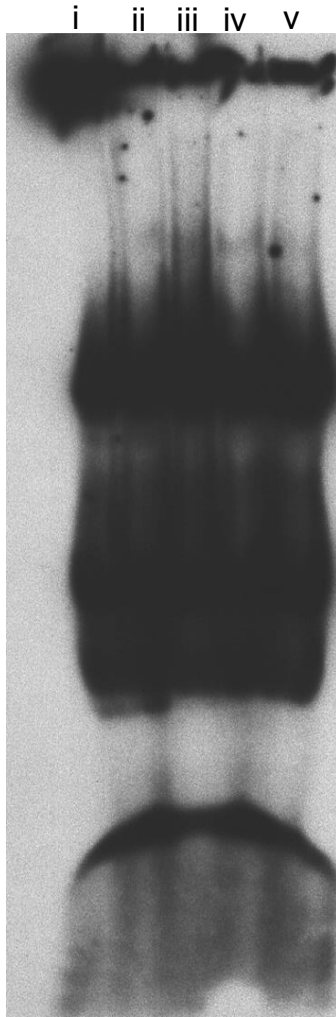


Figure A1: Effect of NaCl on helicase binding. i) Control, DNA probe in buffer, ii) DNA probe in buffer + 150mM NaCl, iii) helicase reaction, iv) helicase reaction with 150mM NaCl added, v) helicase reaction with NP40 and β mercaptoethanol.

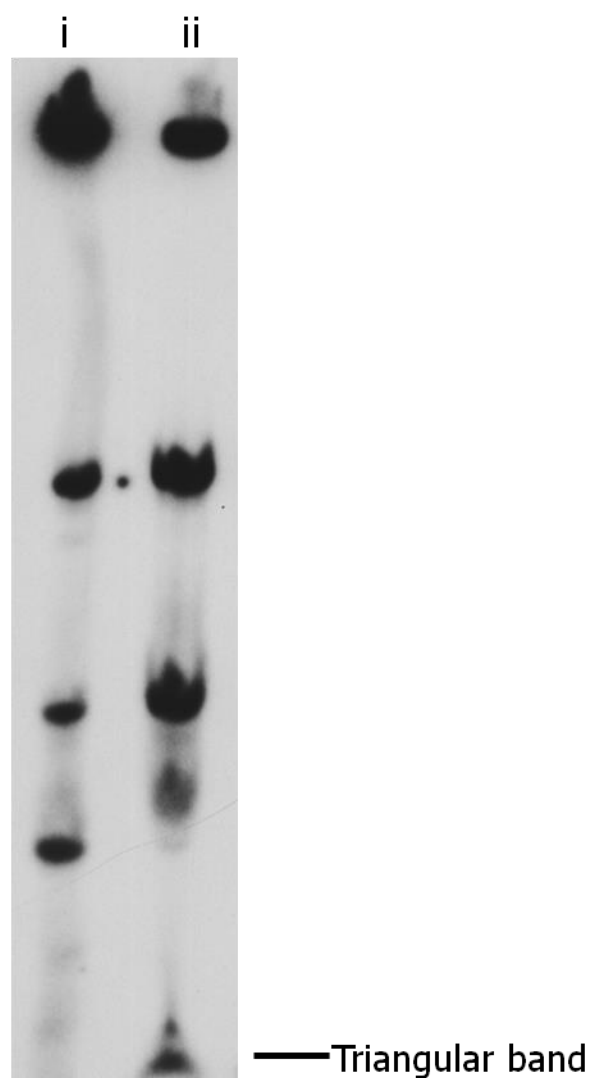


Figure A2: Effect of helicase buffer on DNA probe. i) artificially forked dsDNA probe in loading dye alone, ii) artificially forked dsDNA probe with helicase reaction buffer and loading dye.

