UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEVELOPMENT OF TOOLS FOR THE PREPARATION OF GENOMIC LIBRARIES FOR NEXT-GENERATION SEQUENCING

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioengineering

by

Aric Joneja

Committee in charge:

Professor Xiaohua Huang, Chair
Professor Michael J. Heller
Professor Pamela Itkin-Ansari
Professor Bing Ren
Professor Kun Zhang

2011

The Dissertation of Aric Joneja is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

EPIGRAPH

Why waste time learning, when ignorance is instantaneous?

Bill Watterson

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and assistance I received from my supervisors, colleagues, family, and friends. The group of people who deserve some portion of the credit (or blame) for the quality of this dissertation is too numerous to list exhaustively here; however, this document would be incomplete without the following acknowledgements.

My thesis advisor, Dr. Xiaohua Huang, has provided his full scientific, financial, and personal support for the duration of his mentorship. The breadth of his technical expertise and his willingness to share his experience and knowledge have allowed me to develop disparate skills that will serve me well for the rest of my career. He has always enthusiastically encouraged me to pursue my own research interests and ideas, and has never hesitated to work with me side by side in the lab to help solve problems. It is his leadership and guidance that allowed me to gain the confidence and technical abilities necessary for research success.

I would like to thank the rest of my thesis committee: Dr. Bing Ren, Dr. Pamela Itkin-Ansari, Dr. Kun Zhang, and Dr. Michael Heller, for their input, advice, and encouragement. They have also kindly volunteered the services of their equipment and the expertise of the senior members of their labs. I would like to make special mention of Dr. Sam Chiang from Dr. Zhang's lab, who greatly facilitated my real-time DNA amplification detection.

I would like to thank Eric Roller, who has been both a colleague and close friend for many years. I have relied on him repeatedly to help me troubleshoot problems or more fully understand concepts and fundamental principles related to our work. I am always impressed by his ability to provide insight on projects in the lab that he may be only peripherally involved in. He is the first person I turn to for assistance and he has never hesitated to dedicate himself to finding a solution.

I would like thank Dr. Kris Barbee for recruiting me to Dr. Huang's group, and for his leadership and initiative in all aspects of lab construction, management, and maintenance. My research benefited from his technical expertise, and he was always willing to devote time and energy to assisting me in the lab without thought of compensation. His indefatigable work ethic was both daunting and an inspiration.

I would like to thank Dr. Ying-Ja Chen. As the most senior member of the lab when I began, I learned many of the most essential laboratory techniques from her. She has continued to be a source of expertise for me whenever I had questions regarding imaging or biochemistry.

I would like to thank some other members of the lab, each of whom has excelled in their particular field and has shared their discoveries with me, in many cases directly impacting my research and in all cases imparting valuable knowledge that furthers my capabilities and credibility as a scientist. I am grateful to Nora Theilacker for sharing her expert advice on bioconjugate techniques, particularly the use of magnetic beads and biotin-streptavidin binding and elution; Alex Hsiao and Matt Chandrangsu for their knowledge of microfabrication and microfluidics; Robert

Mifflin for his intimate knowledge of all things electronic and mechanical; Zhixia Liu for her pioneering work on protein mutation, expression, and purification; Kwang-Seuk Ko for sharing his knowledge of organic chemistry and compound synthesis, and Matt Walsh and Geuntak Lee for their work towards single-molecule imaging and sequencing.

I would like to thank Dr. Wayne Giles and Dr. Amy Sung for their guidance during my first years of this program. They generously and without obligation donated their time and experience to help me choose a research group and define my projects.

I would like to thank Karen Wei for her unwavering support both personally and professionally. She has patiently and tirelessly encouraged me, wordlessly inspired me, and I would not have had the endurance to complete this degree without her misguided faith in me.

Lastly, I would like to thank my parents and brother for motivating me to set and achieve lofty goals. They encouraged me to choose my own path and believed in my ability to succeed in any endeavor. They would also occasionally provide money so that I could afford vegetables.

Chapter 2, in part, is a reproduction of the material as it appears in: Joneja, A. and Huang, X. (2009) A device for automated hydrodynamic shearing of genomic DNA. *Biotechniques*, **46**, 553-556. Used with permission. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is a reproduction of the material as it appears in: Joneja, A. and Huang, X. (2011) Linear nicking endonuclease-mediated strand-displacement

DNA amplification. *Anal Biochem*, **414**, 58-69. Used with permission. The dissertation author was the primary investigator and author of this paper.

CURRICULUM VITA

EDUCATION

University of California, San Diego. La Jolla, CA USA
Ph.D., Bioengineering, September 2011
M.S., Bioengineering, December 2006

Sydney University. Sydney, Australia.
January - June 2003

University of Maryland. College Park, MD USA
B.S., Biological Engineering, May 2004.

RESEARCH EXPERIENCE

University of California, San Diego.
Genomics and Systems Biotechnology Lab, La Jolla, CA.
Graduate Student Researcher **January 2006 - 2011**

- Research Advisor: Dr. Xiaohua Huang
- Developed a device for the automated mechanical fragmentation of genomic DNA to assist in genomic library construction. Facilitated the exportation of the device to several academic labs and the commercialization of this instrument by Diagenode.
- Established a method to isothermally and linearly amplify a heterogeneous mixture of DNA fragments of length up to 5 kilobases, including small genomic libraries.
- Research towards a platform for using a directed evolution approach to engineer a restriction enzyme for use in preparing genomic libraries for DNA sequencing. Billions of protein variants are assayed in parallel using an *in vitro* system to identify mutants with desired characteristics.

Burnham Institute for Medical Research.
Stem Cells and Regenerative Biology Lab, La Jolla, CA.
Research Assistant **June - December 2006**

- Research Advisor: Dr. Fred Levine
- Used immunohistological staining and fluorescent microscopy to evaluate the efficacy of macroencapsulation devices for the implantation of harvested tissue and engineered cell lines.

University of California, Irvine.
Cardiovascular Engineering Lab, Irvine CA.
Research Assistant **May - August 2002**

- Assisted in measurements of the residual stress of a porcine aorta and coronary arterial tree to establish a relationship between residual stress and blood vessel diameter. Also measured the mean circumferential stress and midwall stretch ratio to demonstrate the homeostasis of strain in the cardiovascular system.

HONORS AND AWARDS

- Graduated *cum laude* from Honors Program at University of Maryland, 2004
- Phi Sigma Pi National Honor Fraternity, 2002-present
- Recipient of President's Scholarship from University of Maryland, 2000-2004
- Candidate for Presidential Scholars Program, 2000
- Recipient of National Merit Scholarship, 2000

TEACHING EXPERIENCE

Teaching Assistant
Shared responsibility for discussion sections, homework assignments, exams, and grades. Assisted the instruction of courses at both the undergraduate and graduate level including

- BENG 112B, Biomechanics
- BENG 230A, Biochemistry
- BENG 230B, Cell Biology
- BENG 230C, Cardiovascular Physiology

PUBLICATIONS

Joneja A, Huang X. A device for automated hydrodynamic shearing of genomic DNA. *Biotechniques*. 2009 Jun;46(7):553-6.
Joneja A, Huang X. Linear nicking endonuclease-mediated strand-displacement DNA amplification. 2011 *Anal Biochem*, **414**, 58-69.

PROFESSIONAL SKILLS

Instrumentation

- Fluorescent Microscopy
- Real-time PCR
- Spectrophotometry
- Lab Automation (Fluidics)

Cell and Tissue

- Immunohistology

- ELISA
- Bacterial Cell Culture

Biochemistry

- Gel Electrophoresis of DNA (agarose, alkaline, PAGE)
- Gel Electrophoresis of Proteins (SDS-PAGE, native PAGE)
- Protein Expression (*in vivo* and *in vitro*)
- Protein Purification
- Protein-DNA Interactions
- DNA Amplification and Purification
- Bioconjugate techniques

Genetic Engineering

- Gene Cloning and Mutation
- Recombinant DNA Techniques
- Protein Engineering and Directed Evolution

Languages

- Competent in French

ABSTRACT OF THE DISSERTATION

DEVELOPMENT OF TOOLS FOR THE PREPARATION OF GENOMIC LIBRARIES FOR NEXT-GENERATION SEQUENCING

by

Aric Joneja

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2011

Professor Xiaohua Huang, Chair

Recent advances in DNA sequencing technologies have allowed researchers to decrease the cost and time requirements for genomic sequencing by orders of magnitude. Investments in novel sequencing methods and improvements to existing

next-generation sequencing platforms have resulted in much higher accuracy and lower cost-per-base for genomic sequencing using a variety of chemistries. However, one of the bottlenecks for all next-generation sequencing methods is the amount of time and resources required for template and library preparation. This process typically results in considerable sample loss and low throughput. Two of the steps that are in need of improvement are the fragmentation of long DNA strands and the amplification of fragmented genomes. To enable the creation of high quality genomic libraries, I have developed an automated device, based on a syringe pump, for the random fragmentation of genomic DNA. The length of the resulting fragments is tunable using a single parameter and the ends are easily repaired for the efficient ligation of adapters. I have also developed a method for the unbiased linear amplification of long DNA fragments using the concerted activities of a nicking endonuclease and a polymerase. The optimization of reaction conditions resulted in markedly better performance than existing similar protocols and I have demonstrated the utility of this method in amplifying a fragmented phage genome. I explored methods of using a highly specific homing endonuclease for use with this technique, including the incorporation of non-native nucleotides and the engineering of the enzyme's catalytic site. Lastly, I have made significant progress in efforts towards the engineering of the nuclease's DNA recognition residues in order to create a highly specific nicking enzyme for use in the modification or amplification of large genomes.

Chapter 1: Introduction

1.1 Significance of genomic sequencing

The potential applications of genome sequencing have been recognized and have continued to grow since the first sequencing technologies were developed by Frederick Sanger in 1975 (1,2). They led to the initiation of the Human Genome Project (HGP) in 1990 and increased funding for genomic sequencing technologies in recent decades. The cost of DNA sequencing, prohibitive for most applications at $10/base in 1990, has been reduced by several orders of magnitude. Low-cost sequencing has enabled the investigation of: genotype-phenotype associations, gene-expression profiling, identification of pathogens, and comparative genomics, amongst many others (3). Indirectly, DNA sequencing technologies have provided information on transcriptomes at the cell, tissue and organism level via RNA-seq (4) and epigenetic marks via ChIP-seq (5).Miniaturization and multiplexing of sequencing devices as well as the development of novel sequencing chemistries promise to reduce the cost much further.

Perhaps the most compelling potential application of DNA sequencing is health care. In 2003 the X Prize Foundation offered an incentive of $10 million dollars to the first team to accurately sequence a human genome for less than $1000. The National Human Genome Research Institute (NHGRI) then awarded grants in excess of $32 million dollars towards the goal of this “$1000 dollar genome”. The assumption of both institutions was that at this cost, it becomes viable for “health care

professionals to tailor diagnosis, treatment, and prevention to each person's unique genetic profile" (6). The detection and identification of single nucleotide polymorphisms (SNPs) in an individual genome from a growing library of SNPs can assist a doctor in predicting a disease long before the symptoms are manifested. The era of personalized medicine will allow unprecedented accuracy and efficiency in disease treatment and prevention. A disease of particular interest is cancer. The ability to sequence and compare the genomes of normal, neoplastic, and malignant cells would greatly help elucidate the mutations that lead to cancerous cells (3). A greater understanding of the underlying genetic basis of the disease will suggest new ways to combat its development and spread. In short, the motivations for developing faster and cheaper sequencing platforms are well established.

1.2 Next Generation Sequencing

Despite many technical improvements to Sanger sequencing during the last two decades, its inherent limitations, particularly in regard to miniaturization and parallelization, exposed a need for new and improved technologies for sequencing large numbers of human genomes. These massively parallel instruments and chemistries were broadly termed "next-generation sequencing" methods and this class of technologies includes a diverse portfolio of approaches to querying the identity of DNA bases.

Many iterations of sequencing by synthesis (SBS) have been developed and currently exist as commercial platforms through Roche/454 FLX (pyrosequencing) (7), the Illumina/Solexa HiSeq and Genome Analyzer (8), and the Heliscope from Helicos Bio. In the near future two additional SBS platforms are expected to make their way to market: the Pacific Biosciences SMRT instrument and the Ion Torrent Personal Genome Machine. A competing chemistry that has found success via the Applied Biosystems/Life Technologies SOLiD System is sequencing by ligation (SBL). Lastly, a promising but relatively very immature approach is nanopore sequencing (9). This method uses a nanosensor to measure a change in some physical or chemical property as a single DNA molecule passes through a nanopore. This concept has received funding from the NIH in recent years and shows great potential, but significant advances in pore engineering and nanofabrication will required to make this a useful technology.

The figure below illustrates the steps common to most modern commercial cyclic array sequencing strategies. Chromosomal DNA is first extracted from the cell and isolated from cellular debris. Even relatively small bacterial genomes can be millions of base pairs long, so the purified DNA must be fragmented into much shorter strands in order to be more easily manipulated (10). These fragments are then attached to beads or a surface and clonally amplified in order to provide greater signal during the sequencing reaction. While most groups randomly disperse DNA fragments or beads onto a surface, methods of assembling high-density DNA arrays are being developed to improve throughput and imaging efficiency (11). When the library is

complete, sequencing is initiated and the resulting signal is collected via CCD camera or other detection device. The resulting raw data consists of the individual sequences of millions of small DNA fragments and these must be aligned into the complete genome using assembly algorithms. This is referred to as "shotgun sequencing".

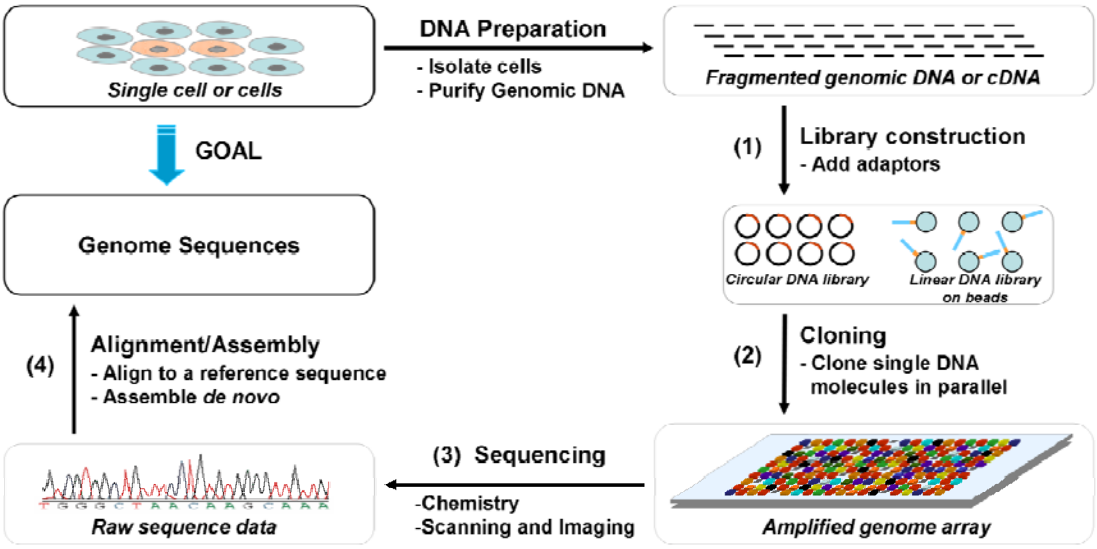

Figure 1.1. The Sequencing Pipeline.

Not all strategies follow this pipeline strictly. In particular, it is important to note that the Pacific Biosciences and Helicos instruments are "single-molecule" sequencers and do not require the amplification of DNA fragments prior to sequencing.

1.3 Library Construction

The need for robust methods that produce a representative, non-biased source of nucleic acid material from the genome under investigation cannot be overemphasized (12). It should be recognized that the "pre-processing" protocols (e.g., library construction) for new sequencing technology platforms are not nearly as

mature as for conventional Sanger dideoxy sequencing. The development of robust, straightforward protocols for *in vitro* library construction for various applications is a critical challenge that must be addressed if investigators are to make the most of next-generation sequencing technologies (13). Two of the steps shared by most sequencing strategies that are in need of improvement are genome fragmentation and genome amplification.

1.3.1 DNA Fragmentation

The first step to library construction is the fragmentation of chromosomal DNA. Currently there are five methods that are used: enzymatic digestion, sonication, ultrasonication, nebulization, and hydrodynamic shearing. They have all been successfully used to make sequencing libraries, but each has its limitations and therefore none of them has emerged as the preferred method for all applications.

During an enzymatic digestion, restriction endonucleases(14,15) or DNAse I (16) are added to the chromosomal DNA and the rate and scope of the digestion is controlled to achieve a desired fragment size. There are many modes of control available including methylation of target DNA, inclusion of competing divalent cations, using a time course for digestion, or by histone protection (10). Unfortunately the use of specific endonucleases inherently carries a sequence bias, and this is unacceptable for sequencing applications. Regions of DNA with few restriction sites are underrepresented or absent from the DNA library, and a lack of overlapping fragments complicates sequence analysis and genome assembly (17). Even the use of a

nonspecific nuclease such as DNAse I, controlled by the presence of Manganese ions(16) or histones (10) exhibits sequence bias, as DNase I has been shown to have both a sequence preference (18,19) and DNA flexibility preference (20), while nucleosome phasing is non-random (21,22).

Fragmenting DNA using physical stress induced by sonication has been suggested as an alternative method of randomly breaking DNA strands (23). Sonication is fast and easy and the target size can be controlled to an extent by varying the amplitude and time of sonication. However, results from sonication are difficult to reliably reproduce and the size distribution of the resulting DNA fragments is large. Many of the fragments produced are unclonable due to damage caused by hydroxyl radicals produced by thermal dissociation of water under the pressures of ultrasonic cavitation (17,24). For these reasons, sonication is not always the preferred method to produce DNA fragments for cloning. A similar method pioneered by Covaris Inc is "Adaptive Focused Acoustics", a form of ultrasonication that harnesses cavitation to fragment DNA. This has several advantages over traditional sonication, including isothermal operation and the ability to focus the acoustic energy into a localized area of the sample due to the much shorter wavelengths used. This approach has become very popular in recent years due to the ease of multiplexing and non-contact nature of the technology.

Nebulization of DNA has become the method of choice for many labs, and kits are commercially available (from Illumina and Invitrogen) to assist in library construction utilizing nebulizers. Nebulizers can break genomic DNA into strands

about 1kb long that have much higher cloning efficiency than those obtained by sonication (25,26). However, the resulting fragments have a very wide size range. Nebulization requires a large volume of DNA solution so dilution of the sample is often necessary. Additionally, to avoid contamination a new disposable nebulizer must be used for each sample to be fragmented. The cost of these nebulizers is nontrivial (a pack of 5 from Invitrogen is over $100).

Most recently, methods of hydrodynamically shearing genomic DNA have been developed. First discovered by Oefner et. al in 1996 (17), a device is now available commercially (Genomic Solutions® Hydroshear). To hydrodynamically shear DNA, a syringe pump is used to rapidly force a DNA solution through a small orifice or length of small-bore tubing. The mode of breakage is two-fold: shear forces within the orifice and extensional strain as the fluid approaches the orifice combine to randomly break long DNA strands (27). The length of product is dependent on the velocity of the solution and the diameter of the orifice and is largely independent of DNA concentration, initial DNA size, temperature, and salt concentration (27). The distribution of sizes is relatively tight, with the longest fragments typically 2-3 times the length of the smallest fragments. This method is the easiest to automate and cloning efficiency of the product is very high. However, there are some drawbacks to this method. Even when using very pure reagents it is easy for the small orifice to get clogged, causing the syringe pump to stall. Unclogging the orifice requires a lengthy procedure where the shearing device is washed and sonicated. It is also very difficult to ensure that the device has been thoroughly washed, leading to the risk of

contamination between runs. Finally, the cost of purchasing the commercial Hydroshear device (in excess of $15,000) is excessive for many researchers.

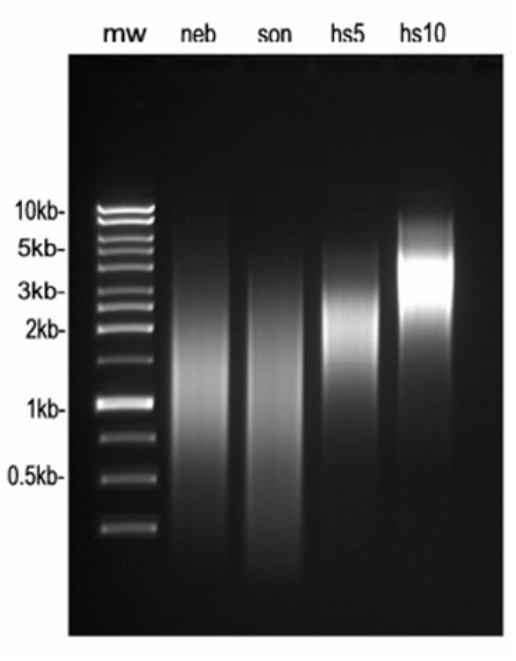

Figure 1.2. Comparison of methods of physically breaking DNA. Lane 1: DNA standards. Lane 2: Typical product of nebulization. Lane 3: Typical product of sonication. Lane 4: Hydrodynamic shearing at high speed. Lane 5: Hydrodynamic shearing at lower speed. (28)

1.3.2 DNA Amplification

After the genomic DNA is fragmented, the next step is often the amplification of the pieces of DNA. This is required by most sequencing chemistries in order to increase the signal-to-noise ratio so that the recorded sequence is accurate. For the amplified DNA to be useful during imaging, it must be somehow segmented so that each template fragment and its amplified products are physically separated from the

other templates in solution or on a surface. For this reason, the templates must be "clonally amplified".

Cell cloning is one of the earliest methods of amplifying DNA and continues to be used today. It was first demonstrated in 1973 that artificial, biologically functional plasmids could be created and inserted into bacteria, and that these plasmids would be replicated along with the cells (29). Thus, single molecules could be amplified millions of times in a manner that was inherently clonal. Once the bacteria received the plasmid, or "Bacterial Artificial Chromosome" (BAC), the DNA could be stored in the cells or a large supply could be created by growing the cells. The Human Genome Project was completed almost entirely using DNA amplified by bacteria. To create a BAC, DNA of interest is digested with a restriction enzyme that produces single-stranded overhangs. When combined with a plasmid that has been similarly digested and a DNA ligase, the DNA is incorporated into the plasmid. This new plasmid can be introduced into a competent cell by electroporation or heat shock (30). While cell cloning is a very useful method, it is also very costly in terms of time and reagents. Improvements in the quality and availability of restriction enzymes and cloning vectors have helped simplify the process. However, cell cloning will always be more laborious, expensive, and complicated to than cell-free methods of DNA amplification.

This is typically done using adaptations of the powerful polymerase chain reaction (PCR) (31,32), although alternate amplification techniques such as rolling circle amplification (RCA) (33,34) have also been used (35). Unlike cell cloning,

however, if a library of templates is amplified with the same primers, products from all templates are mixed in a single volume. This limits its utility for massively parallel sequencing applications. Variations of the standard PCR reaction have been developed to allow clonal amplification, where the amplicons arising from each template in a complex library remain locally clustered. The two most common PCR methods, illustrated in Figure 1.3, are emulsion PCR (36) and solid-phase amplification or "bridge" PCR (37).

Emulsion PCR, favored by Roche/454 and Life Technologies/Applied Biosystems, is carried out in aqueous solutions dispersed in an oil phase to form microscopic compartments (36,38-40). A complex library flanked by universal adaptors can be amplified using a single pair of primers. When the library is dispersed in the emulsion such that only one DNA template exists in each chamber, billions of molecules can be clonally amplified simultaneously. Up to 10^{10} bacterial-sized droplets can be created in a volume of 1 mL (38). Further, streptavidin coated microbeads covered in biotinylated primers can be added to the mix. This enables the solid-phase capture of PCR amplicons within individual aqueous compartments (13). Upon PCR amplification, each bead carries 10^4 to 10^7 copies of a unique DNA template (38). When the emulsion is broken, the beads can be distributed onto a surface, into an array of wells, or immobilized in a polyacrylamide gel. However, it has been shown that the efficiency of emulsion PCR decreases drastically with increasing amplicon length. Efficiency of a 500 bp template is 10% that of a 100 bp template (41).

Bridge PCR, favored by Illumina/Solexa, similarly creates spatially separate clonally amplified product. Universal adaptors are ligated to a library of sequencing templates. Forward and reverse primers are designed to specifically target these adaptors and are immobilized on a glass slide by their 5'-ends using any of several chemical methods (37,42). All other required reagents for PCR are added in the aqueous phase and thermal cycling is performed on the slide. Since all primers are immobilized the copies of each template remain localized, and the result is that each template molecule generates a tight cluster of about 1000 copies (13). Bridge PCR can achieve extremely high density surfaces of up to 10 million colonies/cm^2 (37). Cyclic sequencing can then be performed on this amplified product.

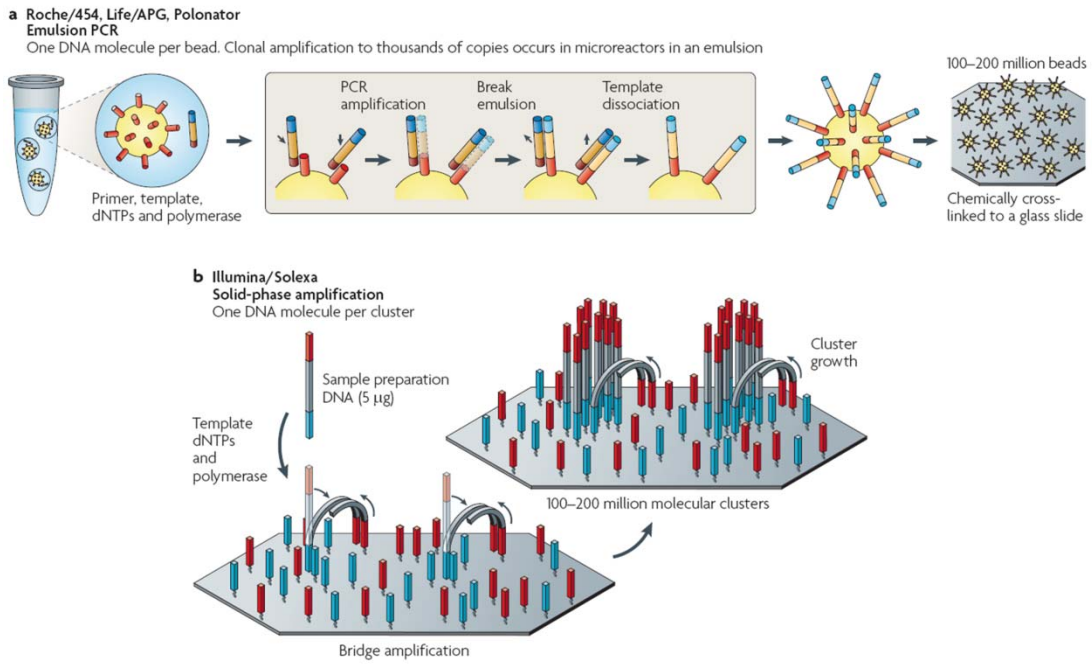

Figure 1.3. Clonal amplification of DNA fragments. (a) Emulsion PCR (b) Solid-phase PCR amplification (12)

These PCR protocols, while effective, do have substantial drawbacks. All thermostable polymerases have nontrivial error rates, introducing mutations into the PCR product. Also, PCR exhibits bias towards AT-rich and GC-rich stretches of sequence, resulting in the overrepresentation of some targets and underrepresentation of others (43,44). This is undesirable when aligning genomes after sequencing and is particularly detrimental to quantitative applications such as RNA-seq (12).

1.4 Homing Endonucleases

Homing is the transfer of a mobile genetic element to a cognate allele that lacks that element resulting in its duplication. Homing endonucleases are rare-cutting enzymes that are encoded by open reading frames embedded within introns or inteins. They promote the homing of their mobile genetic elements by generating strand breaks in the cognate alleles that lack the intervening sequence (45). Homing enzymes can be divided into several unique families based on their structure, including HNH, His-Cys box, GIY-YIG, and LAGLIDAG. As a class they exhibit some properties that are unique to endonucleases. They display extremely high DNA-binding specificities which arise from target sites that can approach 40 base pairs (46). They are tolerant of a range of sequence degeneracy within these sites and display disparate DNA cleavage mechanisms (47). Most homing enzymes create double-stranded breaks in their DNA target but two homing nicking enzymes have been identified (48-50). Homing enzymes have garnered considerable interest in the field of genomics because their unusually long recognition sites and specificity make them uniquely suitable for

genomic research. Many research groups have targeted homing enzymes as substrates for protein engineering and altered specificity investigations (51,52)due to their potential applications in gene therapy.