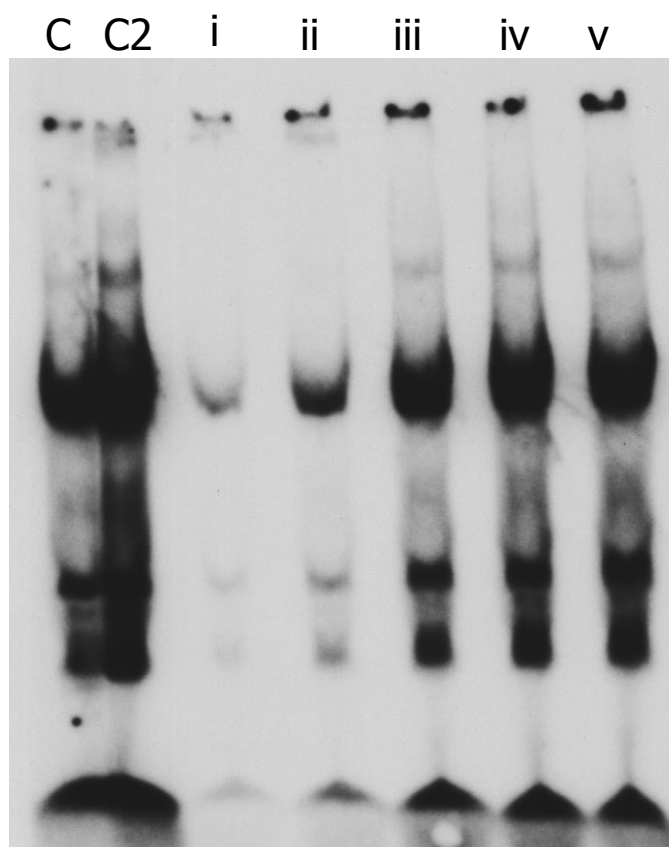


Figure A3: Helicase reaction over time with Mg^{2+} . Reactions were pre-loaded the Mg^{2+} for 5 minutes at 37°C. Mg^{2+} was added for i) 15s, ii) 30s, iii) 45s, iv) 60s, v) 75s. C is the control reaction without helicase, C2 reaction was preloaded without Mg^{2+} for the duration of the experiment.

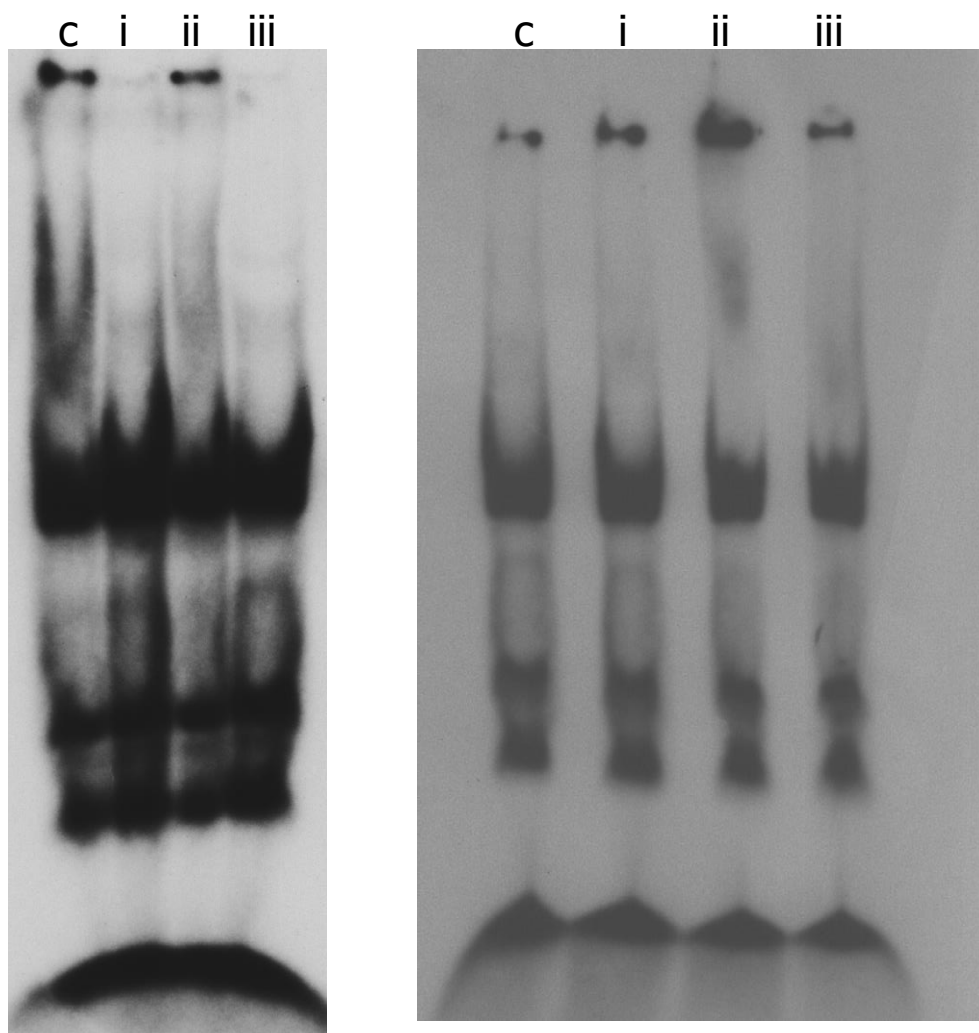
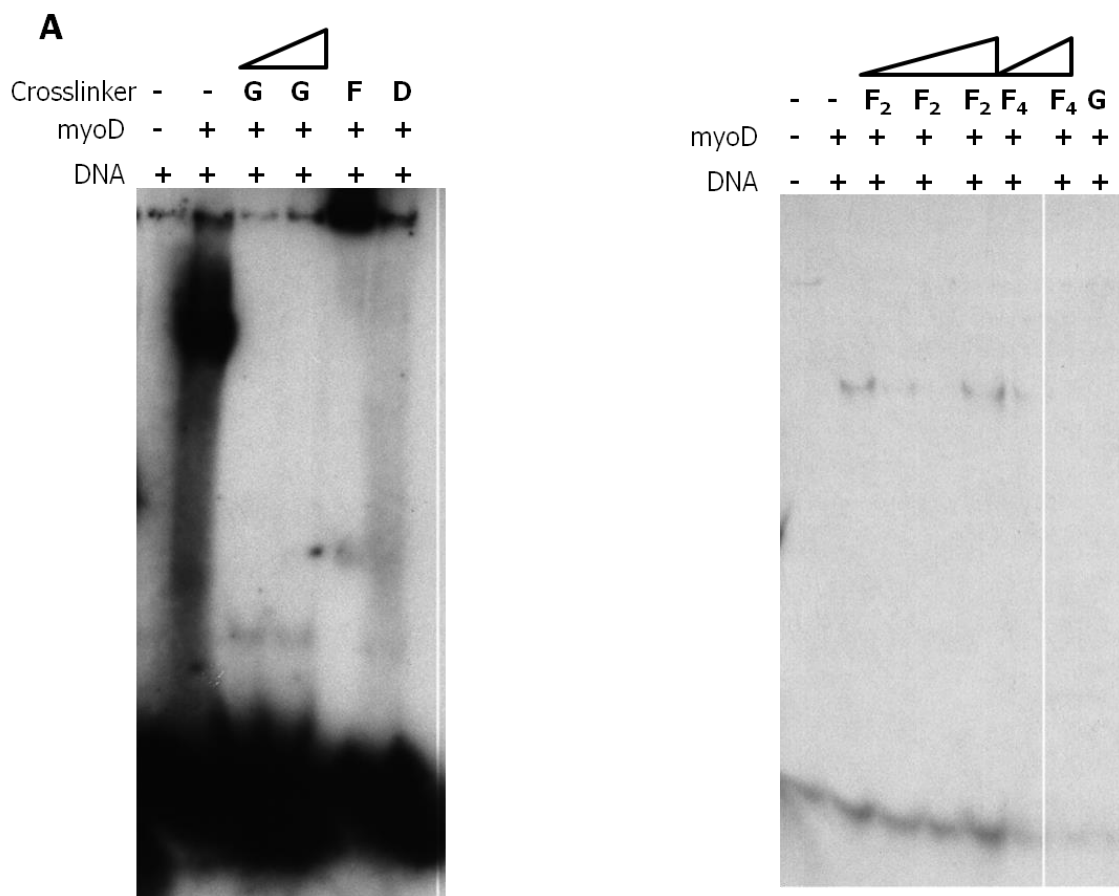


Figure A4: Helicase-Avidin preloading experiment. Control (c) and 0.1 (i), 0.2 (ii) and 0.3 (ii) μ g of helicase-avidin protein regularly produce smears (two experiments shown).