1.5 Engineering Novel Enzymes

All of the enzymes that exist in biology arose through a Darwinian evolution system over millions of years and through countless cycles of polypeptide mutations. As molecular biology has progressed, we have found many commercial and research uses for these effective enzymes designed by nature. The diversity of living beings has provided a library of unique enzymes with valuable properties. The PCR reaction became a viable method of amplifying DNA only after *Taq* polymerase was discovered and isolated from a thermophilic organism (32). As the relationship between proteomic genotype and phenotype was further elucidated, methods were developed to tailor enzymes to a particular need or condition. A greater understanding of protein structure and function led to the ability to engineer enzymes with altered activity, stability, and specificity (53). These artificial enzymes have found uses in the laboratory as well as industrially, where environments of extreme temperature, pH, or salinity preclude the use of natural enzymes (54). Significant progress in computational models, capable of predicting tertiary structure and protein or nucleotide interactions from an amino acid sequence, has also facilitated the design of novel enzymes. Currently there are two well-established approaches for engineering proteins: rational design and directed evolution.

In rational design, the structure of a protein and its amino acid sequence are analyzed to determine which residues play key roles in the function of the enzyme. Educated guesses can then be made to predict the effect of mutating these key amino acids. Site-directed mutagenesis, utilizing either PCR or cell cloning, is used to introduce the desired mutant base pairs into the gene for the protein. Researchers have successfully used rational design to alter the specificity of enzymes(55-57), enhance an enzyme's thermal properties(58,59), tailor a protein's nanomechanical properties (60),and increase resistance to pH denaturation or proteolytic degradation(61). This approach is only possible when detailed knowledge of the structure and function of the protein are known. However, even if the protein in question has not been studied extensively, its properties can often be inferred by comparing it to well-documented homologous proteins (62) since many classes of enzymes have similar motifs or even segments of identical sequence.

If the enzyme structure or knowledge about the relationship between its sequence, structure, and function are unknown, directed evolution can be applied. In directed evolution, the gene for a protein is subjected to random mutagenesis. A library of mutants is then screened either *in vivo* (63)or using a cell-free translation system *in vitro* (64). Directed evolution implements an iterative Darwinian process whereby the fittest variants are selected from a library of mutants. Unlike natural selection, in which proteins evolve under multiple selection pressures, directed evolution employs a single controlled selection pressure for predetermined functions. Further, directed evolution can obtain non-natural functions of practical use while

natural selection must choose those whose function is most advantageous to a living organism (65). Typically, several rounds of selection are carried out where each round is more stringent than the last to obtain the best possible variant. *In vitro* evolution requires the continued introduction of genetic mutations to offset the loss of diversity due to selection, thereby allowing a more extensive exploration of potentially advantageous sequences (66). This is usually achieved using error-prone PCR (epPCR) or genetic recombination (gene shuffling).

Currently many researchers have turned to a combination of the two methods to achieve the most efficient selection of enzymes with desired properties. This combined approach has been called "semi-rational design" (67) or "focused directed evolution" (68). Once target amino acids have been identified, as in rational design, they are subjected to saturation mutagenesis where every residue is mutated to each of the other 19 amino acids. This results in a large library of variants where the mutations are focused into a desired area, often the substrate-binding site. Some evolutionary pressure is then applied to the library to select the best proteins. A major breakthrough that has assisted all directed evolution studies is the use of *in vitro* compartmentalization (IVC). Developed in 1998 (69), these water-in-oil emulsions were originally designed to serve as compartments for molecular evolution long before the technology was adopted by emulsion PCR. These emulsions miniaturized and parallelized biochemical reactions in which it was desirable to maintain a link between genotype and phenotype. The emulsions are stable across a broad range of temperatures, pH, and salt concentrations, making them perfect for the directed

evolution of billions of molecules or proteins simultaneously. They have been shown to be an effective tool for the selection of endonuclease genes from a library of DNA fragments (70,71). Many modern protein engineering studies make use of both focused directed evolution and IVC (72-74).

1.6 Scope of the Dissertation

The objective of this dissertation work was to develop methods to improve the techniques currently used to generate genomic libraries for DNA sequencing. The pursuit of this goal led to a diverse group of projects that address steps in the genomic library pipeline.

In Chapter 2, we describe the design and assembly of device for the automated fragmentation of genomic DNA. We show that the length of the fragments can be easily controlled by varying the velocity of the DNA in solution as it passes through a filter screen. This device represents an improvement over similar commercially available models.

In Chapter 3, we report the development of protocols for the unbiased linear amplification of long DNA fragments. A DNA polymerase and a nicking endonuclease act in concert in a custom buffer to replicate DNA via a strand displacing mechanism. This optimized method is superior to previous iterations of the technique due to the exclusive use of native nucleotides and the length of fragments that can be amplified.

Lastly, in Chapter 4 we present strategies for the use of a highly specific homing endonuclease in strand displacement amplification reactions. Since the method typically requires a nicking enzyme, methods of protecting one DNA strand from cleavage were necessary. These included the use of non-native nucleotides and the inactivation of the enzyme's cutting site. We also present work towards a platform for the engineering of an endonuclease's recognition sequence. Strategies for mutating targeted amino acids, expressing millions of mutant variants in parallel, and selecting desirable genes from the pool using biotin-streptavidin binding are explored. Progress made using the homing endonuclease I-PpoI is described. The platform may be useful for engineering other endonucleases or investigating the plasticity of the recognition sequences of existing nucleases.

1.7 References

1. Sanger, F. and Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*, **94**, 441-448.

2. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463-5467.

3. Shendure, J., Mitra, R.D., Varma, C. and Church, G.M. (2004) Advanced sequencing technologies: methods and goals. *Nature reviews*, **5**, 335-344.

4. Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews*, **10**, 57-63.

5. Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews*, **10**, 669-680.

6. Spencer, G. (2005) NIH News release

7. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376-380.

8. Bentley, D.R. (2006) Whole-genome re-sequencing. *Current opinion in genetics & development*, **16**, 545-552.

9. Deamer, D.W. and Akeson, M. (2000) Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in biotechnology*, **18**, 147-151.

10. Azzoni, E., Sblattero, D., Licciulli, M., Marzari, R. and Edomi, P. (2007) Using archaeal histones for precise DNA fragmentation. *Protein Eng Des Sel*, **20**, 267-271.

11. Barbee, K.D. and Huang, X. (2008) Magnetic assembly of high-density DNA arrays for genomic analyses. *Analytical chemistry*, **80**, 2149-2154.

12. Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nature reviews*, **11**, 31-46.

13. Shendure, J.A., Porreca, G.J. and Church, G.M. (2008) Overview of DNA sequencing strategies. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al*, **Chapter 7**, Unit 7 1.

14. Hoheisel, J.D., Nizetic, D. and Lehrach, H. (1989) Control of partial digestion combining the enzymes dam methylase and MboI. *Nucleic acids research*, **17**, 9571-9582.

15. Wong, K.K., Markillie, L.M. and Saffer, J.D. (1997) A novel method for producing partial restriction digestion of DNA fragments by PCR with 5-methyl-CTP. *Nucleic acids research*, **25**, 4169-4171.

16. Anderson, S. (1981) Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic acids research*, **9**, 3015-3027.

17. Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J. and Davis, R.W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic acids research*, **24**, 3879-3886.

18. Fedor, M.J. and Daniell, E. (1983) DNase I cleavage of adenoviral nucleoprotein. *Nucleic Acids Res*, **11**, 4417-4434.

19. Lehlanc, B. and Moss, T. (1994), *DNA-Protein Interactions*, Vol. 30.

20. Hogan, M.E., Roberson, M.W. and Austin, R.H. (1989) DNA flexibility variation may dominate DNase I cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 9273-9277.

21. Blank, T.A. and Becker, P.B. (1995) Electrostatic mechanism of nucleosome spacing. *J Mol Biol*, **252**, 305-313.

22. Blank, T.A. and Becker, P.B. (1996) The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing. *J Mol Biol*, **260**, 1-8.

23. Deininger, P.L. (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal Biochem*, **129**, 216-223.

24. McKee, J.R., Christman, C.L., O'Brien, W.D., Jr. and Wang, S.Y. (1977) Effects of ultrasound on nucleic acid bases. *Biochemistry*, **16**, 4651-4654.

25. Bodenteich, A., Chissoe, S., Wang, Y.F. and Roe, B.A. (1993) In Venter, J. C. (ed.), *Automated DNA Sequencing and Analysis Techniques*. Academic, London.

26. Roe, B.A. (2004) Shotgun library construction for DNA sequencing. *Methods Mol. Biol.*, **255**, 171-187.

27. Thorstenson, Y.R., Hunicke-Smith, S.P., Oefner, P.J. and Davis, R.W. (1998) An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Res.*, **8**, 848-855.

28. Schoenfeld, T., Godiska, R., Patterson, M. and Meade, D., *eLucidations*. Lucigen Corporation.

29. Cohen, S.N., Chang, A.C., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA*, **70**, 3240-3244.

30. Osoegawa, K., Jong, P.J.d., Frengen, E. and Ioannou, P.A. (2003), *Current Protocols in Molecular Biology*.

31. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, **51 Pt 1**, 263-273.

32. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

33. Fire, A. and Xu, S.Q. (1995) Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA*, **92**, 4641-4645.

34. Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C. and Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.*, **19**, 225-232.

35. Pihlak, A., Bauren, G., Hersoug, E., Lonnerberg, P., Metsis, A. and Linnarsson, S. (2008) Rapid genome sequencing with short universal tiling probes. *Nature biotechnology*, **26**, 676-684.

36. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. and Vogelstein, B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8817-8822.

37. Fedurco, M., Romieu, A., Williams, S., Lawrence, I. and Turcatti, G. (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*, **34**, e22.

38. Griffiths, A.D. and Tawfik, D.S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends in biotechnology*, **24**, 395-402.

39. Diehl, F., Li, M., He, Y., Kinzler, K.W., Vogelstein, B. and Dressman, D. (2006) BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nature methods*, **3**, 551-559.

40. Nakano, M., Komatsu, J., Matsuura, S., Takashima, K., Katsura, S. and Mizuno, A. (2003) Single-molecule PCR using water-in-oil emulsion. *Journal of biotechnology*, **102**, 117-124.

41. Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **309**, 1728-1732.

42. Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P. and Kawashima, E. (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res*, **28**, E87.

43. Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W. *et al.* (2008) Whole-genome sequencing and variant discovery in C. elegans. *Nature methods*, **5**, 183-188.

44. Dohm, J.C., Lottaz, C., Borodina, T. and Himmelbauer, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*, **36**, e105.

45. Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res*, **25**, 3379-3388.

46. Gimble, F.S., Moure, C.M. and Posey, K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol*, **334**, 993-1008.

47. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Quarterly reviews of biophysics*, **38**, 49-95.

48. Goodrich-Blair, H., Scarlato, V., Gott, J.M., Xu, M.Q. and Shub, D.A. (1990) A self-splicing group I intron in the DNA polymerase gene of Bacillus subtilis bacteriophage SPO1. *Cell*, **63**, 417-424.

49. Landthaler, M., Shen, B.W., Stoddard, B.L. and Shub, D.A. (2006) I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J Mol Biol*, **358**, 1137-1151.

50. Shen, B.W., Landthaler, M., Shub, D.A. and Stoddard, B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.

51. Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Molecular cell*, **10**, 895-905.

52. Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*, **31**, 2952-2962.

53. Hibbert, E.G. and Dalby, P.A. (2005) Directed evolution strategies for improved enzymatic performance. *Microbial cell factories*, **4**, 29.

54. Lehmann, M. and Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current opinion in biotechnology*, **12**, 371-375.

55. Scrutton, N.S., Berry, A. and Perham, R.N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature*, **343**, 38-43.

56. Scrutton, N.S., Berry, A., Deonarain, M.P. and Perham, R.N. (1990) Active site complementation in engineered heterodimers of Escherichia coli glutathione reductase created in vivo. *Proceedings*, **242**, 217-224.

57. Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R. and Rutter, W.J. (1985) Redesigning trypsin: alteration of substrate specificity. *Science*, **228**, 291-297.

58. Shih, P. and Kirsch, J.F. (1995) Design and structural analysis of an engineered thermostable chicken lysozyme. *Protein Sci*, **4**, 2063-2072.

59. Haney, P.J., Badger, J.H., Buldak, G.L., Reich, C.I., Woese, C.R. and Olsen, G.J. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 3578-3583.

60. Li, H. (2007) Engineering proteins with tailored nanomechanical properties: a single molecule approach. *Organic & biomolecular chemistry*, **5**, 3399-3406.

61. Akasako, A., Haruki, M., Oobatake, M. and Kanaya, S. (1995) High resistance of Escherichia coli ribonuclease HI variant with quintuple thermostabilizing mutations to thermal denaturation, acid denaturation, and proteolytic degradation. *Biochemistry*, **34**, 8115-8122.

62. Davidson, A.R. (2006) In Guerois, R. (ed.), *Protein Design Methods and Applications*. Humana Press, Vol. 340.

63. Cherry, J.R., Lamsa, M.H., Schneider, P., Vind, J., Svendsen, A., Jones, A. and Pedersen, A.H. (1999) Directed evolution of a fungal peroxidase. *Nature biotechnology*, **17**, 379-384.

64. Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389-391.

65. Sen, S., Venkata Dasu, V. and Mandal, B. (2007) Developments in directed evolution for improving enzyme functions. *Applied biochemistry and biotechnology*, **143**, 212-223.

66. Joyce, G.F. (2004) Directed evolution of nucleic acid enzymes. *Annual review of biochemistry*, **73**, 791-836.

67. Chica, R.A., Doucet, N. and Pelletier, J.N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current opinion in biotechnology*, **16**, 378-384.

68. Park, S., Morley, K.L., Horsman, G.P., Holmquist, M., Hult, K. and Kazlauskas, R.J. (2005) Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. *Chemistry & biology*, **12**, 45-54.

69. Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nature biotechnology*, **16**, 652-656.

70. Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N. and Yanagawa, H. (2004) In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Res*, **32**, e95.

71. Zheng, Y. and Roberts, R.J. (2007) Selection of restriction endonucleases using artificial cells. *Nucleic Acids Res*, **35**, e83.

72. Rothe, A., Surjadi, R.N. and Power, B.E. (2006) Novel proteins in emulsions using in vitro compartmentalization. *Trends in biotechnology*, **24**, 587-592.

73. Mastrobattista, E., Taly, V., Chanudet, E., Treacy, P., Kelly, B.T. and Griffiths, A.D. (2005) High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & biology*, **12**, 1291-1300.

74. Sepp, A. and Choo, Y. (2005) Cell-free selection of zinc finger DNA-binding proteins using in vitro compartmentalization. *J Mol Biol*, **354**, 212-219.

Chapter 2: A Device for Automated Hydrodynamic Shearing of Genomic DNA

2.1 Abstract

We describe a device for automated fragmentation of genomic DNA by hydrodynamic shearing using a filter screen with uniform pores. Human genomic DNA can be fragmented reproducibly to 2 to 12 thousand base pairs by using various fluid flow rates and screens with 0.5 μm to 10 μm pores. The utilization of disposable screens eliminates sample cross-contamination and the tendency of device clogging commonly encountered in single-orifice shearing devices.

2.2 Introduction

The fragmentation of chromosomal DNA and production of unbiased genomic DNA libraries is a critical and sometimes rate-limiting step in the DNA sequencing pipeline, especially for many next-generation sequencing platforms. Currently four methods are commonly used for DNA fragmentation: enzymatic digestion (1-8), sonication (9,10), nebulization (11-13) and hydrodynamic shearing (14,15). Even though all have been used in library construction, these methods have limitations; as examples, the biases produced by endonuclease digestion, low cloning efficiency due to DNA damage by sonication (14,16,17), large size distribution and difficulty in automation by nebulization, and the high cost and clogging issues by hydrodynamic shearing using instruments with single-orifice devices.

The hydrodynamic shearing method is widely used for shearing genomic DNA (14,15). The DNA molecules are fragmented by hydrodynamic shear stress when they are forced through a point-sink such as a small orifice or the bore of small-diameter tubing at high velocity. Genomic DNA can be sheared randomly with a tight 2-3 fold size distribution and the cloning efficiency of the sheared DNA fragments is very high (14,15). The length of the fragments is determined by the flow velocity of the solution and the diameter of the orifice, independent of the initial DNA size and concentration, and salt concentration (15). A commercial instrument and various shearing assemblies containing a single laser-drilled orifice on a ruby jewel (Hydroshear® by Genomic Solutions Inc., Ann Arbor, MI, USA) are available for performing the process semi-automatically. However, the instrument is expensive for many investigators. The shearing assemblies also tend to get clogged and are too expensive to be disposed of after every use.

We describe an inexpensive device for automated fragmentation of genomic DNA using hydrodynamic shearing. A schematic of the device is illustrated in Figure 2.1. The system consists of a computer-controlled syringe pump with a built-in 9-port valve and a filter screen housed in an adaptor connected to the valve via rigid tubing. The key component of the device is a disposable stainless steel filter screen which contains thousands of uniform pores with micron or sub-micron dimensions, in contrast to the single-orifice devices used in the shearing assemblies of the Hydroshear instrument. Similar to the Hydroshear instrument, hydrodynamic shear stress is utilized to break long DNA molecules into smaller fragments.

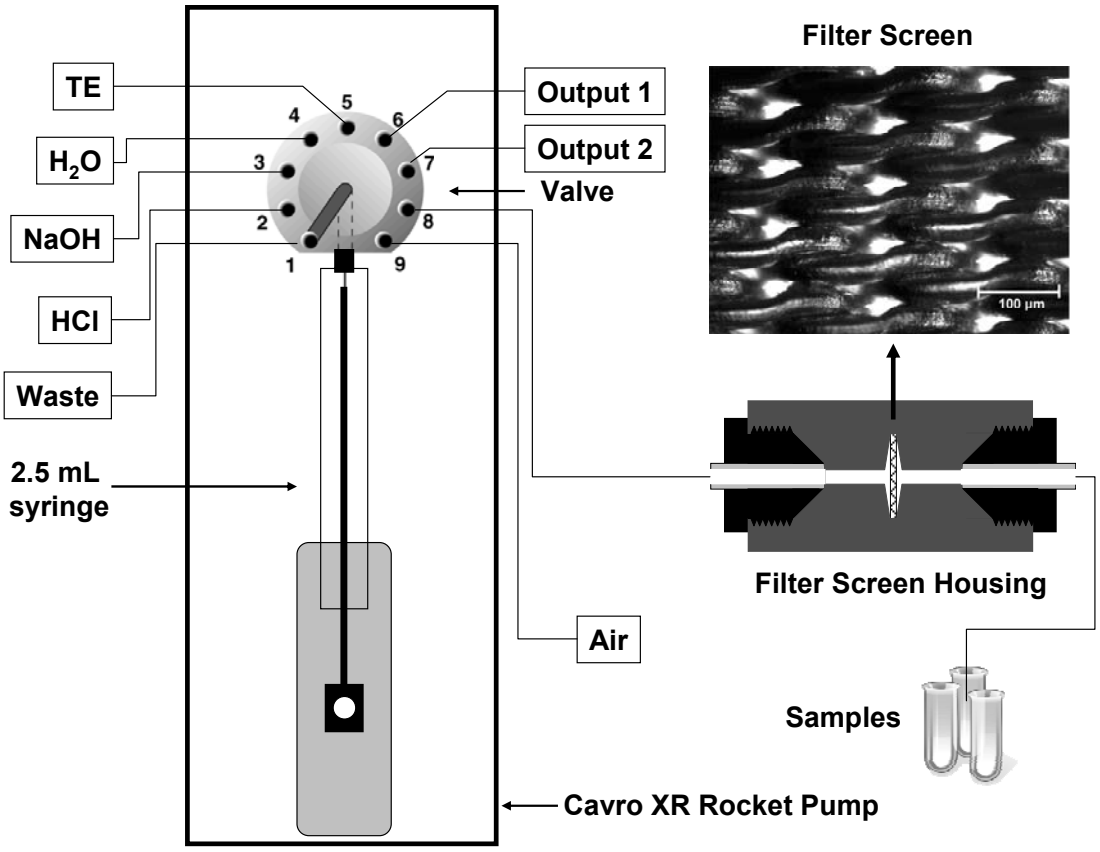

Figure 2.1. Schematic of a device for hydrodynamic shearing of DNA. A Cavro XR pump with a built-in 9-port valve is used. A cross-section of the filter screen housing is also illustrated. Shown at the upper right corner is a transmitted-light micrograph of a portion of a 10 µm screen. The computer used to control the instrument is not shown here.

2.3 Materials and Methods

2.3.1 Materials

Our shearing device consists of a syringe pump with a built-in 9-port valve (Cavro XR syringe pump, part No. 729848, $2073, Tecan Systems, Inc., Männedorf, Switzerland) and a syringe (Cavro XLP syringe, 2.5 mL, part No. 734806, $60, Tecan) with a ¼-28 fitting to the bottom port of the valve. The syringe pump is controlled by a computer via an RS232 interface using the PumpLink program provided by Tecan.

The syringe pump plunger speed can be varied from 1.2 to 1200 seconds per stroke, corresponding to 125 mL/min to 0.125 mL/min with a 2.5 mL syringe. All flow rates tested between 15 mL/min and 125 mL/min produced sheared fragments of predictable length. Shearing is accomplished with a removable stainless steel filter screen of 1/8" diameter and of various pore sizes (0.5 μm, part No. 5SR2-10, $2 each; or 10 μm, part No. 10SR2-10, $1 each; from VICI Valco Instruments Co. Inc., Houston, TX, USA) housed in a standard filter adaptor (Part No. ZUFR1C, $40, VICI Valco). The thickness of each screen varies depending on the diameter of the type 316 stainless steel fibers used to construct the screen. The thickness of the 0.5 μm screen is 40 μm while that of the 10 μm screen is 125 μm. All tubing used in the system is high purity PFA Teflon tubing with 1/16" outer diameter and 0.030" inner diameter (Cat. No. 1513, $1.80/foot, Upchurch Scientific, Oak Harbor, WA, USA). Tubing lengths between the sample and valve are generally less than 5 cm to minimize void volume. We recommend that users concerned about potential rust formation on stainless steel components consider an all-PEEK filter housing (Cheminert® Part No. ZU1FPK.5, $44, VICI Valco) and replaceable PEEK-encapsulated titanium filter elements (Cheminert® Part No. C-F1.5TI, $4.50 each, VICI Valco). The cost of our instrument is about $2200 (computer not included):

Part	Cost
Cavro XR Syringe Pump	$2073.00
Cavro XLP Syringe	$60.00
Filter Adaptor	$40.00
PFA Teflon Tubing	$1.80
0.5 μm filter	$2.00
10 μm filter	$1.00
Total	**$2177.80**

The source DNA was human genomic DNA (Cat. No. G304A, Promega Corp., Madison, WI, USA) diluted to 30 μg/mL in a TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8). Greater than 90% of the DNA is longer than 50 kb in size. All solutions were prepared with 18.2 MΩ-cm H_2O (Milli-Q, Millipore Corp., Billerica, MA, USA) and filtered through a 0.22 μm cellulose nitrate filter (Part No. 430758, $1.50 each, Corning Inc., Corning, NY USA) to prevent potential clogging by impurities.

2.3.2 Hydrodynamic Shearing of Genomic DNA with Screens

DNA shearing with our device is a relatively simple 2-step process that can be carried out automatically in 20 minutes:

(1) Washing. After the chosen screen is inserted into the filter screen housing, the assembly was washed extensively, 6 times with 1.5 mL of 0.5 M HCl followed by 6 times with 1.5 mL of 0.5 M NaOH. 500 μL of air was drawn into the syringe to purge any wash solution remaining in the valve that may damage the DNA. To rinse the

wash solution from the syringe and tubing, the system is washed 8 times with 2 mL of dH_2O and then with TE buffer. About 50 μL TE buffer is left in the tubing so that no air will be present in the system which could cause splashing of the sample onto the tube wall. The entire wash sequence is automated.

(2) DNA Shearing. The DNA was sheared by pulling the sample into the syringe and then forced back into its original tube through the screen for a determined number of iterations at the desired speed. After the shearing, air is pulled into the syringe and used to purge the entire fluid sample into the collection tube. To produce DNA of various fragment lengths from the same genomic DNA, portions of the sample were sheared at incrementally higher speeds and output into tubes connected to the different ports of the valve. The plunger depth was adjusted to account for the decrease in sample volume, and the remainder of the sample was then sheared at a higher speed. After each shearing experiment, the system was washed with the procedure described above.

2.3.3 Characterization of fragment lengths and distributions

To quantify the effect of screen pore sizes and flow rates on fragment lengths and distributions, we sheared human genomic DNA using screens of two different pore sizes (0.5 μm and 10 μm) at six flow rates (15, 25, 35, 45, 79, and 125 mL/min). The initial sample used was 165 μL of 30 μg/mL human genomic DNA in TE buffer. Shearing was performed as described above. A 15 μL aliquot was removed after shearing at each of the six speeds. To determine the potential variation in performance

with different batches of screens, three different screens of both 0.5 and 10 μm pore sizes were used to perform the shearing on three different days.

2.3.4 Elimination of Contamination

To show that the system can be washed sufficiently to eliminate any residual DNA in the tubing and the syringe, we sheared a sample that had been spiked with a 1600 bp known DNA fragment. Forty microliters of 500 nM DNA was added to 10 μL of 163 ng/μL genomic DNA and 70 μL TE buffer. After shearing at 79 mL/min through a 0.5 μm screen, the sample was collected. The system was washed with NaOH and HCl solutions according the procedure described above. Fifteen microliters of the final TE buffer rinse was saved and used as template in a PCR reaction with primers specific for the amplification of the spiked DNA. As a positive control, 0.2 μL of the sample was also used as a template for a PCR reaction. To see if the screen could be a source of contamination, the used screen was discarded and a new one placed into the housing. The system was rinsed with TE buffer again, and 15 μL of this final wash solution was used in the PCR reaction as well. Each PCR was carried out for 30 cycles using Phusion™ High-Fidelity Polymerase (New England Biolabs, Cat. No. F-530L) according to the manufacturer's instructions.

2.3.5 Analysis

The average size and range of the DNA fragments were analyzed by electrophoresis using 0.8% agarose gels. 250 ng of 1 kb Plus DNA ladder (Cat. No. 10787018, Invitrogen, Carlsbad, CA, USA) was also loaded onto the gel for

quantification. Approximately 225 ng of sheared DNA was loaded in each lane and electrophoresis was carried out at 5V/cm for 60 minutes in TAE buffer (40 mM Tris, 20 mM acetic acid and 2 mM EDTA, adjusted to pH 8.0 with NaOH). The gels were stained with SYBR Gold nucleic acid stain (Cat. No. S-11494, Invitrogen) in TAE buffer, and then imaged with a Gel Doc XR and a 12-bit camera system using the Quantity One 1-D Analysis Software (Bio-Rad Laboratories, Hercules, CA, USA). After background subtraction, the raw intensity values at all the points in each lane were exported to Microsoft Excel. To characterize each screen, data from 3 separate shearing runs with the same screen were averaged. An exponential equation relating fragment length and migration distance on the gel was established from the bands of the 1kb Plus standard. That equation was used to assign a base pair value to each data point. By dividing the raw intensity values with their corresponding base pair number, the relative population of each fragment length was calculated. The average fragment length was calculated by dividing the sum of the raw intensity values by the sum of the relative population values. The relative population data was normalized by dividing each point by the maximum value in the lane and then plotted as a function of fragment length using Kaleidagraph (Synergy Software, Reading, PA, USA).

The range of fragment sizes for each screen (0.5 μm and 10 μm) at each speed (15, 25, 35, 45, 79, and 125 mL/min) was defined as the range of base pairs that encompasses 80% of the total fragments, with approximately 10% of fragments falling above and below this range. This was determined by calculating the area underneath the curve of the relative population vs. length to find the lower and higher cut-offs.

The percent range in Table 2.1 is the coefficient of variation from three shearing runs. The coefficient is calculated by dividing the standard deviation by the mean value.

2.4 Results and Discussion

Shearing of genomic DNA with our device entails a washing step and a shearing step that can be carried out automatically in about 20 minutes. First, the desired screen is inserted into the adaptor and the fluidic system is washed extensively with HCl, NaOH, water, and a buffer. Second, the DNA is loaded and sheared by running the solution through the screen 20 times. A computer-controlled syringe pump with a 9-port valve is used to fully automate the entire shearing process. We have shown that human genomic DNA can be sheared using screens with pore sizes of 0.5 μm, 1 μm, 2 μm and 10 μm at fluid flow rates of 15 mL/min to 125 mL/min. Table 2.1 lists the average lengths and distributions of the fragments and the coefficients of variation from three independent runs using three different 0.5 μm and 10 μm screens at six different flow rates. Typical gel images of the sheared products are shown in Figures 2.2A and 2.2B. Shearing with screens of pore sizes 1 μm and 2 μm gave results similar to those obtained with the 0.5 μm screen.

Table 2.1. Flow rates and fragment lengths – Reproducibility between different batches of screens. The average size and range of the DNA fragments produced by shearing with 0.5 μm and 10 μm screens at 15, 25, 35, 45, 79, and 125 mL/min flow rate are listed. For each pore size and flow rate, the averages and coefficients of variation were calculated from data obtained from three runs, each time with a different screen. 80% of the DNA fragments have lengths between the low and high value while 10% of DNA fragments have lengths below and 10% above.

Flow Rate (mL/min.)	Size and Distribution (Kbp)					
	0.5 μm Screen			10 μm Screen		
	Low	High	Mean	Low	High	Mean
15	4.7 ± 17%	10.4 ± 10%	6.2 ± 9%	9.8 ± 7%	17.6 ± 8%	11.7 ± 9%
25	2.9 ± 11%	7.4 ± 4%	4.2 ± 11%	7.0 ± 16%	15.8 ± 5%	9.2 ± 13%
35	2.0 ± 7%	5.6 ± 6%	3.1 ± 11%	5.5 ± 9%	14.1 ± 9%	7.3 ± 9%
45	1.7 ± 3%	4.2 ± 7%	2.4 ± 6%	4.1 ± 4%	12.8 ± 4%	6.0 ± 2%
79	1.3 ± 10%	3.7 ± 2%	2.0 ± 4%	3.7 ± 6%	11.0 ± 13%	5.2 ± 3%
125	1.2 ± 15%	3.3 ± 8%	1.9 ± 7%	3.4 ± 1%	10.7 ± 10%	4.8 ± 4%

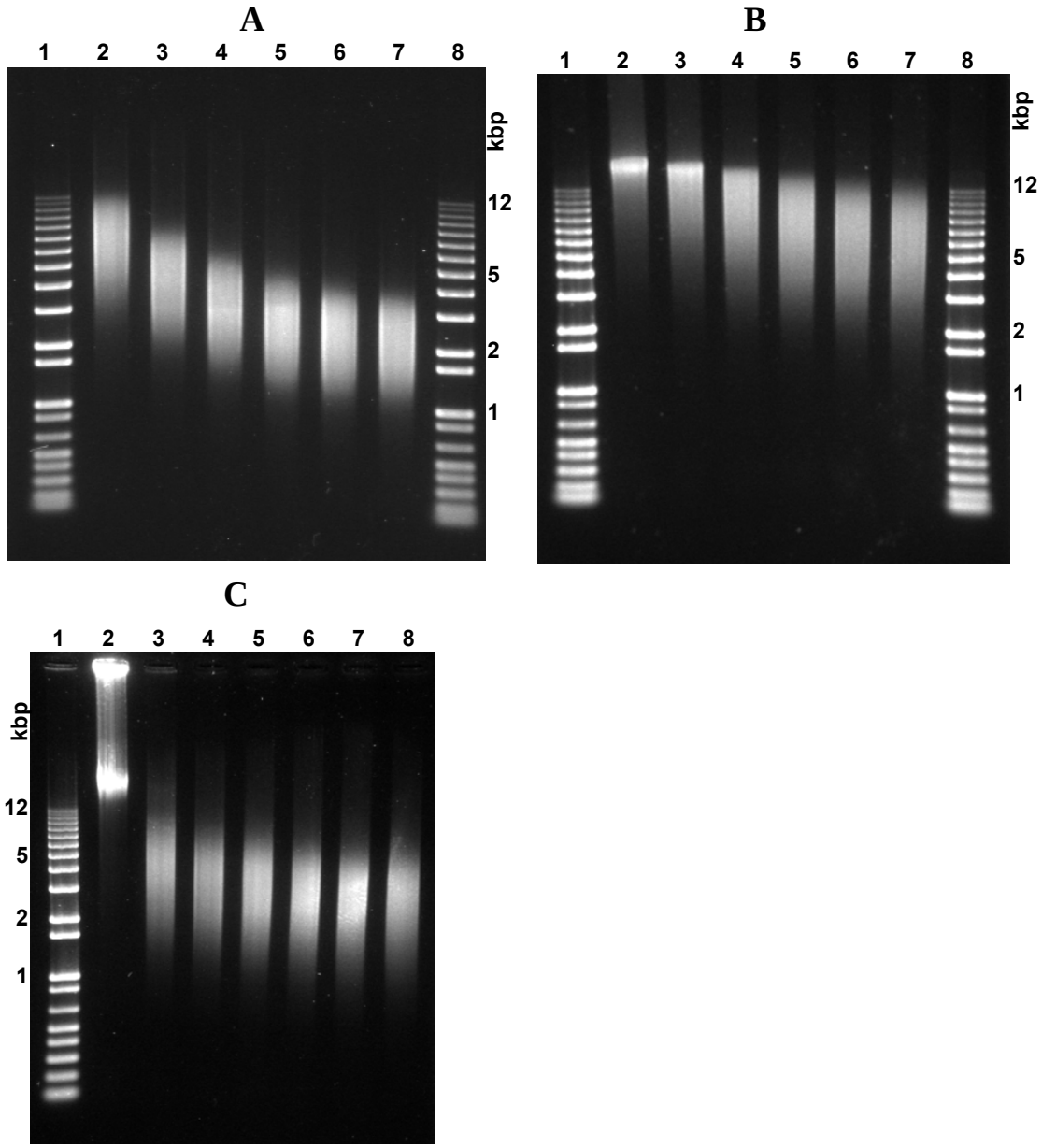

Figure 2.2. Lengths and size distributions of the DNA fragments as a function of screen pore size, flow rate and number of iterations. (A) and **(B)** Gel images of human genomic DNA sheared with 0.5 µm **(A)** and 10 µm **(B)** screens at various flow rates. Lanes 1 and 8: 250 ng of 1 kb Plus ladder. Lane 2-7: DNA sheared with a flow rate of 15, 25, 35, 45, 79, and 125 mL/min respectively. **(C)** Effect of the number iterations**.** Human genomic DNA was sheared by increased number of iterations using a screen with pores of 0.5 µm diameter. Lane 1: 250 ng of 1 kb Plus ladder. Lane 2: 250 ng of genomic DNA. Lanes 3-7: DNA sheared with1, 3, 5, 10, 15 and 20 iterations respectively.

The average length of the fragments is controlled primarily by the fluid flow rate, the size of the pores in the screen, and the iterations of sample passing through the screen. As shown in Figure 2.2C, the majority of DNA molecules were sheared to the target size in just one iteration. Almost all DNA fragments approached the minimal length after 5-10 iterations. No further change in length and distribution was observed after 20 iterations. The more important parameters are the flow rate and screen pore size. The fluid flow rates given in the text and the figures are the speeds at which the syringe pump delivers the fluid, not the actual flow rates of the DNA solution through the pores. The screen is constructed of woven stainless steel fibers. The pores are of irregular geometry but are very uniform in shape and size across the entire screen (Figure 2.3). Due to the complex geometry of the pores, the flow rate of the fluid through each pore increases, reaches the maximum and then decreases as the fluid enters, reaches the smallest cross section of, and leaves the pore. We have estimated that the pores account for approximately 2% and 10% of the cross section of the 0.5 μm and 10 μm screens respectively, so the actual flow rate for the 0.5 μm screen and 10 μm screens is 50 and 10 times of that delivered by the syringe pump. The size of the pores is vast (≥500 nm) compared to the diameter of the DNA molecule (2 nm), so the effect of pore geometry on the shear stresses on the DNA molecules is expected to be minimal. With the exceptional uniformity in size, geometry, and distribution across each screen and between different batches of screens, a narrow range of fragment sizes with over 80% of fragments within a 2- to 3-

fold range can be achieved reproducibly. In all cases a higher flow rate results in smaller DNA fragments and a smaller pore size produces shorter fragments at all fluid flow velocities. The coefficient of variation is usually small, less than 10% between three runs with the same screen or three different batches of screens.

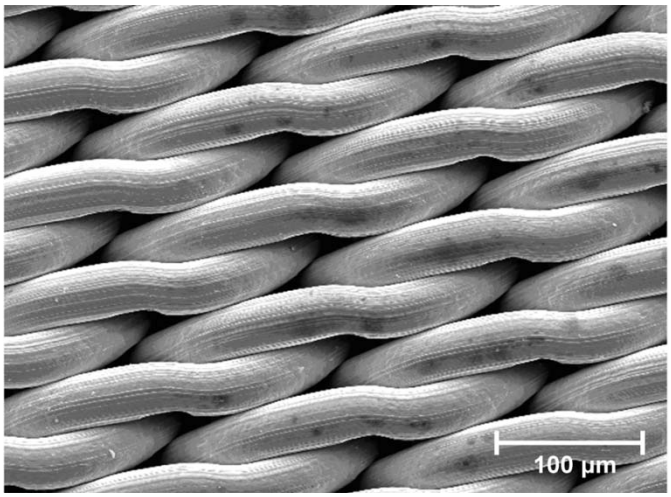

Figure 2.3. SEM image of a 10 µm screen.

Figures 2.4A and 2.4B show the relationships between the flow rates and lengths of the DNA fragments produced by shearing with 0.5 µm and 10 µm screens. Fragment sizes from 2 to 12 kbp can be achieved using screens of these two pore sizes by varying the flow rates. The minimum average size that can be produced with our current device is about 2000 bp using a 0.5 µm screen (the smallest pore size currently available from VICI Valco) with a flow rate of 125 mL/min (maximum flow rate attainable with our current system). As shown in Figure 2.4C, the non-linear relationship between flow rate and fragment size follows a power law. The increase in

flow rate from 40 to 125 mL/min results in only a small decrease in fragment size. To shear DNA to fragments shorter than 2 thousand base pairs will require screens of smaller pore sizes and a syringe system that can deliver higher flow rates. Very little heat is generated in our shearing process. However, as a precaution against any potential GC bias, the shearing assembly can be cooled on ice as usually practiced in DNA shearing process (18).

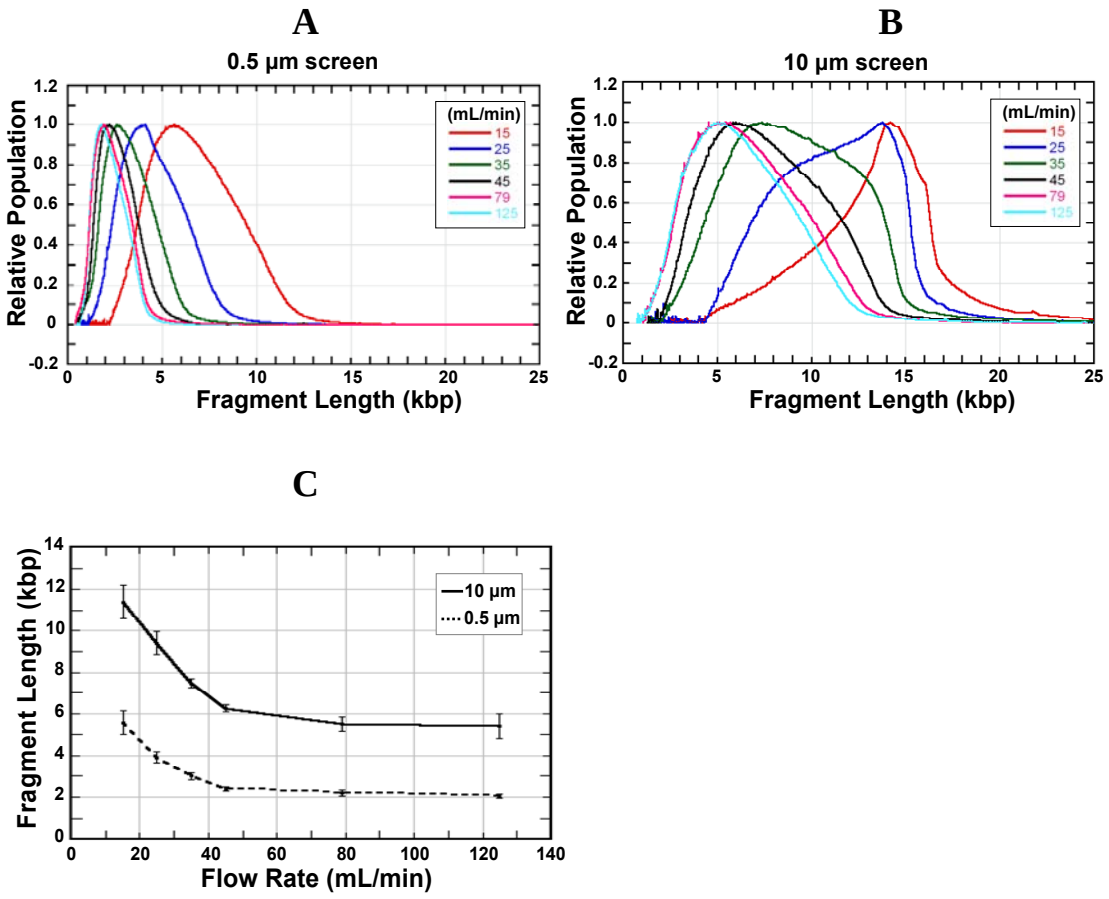

Figure 2.4. Lengths and size distributions of the DNA fragments as a function of screen pore size and flow rate. (A) and **(B)** Distribution of DNA fragments produced by shearing with screens of 0.5 µm **(A)** and 10 µm **(B)** pore sizes at various flow rates, plotted as relative population of the DNA fragments as a function of fragment size. The slight distortion of the curves at larger fragment sizes seen in **(B)** could be due to artifacts in the gel analysis of large DNA fragments. **(C)** Fragment lengths vs. fluid flow rates. Human genomic DNA was sheared at six different speeds using screens with pore sizes of 0.5 µm and 10 µm. The products were separated by electrophoresis with 0.8% agarose gel and the average fragment size was computed as described in the Methods in the Supplementary Materials. Data from three shearing runs with the same screen were averaged to obtain the data. Each error bar is one standard deviation from the mean and is intended to demonstrate reproducibility across different batches of screens.

A major concern for hydrodynamic shearing instruments that employ single-orifice devices is the frequent clogging of the small orifice. Our attempts to design a custom shearing instrument were prompted by frustration following many attempts to shear DNA with the HydroShear-Custom Shearing Assembly - small (650b-5Kb). Despite thorough filtering of all reagents, clogging of the orifice frequently terminated our experiments prematurely. With the use of a screen, however, we have never experienced any clogging. The screens can be used repeatedly without any change in performance. This makes our system more robust and user-friendly. No noticeable evidence of rust was observed on the stainless steel screens even after several weeks of usage. If rust is a concern, filters made of PEEK-encapsulated titanium (available from VICI Valco) can be used. Another concern is potential sample carryover contamination. The shearing assemblies used in the Hydroshear instrument are too expensive to be disposed of after every use. We found that with our standard wash procedure re-using the screens increases the potential of sample carryover which can be detected by PCR amplification (Figure 2.5). Fortunately, the screens used in our device are very inexpensive (2 US dollars at time of writing) and can be replaced easily. Cross-contamination can be eliminated entirely by replacing the screens between samples.

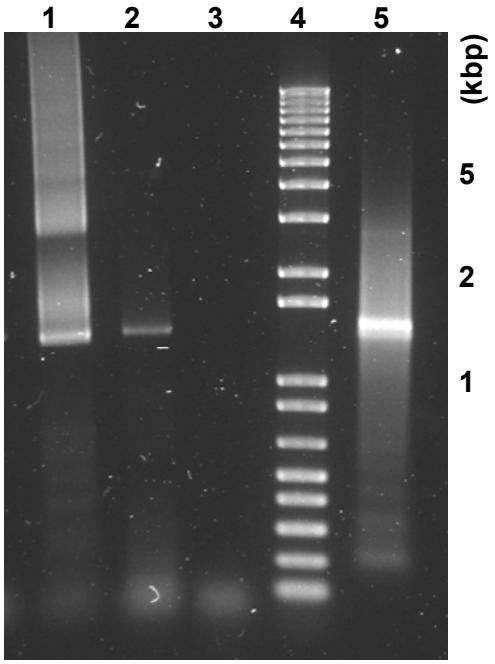

Figure 2.5. Elimination of cross-contamination. After washing the syringe and tubing and replacing the screen, no contamination is detected by PCR. Lane 1: PCR reaction with sheared DNA as template. Lane 2: PCR reaction with TE from tubing after washes. Lane 3: PCR reaction with TE from tubing after the washes and the replacement of the screen. Lane 4: 250 ng of 1 kb Plus Ladder. Lane 5: Sheared DNA spiked with 1600-bp fragment.

It has been shown that DNA fragments obtained by hydrodynamic shearing are random and have high cloning efficiency, suitable for genomic library construction (14,15). Since our device utilizes the same hydrodynamic shearing mechanism, it is reasonable to expect that the fragments produced using our device are random and have similar high cloning efficiency.

2.5 Conclusion

In summary, we have discovered that stainless steel filter screens with very uniform pores can be used in place of single-orifice devices to shear genomic DNA.

We have developed an inexpensive device and the procedure for fully automated hydrodynamic shearing of genomic DNA with performance comparable to that of the Hydroshear instrument (14,15). The utilization of the inexpensive disposable screens eliminates potential sample cross-contamination and the tendency of device clogging. Our device can be assembled easily from very inexpensive commercially available parts. We believe our significantly improved device and method for hydrodynamic shearing of DNA will be of great utility in producing genomic DNA libraries for genome sequencing using next-generation sequencing platforms.

2.6 Acknowledgements

We thank Kristopher D. Barbee for taking the SEM images of the stainless steel filter screens. This work was supported in part by the National Institute of Health (HG003587 and HG004130) and the National Science Foundation (BES-0547193, a CAREER Award to X.H).

2.7 References

1. Anderson, S. (1981) Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic acids research*, **9**, 3015-3027.

2. Roe, B.A. (2004) Shotgun library construction for DNA sequencing. *Methods Mol. Biol.*, **255**, 171-187.

3. Seed, B., Parker, R.C. and Davidson, N. (1982) Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene*, **19**, 201-209.

4. Gingrich, J.C., Boehrer, D.M. and Basu, S.B. (1996) Partial CviJI digestion as an alternative approach to generate cosmid sublibraries for large-scale sequencing projects. *BioTechniques*, **21**, 99-104.

5. Hoheisel, J.D., Nizetic, D. and Lehrach, H. (1989) Control of partial digestion combining the enzymes dam methylase and MboI. *Nucleic acids research*, **17**, 9571-9582.

6. Wong, K.K., Markillie, L.M. and Saffer, J.D. (1997) A novel method for producing partial restriction digestion of DNA fragments by PCR with 5-methyl-CTP. *Nucleic acids research*, **25**, 4169-4171.

7. Osoegawa, K., Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J. and de Jong, P.J. (2001) A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.*, **11**, 483-496.

8. Roe, B.A. (2004), *Bacterial Artificial Chromosomes*. Humana Press, Vol. 255.

9. Deininger, P.L. (1983) Approaches to rapid DNA sequence analysis. *Analytical biochemistry*, **135**, 247-263.

10. Deininger, P.L. (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal Biochem*, **129**, 216-223.

11. Okpodu, C.M., Robertson, D., Boss, W.F., Togasaki, R.K. and Surzycki, S.J. (1994) Rapid isolation of nuclei from carrot suspension culture cells using a BioNebulizer. *BioTechniques*, **16**, 154-159.

12. Hengen, P.N. (1997) Shearing DNA for genomic library construction. *Trends Biochem. Sci.*, **22**, 273-274.

13. Burger, G., Lavrov, D.V., Forget, L. and Lang, B.F. (2007) Sequencing complete mitochondrial and plastid genomes. *Nat. Protoc.*, **2**, 603-614.

14. Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J. and Davis, R.W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic acids research*, **24**, 3879-3886.

15. Thorstenson, Y.R., Hunicke-Smith, S.P., Oefner, P.J. and Davis, R.W. (1998) An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Res.*, **8**, 848-855.

16. Bertazzoni, U. (1975) Analysis of the breaking sites in the physical degradation of DNA. *Biochim. Biophys. Acta.*, **395**, 239-245.

17. McKee, J.R., Christman, C.L., O'Brien, W.D., Jr. and Wang, S.Y. (1977) Effects of ultrasound on nucleic acid bases. *Biochemistry*, **16**, 4651-4654.

18. Roe, B.A., Crabtree, J.S. and Khan, A.S. (1996) *DNA isolation and sequencing*. John Wiley & Sons, New York, New York, U.S.A.

This chapter, in part, is a reproduction of the material as it appears in: Joneja, A. and Huang, X. (2009) A device for automated hydrodynamic shearing of genomic DNA. *Biotechniques*, **46**, 553-556. Used with permission. The dissertation author was the primary investigator and author of this paper.

Chapter 3: Linear Nicking Endonuclease-Mediated Strand Displacement DNA Amplification

3.1 Abstract

We describe a method for linear isothermal DNA amplification using nicking endonuclease-mediated strand displacement by a DNA polymerase. The nicking of one strand of a DNA target by the endonuclease produces a primer for the polymerase to initiate synthesis. As the polymerization proceeds, the downstream strand is displaced into a single-stranded form while the nicking site is also regenerated. The combined continuous repetitive action of nicking by the endonuclease and strand displacement synthesis by the polymerase results in linear amplification of one strand of the DNA molecule. We demonstrate that DNA templates up to five thousand nucleotides can be linearly amplified using a nicking endonuclease with seven base-pair recognition sequence and Sequenase version 2.0 in the presence of single-stranded DNA binding proteins. We also show that a mixture of three templates of 500, 1000, and 5000 nucleotides in length are linearly amplified with the original molar ratios of the templates preserved. Moreover, we demonstrate that a complex library of hydrodynamically sheared genomic DNA from bacteriophage lambda can be amplified linearly.

3.2 Introduction

Many recent advances in biomedical research and applications such as genome sequencing and genetic diagnosis can be attributed, to a large extent, to the invention of many ingenious methods for DNA amplification, such as the cloning of plasmid DNA in bacteria (1) and the polymerase chain reaction (PCR) (2,3). In addition to the revolutionary PCR and the ligase chain reaction methods (4) which are based on thermal cycling, a variety of other methods have been developed for isothermal amplification (5). Notable ones include strand displacement amplification (SDA) with linear templates (6) or rolling circle amplification (RCA) with circular templates (7,8), transcription-mediated amplification (TMA) (5,9,10), multiple-displacement amplification (MDA) (11,12), helicase-dependent amplification (HDA) (13,14) and primase-based amplification (pWGA) (15). SDA, RCA, TMA and HDA can be used for both linear and exponential amplification with sequence-specific primers. MDA and pWGA can only be used for exponential amplification. Short degenerate oligonucleotide (6-8 nt) primers are required for MDA while no primers are needed for pWGA. The powerful MDA has been the method of choice for whole genome amplification from limited amount of genomic DNA (16) and has been applied to the amplification of genomic DNA from single cells for genome sequencing (17-20). Unfortunately, a high degree of amplification by MDA could produce significant amplification bias and artifactual chimeras, which are very likely the results of the stochastic random priming events and the formation of primer-dimers due to the use of the short degenerate primers (17,20,21).

The nicking-mediated SDA method initially described by Walker *et al*. utilizes a polymerase working in concert with a restriction enzyme and a set of specific primers to amplify target DNA molecules isothermally (6,22,23). However, there are substantial drawbacks. First, the method relies on the incorporation of an α-phosphorothioate into one strand of the partially palindromic recognition site of a double-stranded cutting restriction enzyme to prevent the hydrolysis action of the enzyme on the strand containing the α-phosphorothioate. The incorporation of the α-phosphorothioate essentially transforms the cutting site introduced by the primers into a nicking site for the restriction enzyme. Only a very limited number of restriction enzymes (usually HincII or BsoBI) can be used for SDA. Second, a DNA product containing a significant fraction of bases with α-phosphorothioate may not be desirable for certain downstream applications since the non-native nucleotide may interfere with further manipulations of the DNA such as digestion by nucleases (24-27). Third, it has not been demonstrated that DNA molecules with length greater than 100-200 nucleotides (nt) can be amplified with the method (22,28). This may have been due to use of DNA polymerases that do not possess both high processivity and strand-displacement capability in the earlier SDA experiments. In addition, the restriction enzymes used have recognition sequences equivalent to only five base pairs. Therefore, it is expected that in general the method cannot be employed to amplify targets over a thousand nucleotides since there exists an average of one cutting site per 1024 nt assuming the template sequence is random. While SDA with endonucleases and phosphorothioate nucleotides has proven to be very useful for

signal amplification and sequence detection (29,30), so far its utility for amplifying longer DNA sequences has yet to be demonstrated.

To circumvent the use of α-phosphorothioates and double-stranded cutting restriction enzymes in SDA, two groups have reported attempts at utilizing nicking enzymes for SDA (31,32). In a method called nicking-endonuclease mediated DNA amplification (NEMDA) reported by Chan *et al.*, an engineered nicking nuclease with only three-base recognition sequence was used in combination with a DNA polymerase for the amplification of genomic DNA (31). Through extensive optimization, Ehses *et al.* also demonstrated some success in the amplification of a 93-nt fragment by SDA with engineered nicking endonuclease Nt.BstNBI, which has a five-base recognition sequence (32). However, so far there has not been any report of success in the amplification of DNA molecules greater than 200 bases by SDA (28,32). Other methods such as NESA (33) and EXPAR (34,35) that employ polymerases and nicking enzymes are limited to even shorter targets and rely on the spontaneous dissociation of DNA strands following nicking rather than the strand displacement activity of the polymerase. Techniques that use cycles of nicking and polymerization to stimulate the aggregation of nanoparticles or light emission have also been developed (36,37). Much like SDA with phosphorothioate nucleotides, these methods are excellent for signal amplification but cannot be used for amplifying long DNA strands.

Other similar approaches have been developed for the isothermal linear amplification of RNA and DNA as well. Small quantities of RNA can be amplified

with little bias by T7 transcriptional amplification (38-40), and adaptations of this method for DNA amplification (41) have demonstrated replication of genomic DNA. Recently, a method called circular nicking endonuclease-dependent amplification (cNDA) demonstrated the ability to combine a nicking enzyme with the T4 replisome to amplify plasmid DNA (42). However, these methods require the concerted action of several enzymes. There is still a potential for bias introduced by the T7 primer sequence (43) and RNA intermediates or final products are susceptible to degradation (44), while cNDA requires a circular template. Another group has exploited the nicking activity of the DNA mismatch-repair enzyme endonuclease V to enable linear SDA of target molecules (45), but did not demonstrate the ability to amplify long DNA molecules.

Very long DNA molecules in a complex mixture such as a whole genome can be amplified by RCA (8,46) and MDA presumably due to the use of ϕ29 DNA polymerase which has extremely high processivity and strong strand-displacement capability (16,47). However, the construction of circular DNA templates for RCA could be cumbersome or not practical and large DNA molecules may not be amplified efficiently by RCA, while MDA may still produce significant amplification bias and chimeras due to the use of short degenerate primers (17,20,21). Methods for unbiased linear or exponential amplification of long DNA molecules in a complex mixture are useful for many applications. SDA appears to be an ideal method because of its unique mechanism of amplification.

In this study, we investigated the use of DNA polymerases with high processivity and strong strand displacement capability in combination with nicking endonucleases with long recognition sequences for linear amplification of long DNA molecules by SDA. A number of DNA polymerases and nicking endonucleases were examined. The DNA polymerases include Bst (large fragment), ϕ29, and Sequenase 2.0, all of which seem to possess the desired characteristics for SDA. To enable the specific amplification of long DNA targets, nicking endonucleases with long recognition sequences are essential. Fortunately, several nicking endonucleases have been recently engineered and are commercially available (31,48-54). The engineered nicking enzymes Nt.BspQI and Nt.BbvCI both have seven bp recognition sequences (52-54), which are 16X more specific than an enzyme with a five bp recognition sequence. On average, they would nick only once every 16,000 base pairs (4^7 = 16,384) in a DNA molecule with random sequence, ensuring that for most templates amplification only occurs at nicking sites introduced by the primers. We report the use of these commercially available nicking endonucleases for linear amplification of DNA molecules by SDA. The basic principle of linear strand displacement amplification (LSDA) is illustrated in Figure 3.1. We have demonstrated for the first time that a mixture of DNA molecules from 500 to 5000 nucleotides can be amplified in a linear fashion independent of the lengths and sequences of the DNA molecules. We have also demonstrated that a complex library of bacteriophage lambda genomic DNA can be amplified linearly with the original distribution of the fragments largely preserved.