Figure

A5: Alternate exposures for Figure 5.1: Effect of crosslinkers on MyoD binding. A) MyoD protein bound to E-box DNA sequence, with or without MyoD and crosslinking agent. G represents 0.02% glutaraldehyde at 20s (left) and 60s (right), F represents 0.365% formaldehyde for 4 minutes and D represents 2.5 mM DSS for 4 minutes. B) Further investigation on effects of formaldehyde. F2 represents formaldehyde at 0.02% with times of 20s, 60s and 4 minutes (left to right), F4 represents formaldehyde at 0.04% with times at 20s (left) and 60s (right). G represents 0.02% glutaraldehyde for 30 seconds.

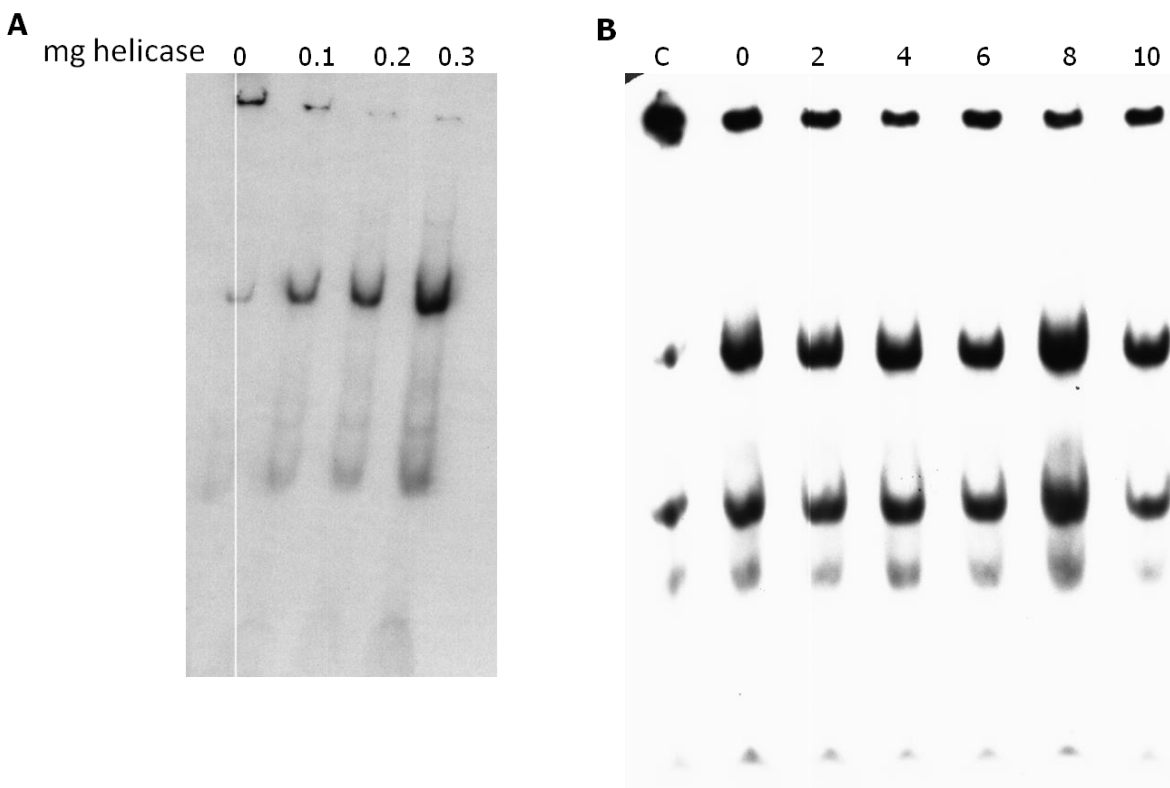


Figure A6: Alternate exposures for Figure 5.2: Helicase binding to DNA. A) 0.1, 0.2 and 0.3 mg helicase protein were incubated with artificially forked DNA. B) 2 μ L of helicase was incubated with 0.75 μ L of radioactively labeled artificially forked dsDNA. C represents control which was dsDNA in the buffer without protein added. Numbered wells represent the time of incubation in minutes. The reactions were crosslinked with formaldehyde, loaded onto a 5% native acrylamide gel electrophoresis and exposed to X-ray film for visualization.