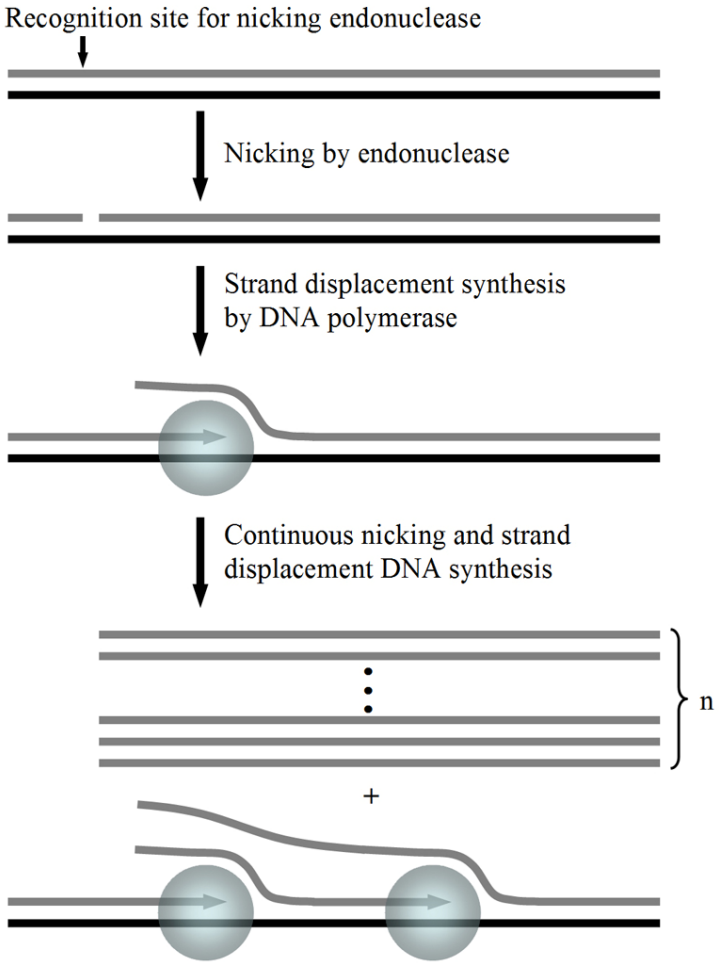

Figure 3.1. Linear DNA amplification by nicking endonuclease-mediated strand displacement DNA synthesis. After a nicking endonuclease cleaves a phosphodiester bond in the recognition sequence in one strand of the double-stranded DNA, a DNA polymerase binds to the nicking site and extends from the 3' OH group, displacing the downstream strand. The extension by the DNA polymerase from the nick regenerates the double-stranded recognition site for the nicking enzyme. The continuous combined actions of the nicking endonuclease and DNA polymerase result in the linear amplification of one strand of the DNA. The recognition site for a nicking enzyme is either an endogenous site on the target DNA, or a site added to the end of the target DNA by the ligation of an oligonucleotide duplex containing the recognition sequence or by PCR using a primer containing the recognition sequence.

3.3 Materials and Methods

3.3.1 Oligonucleotides, enzymes and other reagents

All oligonucleotides were purchased from Integrated DNA Technologies. All nicking enzymes and several polymerases, including Nt.AlwI, Nt.BbvCI, Nb.BsmI, Nt.BspQI, Nb.BsrDI and Nt.BstNBI, Exo- Klenow enzyme (*E. coli*. DNA polymerase large fragment), Bst DNA polymerase large fragment, ϕ29 DNA polymerase, 9°Nm DNA polymerase, Vent exo- DNA polymerase, and BSA (bovine serum albumin) were acquired from New England Biolabs (NEB). T4 DNA ligase was also obtained from NEB. Sequenase version 2.0, an engineered T7 DNA polymerase (55), and *E. coli*. single-stranded DNA binding protein (SSB) were purchased from United States Biochemicals. PCR was performed using a Phusion PCR kit from Finnzymes. Nucleotides were from Sigma Aldrich. Plasmid pET-24(a) was acquired from Novagen, now part of Merck Biosciences. The PCR Purification kit (Cat. No. K3100-01) and nucleic acid dyes SYBR Gold, SYBR Green I, and SYBR Green II were obtained from Invitrogen. The PerfeCTa SYBR Green FastMix from Quanta Biosciences was used for the quantitative PCR (qPCR) experiments. Sequencing was provided by Eton Bioscience Inc.

3.3.2 DNA templates

The 0.5, 1, and 5-kilonucleotide (knt) DNA templates were amplified from plasmid pET-24(a) by PCR. The same forward primer was used for the amplification of all the templates. In addition to 21 bases that anneal to the plasmid DNA, this

primer contained at its 5’ end the recognition sequences for both Nt.BspQI (GCTCTTCN^) and Nt.BbvCI (CC^TCAGC), and an additional 23 nucleotides of random sequence to increase the melting temperature of the upstream fragment following nicking activity. Each fragment produced from a template created with this forward primer will be 38 nucleotides shorter than the initial template. The forward primer had this sequence: 5’-CTG GAG TCA ACG CAT CGA GCA TAC CTC AGC GCT CTT CCG CTT CCT CGC TCA-3’. The reverse primers were 20-21 base long oligonucleotides designed to bind at the appropriate locations on the plasmid to provide the desired template lengths. The following primers were used: 5’-CGG GTT GGA CTC AAG ACG ATA-3’ for the 500 nt template, 5’-GAC ATT ATC GCG AGC CCA TT-3’ for the 1 knt template, and 5’-CTG TTC ATC CGC GTC CAG CTC-3’ for the 5 knt template. The GC content of the fragments varied from 52% for the 1 and 5 knt templates to 58% for the 0.5 knt template. All of the templates were amplified by 33 cycles of PCR. Each cycle consisted of 10 s at 98 ºC, 15 s at 78 ºC, and 15 s/knt at 75 ºC. The amplification was finished by a final reaction at 72 ºC for five min and cooling to 4 ºC. The amplified products were purified with a PCR purification kit and quantified by absorption measurement at 260 nm using a spectrophotometer (Nanodrop Model ND-1000 and software version 3.5.2, Thermo Scientific). Templates to test other nicking endonucleases were obtained similarly.

3.3.3 Genomic DNA library construction

A library of randomly fragmented bacteriophage lambda genomic DNA was prepared and used as a model system to demonstrate the ability to amplify a complex library linearly by SDA. The lambda genomic DNA was fragmented by hydrodynamic shearing using a custom-built device (56). Briefly, linear lambda genomic DNA (48.5 kbp, NEB) at a concentration of 30 ng/μL was passed through a filter screen (1-μm pore size and 1/16" in diameter, Product No. 1SR1-10, Valco Instruments Co. Inc.) housed in an internal stainless steel union with 0.25 mm diameter bore (Product No. ZU1C, Valco Instruments Co. Inc.). The DNA solution was pumped through the screen 20 times at a flow rate of 50 mL/min. The resulting DNA fragments were blunt-ended and 5' phosphorylated using the NEBNext™ End Repair Module (Cat. No. E6050S, NEB). Sheared DNA (85 μL) was mixed with 10 μL 10X End-Repair Buffer and 5 μL Enzyme Mix (T4 polymerase and T4 Polynucleotide Kinase) and incubated at 20 °C for one hour. The enzymes were removed by membrane purification using the PCR Clean-up kit. An 8 μL aliquot of the resulting library was run on a 1% agarose gel to determine the mean size of the fragments, and the molar concentration of the fragments was calculated. Next, both ends of the genomic DNA fragments were ligated to a duplex adapter containing the recognition site for nicking endonucleases Nt.BbvCI and Nt.BspQI. To ensure that the adapter ligation is uni-directional and to avoid self-ligation between the adapters, the duplex adapter is designed to have one blunt end without a 5' phosphate group and one end with a 1-base 5' overhang. The duplex adapter consists of a 51-base oligonucleotide with

sequence 5'-CTG GAG TCA ACG CAT CGA GCA TAC CTC AGC GCT CTT CCG CTT CCT CGC TCA-3', and a 50-base complement from the 3' end. The nicking sites for Nt.BbvCI and Nt.BspQI are located at 25 and 38 nucleotides respectively from the 5' end of the top strand. An 80-fold molar excess of adapters over genomic DNA was used to prevent the ligation between genomic DNA fragments. The ligation reaction contained 2 μM of adapters and 25 nM of end-repaired genomic DNA in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 20 mM KCl, 1 mM ATP, and 10 mM dithiothreitol (DTT). T4 ligase was added to a final concentration of 15 cohesive end units/μL, and ligation was carried out at 26 °C for 12 hours. The reaction was stopped by heating at 65 °C for 20 minutes. Since the adapters do not have a 5' phosphate, only one strand of the duplex is ligated to the genomic DNA, leaving a nick on the other strand. The nicks were removed by performing a short strand displacement reaction using the nicks as the priming sites. Nucleotides, BSA and Bst DNA polymerase large fragment were added to a final concentration of 150 μM, 100 μg/mL, and 35 units/mL respectively and the mix was heated to 37 °C for 30 minutes. The polymerase and excess adapters were removed by two-time purification using a PCR Clean-Up kit. The final concentration of the DNA library was determined by absorption measurement using the Nanodrop spectrophotometer. The mean fragment size and the molar concentration of the resultant DNA fragments with the adapters were obtained by gel electrophoresis analysis of a small aliquot of the final library.

3.3.4 Strand displacement amplification reactions

For SDA reactions with only one DNA template, each 25 μL reaction volume was assembled on ice and contained 2 nM DNA template, 40 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 5 mM DTT, 100 μg/mL BSA, 500 μM each of all four dNTPs, 5 μM SSB, 400 nM (0.45 U/μL) Sequenase 2.0, and 3 nM (0.04 U/μL) Nt.BspQI (the molar concentration of Nt.BspQI was calculated using a specific activity of 370,000 U/mg and a molecular weight of 50 kDa, personal communication from New England Biolabs). The solution was divided into 5 μL aliquots in five microtubes. Immediately, 15 μL of PAGE gel stop buffer (80% formamide, 20 mM EDTA pH 8.0, 10 mM Tris-borate, 0.0125% bromophenol blue and 0.0125% xylene cyanol FF) was added to one tube for the zero time point reaction. The remaining tubes were incubated at 37 ºC. After 5 min, 10 min, 20 min, and 40 min incubation, one tube was removed and 15 μL of stop buffer was added to terminate the reaction.

For simultaneous amplification of a mixture of three templates at equal molar concentrations, the reaction conditions were the same except that 2 nM of each of the 0.5, 1 and 5 knt templates was used. For simultaneous amplification of a mixture of three templates at equal mass or nucleotide concentrations, all conditions were also the same except that the concentrations of the 0.5, 1, and 5 knt templates were 3 nM, 1.5 nM, and 0.3 nM, respectively. When expressed in terms of mass or nucleotide, all three templates have the same concentration, 1.5 μM in nucleotide. To investigate the important contribution of SSB to the amplification reaction, two SDA reactions were performed in parallel on 1 nM 0.5 knt template using the conditions described above

without SSB or with 5 μM of SSB. Time points were taken at 0 min, 5 min, 10 min, and 20 min.

For the amplification of the sheared genomic DNA library from bacteriophage lambda, a 15 μL reaction was assembled on ice. The reaction mix contained 2 nM DNA templates, 40 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 5 mM DTT, 100 μg/mL BSA, 500 μM each of all four dNTPs, 5 μM SSB, 400 nM (0.45 U/μL) Sequenase 2.0, and 3 nM (0.04 U/μL) Nt.BspQI. The solution was divided into 5 μL aliquots in three microtubes. For the zero reaction time point, an agarose gel stop buffer (1.2% SDS, 7% Ficoll-400, 25 mM Tris-Cl pH 7.5, 25 mM EDTA, 0.0025% bromophenol blue, and 0.0025% xylene cyanol FF at 1X) was added immediately to one tube. The remaining tubes were incubated at 37 °C. After 15 min and 30 min, one tube was removed and the stop buffer was added to terminate the reaction.

To observe the effect of nicking enzyme concentration on the SDA reaction, four different molar concentrations of Nt.BspQI were used in parallel SDA reactions on 2 nM 0.5 knt template. The concentrations of the enzyme tested were 0.6 nM, 3 nM, 15 nM, and 75 nM. In addition to a zero min time point, samples were taken from each reaction at 10 min and 20 min and added to PAGE gel stop buffer.

3.3.5 Quantification by gel analysis

The rates and linearity of strand displacement amplification of the DNA templates prepared by PCR were quantified by electrophoretic analysis with 5% denaturing polyacrylamide gel in 0.5X TBE buffer (45 mM Tris-borate and 1 mM

EDTA, pH 8.4). For the 0.5 knt reaction, the entire volume of each time point was loaded onto the gel. For the other templates, half of the sample for each time point was loaded. Additionally, 250 ng of 1 kb DNA ladder (NEB) was run for size comparison and four standard lanes were loaded on each gel for mass quantification. The standards contained known amounts of the template being analyzed and represented 1X, 4X, 16X, and 32X the mass of initial template in the reaction. Each gel was run for 35 minutes at 400V constant voltage with a Bio-Rad Mini PROTEAN 3 gel setup. The gels were briefly soaked in deionized water before staining with SYBR Gold in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 2 mM EDTA, pH 8.0). The gels were imaged with a Gel-Doc XR and 8-bit camera system using Quantity One 1-D analysis software (Bio-Rad Laboratories). After background subtraction, the lanes of known mass were used to create a standard curve so that the amount of DNA in each sample band could be calculated. Following quantification of each sample band, the mass of the product at each time point was plotted vs. time using Microsoft Excel 2007. A linear equation was fitted to the data and the R-squared value for each time series was calculated.

Alkaline agarose gel electrophoresis was used to analyze the strand displacement amplification of the bacteriophage lambda genomic DNA library. Prior to loading onto the gel, the samples were heated to 95 °C for two minutes and then cooled on ice. After NaOH was added to a final concentration of 30 mM, the entire volume of sample for each time point was loaded into a well. Samples from three separate SDA reactions using the same library were loaded onto one gel. For size

comparison, one lane was loaded with DNA fragments from the genomic library at 15 times of the initial amount used in the amplification. Along with the samples to be analyzed, 400 ng O'Gene Ruler^{TM} Express DNA Ladder (Fermentas) was loaded onto each gel. The alkaline gel (0.8% agarose, 30 mM NaOH, 2 mM EDTA) was run on ice at 3.5 V/cm for 135 min in 30 mM NaOH and 2 mM EDTA. The gel was briefly soaked in deionized water, neutralized by soaking in 2X SSC (300 mM NaCl, 30 mM sodium citrate pH 7.0) for 20 minutes, stained with SYBR Gold in 1X SSC for 15 minutes, and destained in TAE for 30 minutes. The gel was imaged as described above. An exponential equation relating fragment length and migration distance on the gel was established from the bands of the DNA Ladder standard. After background subtraction, the raw intensity values of each pixel along the length of each lane were exported to Microsoft Excel, and the average pixel value for each migration distance was calculated from the three samples at each time point. The migration distance was converted to base pair values using the exponential equation, and we plotted the average raw intensity vs. fragment length for each of the time points. By dividing the raw intensity values with their corresponding base pair number, the relative population of each fragment length was calculated. The average fragment length was found by dividing the sum of the raw intensity values by the sum of the relative population values. Using three sets of samples on the gel, we calculated the average mean fragment length for each time point and the standard deviation. The amplification of PCR-generated fragments with hairpin-forming adapters was also visualized by alkaline gel electrophoresis using the same conditions.

3.3.6 Real-time monitoring of strand displacement amplification

For the real-time detection experiments, the reaction mix contained the DNA template (1 nM and 3 nM for the 0.5 knt fragment, 0.33 nM and 1 nM for the 5 knt fragment), 100 nM reverse primer (same primer used to generate the fragment by PCR), 40 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 5 mM DTT, 100 μg/mL BSA, 500 μM each of all four dNTPs, 5 μM SSB, and 0.5X SYBR Green II. The reaction mix was incubated without enzymes at room temperature for 30 minutes. After placing the mix on ice, Nt.BspQI and Sequenase 2.0 were added to a final concentration of 3 nM and 400 nM, respectively. Each template concentration was loaded in triplicate into adjacent wells of a Real-Time PCR Detection System (MiniOpticon, Bio-Rad Laboratories). The samples were pre-incubated for 2 minutes at 37 °C, and a reading was then taken every 40 seconds for 60 minutes with the temperature maintained at 37 °C. Data was collected using the Bio-Rad CFX Manager Software Version 1.5.

At the conclusion of the run, all data was exported to Microsoft Excel 2007. Baseline subtraction was done by subtracting the relative fluorescent unit (RFU) value of the first read of each well. The resulting data for each template concentration were averaged and the standard deviation calculated. The mean values were then plotted against time with vertical error bars indicating one standard deviation from the mean. A linear equation was fitted to each set of data and the R-squared value for each time series was calculated.

3.3.7 Analysis of amplification bias by qPCR

Three primer pairs were designed to amplify three sequences of lengths 129, 152, and 152 nt from the SDA product. The sequences are located at approximately 5 knt, 30 knt, and 40 knt from one end of the 48.5 knt linear bacteriophage lambda genome. The primer pairs used are 5'-ACA CTG CAG TCC CGG ATG GA-3' and 5'-ATC AAT GGC CTC CTG ACC GC-3', 5'-CGC GTC ACC CAC ATG CTG TA-3' and 5'-TGC TCT CCC GAT GGT TTA TGC A-3', and 5'-TAC CGC TCA CCG TAT TGC AGG TTG-3' and 5'-GCC GAC GTA TGG AGT GCC ATA TTT-3', respectively. The sheared genomic DNA library from was amplified for 30 min by SDA as described above and the reaction was stopped by heating at 80° C for 20 min. qPCR reactions were performed to quantify the three sequences in the SDA product. Each qPCR reaction mix contained 1X qPCR mix, 125 nM of a primer pair, and a small volume of product from the SDA reaction. The PCR reaction consisted of one initial denaturation/activation cycle of five min incubation at 95° C followed by 40 cycles of 10 s denaturation at 95° C and 20 s hybridization/synthesis at 68° C.

3.3.8 Hairpin adapter to produce double-stranded products

The 3'-end of one molecule of the amplified single-stranded products could partially hybridize to certain regions of the same molecule or other molecules, resulting in chimeric products. To prevent this, we designed an adapter oligonucleotide with self-complementary regions to form a hairpin at the end of the single-stranded SDA product. Once a product is fully synthesized and released from

the polymerase, the hairpin forms, allowing the 3' end to serve as a primer for the synthesis of a double-stranded product. The adapter contains the sequence used in the lambda phage genome library followed by five random nucleotides and 12 additional bases added at the 3' end that are the reverse complement of the 12 bases immediately following the Nt.BspQI nicking site. One strand of the duplex adapter has a 68-base sequence of 5'-CTG GAG TCA ACG CAT CGA GCA TAC CTC AGC GCT CTT CCG CTT CCT CGC TCA ATC CAG AGC GAG GAA GC-3' while the complement strand has a two base overhang (AA) at the 3' end. The oligonucleotides are not phosphorylated at the 5' end. We used a 1.5 knt fragment without any Nt.BspQI nicking sites amplified by PCR from the *Saccharomyces cerevisiae* genome as a test template. The forward and reverse PCR primers have the sequences 5'-phosphate-TGC TTT GCC AAG GGT ACC AAT GTT T-3', and 5'-phosphate-GCA ATT ATG GAC GAC AAC CTG GTT G-3', respectively. The PCR was performed by denaturation of the genomic DNA at 98 °C for two minutes followed by 30 cycles of PCR, each consisting of 10 s at 98 ºC, 20 s at 63 ºC, and 25 sec at 68 ºC. The amplification was completed by a final reaction at 72 ºC for five min and cooling to 4 ºC. The amplified products were purified with a PCR purification kit and quantified by absorption measurement at 260 nm using a Nanodrop spectrophotometer. The duplex adapters were ligated to the PCR product, and SDA reactions were performed by following the same procedures used for genomic DNA library construction and amplification.

3.4 Results

We present an approach for isothermal linear amplification of DNA. In order to demonstrate the linear kinetics of this method and to illustrate its ability to amplify long DNA molecules, DNA templates of various lengths were obtained via PCR using primers containing restriction sites and plasmid pET-24(a) as source material. Fragments of length 0.5, 1, and 5 knt were then amplified using Sequenase 2.0 and Nt.BspQI. The results are shown in Figure 3.2. First, each fragment was amplified separately and time points were taken at 5 min, 10 min, 20 min, and 40 min. The products were quantified and plotted, and a linear regression line was fitted to each set of data. For each template length, the SDA product is 38 bases shorter than the original molecule since the nicking endonuclease creates a single-stranded break in the DNA 38 base pairs from the 5' end of the template. This size difference is seen most clearly in the amplification of the 0.5 knt template, but the polyacrylamide gel is unable to resolve the two bands of higher molecular weight templates. We designed the templates to have a long 38-base "primer" upstream of the nicking site so that the nicked primer remains hybridized even at 70 °C for testing amplification at elevated temperatures. At 37 °C, the temperature at which SDA appeared to be most successful, a shorter primer (e.g. 12-14 bases long) would be sufficient. To verify the identity of the amplified fragments, we sequenced the SDA products of the 500 nt and 1 knt fragments by the conventional Sanger dideoxy method. The results confirmed that each product has the same sequence as the initial template.

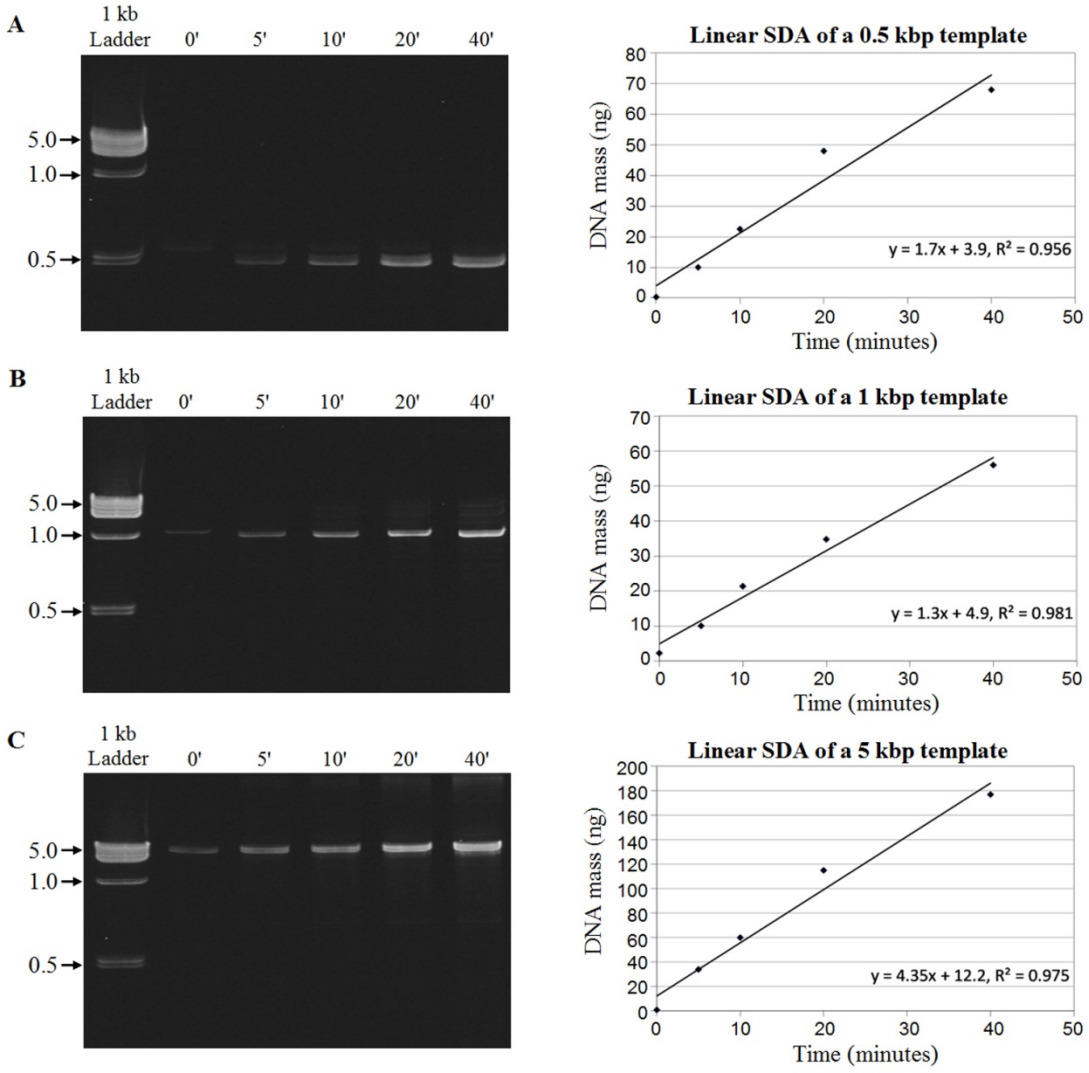

Figure 3.2. Linear amplification of 0.5 to 5 knt DNA templates by nicking endonuclease-mediated SDA. (A) – (C) Linear SDA amplification of 0.5, 1, and 5 knt templates respectively. A 1 kb ladder is included for size reference with the 0.5, 1 and 5 kb bands indicated by arrows. The SDA product for each template should be 38 nucleotides shorter than the original molecules. Graphs beside each gel are the calculated masses of each sample band plotted vs. time. In each case, a regression line was fit to the data, and the resulting equation and R-squared value are shown.

The amplification reaction could be modeled very well by a linear equation (R^2 > 0.95). Additionally, it appeared that each template was amplified about 15-20 times

during the 40 min incubation. Unlike in previous iterations of SDA, an increase in target length does not result in an observable decrease in the amplification factor. Furthermore, the doubling time of this reaction is approximately 2-2.5 minutes. This compares favorably to the original mesophilic SDA using HincII and Klenow exo-, which had a doubling time of 3-5 minutes (28,57). Due to the limited dynamic range of the gel imaging instrument and the difficulty in quantifying very low concentrations of DNA, the first data point in each series appears close to zero.

In order to show that our method can be used for linear amplification of a library of targets with fragments of disparate lengths, three target templates with lengths of 0.5, 1, and 5 knt were amplified simultaneously in the same solution. In the first experiment, as illustrated in Figure 3.3A, the three templates were initially present at equal molar concentrations. The fluorescent intensity of each stained product band is proportional to the mass or the number of nucleotides in the amplified product. Therefore, if the reaction is linear across the templates of various lengths, the product of each template should increase linearly and the slope of the equation fitted to the 5 knt template should be about 10 times that of the 0.5 knt template and 5 times that of the 1 knt template. As shown, the amplification of each target is linear ($R^2 > 0.99$) and the slope of the trend lines for both the 5 knt and 1 knt fragments are within 10-15% of the expected values.

In the second experiment, the three templates are initially present at equal mass concentrations. In this case, the amount (in mass or total number of nucleotides) of DNA in all three bands should increase linearly at the same rate. As shown in Figure

3.3B, this is what is observed. The amplification of each template is linear and the slopes of each trend line are within 15-20% of each other, indicating that all three templates are amplified equally in mass.

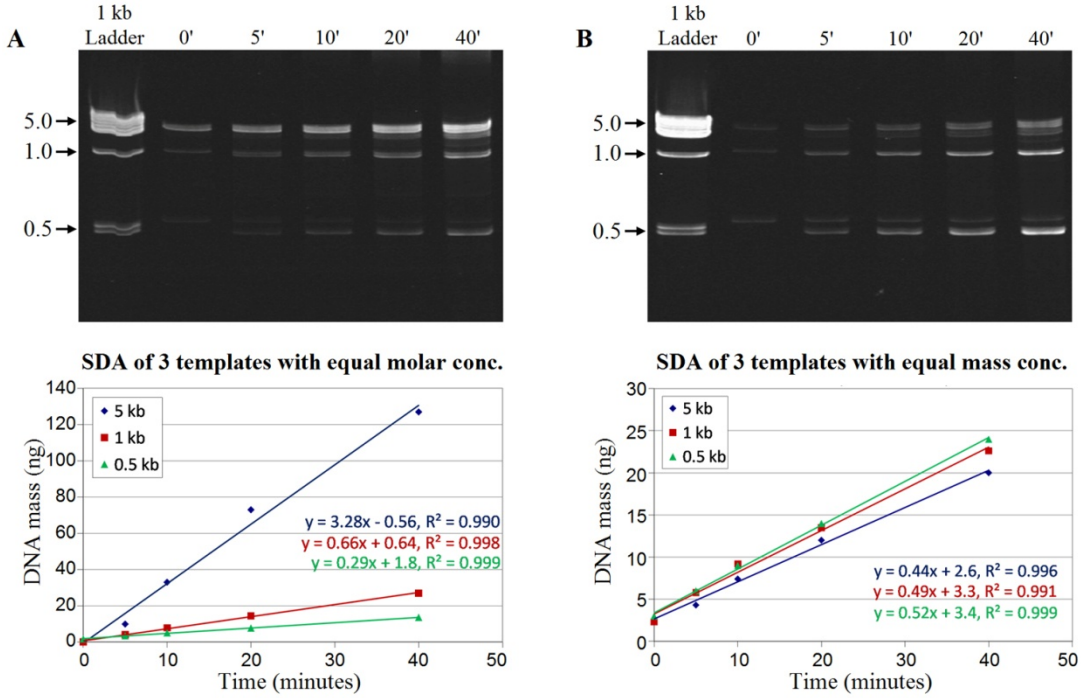

Figure 3.3. Simultaneous linear amplification of 3 templates of various lengths. (A) Amplification of three templates of lengths 0.5, 1, and 5 knt with equal initial molar concentration of 2 nM. A 1 kb ladder is included for size reference with the 0.5, 1 and 5 kb bands indicated by arrows. A plot of DNA mass vs. time for each template length is shown below each gel image. As anticipated, in mass, the 5 knt product increases 10 times faster than the 0.5 knt product and 5 times faster than the 1 knt product. **(B)** Amplification of three templates of lengths 0.5, 1, and 5 knt with equal initial concentration in mass or total nucleotides of 1.5 μM. As expected, the mass or the total number of nucleotides of all three templates increases at the same rate.

To prove the linearity of the SDA reaction, the 0.5 knt and 5 knt templates were amplified using a real-time PCR machine to monitor the increase in DNA mass in real time. Various concentrations of SYBR Gold, SYBR Green I, and SYBR Green

II were investigated to find the optimal conditions for the best fluorescent signal. We found that concentrations of SYBR Gold or SYBR Green I above 0.5X significantly inhibited the SDA reaction. Additionally, it had been previously reported that although SYBR Green II is commonly considered a single-strand DNA or RNA binding dye, it exhibits increased fluorescence when bound to double-stranded DNA as opposed to single-stranded DNA (34). We also found this to be the case and included the reverse primers in the reaction mix to make the SDA product double-stranded.

Plots of relative fluorescent units vs. time for the 500 nt and 5 knt templates are shown in Figure 3.4. A straight line is a good fit for each data set ($R^2 > 0.99$). In Figure 3.4A, it was anticipated that the slope of the trend line for the 3 nM sample wells would be 3-fold higher than the slope of the trend line for the 1 nM samples. Likewise, in Figure 3.4B, it would be expected that the slope of the 1 nM template would be three-fold higher than the slope of the 0.33 nM template samples. In each case, the magnitude of the slopes are within 10% and 20% of the expected values, respectively. It should be noted that the vertical axes in Figures 3.2 and 3.3 reflect the total amount of DNA in mass or nucleotide while the vertical axes in Figure 3.4 represent the increase in relative fluorescent units from the samples.

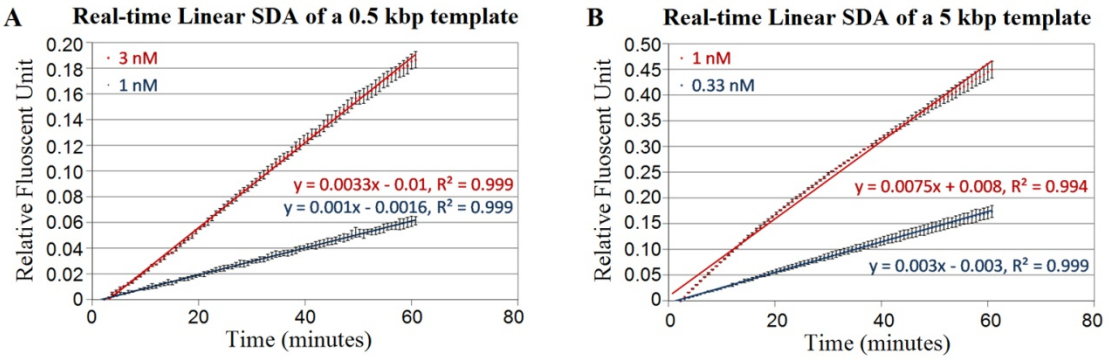

Figure 3.4. Real-time monitoring of SDA. Linear SDA of 500 nt **(A)** and 5 knt **(B)** templates. Relative fluorescent units (RFU) were measured every 40 seconds for 60 minutes. After background subtraction, a linear regression line was fitted to each data set. The error bars indicate the standard deviation of measurements from three wells. For both templates, a three-fold increase in template concentration, from 1 nM to 3 nM for the 0.5 knt template and from 0.33 nM to 1 nM for the 5 knt template, resulted in three times faster increase in the measured RFU.

The data curve for the 5 knt template at 3 nM concentration appears to be slightly less linear than those of the other samples. This moderate decrease in reaction rate is likely due to the depletion of reagents in the reaction, which causes the slope of the regression line to be lower than expected. It should be emphasized, however, that in a mixture of templates with different lengths, each template would experience the same reduction in available reagent concentrations. While the overall reaction rate may diminish somewhat for a long reaction, this would not result in longer templates being underrepresented in the final amplified DNA mix.

After demonstrating that SDA can be used to amplify a mixture of templates of disparate lengths in an unbiased manner, we investigated linear SDA of more complex templates. We used a genomic library constructed from the relatively small genome (48.5 kbp) of the bacteriophage lambda as a model system. A genomic library with a

mean fragment length of 1.15 knt was constructed. Figure 3.5 shows some typical results from the amplification of the library by SDA using the conditions we established for the PCR-generated templates. After 30 min of incubation with Nt.BspQI and Sequenase 2.0, the template has been replicated about 15 times. When the mean raw intensity from triplicate experiments is plotted as a function of fragment length (Figure 3.5B), it is observed that the overall curves of the SDA products shift toward a shorter length relative to the initial template library. The amplified product from each template is expected to be 76-base shorter than the original template since Nt.BspQI nicks at a site 38 bases from both ends of the adapters. However, the mean fragment length is 810 nt with a standard deviation of 30 nt from triplicate samples, about 200 nt shorter than the expected full-length product. This discrepancy can be explained by the presence of 10 "native" Nt.BspQI nicking sites within the phage genomic library. Amplification from an internal nicking site results in a product shorter than the full-length fragment. The intensity of the bands in the gel image in Figure 3.5A and the vertical axis in Figure 3.5B depict the raw intensity which is proportional to the total mass of the DNA, not the relative population of each fragment. Therefore, the centers of the bands on the gel and the peaks of the curves in Figure 3.5B do not correspond directly to the true mean fragment sizes reported above. In order to show that the library amplification was unbiased, we selected three short sequences located approximately in the middle and at both ends of the lambda phage genome. We used qPCR to determine the number of molecules in the initial

unamplified library and library amplified by SDA. The three fragments were found to be amplified by about 10-, 10-, and 12-fold respectively.

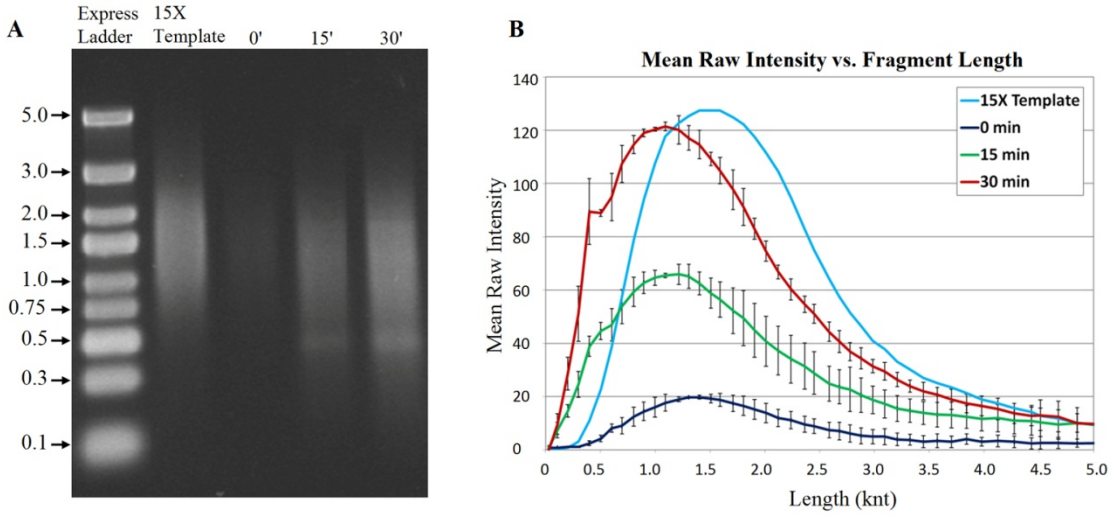

Figure 3.5. Amplification of a genomic library from bacteriophage lambda. (A) Analysis of the linear amplification of a sheared genome on a denaturing alkaline gel. An Express DNA Ladder is included for size reference. A lane containing 15 times the amount of starting material used for the amplification is included for comparison. **(B)** Average raw intensity of three reactions at each reaction time plotted as a function of DNA fragment length. Error bars indicate one standard deviation from the mean, calculated from a triplicate data set. The overall shift of the SDA product towards shorter lengths relative to the initial library is likely due to the presence of internal nicking sites within the genomic DNA fragments and the placement of the Nt.BspQI nicking site in the adapter.

The amplified products of a complex library by SDA are single-stranded. The 3'-end of one product molecule could partially hybridize to certain regions of itself or other molecules, resulting in chimeric products. The use of an adapter sequence that results in a product with a self-priming hairpin structure could alleviate this potential problem. The strategy is illustrated in Figure 3.6A. A DNA fragment amplified from the *Saccharomyces cerevisiae* genome was used as a test template. As shown in Figure

3.6B, the amplification of a 1.5 knt template results in a product of twice that length on a denaturing gel. This indicates that the product is converted to a double strand with a hairpin at one end with 100% efficiency.

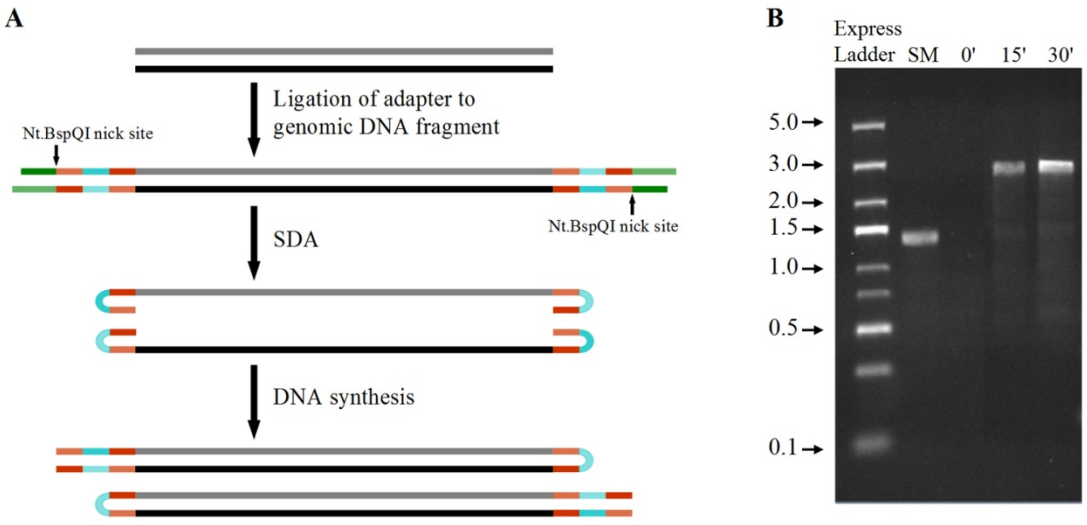

Figure 3.6. Use hairpin adapters to convert SDA product to double-stranded DNA. (A) Both ends of the genomic DNA fragment are ligated to an adapter oligonucleotide with two self-complementary regions to form a hairpin at the end of the single-stranded SDA product. Once a product is fully synthesized and released from the polymerase, the 3' end of each single-stranded SDA product folds into a hairpin and serves as a primer for the synthesis of the double-stranded product with a hairpin at one end. This helps to prevent the non-specific hybridization of the 3' end of the single-stranded molecules to regions of itself or other molecules. **(B)** Strand displacement amplification of a 1.4 knt template from lambda phage genome. With the 60-base hairpin adaptors, the product appears as a 3 knt fragment on a denaturing alkaline gel. An Express DNA Ladder is included for size reference. Lane "SM" contains 30 times the amount of starting material used in the amplification. The lane with the small amount of starting materials (zero min lane) is not visible on the gel.

3.5 Discussion

We have established the reaction conditions for strand-displacement amplification of DNA molecules up to 5 knt in length. These conditions differ from previously reported SDA techniques in several important aspects. Past publications

have advocated the use of very high concentrations of endonuclease and suggested that a linear relationship exists between endonuclease concentration and reaction rate (22,28). In our initial experiments, very high concentrations of nicking enzymes were used. Much to our surprise, our numerous attempts at SDA using templates of various lengths and concentrations, and many combinations of DNA polymerases (including Klenow exo-, ϕ29, Bst large fragment, 9°Nm and Vent exo-) and nicking enzymes (including Nt.AlwI, Nt.BstNBI, Nb.BsrDI, Nb.BsmI, Nt.BbvCI, and Nt.BspQI) were not successful. Eventually, we discovered that, in contrast to what was previously reported (28), high concentrations of nicking enzyme in the solution could severely inhibit the SDA reaction. Tan *et al.* also reported that high concentrations of nicking endonucleases inhibit DNA amplification by EXPAR (34).

We found that the SDA reaction is highly sensitive to nicking enzyme concentration. There exists an optimal concentration unique to each nicking endonuclease. Higher concentrations inhibit the reaction while lower concentrations may not be sufficient to sustain the reaction rate. It has been suggested that some nicking enzymes may bind tightly to the DNA templates at high concentrations (34) and certain nicking endonucleases such as Nt.BbvCI are known to have a tendency to aggregate into higher-order species that inhibit the activity of the enzymes (53,54) . However, the mechanism by which high nicking enzyme concentrations inhibit SDA is still not very clear. We found that it is not due to nonspecific nicking by the nuclease or stifling levels of glycerol or BSA contributed by the enzyme storage buffer. The phenomenon was observed for all combinations of nicking enzymes and

polymerases we have investigated. The effect is demonstrated in Figure 3.7A. A 500 nt template with a starting concentration of 2 nM was amplified by SDA with Sequenase 2.0 and four different concentrations of Nt.BspQI. An increase in the reaction rate is observed when the Nt.BspQI is increased from 0.6 nM to 3 nM. However, at 15 nM the reaction rate decreases considerably, and at 75 nM the reaction stops completely even though the DNA templates are still nicked by the endonuclease. Similar experiments were repeated to identify the optimal concentration of each nicking enzyme. The optimal concentration for the nicking enzymes we tested varies by nearly two orders of magnitude. We found that if the optimal concentration of the nicking endonuclease was used, SDA of templates up to 250 nt in length was successful with all the nicking endonucleases, including Nt.AlwI, Nb.BsrDI, Nb.BsmI, Nt.BbvCI, Nt.BstNBI, and Nt.BspQI. It had been previously reported that the ratio of polymerase to nicking enzyme was important (28). However, this does not seem to be the case for linear SDA under our conditions. We found that the optimal concentration of each nicking endonuclease remains the same regardless of polymerase concentrations. Moreover, the optimal concentration of nicking endonuclease was the same regardless of template concentration. Increasing the template concentration does increase the overall reaction rate. It is quite possible that template concentrations in the 1-15 nM range are below the K_M of the nicking endonuclease; therefore, any change in substrate concentration will have a large effect on the turnover rate of each enzyme.

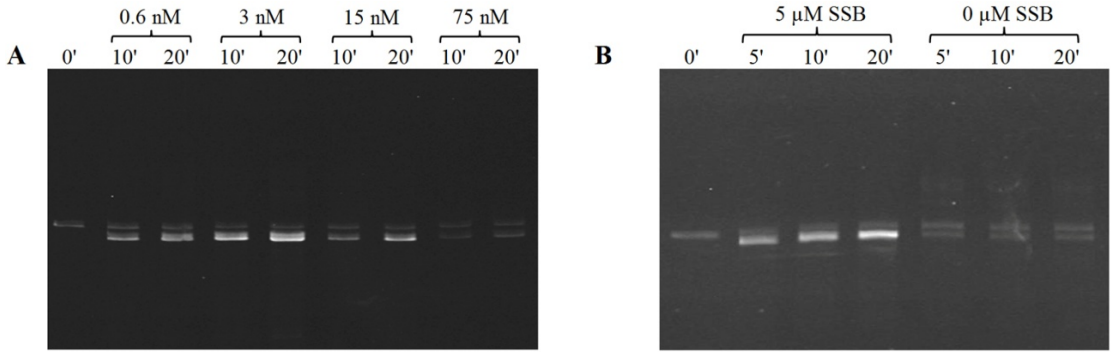

Figure 3.7. Critical parameters for successful strand displacement amplification. (A) There exists an optimal concentration for each nicking enzyme for SDA. Shown is a gel image of SDA of a 500 nt template using varying concentrations of Nt.BspQI. Maximal amplification of a 500 nt fragment at a template concentration of 2 nM is achieved with 3 nM of Nt.BspQI. This concentration of nuclease is also optimal for other templates of varying lengths and concentration. Amplification yield drops sharply at concentrations above this value. **(B)** The presence of a high concentration of *E. coli* single-strand DNA binding protein (SSB) is essential for efficient SDA by Sequenase 2.0 DNA polymerase.

Once the optimal concentrations for the nicking endonucleases were established, SDA of short templates was successful with all of the nicking enzymes and most of the polymerases. The surprise exception was ϕ29 DNA polymerase. With its high processivity, strong strand displacement activity and high fidelity, it would seem to be an ideal DNA polymerase for SDA. Unfortunately, ϕ29 DNA polymerase is unable to initiate strand displacement DNA synthesis from an endonuclease-mediated nick in a double-stranded DNA molecule. Similar to what was reported in the original SDA method, SDA with Klenow exo- resulted in specific amplification of short templates, but the amplification factor was highly dependent on target length. SDA with Bst DNA polymerase large fragment gave mixed results. While short templates (<100 nt) were amplified efficiently by Bst DNA polymerase with fast kinetics at high reaction temperatures (50-60 °C), for longer templates we observed

very high levels of nonspecific products and the yields of the specific products were much lower. Similar phenomena were observed for other thermophilic polymerases including Vent exo- and 9°Nm. In some cases, a brief period of specific amplification was followed by a rapid increase in nonspecific amplification. Thermophilic polymerases are known to be able to initiate template-independent *de novo* DNA synthesis and amplification (58,59), particularly in the presence of endonucleases (34,60,61). We chose Sequenase 2.0 as the best polymerase candidate for further investigation because it has a far superior synthesis rate and processivity (55) compared to Klenow exo- and does not exhibit the nonspecific background amplification observed in most thermophilic polymerases. Nt.BspQI was selected as the nicking endonuclease because it has a long seven bp recognition sequence and has a faster turnover rate than Nt.BbvCI. Table 3.1 lists the concentrations of the enzymes we have optimized for linear SDA.

Table 3.1. The concentrations of enzymes for optimal isothermal linear SDA. For enzymes whose specific activities were not available, the concentrations are given in units per microliter. The usage of higher concentrations of thermophilic polymerases usually leads to nonspecific amplification.

Optimal Concentrations of Enzymes for SDA				
Polymerase	Concentration	Nicking Enzyme	Restriction Site	Concentration
Klenow exo-	25 nM	Nt.AlwI	GGATCNNNN^	20 pM
Bst Large Fragment	5 nM	Nt.BstNBI	GAGTCNNNN^	10 nM
Sequenase 2.0	400 nM	Nt.BspQI	GCTCTTCN^	3 nM
Vent exo-	100 nM	Nt.BbvCI	CC^TCAGC	50 nM
9° Nm	5 nM	Nb.BsrDI	^CATTGC	0.08 U/μL
		Nb.BsmI	G^CATTC	0.25 U/μL

It is known that the addition of single-stranded DNA binding proteins can significantly facilitate DNA synthesis by many DNA polymerases (62-67). We found that single-stranded DNA binding proteins are essential for SDA with Sequenase 2.0. Even though both T4 gene 32 protein and *E. coli* SSB can be used, amplification with the latter and Sequenase 2.0 gives cleaner bands on polyacrylamide gels. In addition, the use of SSB enhances SDA amplification by Klenow exo- and significantly suppresses background amplification by Bst DNA polymerase. The binding of the displaced DNA strand by SSB facilitates the strand displacement synthesis by the polymerase and suppresses nonspecific amplification by preventing non-specific hybridization of the 3'-ends of the single-stranded products. The profound effect of SSB on SDA with Sequenase 2.0 is shown in Figure 3.7B. With 5 μM SSB, long DNA

molecules can be amplified by Sequenase 2.0. In the absence of SSB, however, no observable amplification occurs. It can be seen that the nicking enzyme did cut one strand of the target, but Sequenase 2.0 was unable to proceed with strand displacement synthesis in the absence of SSB. As shown in Figure 3.7B, the nicked strand appears as a band 38 bases below the original template.

If the nicking site is absent in the internal sequences of the target DNA molecules, all of the target molecules in a complex mixture can be amplified by nicking-endonuclease mediated strand displacement in a linear fashion regardless of the length or sequence of the molecules. In other words, the molecules are multiplied to the same fold factor with the original ratio among the molecules in the mix maintained. This is because each template molecule contains an identical adaptor sequence that is nicked at an equal rate by the endonuclease with the nicked site in turn serving as the primer for strand-displacement synthesis by the DNA polymerase. The amplification reaction rate could be limited either by the nicking rate of the endonuclease or by the rate of strand displacement synthesis from the nicks.

The rate-limiting step for SDA under our conditions appears to be the rate of nicking by the endonuclease. This is evidenced by the absence of visible unfinished or truncated products on the gel following the immediate termination of an SDA reaction with stop buffer containing formamide and EDTA. This implies that Sequenase 2.0 can initiate the SDA reaction immediately after each nicking event and can complete the SDA along the entire template in a short amount of time before another nick is made again in the template. A quick calculation supports this theory. The rate of

synthesis for Sequenase 2.0 is about 200 nucleotides per second, and the polymerase can incorporate an average of 800 nucleotides without dissociating from the template (55,68). The presence of at least 100-fold molar excess of polymerase to template and a high concentration of the polymerase (400 nM) ensure that if an enzyme does fall off the template, it is quickly replaced. Therefore, we can expect a 5 knt long fragment to be completely synthesized in less than 30 seconds. In contrast, the turnover rate for a mesophilic nicking enzyme appears to range from 30 seconds for the very active Nt.AlwI (48) to 4.5 minutes for Nt.BbvCI (53). Nt.BspQI is considered a thermophilic enzyme with an ideal operating temperature of 50 °C. The K_M and intrinsic turnover rate of the enzyme have not been reported. We determined the intrinsic turnover rate of substrate-saturated Nt.BspQI at 37 °C to be about 45 seconds under our reaction conditions (data not shown). Interestingly, the enzyme was three times more active in our SDA buffer than in the buffer recommended by New England Biolabs. At a template concentration below 3 nM, the amplification may be operating below the K_M of the nicking enzyme. Therefore, the nicking enzyme is not saturated with substrate and Nt.BspQI would have a lower effective turnover rate. Our observation that 20-fold amplification seems to occur in 40 minutes indicates that the effective turnover rate under our reaction conditions is about two minutes.

The low turnover rate of the nicking enzyme essentially acts as a method of local isothermal cycling. For each nicking event, each template is replicated once regardless of length or sequence. Complications could arise during SDA of longer DNA molecules if the template was nicked multiple times before the initial

polymerization of the strand was complete. This would lead to multiple polymerases performing SDA reactions simultaneously at different sites on the same template. While the amplification would still remain linear, the template would be susceptible to a traffic jam of polymerases if a region of unusual sequence or complexity were encountered by the first polymerase. Additionally, a fraction of truncated products would be present at the conclusion of the reaction. However, in our SDA experiments, very little unfinished product is observed (Figures 3.2 and 3.3). This is due to the rapid kinetics, high processivity, and strong strand displacement properties of the Sequenase 2.0 DNA polymerase and the slower turnover rate of the nicking enzyme Nt.BspQI used for the SDA reaction.

It is likely that the nicking rate can be increased somewhat without jeopardizing the linearity of the amplification. This would be desirable to obtain greater product yields. Although the optimal temperature for Nt.BspQI is 50 °C, the reaction rate cannot be increased by raising the reaction temperature because Sequenase 2.0 DNA polymerase is not stable above 37 °C. Faster nicking enzymes will be required to increase the speed of linear SDA.

The amplification of a complex library constructed from the lambda phage genome was successful and commensurate to the amplification of single templates. Approximately 10-15 fold amplification was achieved in 30 minutes with the original length distribution of the fragments remaining largely intact. The larger than expected increase in the number of shorter fragments is likely due to the presence of internal Nt.BspQI nicking sites that were not introduced by our adapter fragments but are

instead native to the lambda phage genome. The Nt.BspQI recognition site is overrepresented in the lambda phage genome with 10 cutting sites. If the distribution of the native nicking sites follows a Poisson distribution, about 18% of the 1.15 knt fragments contain one or more internal nicking sites. Amplification from the internal native cutting sites results in shorter products, introducing some bias toward certain fragments. This underscores the need for the discovery or engineering of novel nicking endonucleases with much greater specificity. For the amplification of longer DNA molecules and larger genomes, a nicking endonuclease with an even longer recognition sequence (e.g. 12-16 bases) would be ideal. Fortunately, there has been an increased effort in engineering such nicking endonucleases for both DNA amplification and genome engineering (69-71).

There is a relatively bright low molecular weight band at around 0.5 knt in the amplified product which is not present in the original library (Figure 3.5A). This band corresponds to the shoulder peak in the curves of Figure 3.5B. The source of this is not known. This could be due to the products generated from the internal nicking sites in the genomic DNA fragments, or some non-specific amplification. Non-specific amplification is commonly observed in SDA with Bst DNA polymerases even with a single template species. It is, however, rarely observed in SDA by Sequenase 2.0 in the presence of high concentration of SSB. The use of hairpin adapters may provide a means to improve the specificity of the amplification by converting the free 3'-ends of initial SDA product into a double-stranded form (Figure 6). An added benefit of using such a design is that the conversion of the single-stranded DNA into a double-strand

form also frees up the bound SSB which is essential for strand displacement replication by Sequenase 2.0.

Given that a 5 knt template can be amplified with ease by SDA using Nt.BspQI and Sequenase 2.0, we expected that longer templates could also be amplified efficiently. Surprisingly, our attempts to amplify a 10 knt fragment from the lambda phage genome were less successful. A large number of shorter DNA fragments are present in the final product. Further optimization of the reaction conditions or the use of a more processive enzyme will be required for the efficient amplification of longer templates.