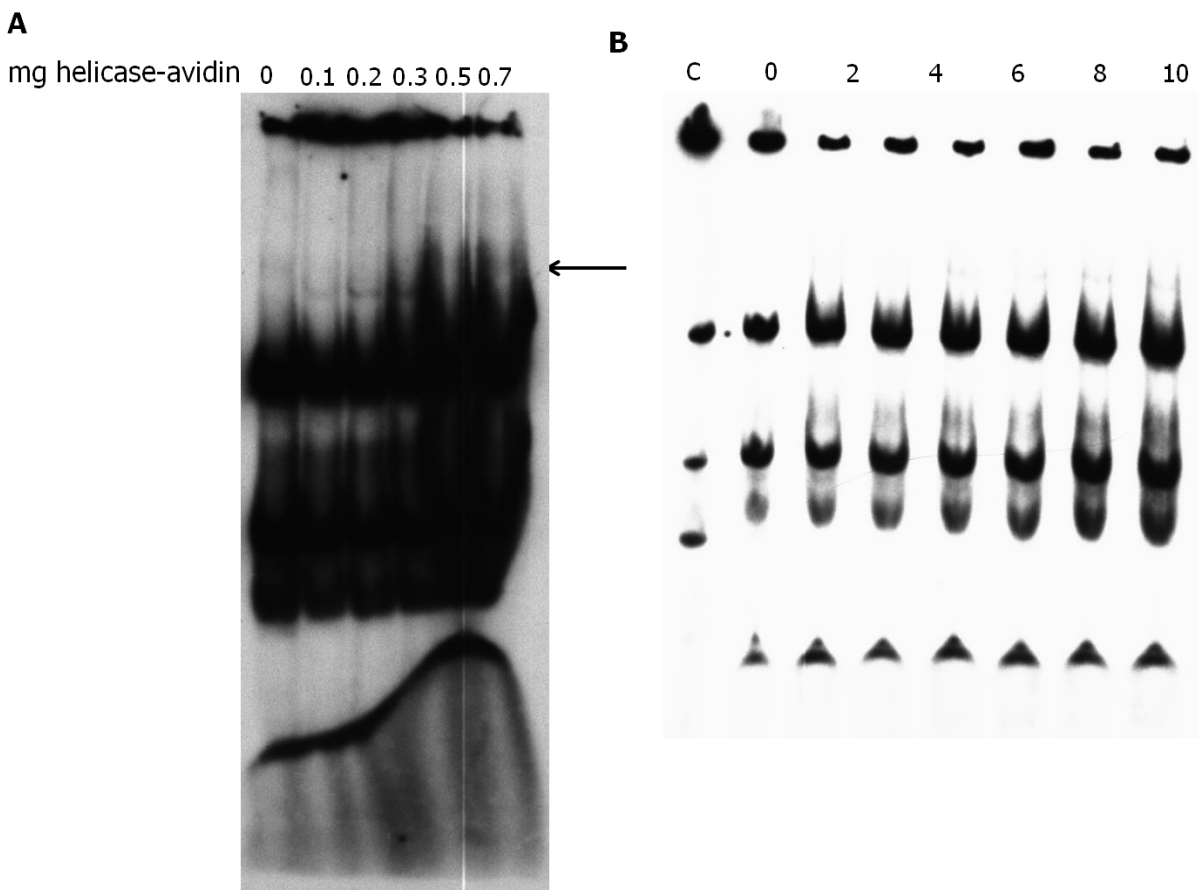


Figure A6: Alternate exposures for Figure 5.3: Helicase-avidin binding to DNA. A) 0.1, 0.2, 0.3, 0.5 and 0.7 mg helicase proteins were incubated with artificially forked DNA, B) 6 μ L of helicase-avidin was incubated with 2 μ L of radioactively labeled artificially forked dsDNA for 5 minutes in a 25°C waterbath. C represents control which was dsDNA in the buffer without protein added. Numbered wells represent the time of incubation in minutes. The reactions were crosslinked with formaldehyde, loaded onto a 5% native acrylamide gel for eletrophoresis, and exposed to X-ray film for visualization.

Appendix V – Helicase Assays

The most suitable technique for further characterization of the helicase enzymes is the reagentless biosensor for inorganic phosphate (P_i) release of the NTPase activity [131, 132]. This provides as little interference as possible with the helicase system as it only requires one additional component which interacts with the P_i byproduct of the dTTPase reaction coupled to helicase's translocation along DNA. By measuring P_i release kinetics, ATP usage and DNA unwinding information is provided.