3.6 Conclusions

In summary, we have developed a method for isothermal DNA amplification by nicking endonuclease-mediated DNA polymerase strand displacement. Our technique differs from the SDA method pioneered by Walker *et al.* in several important aspects. The most significant attribute of our method is the ability to amplify fragments up to 5 knt in length with very little bias. We have shown that 5 knt fragments can be amplified with an efficiency approximately equal to that of 0.5 knt fragments. This is feasible because our method employs a DNA polymerase, Sequenase 2.0, which has very high processivity and strong strand-displacement capability in the presence of single-stranded binding proteins, and a nicking endonuclease, Nt.BspQI, which has a long seven bp recognition sequence. The other attribute is our use of a nicking endonuclease, which obviates the need for α-

phosphorothioate nucleotides in the reaction so the product contains only native nucleotides. We have demonstrated that a mixture of DNA molecules with lengths of 0.5, 1, and 5 knt can be amplified in a linear fashion with the original molar ratio preserved. Using a library of randomly sheared genomic fragments from bacteriophage lambda, we also showed that a complex library was amplified linearly with the original distribution of the fragments largely maintained. Together, these significant improvements open the door to the possibility of using SDA not only for sequence detection but also for linear amplification of long DNA templates and heterogeneous mixtures of templates with little bias and without thermal cycling. Our method is potentially useful for certain applications such as genome sequencing or gene expression profiling from a limited sample source where linear unbiased amplification of a mixture of complex DNA molecules is highly desirable. It is quite possible that our method can be extended to linear amplification of templates longer than five thousand bases, and to exponential SDA amplification by including both the forward and reverse primers containing a recognition site for the nicking endonuclease in the reaction mix. Further work will be required to demonstrate these capabilities and to enable the unbiased linear amplification of large genomes by SDA.

3.7 Acknowledgements

The authors wish to thank Dr. Hsin-I Chiang for assistance with real-time detection of DNA amplification.

3.8 References

1. Cohen, S.N., Chang, A.C., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA*, **70**, 3240-3244.

2. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, **51 Pt 1**, 263-273.

3. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

4. Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl. Acad. Sci. USA*, **88**, 189-193.

5. Gill, P. and Ghaemi, A. (2008) Nucleic acid isothermal amplification technologies: a review. *Nucleosides Nucleotides Nucleic Acids*, **27**, 224-243.

6. Walker, G.T., Little, M.C., Nadeau, J.G. and Shank, D.D. (1992) Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. *Proc. Natl. Acad. Sci. USA*, **89**, 392-396.

7. Fire, A. and Xu, S.Q. (1995) Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA*, **92**, 4641-4645.

8. Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C. and Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.*, **19**, 225-232.

9. Compton, J. (1991) Nucleic acid sequence-based amplification. *Nature*, **350**, 91-92.

10. Kacian, D.L. and Fultz, T.J. (1998) USA patent no. 5,824,518.

11. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA*, **99**, 5261-5266.

12. Lage, J.M., Leamon, J.H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Segraves, R., Vossbrinck, B., Gonzalez, A., Pinkel, D. *et al.* (2003) Whole

genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.*, **13**, 294-307.

13. Vincent, M., Xu, Y. and Kong, H. (2004) Helicase-dependent isothermal DNA amplification. *EMBO Rep.*, **5**, 795-800.

14. Motre, A., Li, Y. and Kong, H. (2008) Enhancing helicase-dependent amplification by fusing the helicase with the DNA polymerase. *Gene*, **420**, 17-22.

15. Li, Y., Kim, H.J., Zheng, C., Chow, W.H., Lim, J., Keenan, B., Pan, X., Lemieux, B. and Kong, H. (2008) Primase-based whole genome amplification. *Nucleic acids research*, **36**, e79.

16. Lasken, R.S. (2009) Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem. Soc. Trans.*, **37**, 450-453.

17. Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W. and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.*, **24**, 680-686.

18. Lasken, R.S. (2007) Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr. Opin. Microbiol.*, **10**, 510-516.

19. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. and Lasken, R.S. (2008) Genomic sequencing of single microbial cells from environmental samples. *Curr. Opin. Microbiol.*, **11**, 198-204.

20. Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R. and Chisholm, S.W. (2009) Whole genome amplification and de novo assembly of single bacterial cells. *PLoS One*, **4**, e6864.

21. Lasken, R.S. and Stockwell, T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol.*, **7**, 19.

22. Spargo, C.A., Fraiser, M.S., Van Cleve, M., Wright, D.J., Nycz, C.M., Spears, P.A. and Walker, G.T. (1996) Detection of M. tuberculosis DNA using thermophilic strand displacement amplification. *Mol. Cell Probes*, **10**, 247-256.

23. Walter, N.G. and Strunk, G. (1994) Strand displacement amplification as an in vitro model for rolling-circle replication: deletion formation and evolution during serial transfer. *Proc. Natl. Acad. Sci. USA*, **91**, 7937-7941.

24. Connolly, B.A., Potter, B.V., Eckstein, F., Pingoud, A. and Grotjahn, L. (1984) Synthesis and characterization of an octanucleotide containing the EcoRI recognition sequence with a phosphorothioate group at the cleavage site. *Biochemistry*, **23**, 3443-3453.

25. Potter, B.V. and Eckstein, F. (1984) Cleavage of phosphorothioate-substituted DNA by restriction endonucleases. *The Journal of Biological Chemistry*, **259**, 14243-14248.

26. Putney, S.D., Benkovic, S.J. and Schimmel, P.R. (1981) A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. *Proc. Natl. Acad. Sci. USA*, **78**, 7350-7354.

27. Vosberg, H.P. and Eckstein, F. (1982) Effect of deoxynucleoside phosphorothioates incorporated in DNA on cleavage by restriction enzymes. *The Journal of Biological Chemistry*, **257**, 6595-6599.

28. Walker, G.T. (1993) Empirical aspects of strand displacement amplification. *PCR Methods Appl.*, **3**, 1-6.

29. Cosentino, L.A., Landers, D.V. and Hillier, S.L. (2003) Detection of Chlamydia trachomatis and Neisseria gonorrhoeae by strand displacement amplification and relevance of the amplification control for use with vaginal swab specimens. *J. Clin. Microbiol.*, **41**, 3592-3596.

30. Little, M.C., Andrews, J., Moore, R., Bustos, S., Jones, L., Embres, C., Durmowicz, G., Harris, J., Berger, D., Yanson, K. *et al.* (1999) Strand displacement amplification and homogeneous real-time detection incorporated in a second-generation DNA probe system, BDProbeTecET. *Clin. Chem.*, **45**, 777-784.

31. Chan, S.H., Zhu, Z., Van Etten, J.L. and Xu, S.Y. (2004) Cloning of CviPII nicking and modification system from chlorella virus NYs-1 and application of Nt.CviPII in random DNA amplification. *Nucleic acids research*, **32**, 6187-6199.

32. Ehses, S., Ackermann, J. and McCaskill, J.S. (2005) Optimization and design of oligonucleotide setup for strand displacement amplification. *Journal of Biochemical and Biophysical Methods*, **63**, 170-186.

33. Kiesling, T., Cox, K., Davidson, E.A., Dretchen, K., Grater, G., Hibbard, S., Lasken, R.S., Leshin, J., Skowronski, E. and Danielsen, M. (2007) Sequence

specific detection of DNA using nicking endonuclease signal amplification (NESA). *Nucleic acids research*, **35**, e117.

34. Tan, E., Erwin, B., Dames, S., Ferguson, T., Buechel, M., Irvine, B., Voelkerding, K. and Niemz, A. (2008) Specific versus nonspecific isothermal DNA amplification through thermophilic polymerase and nicking enzyme activities. *Biochemistry*, **47**, 9987-9999.

35. Van Ness, J., Van Ness, L.K. and Galas, D.J. (2003) Isothermal reactions for the amplification of oligonucleotides. *Proc. Natl. Acad. Sci. USA*, **100**, 4504-4509.

36. Beissenhirtz, M.K., Elnathan, R., Weizmann, Y. and Willner, I. (2007) The aggregation of Au nanoparticles by an autonomous DNA machine detects viruses. *Small*, **3**, 375-379.

37. Weizmann, Y., Beissenhirtz, M.K., Cheglakov, Z., Nowarski, R., Kotler, M. and Willner, I. (2006) A virus spotlighted by an autonomous DNA machine. *Angew. Chem. Intl. Ed.*, **45**, 7384-7388.

38. Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P. (1992) Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA*, **89**, 3010-3014.

39. Stirewalt, D.L., Pogosova-Agadjanyan, E.L., Khalid, N., Hare, D.R., Ladne, P.A., Sala-Torra, O., Zhao, L.P. and Radich, J.P. (2004) Single-stranded linear amplification protocol results in reproducible and reliable microarray data from nanogram amounts of starting RNA. *Genomics*, **83**, 321-331.

40. Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D. and Eberwine, J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA*, **87**, 1663-1667.

41. Liu, C.L., Schreiber, S.L. and Bernstein, B.E. (2003) Development and validation of a T7 based linear amplification for genomic DNA. *BMC genomics*, **4**, 19.

42. Schaerli, Y., Stein, V., Spiering, M.M., Benkovic, S.J., Abell, C. and Hollfelder, F. (2010) Isothermal DNA amplification using the T4 replisome: circular nicking endonuclease-dependent amplification and primase-based whole-genome amplification. *Nucleic acids research*, **38**, e201.

43. Kerkhoven, R.M., Sie, D., Nieuwland, M., Heimerikx, M., De Ronde, J., Brugman, W. and Velds, A. (2008) The T7-primer is a source of experimental

bias and introduces variability between microarray platforms. *PLoS One*, **3**, e1980.

44. Spiess, A.N., Mueller, N. and Ivell, R. (2003) Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. *BMC genomics*, **4**, 44.

45. Turner, D.J., Pingle, M.R. and Barany, F. (2006) Harnessing asymmetrical substrate recognition by thermostable EndoV to achieve balanced linear amplification in multiplexed SNP typing. *Biochemistry and Cell Biology*, **84**, 232-242.

46. Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F. *et al.* (2007) Multiplex amplification of large sets of human exons. *Nat. Methods*, **4**, 931-936.

47. Blanco, L. and Salas, M. (1996) Relating structure to function in phi29 DNA polymerase. *The Journal of Biological Chemistry*, **271**, 8509-8512.

48. Xu, Y., Lunnen, K.D. and Kong, H. (2001) Engineering a nicking endonuclease N.AlwI by domain swapping. *Proc. Natl. Acad. Sci. USA*, **98**, 12990-12995.

49. Samuelson, J.C., Zhu, Z. and Xu, S.Y. (2004) The isolation of strand-specific nicking endonucleases from a randomized SapI expression library. *Nucleic acids research*, **32**, 3661-3671.

50. Chan, S.H., Zhu, Z., Dunigan, D.D., Van Etten, J.L. and Xu, S.Y. (2006) Cloning of Nt.CviQII nicking endonuclease and its cognate methyltransferase: M.CviQII methylates AG sequences. *Protein Expression and Purification*, **49**, 138-150.

51. Xu, S.Y., Zhu, Z., Zhang, P., Chan, S.H., Samuelson, J.C., Xiao, J., Ingalls, D. and Wilson, G.G. (2007) Discovery of natural nicking endonucleases Nb.BsrDI and Nb.BtsI and engineering of top-strand nicking variants from BsrDI and BtsI. *Nucleic acids research*, **35**, 4608-4618.

52. Zhang, P., Too, P.H., Samuelson, J.C., Chan, S.H., Vincze, T., Doucette, S., Backstrom, S., Potamousis, K.D., Schramm, T.M., Forrest, D. *et al.* (2010) Engineering BspQI nicking enzymes and application of N.BspQI in DNA labeling and production of single-strand DNA. *Protein Expression and Purification*, **69**, 226-234.

53. Bellamy, S.R., Milsom, S.E., Scott, D.J., Daniels, L.E., Wilson, G.G. and Halford, S.E. (2005) Cleavage of individual DNA strands by the different

subunits of the heterodimeric restriction endonuclease BbvCI. *Journal of Molecular Biology*, **348**, 641-653.

54. Heiter, D.F., Lunnen, K.D. and Wilson, G.G. (2005) Site-specific DNA-nicking mutants of the heterodimeric restriction endonuclease R.BbvCI. *Journal of Molecular Biology*, **348**, 631-640.

55. Tabor, S. and Richardson, C.C. (1989) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *The Journal of Biological Chemistry*, **264**, 6447-6458.

56. Joneja, A. and Huang, X. (2009) A device for automated hydrodynamic shearing of genomic DNA. *Biotechniques*, **46**, 553-556.

57. Walker, G.T., Fraiser, M.S., Schram, J.L., Little, M.C., Nadeau, J.G. and Malinowski, D.P. (1992) Strand displacement amplification--an isothermal, in vitro DNA amplification technique. *Nucleic Acids Res*, **20**, 1691-1696.

58. Hanaki, K., Odawara, T., Nakajima, N., Shimizu, Y.K., Nozaki, C., Mizuno, K., Muramatsu, T., Kuchino, Y. and Yoshikura, H. (1998) Two different reactions involved in the primer/template-independent polymerization of dATP and dTTP by Taq DNA polymerase. *Biochem. Biophys. Res. Commun.*, **244**, 210-219.

59. Ogata, N. and Miura, T. (1998) Creation of genetic information by DNA polymerase of the thermophilic bacterium Thermus thermophilus. *Nucleic acids research*, **26**, 4657-4661.

60. Liang, X., Jensen, K. and Frank-Kamenetskii, M.D. (2004) Very efficient template/primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease. *Biochemistry*, **43**, 13459-13466.

61. Zyrina, N.V., Zheleznaya, L.A., Dvoretsky, E.V., Vasiliev, V.D., Chernov, A. and Matvienko, N.I. (2007) N.BspD6I DNA nickase strongly stimulates template-independent synthesis of non-palindromic repetitive DNA by Bst DNA polymerase. *Biol. Chem.*, **388**, 367-372.

62. Fu, D.J., Koster, H., Smith, C.L. and Cantor, C.R. (1997) Sequencing double-stranded DNA by strand displacement. *Nucleic acids research*, **25**, 677-679.

63. Myers, T.W. and Romano, L.J. (1988) Mechanism of stimulation of T7 DNA polymerase by Escherichia coli single-stranded DNA binding protein (SSB). *The Journal of Biological Chemistry*, **263**, 17006-17015.

64. Nakai, H. and Richardson, C.C. (1988) The effect of the T7 and Escherichia coli DNA-binding proteins at the replication fork of bacteriophage T7. *The Journal of Biological Chemistry*, **263**, 9831-9839.

65. Rapley, R. (1994) Enhancing PCR amplification and sequencing using DNA-binding proteins. *Molecular Biotechnology*, **2**, 295-298.

66. Sigal, N., Delius, H., Kornberg, T., Gefter, M.L. and Alberts, B. (1972) A DNA-unwinding protein isolated from Escherichia coli: its interaction with DNA and with DNA polymerases. *Proc. Natl. Acad. Sci. USA*, **69**, 3537-3541.

67. Fuller, C.W., McArdle, B.F., Griffin, A.M. and Griffin, H.G. (1996) DNA sequencing using sequenase version 2.0 T7 DNA polymerase. *Methods Mol. Biol.*, **58**, 373-387.

68. Lee, J.B., Hite, R.K., Hamdan, S.M., Xie, X.S., Richardson, C.C. and van Oijen, A.M. (2006) DNA primase acts as a molecular brake in DNA replication. *Nature*, **439**, 621-624.

69. Eklund, J.L., Ulge, U.Y., Eastberg, J. and Monnat, R.J., Jr. (2007) Altered target site specificity variants of the I-PpoI His-Cys box homing endonuclease. *Nucleic acids research*, **35**, 5839-5850.

70. Niu, Y., Tenney, K., Li, H. and Gimble, F.S. (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *Journal of Molecular Biology*, **382**, 188-202.

71. Smith, A.M., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J., Jr. and Stoddard, B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. USA*, **106**, 5099-5104.

Chapter 3, in part, is a reproduction of the material as it appears in: Joneja, A. and Huang, X. (2011) Linear nicking endonuclease-mediated strand-displacement DNA amplification. *Anal Biochem*, **414**, 58-69. Used with permission. The dissertation author was the primary investigator and author of this paper.

Chapter 4: Methods and Strategies for Using Homing Endonuclease I-PpoI in Strand Displacement Amplification Reactions

4.1 Abstract

We describe efforts towards the use of homing endonuclease I-PpoI in Strand Displacement Amplification (SDA) reactions. Since the wild type of this enzyme cuts both strands of its DNA template, in order to make it compatible with SDA it is necessary to block the cleavage of one strand. We have pursued several strategies to achieve this goal. In the first, we explored the possibility of using nuclease-resistant nucleotide analogues to protect one strand of DNA. In the second, we attempted to find variants of the I-PpoI recognition sequence that allow it to cut only one strand of the template. Lastly, we devised a strategy to engineer a library of mutant enzymes and a method of selection that allows us to enrich the coding genes for desirable variants. Here, we report limited success in amplifying DNA containing non-native nucleotides. We also report techniques for the selection of biotin-labeled oligos from a library of unlabeled DNA using biotin-streptavidin binding. We demonstrate limited success performing SDA with an engineered version of I-PpoI that has been partially inactivated. Finally, we describe methods for the expression of billions of mutant enzymes in parallel using water-in-oil emulsions and an *in vitro* transcription/translation kit based on the E. Coli system.

4.2 Introduction

Nicking enzymes are widely used in research applications such as replacement DNA synthesis, strand-displacement amplification, exonucleolytic degradation or the creation of small gaps in DNA for the study of DNA mismatch repair. However, there is a dearth of nicking enzymes available. While New England Biolabs lists over 225 different restriction enzymes with a range of recognition sites and optimum temperatures and buffers, only 9 nicking enzymes are available for purchase. Further, only one of these 9 is a naturally occurring enzyme while the others were engineered from endonucleases. This underscores a need for more nicking enzymes and the necessity of engineering them. In particular, it would be advantageous to have a selection of nicking endonucleases available with highly specific recognition sequences for the manipulation of large genomes. Homing endonucleases are a class of restriction enzymes whose extremely long recognition sequences make them excellent candidates for use with genomic DNA.

The homing endonuclease I-PpoI is the best characterized enzyme from the His-Cys box family and one of the most thoroughly examined homing enzymes. It has been isolated from the slime mold *Physarum Polycephalum* and is now commercially available from Promega. There are many features of this enzyme that make it a particularly attractive target for engineering and use in genomic applications. Details of DNA recognition and the mechanism of cleavage have been ascertained from both activity assays as well as from extensive X-ray crystallographic structural analyses of the enzyme bound to its DNA substrate (1). Its catalytic activity (turnover rate of 2.6

min^{-1}) is similar to that of Type II restriction enzymes and far superior to other homing endonucleases, and it is completely insensitive to DNA methylation (2,3). I-PpoI has a 15-bp recognition sequence but has been shown to cut sequences with up to 3 degenerate base pairs (4). Lastly, it is stable for long periods of time at 37º C.

Figure 4.1. The consensus recognition sequence of I-PpoI. Cutting sites in the pseudo-palindromic sequence are indicated with arrows.

Given the success that was achieved with using nicking endonuclease-mediated strand displacement amplification with unbiased replication of long stretches of DNA (5), it is natural to seek nicking enzymes with longer recognition sequences to make the technique more compatible with large genomes. It is important that the nicking or restriction enzyme is not able to cut within the template itself as this will create myriad spurious products. Most enzymes will cut many times within a genome and are unsatisfactory for genomic amplification. Naturally occurring nicking homing endonucleases, such as I-HmuI, are very rare. I-HmuI has previously been isolated from the *Bacillus subtilus* bacteriophage SPO1 (6,7). The crystal structure of the substrate-bound enzyme has been determined and a 25 bp recognition site has been suggested. However, it is hypothesized that the actual specificity of I-HmuI is far less than 25 base pairs as the observed contacts between enzyme and DNA indicate that no more than 14 sequence specific contacts exist (8).

Thus, it is desirable to adapt a well-studied homing endonuclease such as I-PpoI for use with SDA. In order to convert I-PpoI to a nicking enzyme, 4 strategies were pursued.

The first strategy was to use non-native nucleotides to protect one strand of the DNA template from digestion. Phosphorothioate-modified nucleotides have been shown to be resistant to enzyme digestion (9-12) and the original experiments demonstrating successful strand displacement amplification used hemiphosphorothioate recognition sites (13-15). Another promising nucleotide analogue is the locked nucleic acid (LNA). These have demonstrated exceptional resistance to nucleases and are often used *in vivo* due to their extended lifetimes in cells (16-18). If either of these non-native nucleotides is successful in blocking the cleavage of one strand of a recognition sequence by I-PpoI, they could be used to force I-PpoI into acting as nicking enzyme and linear SDA would be possible. However, while it may be possible to perform SDA using I-PpoI and dNTP with an alpha-thiophosphotate (α–S-dNTP), this approach has several disadvantages. As α–S-dNTP is not a natural substrate of the polymerase, it likely produces some decay in the processivity of the enzyme. For example, Taq polymerase has a 10-fold higher affinity for normal nucleotides over thiol nucleotides (19). It also adds complexity and cost to the reagent mix (20). It may also interfere in the sequencing chemistry itself. It would be preferable to use a nicking enzyme.

The second strategy was to convert I-PpoI into a nicking endonuclease by inactivating one of its two cutting sites. Since I-PpoI is a dimer, this required us to

express the protein as a monomer by tethering two subunits together with an amino acid chain. If one subunit was active and the other a knock-out, only one strand of DNA template could be cut for each binding event. If the polymerase was able to regenerate the nicked strand before the restriction enzyme disengaged, bound to the template again, and cut the second strand, I-PpoI could act as a nicking enzyme. We theorized that this could allow I-PpoI to be used for SDA.

The third strategy was to analyze the entire sequence space of the enzyme's recognition site. Since it is known that I-PpoI has an elastic recognition sequence, it is possible that there exists a sequence context that allows I-PpoI to cut one strand of template but not the other. In order to identify any sequences that satisfy this requirement, we optimized a platform for the targeted labeling of nicked templates by biotinylated nucleotides, and the recovery of the labeled oligonucleotides from a library of randomized sequences. It has been shown that biotin labeling is an effective method of separating target DNA molecules (21).

The final and most ambitious strategy was to find mutations in the protein sequence that allow the enzyme to recognize and cleave a non-palindromic sequence. Doi et al (22) and Zheng and Roberts (23) recently showed that it is possible to select restriction endonuclease genes from a randomized library or bacterial genome using *in vitro* compartmentalization (IVC). We intended to modify and extend this technology for the selection of nicking endonucleases and select an active nicking variant of homing enzyme I-PpoI. The expression of a library of mutated genes would occur in artificial cells. The artificial cells were water-in-oil emulsions containing recombinant

transcription/translation reagents and one copy of a mutated open reading frame (ORF). This provided a link between phenotype and genotype such that an enzyme that modifies its own DNA template will be selected. Only nicking enzymes enable SDA to occur on their own gene, and these newly synthesized strands could be identified using biotin-labeled nucleotides.

4.3 Materials and Methods

4.3.1 Strand displacement amplification with non-native nucleotides

All enzymes except I-PpoI were obtained from NEB. I-PpoI was obtained from Promega. All oligos were ordered from Integrated DNA Technologies. The templates for testing SDA with phosphorothioate nucleotides were 5' – ATG GTC AGC TTA CAT GGA TCG ACT CGA TCC GCT GAG GCT *GCT ACC TTA A*GA GAG* CTA CCA G – 3', 5' – CAG CTT ACA TGG ATC GAC TCG ATC CGC TGA GGC *TGC TAC CTT A*A*G* AGA G*CT ACC AGC TGG T – 3', and their respective reverse complements. The * indicates a phosphorothioate bond replacing the non-bridging oxygen in the oligo phosphate backbone. The recognition sequence for I-PpoI is in italics. The templates for testing SDA with locked nucleic acids (LNA) were 5' – CAG CTT ACA TGG ATC GAC TCG ATC CGC TGA GGC T*GC TAC CTT A+A+G AGA G*CT ACC AGC TGG T – 3', 5' – CAG CTT ACA TGG ATC GAC TCG ATC CGC TGA GGC T*GC TAC CTT +A+A+G +AGA G*CT ACC AGC TGG T – 3', and their respective reverse complements. In accordance with IDT notation, the LNA bases are indicated by placing a + sign prior to the target base. The recognition

sequence for I-PpoI is in italics. The template for testing SDA with RNA bases was 5' – CAG ATC CTG TAA CG*G CTA CCT TrArA* r*GrAG AG*A TGC ATT GAC T – 3' and its reverse complement in DNA. RNA bases in the DNA oligo are indicated by a lowercase "r" prior to the RNA base.

Each SDA reaction with non-native nucleotides contained 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol (DTT), 330 μM each dTNP, 12 nM DNA template, 100 μg/mL BSA, 5 μM SSB, between 5 nM and 30 nM I-PpoI, and either 125 nM Klenow exo- or 30 nM Bst Large Fragment. The reaction was allowed to proceed at 37 °C for various amounts of time between five min and five hours. Time points were taken by removing 5 μL of the reaction volume and adding it to 15 μL of PAGE gel stop buffer (80% formamide, 20 mM EDTA pH 8.0, 10 mM Tris-borate, 0.0125% bromophenol blue and 0.0125% xylene cyanol FF).

The products were visualized using 20% PAGE. Gels were run at 300V for 40 min, stained with SYBR Gold, and imaged using a BioRad Gel-Doc.

4.3.2 Design, expression, purification, and use of dimeric I-PpoI

A plasmid containing the gene for I-PpoI was obtained with permission from the manufacturer, Promega. The gene was extracted via PCR using primers 5' – GAG TCC TAC ATA TGG GCA GCA GCC ATC ATC ATC ATC ACA GCA GCG GCC TGG TGC CGC GCG GCA GCA TGG CGC TCA CCA ATG CTC – 3' and 5' – CCT AAC GTC TCG AGT TAT ACC ACA AAG TGA CTG CCC C – 3'. All oligos were ordered from Integrated DNA Technologies. These primers also added a

6x His tag to the N-terminus of the gene, and a stop codon and XhoI cutting site to the C-terminus of the gene.

One aliquot of I-PpoI gene was transformed into an H98A mutant. It was first digested with restriction enzymes BtgZI and SfaNI (from NEB). The fragment between the two cuts was removed with a PCR Clean-up kit from NEB. It was replaced via ligation with pre-annealed oligos 5' – P – CAA AAC CTG CAC AGC ATC GGC GCT ATG TC – 3' and 5' – P – TTA TGA CAT AGC GCC GAT GCT GTG CAG GT – 3'.

A tether was designed to join two gene monomers, one the native protein and one with the H98A mutation. It had amino acid sequence N-terminus – NSP GSG NSP TSS GNT GPS SGT NTP AA – C-terminus. The mutant gene was digested with DraIII and the wild type gene with HaeII, creating sticky ends on each. The DNA oligos for the tether, using the most common codons for E. Coli and including the requisite sticky ends, were 5' – P – GTG GTA AAC AGC CCG GGC AGC GGC AAC AGC CCG ACC AGC AGC GGC AAC ACC GGC CCG AGC AGC GGC ACC AAC ACC CCG GCG GCG ATG GCG C – 3' and 5' – P – CAT CGC CGC CGG GGT GTT GGT GCC GCT GCT CGG GCC GGT GTT GCC GCT GCT GGT CGG GCT GTT GCC GCT GCC CGG GCT GTT TAC CAC AAA – 3'. When ligated with T4 ligase (USB), a dimeric protein was encoded by a single gene (6His-H98A-WT). This gene was spliced into plasmid pET-24(a) (Merck Biosciences) and transformed into DH5α cells (Life Technologies). Following plating, one colony was chosen and grown in 2X YT media overnight. The plasmid was isolated with a

miniprep kit from Life Technologies and the sequence was confirmed by Eton Biosciences. The plasmid was then transformed into BL21(DE3) cells (NEB).

Approximately 3 mL of LB medium with 50 µg/mL kanamycin was inoculated with a single fresh colony and grown at 37 °C overnight. An aliquot (600 µL) was then added to 100 mL of LB medium with 50 µg/mL kanamycin and 1 mM zinc acetate. The broth was incubated at 37 °C with shaking for 3.5 hours. Then IPTG was added to a final concentration of 1 mM and the culture was grown at 37 °C with shaking for 4 hours. The cells were harvested by centrifugation at 4000 x g for 20 min.

A series of imidazole buffers were made containing 50 mM Sodium Phosphate Buffer pH 8.0, 300 mM NaCl and either 1 mM, 20 mM, 50 mM, or 250 mM imidazole. The cells were resuspended in 1 mL of the 1 mM imidazole buffer. Lysozyme was added to 1 mg/mL and the cells were incubated on ice for 30 minutes. The cells were then lysed by sonicating in an ice bath, six times for 10 seconds each time with 5 seconds pauses in between, at max power. DNase I and RNase A (Roche) were added to final concentrations of 100 units/mL and 5 µg/mL respectively. The cells were frozen and thawed three times, and the lysate was centrifuged at 10,000 x g for 30 minutes at 4 °C. The supernatant was collected and 150 µL of Ni-NTA resin (Qiagen) was added. The mixture was incubated on a roller for 2 hours at 4 °C. It was then centrifuged at 16,000 x g for 5 min and the supernatant was collected. The resin was washed with 1 mL 20 mM imidazole once and 50 mM imidazole twice. Each wash consisted of 5 min on a roller at 4 °C and a centrifugation step. The protein was eluted with 10 washes of 100 µL 250 mM imidazole. The purity of the elutions was

confirmed by denaturing SDS-PAGE. The gels were 12% pre-cast from Bio-Rad and stained with Gel-Code Blue protein stain. They were then assayed for activity using a test template containing the I-PpoI recognition sequence.