Coupled enzyme assays measure the formation of byproducts of the reaction such as ADP and P_i using a series of many components [102, 132, 145]. Such components may interfere with the helicase system and must be fast relative to helicase's own action whereas reagentless biosensors directly measure the biproducts formed by the helicase reaction. Reagentless biosensors exist for P_i , ADP, and ssDNA. ADP is not a suitable choice for this research undertaking due to bacteriophage T7gp4a preferring dTTP as its nucleotide base, and ssDNA is not suitable as translocation along ssDNA will also be studied.

Techniques which analyze helicase function by modifications of the various molecules involved are not desirable because of the possibility of skewing the results and thus negating their applicability to the DNA sequencing technique. Such modifications include:

1. Modification of the nucleotide base [132] with a fluorescent analogue, often at the ribose sugar, to detect binding and unbinding of the base from the enzyme. This does not affect the reaction but does not provide any sort of kinetic information, and would be more costly to include in the final device.
2. Modification of the DNA strands by fluorescence resonance energy transfer (FRET) [132, 146] or a fluorophore-quencher pair [132, 147] have been used to determine unwinding of double stranded DNA as an all-or-none response. This technique involves labeling one strand of a nick in DNA with a fluorescent donor and the other with a fluorescent acceptor. This technique is not suitable because it relies absolutely on dsDNA being used, the biotin oligonucleotide hypothesis could not

therefore be tested, and it is an all-or-none response therefore the ongoing action of helicase is not observed [148]. Generally, this technique is used technique to investigate nicks [149].

3. End-labeling the 5' end of ssDNA with a fluorescent nucleotide as the terminal base [132, 150], or with fluorophore sensitive to its environment [132, 151] such as fluorescein.

Fluorescence anisotropy [152] is a technique which uses fluorescent labeling of DNA molecules and the photoselective excitation of such fluorophores by polarized light. This gives information about rotational motion as well as equilibrium and kinetic studies. While it requires modifications of the basis, it is also overly complex for requirements of this project.

Fluorescence cross-correlation spectroscopy (FCCS)) [152] labels dsDNA with different dual-colour two-photon excitation fluorophores to detect the separation of strands. This method gives information on unwinding activity as well as real-time kinetics. This method is overly complex for the purposes of this research undertaking and requires complex equipment.

Intercalating dyes which bind dsDNA [153] are used to detect unwinding. This is a more simplistic method, however is not suitable for the purpose of this research that may also look at the translocation of helicase along ssDNA. Dyes used for this method include Hoechst 33258 (absorbance = 344 nm, emission = 487 nm, greatest specificity for dsDNA over ssDNA), Thiazole Orange (absorbance = 504nm, emission = 528 nm), Bis-Benzimide (DAPI) (absorbance = 345 nm, emission = 467 nm),

Triplex displacement assays bind a 3rd strand of DNA at some parts of the DNA substrate. Fluorescence occurs upon the release of this 3rd strand of DNA during unwinding. This provides all or none data at that time point [133] simply indicating whether or not the helicase has translocated and unwound the DNA probe. The third strand may disrupt the interaction of helicase with the substrate DNA, and does not allow for the study of ssDNA. For these reasons it is not suitable for this research undertaking.

Reagentless Biosensing of Inorganic Phosphate Release

This reagentless biosensing of helicase activity is based on the *Escherichia coli*'s phosphate binding protein (PBP) [131]. PBP is a periplasmic protein that scavenges phosphate for importation into the cell when P_i concentrations within the cell are low [154]. In 1994 a fluorophore was developed, N- [2-(l-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC), which was attached to a mutated A197C residue in PBP [155]. MDCC in free solution has a maximal excitation at a wavelength of 425 nm with maximal emission at 474 nm, but in the presence of P_i the maximal emission goes to 464 nm with a 13 fold increase in fluorescence intensity [154]. The absorbance spectra to track this phenomenon and the kinetic reaction could be obtained from a fluorimeter with a xenon lamp [154, 154, 156, 157]

MDCC-PBP can prepared in the laboratory [154, 155] or purchased in solution from Thermo Fisher Scientific™ [158] as it is used to measure the kinetics of P_i release and ATP usage in other enzyme systems [132].