With the functionality of the enzyme confirmed, an aliquot was used in a strand displacement amplification reaction. The reaction contained 40 nM test template (5’ – AGT CAA TGC AT*C TCT CTT AA/G GTA GC*C GTT ACA GGA TCT G – 3’ and its reverse complement, recognition site in italics and cleavage indicated by “/”), 25 mM CHES and 25 mM CAPS pH 10.0, 2 mM $MgCl_2$, 1 mM DTT, 100 µg/mL BSA, 5 µM SSB, 250 µM each dNTP, 125 nM Klenow exo-, and a small volume of eluted H98A-WT I-PpoI dimer. Time points were taken at 0 min, 10 min, 30 min, and 60 min by removing 5 µL of the reaction volume and adding it to 15 µL of PAGE gel stop buffer. The products were visualized using 20% PAGE. Gels were run at 300V for 40 min, stained with SYBR Gold, and imaged using a BioRad Gel-Doc.

4.3.3 Using biotin-streptavidin binding to purify labeled oligos

In order to simulate an SDA product with biotinylated nucleotides incorporated, a test oligo was ordered with an internal biotin. The 40 bp oligo was obtained from Integrated DNT Technologies and had sequence 5' – CAG ACT GAT GAC GAT GAG TG-BiotinT GTG ACG TGA CCT CTC AGC TC - 3' and its reverse complement. Eight conditions were tested for eluting this oligo from streptavidin beads. The yield was estimated by using 20% PAGE at 400V for 30 min, or by qPCR

using the Bio-Rad MiniOpticon Real-Time PCR Detection System. The primers for qPCR were 5’ – CAG ACT GAT GAC GAT GAG TG - 3’ and 5’ – GAG CTG AGA GGT CAC GTC AC - 3’.

Positive and negative control templates for biotin-incorporating SDA were obtained via PCR from plasmid pET-24(a) (Merck Biosciences). A 157 bp nickable fragment was extracted using primers 5’ – CGG TTA CCG ACT CAG CTT GAG CCA GTC ATG GGA TCG A – 3’ and 5’ – CGG CGG GAC CAG AGA AAA ATC ACT C – 3’. The forward primer also introduced the recognition sequence for nicking enzyme Nt.BbvCI at the 5’ end of the fragment. A 198 bp sequence that lacks the Nt.BbvCI nicking site was obtained with primers 5’ – TGG CTT CTG ATA AAG CGG GCC ATG T – 3’ and 5’ – CAT CCA TAC CGC CAG TTG TTT ACC CTC – 3’.

An SDA reaction was performed to label the 157 bp template with biotin, either by itself or in the presence of the 198 bp negative control or “decoy” fragment. In the case of the template being labeled as the only species in the mix, it was first nicked in 10 mM Tris-Cl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT using 200 nM Nt.BbvCI at a concentration of 40 nM. This volume was incubated at 37 °C for 1 hour and then heated to 80 °C for 25 minutes to kill the nicking enzyme. The labeling was done in the same reaction buffer but included 100 μg/mL BSA, 5 μM SSB, 5 uM each dNTP with between 5% and 100% of dTTP replaced by biotin-dUTP (Roche), and 125 nM Klenow exo- at a template concentration of 25 nM. To remove excess biotinylated nucleotides, the reaction volume was run through an Amicon

Ultra-0.5 mL 3K with 400 µL TE twice and recovered in a 50 µL volume of TE buffer. When the reaction was done in the presence of the decoy 198 bp fragment, the protocol was the same except each tube contained a final concentration of 60 pM 198 bp fragment and either 60 pM, 6 pM, 600 fM, 60 fM, 6 fM, or 0 fM of the 157 bp target.

Templates were bound to streptavidin-coated magnetic beads based on carboxylic acid beads (Dynabeads MyOne Streptavidin C1) obtained from Invitrogen. Stock beads (10 mg/mL) were washed by adding 1 µL of beads to 300 µL of 1X Binding and Washing (BW) buffer (5 mM Tris-Cl, pH 7.5, 1 mM EDTA, 1 M NaCl, 0.1% Tween 20) in a DNA LoBind tube (Eppendorf). The beads were pulled down with a magnetic tube rack and resuspended in 25 µL 2X BW buffer. The DNA template of the same volume was added and the tube was incubated at room temperature on a tube rotator (MACSmix) for at least 2 hours. The beads were then washed 6 times with 1X BW. Each wash step consisted of adding 400 µL 1X BW, transferring the sample to a new DNA LoBind tube, vortexing or incubating briefly on the tube rotator, pulling the beads down on the magnetic rack, and removing the supernatant. The biotinylated oligos were recovered from the beads by resuspending them in 20 µL of an elution buffer, typically 95% formamide with 10 mM EDTA, and heating to 98 °C for 10 minutes. The beads were then pulled down and the supernatant quickly removed. If the samples were to be used in downstream reactions such as qPCR, the formamide was washed from the samples by running them through an Amicon Ultra-0.5 mL 3K with 400 µL TE three times. The primers used to extract the

fragments were also used in qPCRs with PerfeCTa Sybr Green Fastmix (Quanta Biosciences) on a Bio-Rad MiniOpticon Real-Time PCR Detection System.

4.3.4 Expression of I-PpoI in artificial cells

A protocol for creating emulsion droplets of the desired size was established using a fluorescein aqueous phase to enhance visibility of the drops for analysis. A fresh oil phase was created daily, consisting of 4.5% Span-80 and 0.5% Triton X-100 in molecular biology grade mineral oil (all from Sigma-Aldrich). The aqueous phase consisted of 10 μM fluorescein, 10 mM Tris-Cl pH 8.0, and 0.1% Triton X-100. First, 450 μL of the oil phase was put into a 2 mL round-bottom cryogenic vial (no. 430661, Corning) positioned on a magnetic stirrer. A magnetic microstir bar was added (8 x 3 mm) and stirring was initiated at 1500 rpm. The aqueous phase was added over the course of 1 minute, 5 x 5 μL aliquots for a total volume of 25 μL. The entire experiment was conducted in a cold room at 4 °C. A few microliters were removed every few minutes for analysis. The samples were placed on a coverslip and sandwiched with an additional coverslip. They were imaged using a 20X objective (200X total magnification) on a Zeiss Axio Observer microscope with a FITC filter cube and Lambda DG-5 light source from Sutter Instruments. ImageJ was used to quantify the size of the emulsion compartments.

For the *in vitro* protein expression, the PURExpress kit from New England Biolabs (NEB) was used. The components were assembled as per the manufacturer's directions and included 1.6 units/μL of murine RNAse inhibitor (NEB) and a quantity

of linear I-PpoI gene, including all E. Coli regulatory sequences, at a concentration of between 3 ng/µL and 0.009 ng/µL. The reaction was assembled on ice. Once again 25 µL of aqueous phase was added over the course of one minute to 475 µl oil phase stirring at 1500 rpm (magnetic microstir bar 8 mm x 3 mm) in a cold room at 4 °C. The stirring was allowed to continue for 30 minutes. At this time, the emulsion was moved to an incubator at 37 °C for three hours. The emulsions were transferred to 1.5 mL tubes and centrifuged at 23,000 x g at 4 °C for 1 hour. The upper (oil) phase was removed and 20 µL of stop buffer (10 mM Tris-Cl pH 8.0, 20 mM EDTA) was added and the tube was vortexed to mix. Two methods were used to break the emulsions. In the first, the emulsion was washed three times with chilled mineral oil. For each wash, 1 mL of mineral oil was added, the emulsion was resuspended by pipeting, and then spun for 20 minutes at 4 °C. A wash with 1 mL hexane could be used to remove the remaining organic phase before the aqueous phase was collected. In the second method, 1 mL of water-saturated ether was added to the emulsion and the tube was vortexed. The upper (solvent) phase was removed and the residual ether was evaporated by heating to 30 °C for 30 minutes.

A small volume (2.5 µL) of the recovered aqueous phase was used in an activity test to determine the enzyme functionality. Each reaction contained 750 nM of a 40 bp oligo containing the I-PpoI recognition sequence, 25 mM CHES and 25 mM CAPS pH 10.0, 3 mM $MgCl_2$, and 100 µg/mL acetylated BSA. The reaction was incubated at 37 °C for three hours. To remove high molecular weight DNA and RNA present in the *in vitro* expression mix, the reactions were run through a Microcon YM-

100 (Millipore) and the flow through was collected for analysis on a 20% polyacrylamide gel.

4.4 Results and Discussion

We have investigated the potential of several strategies to allow the homing endonuclease I-PpoI to perform strand displacement amplification on DNA templates. At the core of each is the necessity of either slowing or entirely preventing the cleavage of one strand of the DNA template so that a polymerase can bind to and extend the primer created by the opposite nick.

We have shown that if phosphorothioate linkages or locked nucleic acids (LNAs) are incorporated in sufficient numbers in one strand of the DNA template, they are capable of sufficiently inhibiting the cutting of that strand so that SDA can proceed. The use of RNA bases in the template did not seem to affect cleavage at all. However, preventing the break in one strand causes the entire reaction to be much slower than would be the case with a template consisting of purely native nucleotides. Additionally, there is a trade-off between strand protection and reaction rate that depends on the number of non-native nucleotides present in the recognition sequence.

Figure 4.1 illustrates SDA using either 1 (A) or 3 (B) phosphorothioate linkages and Figure 4.2 illustrates SDA using either 2 (A) or 4 (B) LNAs in I-PpoI's recognition site. In the case of the phosphorothioate-modified strands, one modified nucleotide was unable to prevent cleavage of the DNA. However, three modified nucleotides provided adequate protection to allow SDA to occur. In the case of the

LNAs, 2 modified nucleotides were sufficient to allow SDA to proceed for 2 hours. The use of 4 LNAs in the strand significantly slowed the reaction.

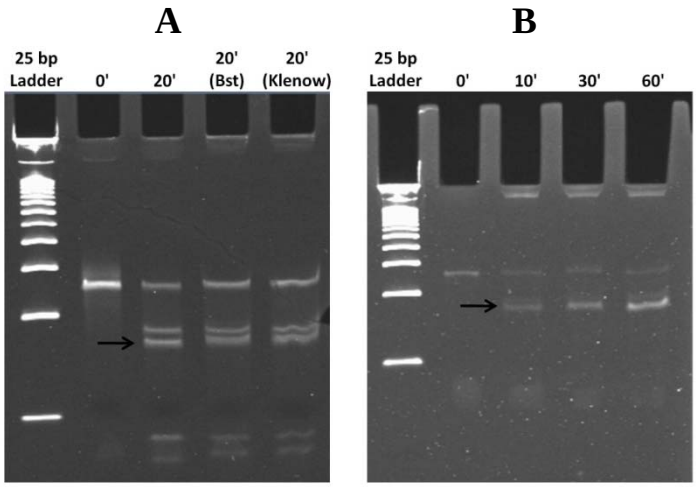

Figure 4.2. Strand Displacement Amplification using phosphorothioate linkages to block the cleavage of one strand of the DNA template. In (A), the DNA template contained one phosphorothioate in the backbone. The four lanes are a zero time point, template incubated with just I-PpoI for 20 min, template incubated with I-PpoI and Bst Large Fragment polymerase for 20 min, and template incubated with I-PpoI and Klenow exo-, respectively. In (B), the DNA template contained three phosphorothioate linkages in the backbone. The template was incubated with I-PpoI and Klenow exo- for one hour.

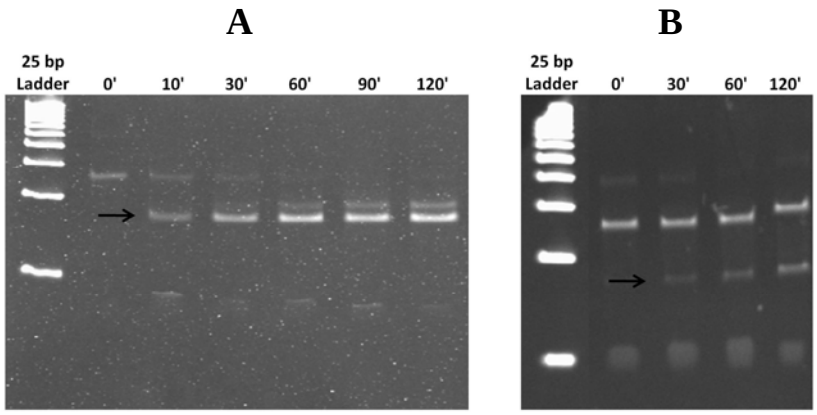

Figure 4.3. Strand Displacement Amplification using locked nucleic acids to block the cleavage of one strand of the DNA template. In **(A)**, the DNA template contained two LNA residues. The template was incubated with I-PpoI and Klenow exo- for two hours. In (B), the DNA template contained four locked nucleic acids. The template was incubated with I-PpoI and Klenow exo- for two hours.

As is evident, in all cases the reaction proceeds very slowly. We believe that since the DNA template is never fully cleaved, the restriction enzyme is reluctant to disengage from its template. In the Michaelis Menten enzyme kinetics equation below, k_1 is the association rate, k_2 is the cleavage rate, and k_3 is the dissociation rate. It has been shown that k_3 is frequently the rate-limiting portion of the reaction, particularly for restriction enzymes that have been mutated to nicking enzymes (24). While the association rate constant may also be affected, it seems likely to be the cleavage and dissociation rates that are most decreased by the modified nucleotides.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} EP \xrightarrow{k_3} E + P$$

We have demonstrated that a tethered dimeric mutant version of I-PpoI can also be used to do SDA. We analyzed the crystal structure of two I-PpoI monomers

complexed to their DNA substrate (25,26) and determined the distance between the C-terminus of one monomer and the N-terminus of the second monomer to be approximately 45 angstroms. Assuming a length of 2.5 angstroms per amino acid, we decided to use a tether of length 19 amino acids. The tether was designed to have a nearly equal proportion of hydrophobic and hydrophilic residues, incorporated a number of scattered prolines to break up any potential secondary structures, and ended with a pair of alanines to encourage the formation of an alpha helix at the N-terminus of the second monomer. After confirming that a dimer containing two wild type I-PpoI monomers connected by this tether was functional, we proceeded to tether a wild type monomer to an inactivated mutant monomer.

The mechanism of cleavage of I-PpoI has been described in detail (25-28), and the key amino acids have been identified: A histidine imidazole ring (His98) serves as a base to deprotonate and activate a water molecule. The water molecule then attacks the scissile phosphate group, cleaving it from the 3'- nucleotide. An asparagine (Asn119) serves to stabilize a metal cation, usually Mg^{2+}, which in turn stabilizes the phosphoanion intermediate and the 3'- hydroxylate leaving group. An arginine (Arg61) donates hydrogen bonds to the scissile phosphate group and stabilizes the final product. It has been suggested that a nearby alternate histidine (His78) can, to an extent, rescue activity of the enzyme in the event that H98 is mutated (28). However, it has been shown, and we have confirmed, that an H98A mutation in I-PpoI completely abolishes catalytic activity without affecting the affinity of the enzyme for its DNA template (27,29). Thus, we chose to use this mutation to inactivate one

monomer of I-PpoI in the tethered dimer. Since the recognition sequence for I-PpoI is pseudopalindromic, we could not control which strand of the DNA template was nicked. There is the possibility that an I-PpoI complex could bind to the target, nick one strand, disassociate from the template, and bind again in the opposite orientation to nick the second strand, creating a full cut. However, since only one strand could be cut at a time, we hoped that using a much lower concentration of I-PpoI vs. polymerase would allow the polymerase to bind to and extend from the nick before the I-PpoI could reorient itself and cut the second strand.

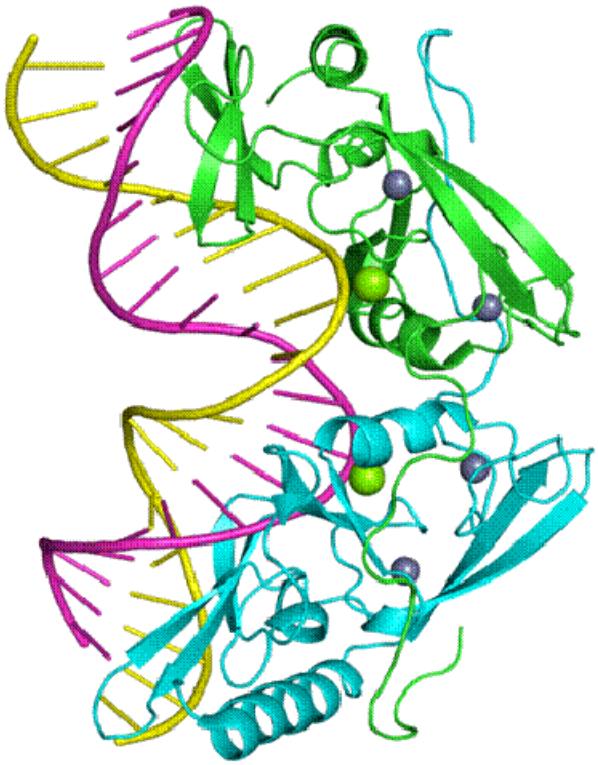

Figure 4.4. An I-PpoI dimer complexed with its DNA substrate (From Protein Data Bank ID 1cyq (26)). The two identical subunits appear in green and blue. Note that the tail end of each subunit wraps around the other, so that the N-terminus of one subunit is only 45 angstroms from the C-terminus of the other. Magnesium ions in green appear close to the scissile groups, in particular His98.

Figure 4.5 illustrates SDA with the H98A-WT dimeric I-PpoI. The 40 bp template can be nicked on either the “top” or “bottom” strand. If the top strand is nicked, a 20 nucleotide (nt) SDA product is formed. If the bottom strand is nicked, a 16 nt fragment is amplified. In the gel image, both products can be seen increasing with time. It is apparent that the 16 nt product is being amplified faster than the 20 nt fragment. This may indicate a slight preference for one side of the pseudopalindromic recognition sequence. Much like the SDA with modified nucleotides, the amplification is overall quite slow. Also, in practice, it may not be desirable to amplify both strands of the template.

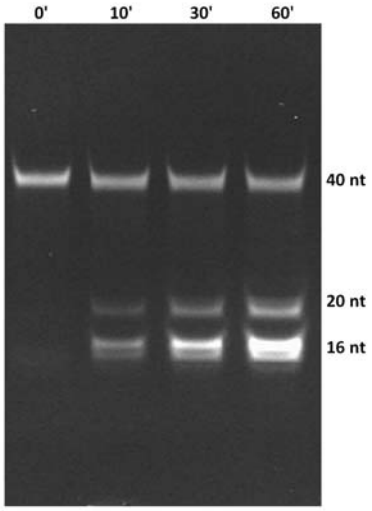

Figure 4.5. SDA with mutant tethered dimer H98A-WT. Since either strand of the template can be nicked, two distinct products are formed in the presence of a polymerase. The enzyme variant was expressed in E. Coli, purified using a 6 x His tag, and incubated with a 40 bp DNA target and Klenow exo- polymerase.

In order to identify potential recognition sequences that would allow I-PpoI to nick, but not cut, at its binding site, we turned to strand displacement amplification

again. A DNA template that is fully cleaved by I-PpoI is not a suitable template for a polymerase because 3' ends are left on each strand, precluding DNA synthesis. However, if either strand is cut and the opposite strand is left intact, a polymerase can incorporate nucleotides via SDA. If labeled nucleotides are used, this could be a mechanism to differentiate templates that had been nicked from those that had been fully cleaved or were left untouched.

If we choose to sample the entire sequence space of the I-PpoI recognition site, that is a 15 bp stretch that must be completely randomized. This would yield 4^{15} variants, or about a billion different sequences. Of these, we can guess that somewhere between 0 and 10 will have the desired properties that allow I-PpoI to nick but not fully cleave the template. Assuming we use a reasonable concentration (1 μM) and volume (100 μL), we are left with the challenge of separating less than 100,000 labeled molecules from a comparatively vast quantity (10 trillion) of unnicked or fully cleaved DNA fragments.

In order to purify target labeled oligos from a library of undesired DNA, we chose to use biotin-streptavidin binding. Streptavidin is a tetrameric protein purified from *Streptomyces avidinii* that has an extremely high affinity for biotin (vitamin B7). The dissociation constant of the biotin-streptavidin complex is between 10^{-15} and 10^{-14} mol/L (30) and the interaction has a half-life at room temperature of 35 hours (31), making it one of the strongest non-covalent bonds in nature. These qualities, along with the commercial availability of biotinylated nucleotides and previous publications illustrating the feasibility of using biotin-streptavidin bonding to purify labeled oligos

(21,22), suggested that this interaction would be an ideal candidate for the selection of target oligos. However, initial experiments did not yield a satisfactory enrichment factor so we sought to optimize our protocol. There are a number of parameters that could be responsible for the poor separation of biotinylated oligos; these include the method of labeling (choice of polymerase, concentration of biotinylated nucleotides, etc.), binding buffer, the incubation time, the washing buffer, and the elution conditions. Realistically, we cannot expect the procedure to be 100% efficient because each step will involve losses. A fraction of target oligos will not be labeled, a fraction of labeled oligos will not bind to the beads, and a fraction of bound oligos will likely be washed away before the elution step. Further, some of the target molecules that do bind and stick during the washes may not be eluted from the beads and will not be recovered and detected. Additionally, even with the use of ultra-pure DNA LoBind polypropylene tubes, it is inevitable that some DNA, both labeled and unlabeled, will coat the surface of the tube and stick during washing. Replacing the tube several times during washing helps to mitigate this particular issue. Unfortunately, another source of contamination is the streptavidin beads themselves. They are known to be "sticky" and nonspecific binding is expected. Since DNA is negatively charged, it is important to use beads that are also negative or near-neutrally charged. Most commercially available recombinant streptavidin has an isoelectric point near 7.0, so the use of a buffer with pH $\geq$7.5 is essential.

First, we attempted to optimize the binding and washing protocols by adjusting the pH, salt concentrations, and detergent concentrations of the buffers and the length

of each incubation. None of the new formulas provided markedly superior capture of biotinylated fragments or suppression of the background unlabeled DNA. We also tried to pre-coat the beads with degraded herring sperm DNA to reduce the nonspecific binding to the streptavidin beads, but this was ineffective. Real-time PCR data indicated that the poor enrichment factor was likely due to insufficient recovery of the target fragment rather than excessive retention of background fragments.

Since modifications to the binding and washing protocols failed to produce better yield of the labeled oligos, we chose to focus on the elution conditions. A literature search provided an array of different methods for the breaking of the biotin-streptavidin bond. These included buffers with SDS or urea to denature the streptavidin and free biotin to compete with the labeled oligos (32) (Novagen's MagPrep protocol), alkaline conditions to denature the double-stranded DNA and cause the dissociation of the biotin-streptavidin complex (33), or formamide and EDTA to disrupt the biotin-streptavidin bond (Life Technologies Dynabeads protocol). Uniquely, one group has reported the efficient elution of the biotin-streptavidin bond using only nonionic water (34). They all incorporated elevated temperatures. Table 4.1 summarizes the elution conditions we surveyed.

Table 4.1. Conditions tested to elute a biotinylated DNA fragment from streptavidin beads

	Buffer	Temperature	Incubation
1	2% SDS, 3 mM biotin in PBS	99 °C	15 min
2	2% SDS, 3 mM biotin, 8 M urea in PBS	99 °C	15 min
3	0.1% SDS	99 °C	15 min
4	Millipore dH_2O	80 °C	5 min
5	10 mM EDTA, 95% formamide	90 °C	10 min
6	10 mM EDTA	90 °C	10 min
7	95% formamide	90 °C	10 min
8	25% Ammonium Hydroxide	70 °C	10 min

To assay the efficiency of each elution condition, a 40-bp oligo with an internal biotinylated nucleotide was incubated with an equal mass quantity of an unlabeled 25 nucleotide and magnetic streptavidin beads. After incubation, the DNA was eluted using each of the eight conditions and the supernatant was run on a gel. Surprisingly, as illustrated in Figure 4.6, all of the elution conditions performed about equally well. Approximately 50% of the original quantity of labeled DNA is recovered, while all of the unbiotinylated DNA has been washed away. The exception is the elution in pure water, which only recovered about half as much oligo as the other methods. Since it is unclear what percent of the labeled oligo ever bound to the beads, the 50% figure is not truly an indictment of the selected elution conditions; however, it is clear that the efficiency is not less than 50%.

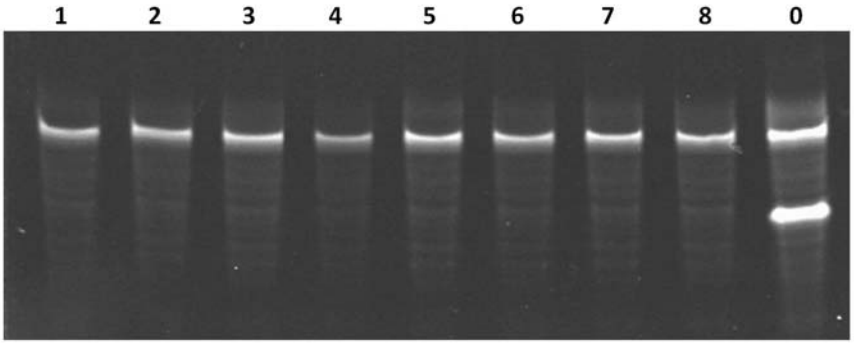

Figure 4.6. Biotin-streptavidin elution efficiencies using various conditions. Lane "0" indicates the quantity of template DNA that was incubated with magnetic streptavidin beads. The higher molecular weight band, 40 nucleotides long, includes an internal biotinylated dTTP. The lower molecular weight band, 25 nucleotides long, is unlabeled. The numbers above each lane correspond to the elution conditions described in Table 4.1.

One other attempt was made to optimize the elution step. It is possible that the elution is inefficient in every buffer because the biotin-streptavidin bond is too strong to be easily broken. Therefore, we tested monomeric avidin magnetic beads in a similar assay. Streptavidin and avidin bind most strongly to biotin when all four subunits are present in the tetramer. An engineered monomeric avidin is unable to crosslink with another subunits of avidin, greatly decreasing its affinity for biotin (K_d = 10^{-7} mol/L). It should therefore be much easier to elute. However, it also may have more difficulty binding the target molecule and, due to its carbohydrate side groups and basic pI of 10, is likely to attract more nonspecific binding. We observed both of these phenomena and concluded that streptavidin beads were the far superior choice.

Another factor that may influence the binding and elution efficiencies is the number of biotins present in the target oligo. In the previous experiment we used a synthetic DNA fragment with one internal biotin. However, fragments labeled via

SDA will be far less uniform and contain a distribution of biotins. This distribution, in turn, will depend on the number of relevant bases being incorporated, the percent of biotinylated vs. non-biotinylated nucleotides, and any preference the polymerase has for native nucleotides over modified, biotinylated nucleotides.

The test fragment we used is 157 bp long, of which 129 bases are downstream of the nick created by Nt.BbvCI. Of these, 32 are dTTP, and are eligible to be replaced by biotin-dUTP during a strand displacement amplification reaction. If no dTTP is included in the nucleotide mix during strand displacement, and the fragment is fully resynthesized, then there must be 32 labels on the new strand. However, if both dTTP and biotin-dUTP are included in the reaction, it is more difficult to predict how many biotins will be incorporated due to potential bias of the polymerase. The biotin-dUTP we used, provided by Roche, has a 16-Carbon linker connecting the biotin to the nucleotide. This is fairly long and should minimize the perturbation of the active site on the polymerase, but there is insufficient data available to confirm this. It has been reported that the incorporation rate of biotinylated nucleotides by Klenow exo- polymerase is about 50% of the rate for native nucleotides (35), but this does not guarantee that there will be twice as many native nucleotides as modified nucleotides in a newly synthesized strand. The protocol for NEB's Phototope kit indicates that when using a nucleotide mix with 16% biotinylated dATP and Klenow exo-, the expected biotin label rate is about one biotinylated dATP per 7-8 dATP bases, implying very little bias towards native nucleotides. Further, they state that using more

than 16% biotinylated dATP leads to crowding on the new strand that inhibits synthesis.

We elected to empirically determine the optimum ratio of native:biotinylated nucleotides. We investigated the efficiency of binding and eluting using increments of 10% biotinylated nucleotide between 0 and 100%. The results are illustrated in Figure 4.7.

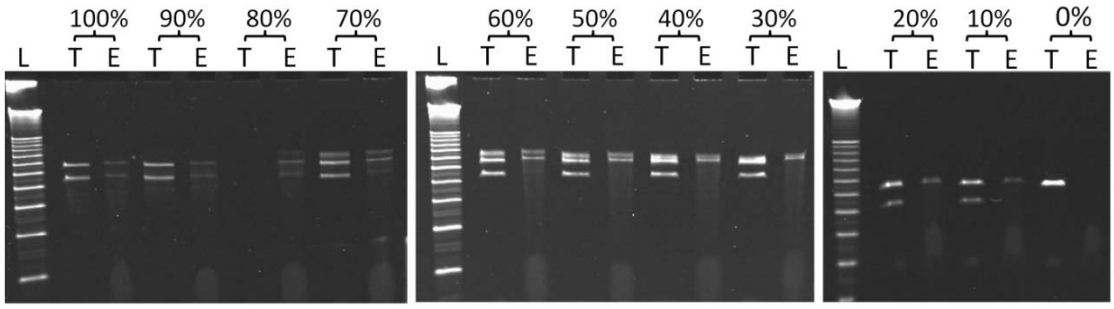

Figure 4.7. Binding and elution efficiency with different percentages of biotinylated nucleotides. Lanes "L" are 25 bp ladder. Lanes "T" are starting template and "E" are eluates. The percent of biotin-dUTP out of the total quantity of dTTP + biotin-dUTP is indicated above each pair of lanes.

There are up to three bands visible for each sample. The highest molecular weight band represents the newly synthesized strand that includes biotinylated nucleotides. Despite being the same length, it is higher on the gel than the middle band, which is the unnicked strand, because each biotinylated nucleotide is 3X the mass of a native nucleotide. Since the binding and washing is done under non-denaturing conditions, both of these bands should be present in the eluate. The lowest band represents the native DNA that was displaced during synthesis, and should not bind to the streptavidin beads.

The results indicate that it is not optimal to use a very high or very low percent of biotinylated dUTP, but rather something in the range of 50% - 60%. It is somewhat surprising that higher concentrations of labeled nucleotide were not more successful. If excessive biotinylated nucleotides were contributing too many bulky residues and inhibiting the polymerase during synthesis of the new strand, we would expect to see some truncated fragments on the gel. However, at 90% and 100% biotin-dUTP, there are two crisp bands on the gel and no indication of DNA with biotin. It appears that synthesis after nicking has not occurred.

Despite having optimized conditions for labeling, washing, and elution, technical difficulties still plague the purification of very small quantities of DNA using biotin-streptavidin binding. We have observed that when a positive control is labeled in the presence of an excess of negative control lacking a nicking site, the enrichment factor does not exceed 10-fold as detected by qPCR. However, if the positive control is pre-labeled prior to being mixed with the negative control and then the mixture is incubated on streptavidin beads, the enrichment factor can exceed 10,000-fold. This strongly suggests that the negative controls, in the presence of nicking enzyme and polymerase with biotinylated nucleotides, are being somehow labeled. Nonspecific nicking of the negative control is not visible by gel electrophoresis even after extended incubations. The use of an exo- polymerase does not allow the polymerase to chew back the 3’ ends of the DNA and then fill in with labeled nucleotides. Further, we have not observed the incorporation of bases to create 3’ overhangs on the DNA template. Thus, while the mechanism of labeling remains

unclear, it is obvious that further changes to the protocol are required in order to suppress nonspecific labeling. As a first step we will utilize a template with 3' dideoxy nucleotides to ensure that no bases are added as overhangs.

Lastly, the ideal method of using I-PpoI in SDA would be to convert it to a nicking enzyme with a non-palindromic recognition sequence. This would require not only a mutation to inactivate a cutting site (H98A) but also mutations to the amino acids that make contact with the DNA template and determine its specificity. It has been shown that the specificity of homing endonucleases I-CreI (36-38) and PI-SceI (39-41) can be modified, that asymmetry can be introduced into the recognition sequence of typically symmetrical target sites for restriction enzymes (42), and that homing endonucleases can be modified to strand-specifically nick their targets (43,44). The interaction of I-PpoI with its DNA substrate is well-studied (25,29,45,46), therefore several residues are known to be good candidates for mutation in order to alter the specificity of I-PpoI. These include R61, N57, Q63, and R74, interacting with bases ±4, 5, 6, and 7 in the recognition sequence, respectively.

Since we are not trying to force I-PpoI to recognize a particular sequence but rather any non-palindromic sequence, we can explore a larger sequence space by randomizing both the amino acid residues that make contact with the DNA and the nucleotides they interact with. However, this leads to a prohibitively large number of variants and conditions to test: $20^4 * 4^4 = 40$ million. In order to process this quantity of conditions in parallel, we chose to pursue a strategy based on *in vitro* compartmentalization (IVC) and the incorporation of biotinylated nucleotides. In this

strategy, illustrated in Figure 4.8, one DNA template containing a copy of a mutant I-PpoI gene and a mutant recognition sequence would be placed into each of billions of emulsion droplets. These water-in-oil emulsions would consist of recombinant transcription/translation reagents, allowing the protein to be expressed. Since the mutant gene and recognition sequence exist on the same DNA strand and are isolated from all other genes and templates, the genotype and phenotype are linked. An active nicking enzyme would interact only with the DNA fragment that created it, creating a single-stranded break in the recognition sequence. After heat-inactivating the enzymes, the emulsion can be broken and a polymerase with biotinylated nucleotides introduced to the mix to label the genes and recognition sequences that were successfully nicked. After recovering these fragments, they can be PCR amplified to do further rounds of selection or spliced into plasmids and transformed into E. Coli in order to separate each unique fragment for sequencing.

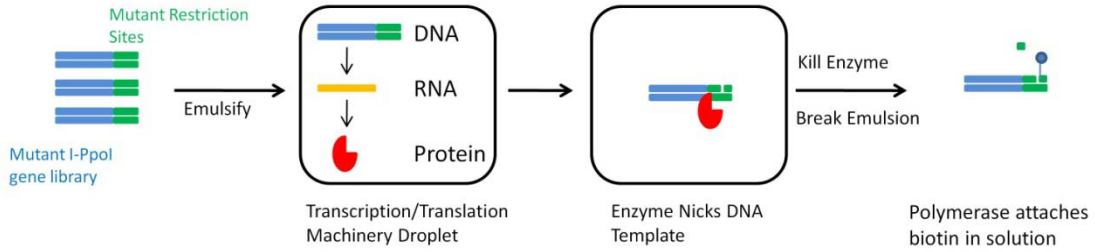

Figure 4.8. Strategy for selecting nicking genes and their substrates from a library. A library of fragments containing both mutant genes and degenerate recognition sequences is placed into an emulsion containing recombinant transcription/translation reagents such that there exists no more than one template per droplet. If an enzyme is produced that recognizes the sequence attached to its gene it may nick one strand. After the enzymes are inactivated and the emulsion is broken, a polymerase and biotinylated nucleotides are introduced to label nicked strands with biotin. These fragments can then be captured and separated from the rest of the library using streptavidin beads. If necessary they can be amplified for subsequent rounds of selection.

The first requirement for this strategy is the ability to create uniform emulsions of tunable size. This is a notoriously fickle and temperamental process that has been referred to as the "black art of emulsions" (47). While microfluidic channels have been touted as a method of creating extremely uniform emulsion droplets (47-52), they suffer from two drawbacks. First, the emulsion droplets created are generally in the range of 10 μm to 500 μm, which is too large for some applications. Second, while some devices can create microdroplets at a rate exceeding 10 kHz (47), this is still too low throughput for a system that requires billions of compartments. An alternative method that sacrifices uniformity for speed is the simple high speed stirring of an aqueous phase into an oil phase using a magnetic stir bar (22,23,47,53,54). We sought to calibrate this method for the *in vitro* expression of our protein.

E. Coli cells are generally on the order of 1.1-1.5 X 2.0-6.0 μm rods and the prokaryotic transcription/translation machinery has evolved to operate in cells of approximately 2 μm diameter (54), at which size the concentration of a single gene is about 400 nM. The use of droplets larger than 2 um with a single gene is inadvisable, as coupled transcription/translation can become inefficient at lower concentrations. Approximately 33 μL of fluid divided into 2 μm diameter compartments (33 fL volume) would provide one billion droplets. We found that when 500 μL of an emulsion containing 5% aqueous phase was stirred at a constant rate, the size of the emulsion droplets produced was dependent on the amount of time stirring was allowed to proceed. As shown in Figure 4.9, approximately 30 minutes of stirring at 1500 rpm with an 8 X 3 mm magnetic stir bar produced droplets of mean size 2 μm with a coefficient of variation of <0.3.

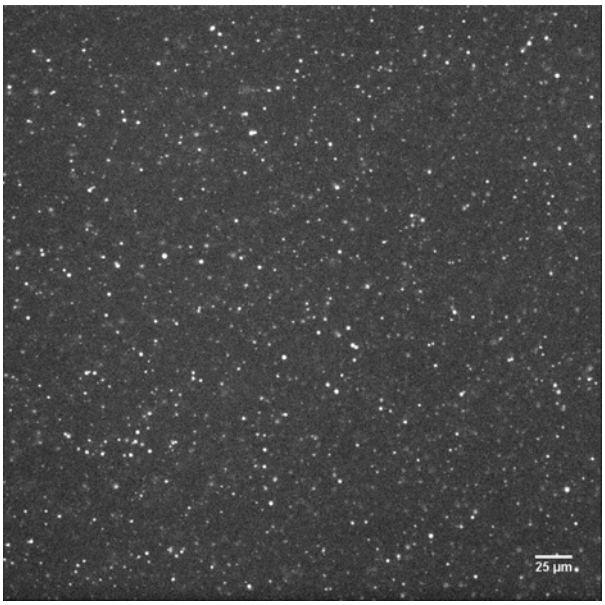

Figure 4.9. Fluorescent micrograph of an emulsion created by rapid stirring. An aqueous phase containing 10 µM fluorescein was slowly added to a stirring oil phase containing detergents and allowed to spin for 30 minutes at 4 °C. The mean drop size is 2 µm with a coefficient of variation less than 0.3.

In order to maximize the chances of selecting the correct gene, it is desirable to use a low ratio of templates:droplets to ensure that each drop only contains one gene. However, using an extremely low ratio will lead to massive waste of the transcription/translation kit and unacceptable cost. The percent of droplets with *x* genes can be predicted using Poisson's distribution:

$$f(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

where λ is the ratio between the number of templates and the number of droplets in the emulsion. Thus, if the ratio of templates:droplets was 1, approximately 37% of drops

would contain a single gene, 37% would contain no templates, and 26% would contain multiple genes.

Zheng and Roberts(23) calculated the theoretical enrichment of desirable genes from a random library. If we assume that every template in a droplet containing a nicking enzyme is selected, the number of nicking enzyme genes (*Nick*) is:

$$Nick = N\sum_{x=1} f(x,\lambda)\times\sum_{k=1}^{x}\left(\frac{x!}{k!(x-k)!}\times p^{k}\times(1-p)^{x-k}\times k\right)$$

where *N* is the total number of aqueous drops in the emulsion, *x* is the number of genes per droplet, and *p* is the percent of nicking enzyme genes in the library. The number of nonspecific genes that are selected (those that do not code for an active nicking enzyme but by chance were in the same droplet as one that does, leading to a false positive, *NonS*) is:

$$NonS = N\sum_{x=1} f(x,\lambda)\times\sum_{k=1}^{x}\frac{x!}{k!(x-k)!}\times p^{k}\times(1-p)^{x-k}\times(x-k)$$

This is also the maximum number of nicking genes that may be lost because they were, by chance, included in the same droplet as an active endonuclease, leading to a false negative. The theoretical enrichment is then:

$$E = \frac{r}{p/(1-p)}$$ where *r* is the ratio $$r = \frac{Nick - NonS}{NonS}$$

Each round should enrich the concentration of desirable genes by E. Depending on p, several rounds may be necessary to isolate a nicking enzyme gene. Selected genes will be amplified with PCR and can be sequenced or used for another round of selection.

In order to demonstrate that the expression of functional protein in an emulsified recombinant transcription/translation environment is feasible, we added the linear gene for I-PpoI with all E.Coli regulatory elements to an *in vitro* transcription/translation mix at concentrations calculated to provide final ratios of genes:droplets of 10:1, 1:1, and 1:10. A negative control lacking any I-PpoI gene was also included. After spinning for 30 minutes at 4 °C and incubating at 37 °C to allow the protein to be expressed, the reaction was stopped and the emulsion was broken using either oil washes or an ether wash. While the ether wash is much faster, some proteins are sensitive to denaturation by ether (53), so both methods were investigated. There was insufficient protein expressed to visualize on an SDS-PAGE gel, so an aliquot of each recovered aqueous phase was incubated with an oligo containing the recognition site for I-PpoI to test for restriction activity. The results are shown in Figure 4.10.

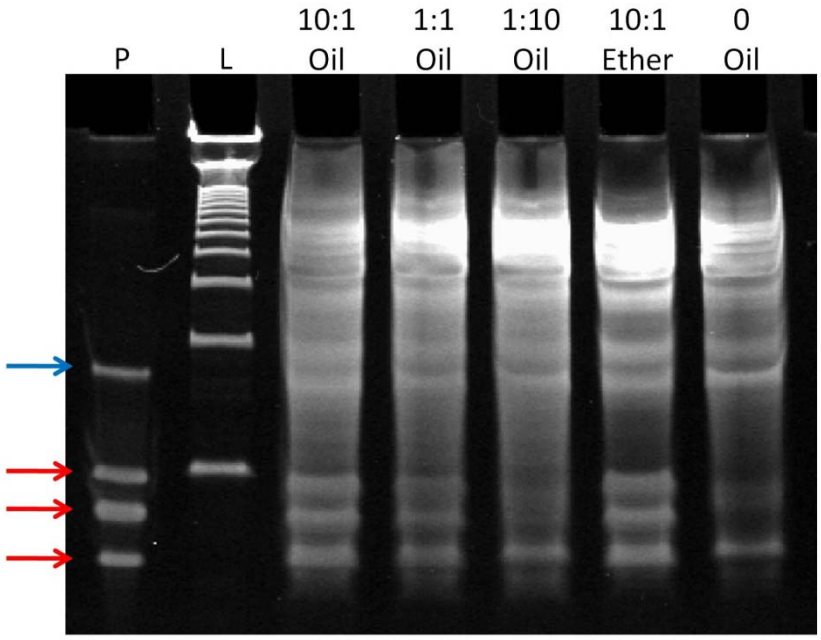

Figure 4.10. Activity of I-PpoI expressed in vitro in a water-in-oil emulsion. I-PpoI genes were added at various concentrations to a water-in-oil emulsion containing recombinant E. Coli transcription/translation machinery. They were then incubated with a 40 bp fragment (blue arrow) containing the I-PpoI recognition sequence. Active enzymes would digest the template into 24, 20, and 16 bp fragments (red arrows). Lane "P" indicates the positive control using enzyme obtained from Promega. Lane "L" is a 25 bp ladder. For each sample lane, the ratio of template:drops is indicated, along with whether the emulsion was broken with oil washes or a water-saturated ether wash. High MW smears are tRNAs, ribosomal RNA, and other nucleic acid mass contained in the transcription/translation mix.

It is evident that at the higher concentrations (10:1, 1:1) of gene templates, functional protein was expressed and it specifically cut its DNA substrate. Also, the protein was not denatured during the ether wash, suggesting that this is the method of choice for breaking the emulsion. However, the lowest ratio of genes (1:10) did not produce enough functional protein to detect using this activity assay. As calculated above, a 1:1 ratio of template:droplet is probably too high, as 26% of drops will

contain multiple genes and lead to false positives or negatives. On the other hand, 1:10 may be excessively low, as over 90% of all drops will be empty. A compromise in the neighborhood of 1:2 to 1:4 will likely produce the best results.

An emulsion-based platform, then, has been established that will allow us to express isolated mutant proteins with the ability to interact with and modify the DNA fragment containing their own gene and a degenerate recognition site. The next step is to complete the optimization of the biotin-streptavidin enrichment protocol so that the desirable labeled fragments can be reliably identified and amplified for subsequent rounds of selection.

4.5 References

1. Galburt, E. and Jurica, M. (2005) In Belfort, M. (ed.), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Vol. 16.

2. Lowery, R., Hung, L., Knoche, K. and Bandziulis, R. (1992) Properties of I-Ppo I: A rare-cutting intronencoded endonuclease. *Promega Notes Magazine*, **38**.

3. Wittmayer, P.K. and Raines, R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry*, **35**, 1076-1083.

4. Argast, G.M., Stephens, K.M., Emond, M.J. and Monnat, R.J., Jr. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol*, **280**, 345-353.

5. Joneja, A. and Huang, X. (2011) Linear nicking endonuclease-mediated strand-displacement DNA amplification. *Anal Biochem*, **414**, 58-69.

6. Goodrich-Blair, H., Scarlato, V., Gott, J.M., Xu, M.Q. and Shub, D.A. (1990) A self-splicing group I intron in the DNA polymerase gene of Bacillus subtilis bacteriophage SPO1. *Cell*, **63**, 417-424.

7. Goodrich-Blair, H. and Shub, D.A. (1996) Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell*, **84**, 211-221.

8. Shen, B.W., Landthaler, M., Shub, D.A. and Stoddard, B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.

9. Eckstein, F. (1985) Nucleoside phosphorothioates. *Annual review of biochemistry*, **54**, 367-402.

10. Potter, B.V. and Eckstein, F. (1984) Cleavage of phosphorothioate-substituted DNA by restriction endonucleases. *J Biol Chem*, **259**, 14243-14248.

11. Putney, S.D., Benkovic, S.J. and Schimmel, P.R. (1981) A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 7350-7354.

12. Shimayama, T., Nishikawa, F., Nishikawa, S. and Taira, K. (1993) Nuclease-resistant chimeric ribozymes containing deoxyribonucleotides and phosphorothioate linkages. *Nucleic Acids Res*, **21**, 2605-2611.

13. Walker, G.T. (1993) Empirical aspects of strand displacement amplification. *PCR Methods Appl.*, **3**, 1-6.

14. Walker, G.T., Fraiser, M.S., Schram, J.L., Little, M.C., Nadeau, J.G. and Malinowski, D.P. (1992) Strand displacement amplification--an isothermal, in vitro DNA amplification technique. *Nucleic Acids Res*, **20**, 1691-1696.

15. Walker, G.T., Little, M.C., Nadeau, J.G. and Shank, D.D. (1992) Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. *Proc. Natl. Acad. Sci. USA*, **89**, 392-396.

16. Crinelli, R., Bianchi, M., Gentilini, L., Palma, L. and Magnani, M. (2004) Locked nucleic acids (LNA): versatile tools for designing oligonucleotide decoys with high stability and affinity. *Curr Drug Targets*, **5**, 745-752.

17. Frieden, M., Hansen, H.F. and Koch, T. (2003) Nuclease stability of LNA oligonucleotides and LNA-DNA chimeras. *Nucleosides Nucleotides Nucleic Acids*, **22**, 1041-1043.

18. Jepsen, J.S., Sorensen, M.D. and Wengel, J. (2004) Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology. *Oligonucleotides*, **14**, 130-146.

19. Chen, J. and Hebert, P.D. (1998) Directed termination PCR: a one-step approach to mutation detection. *Nucleic Acids Res*, **26**, 1546-1547.

20. Sosnowski, R. (1995) A Chip-Based Genetic Detector for Rapid Identification of Individuals Justice, D. o.

21. Xu, J.Y., Xu, G.B. and Chen, S.L. (2009) A new method for SNP discovery. *BioTechniques*, **46**, 201-208.

22. Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N. and Yanagawa, H. (2004) In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Res*, **32**, e95.

23. Zheng, Y. and Roberts, R.J. (2007) Selection of restriction endonucleases using artificial cells. *Nucleic Acids Res*, **35**, e83.

24. Bellamy, S.R., Milsom, S.E., Scott, D.J., Daniels, L.E., Wilson, G.G. and Halford, S.E. (2005) Cleavage of individual DNA strands by the different subunits of the heterodimeric restriction endonuclease BbvCI. *Journal of Molecular Biology*, **348**, 641-653.

25. Flick, K.E., Jurica, M.S., Monnat, R.J., Jr. and Stoddard, B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96-101.

26. Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J., Jr. and Stoddard, B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nature structural biology*, **6**, 1096-1099.

27. Mannino, S.J., Jenkins, C.L. and Raines, R.T. (1999) Chemical mechanism of DNA cleavage by the homing endonuclease I-PpoI. *Biochemistry*, **38**, 16178-16186.

28. Eastberg, J.H., Eklund, J., Monnat, R., Jr. and Stoddard, B.L. (2007) Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry*, **46**, 7215-7225.

29. Eklund, J.L., Ulge, U.Y., Eastberg, J. and Monnat, R.J., Jr. (2007) Altered target site specificity variants of the I-PpoI His-Cys box homing endonuclease. *Nucleic acids research*, **35**, 5839-5850.

30. Green, N.M. (1975) Avidin. *Adv Protein Chem*, **29**, 85-133.

31. Chilkoti, A., Tan, P.H. and Stayton, P.S. (1995) Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 1754-1758.

32. Rybak, J.N., Scheurer, S.B., Neri, D. and Elia, G. (2004) Purification of biotinylated proteins on streptavidin resin: a protocol for quantitative elution. *Proteomics*, **4**, 2296-2299.

33. Jurinke, C., van den Boom, D., Collazo, V., Luchow, A., Jacob, A. and Koster, H. (1997) Recovery of nucleic acids from immobilized biotin-streptavidin complexes using ammonium hydroxide and applications in MALDI-TOF mass spectrometry. *Analytical chemistry*, **69**, 904-910.

34. Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J. and Uhlen, M. (2005) The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis*, **26**, 501-510.

35. Anderson, J.P., Angerer, B. and Loeb, L.A. (2005) Incorporation of reporter-labeled nucleotides by DNA polymerases. *BioTechniques*, **38**, 257-264.

36. Rosen, L.E., Morrison, H.A., Masri, S., Brown, M.J., Springstubb, B., Sussman, D., Stoddard, B.L. and Seligman, L.M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res*, **34**, 4791-4800.

37. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res*, **34**, e149.

38. Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P. and Paques, F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol*, **371**, 49-65.

39. Steuer, S., Pingoud, V., Pingoud, A. and Wende, W. (2004) Chimeras of the homing endonuclease PI-SceI and the homologous Candida tropicalis intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *Chembiochem*, **5**, 206-213.

40. Gimble, F.S., Moure, C.M. and Posey, K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol*, **334**, 993-1008.

41. Chen, Z., Wen, F., Sun, N. and Zhao, H. (2009) Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng Des Sel*, **22**, 249-256.

42. Stahl, F., Wende, W., Jeltsch, A. and Pingoud, A. (1996) Introduction of asymmetry in the naturally symmetric restriction endonuclease EcoRV to investigate intersubunit communication in the homodimeric protein. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6175-6180.

43. Smith, A.M., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J., Jr. and Stoddard, B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. USA*, **106**, 5099-5104.

44. Niu, Y., Tenney, K., Li, H. and Gimble, F.S. (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *Journal of Molecular Biology*, **382**, 188-202.

45. Chevalier, B.S. and Stoddard, B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res*, **29**, 3757-3774.

46. Ellison, E.L. and Vogt, V.M. (1993) Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol Cell Biol*, **13**, 7531-7539.

47. Griffiths, A.D. and Tawfik, D.S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends in biotechnology*, **24**, 395-402.

48. Joanicot, M. and Ajdari, A. (2005) Applied physics. Droplet control for microfluidics. *Science*, **309**, 887-888.

49. Nisisako, T., Torii, T. and Higuchi, T. (2002) Droplet formation in a microchannel network. *Lab Chip*, **2**, 24-26.

50. Xu, J.H., Luo, G.S., Li, S.W. and Chen, G.G. (2006) Shear force induced monodisperse droplet formation in a microfluidic device by controlling wetting properties. *Lab Chip*, **6**, 131-136.

51. Tan, Y.C. and Lee, A.P. (2005) Microfluidic separation of satellite droplets as the basis of a monodispersed micron and submicron emulsification system. *Lab Chip*, **5**, 1178-1183.

52. Umbanhowar, P.B., Prasad, V. and Weitz, D.A. (2000) Monodisperse emulsion generation via drop break off in a coflowing stream. *Langmuir*, **16**, 347-351.

53. Miller, O.J., Bernath, K., Agresti, J.J., Amitai, G., Kelly, B.T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D.S. and Griffiths, A.D. (2006) Directed evolution by in vitro compartmentalization. *Nature methods*, **3**, 561-570.

54. Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nature biotechnology*, **16**, 652-656.

Chapter 5: Conclusions and Future Directions

5.1 Device for Hydrodynamic Shearing of DNA

In this work, we developed an apparatus for the automated shearing of genomic DNA. Compared to the commercial instrument that operates using the same mechanism, our device exhibits similar performance characteristics at less than 20% of the cost. More importantly, it offers several significant improvements. First, the usage of a filter screen with thousands of parallel pores circumvents the clogging issues that have plagued single-orifice shearing devices. Second, the filter screens are cheap and disposable, making it much easier to decontaminate the system between samples. Third, it is more automated, allowing the instrument to be more high-throughput and eliminate sources of user error. In recent years, the read lengths of many DNA sequencing technologies has increased markedly, making long randomly sheared DNA fragments more desirable. Currently our instrument is one of the only methods for creating libraries of long (> 5kb) DNA fragments. We are collaborating with an industrial partner to bring this device to market. The most pressing need to make this a viable product is a more intuitive and GUI-based software suite for controlling the pump and valve.

Since small DNA fragments are still useful, we have made efforts to expand the range of sizes attainable with our system. To that end, we have replaced the filter screen housing with stainless steel internal unions with bore sizes of 150 μm or 250 μm. This allows us to increase the fluid flow rate at the constrictions (pores) and

decrease the size of the resulting fragments. We have demonstrated an ability to reach mean sizes of about 800 bp, however at these pressures the valve begins to fail and we experience sample loss via internal leaking. There is no reason to believe that a more robust pump and valve would not be able to further decrease the sizes achievable. However, as noted in the manuscript, the plot of fragment size vs. velocity is non-linear and approximately follows a power law. Thus, significant increases in flow rate will be required to realize incremental decreases in fragment size.

5.2 Strand Displacement Amplification with a Nicking Endonuclease

In this work, we demonstrated the linear amplification of DNA strands up to 5 kilonucleotides in length. A complex mixture of DNA fragments of disparate length can be amplified simultaneously. A critical feature of the method is that it does not exhibit bias based on length or sequence, therefore the original molar ratio of fragments in a mixture is preserved. The technique is capable of doing this because it uses a nicking enzyme with a relatively long 7 bp recognition sequence, a polymerase with extremely high fidelity and processivity, and single-stranded binding proteins (SSB) to stabilize the displaced DNA. We succeeded in amplifying a small phage genome that had been randomly fragmented using hydrodynamic shearing.

In its current form, the method is only useful when limited amplification is required. Since it is linear amplification, it cannot compete with other exponential methods such as PCR if a large quantity of DNA is desired from a small starting

sample size. A challenge for the future will be to adapt the current protocol to include primers that allow exponential amplification. This has already been demonstrated with traditional SDA. Another challenge will be to improve the hairpin adapters that we have developed. We have shown them to work with single templates, but not with a library of fragments. They may make the reaction simpler and more efficient by eliminating the potential for nonspecific interactions between strands of single-stranded DNA, and by freeing up SSB from newly synthesized strands for use with continued amplification.

5.3 Utilizing Homing Endonuclease I-PpoI in SDA reactions

In this work, we explored a number of approaches to using a well characterized homing endonuclease to perform Strand Displacement Amplification, culminating in an effort to engineer the enzyme to modify its recognition sequence and inactivate one of its cutting sites. Through the use of non-native nucleotides including phosphorothioate-modified nucleotides and locked nucleic acids (LNA), we managed to slow down the cutting of one strand of the DNA template so that SDA could proceed slowly. An attempt to identify sequences that can be nicked but not cleaved led to the development of protocols for separating biotinylated oligos from a library of unmodified oligos but did not yield the target sequence we were searching for. Turning the normally homodimeric enzyme into a tethered monomer with one cutting site inactivated led to an I-PpoI variant that was able to do SDA for a limited time before the recognition site was eventually fully cleaved. We pursued a strategy for

expressing and evaluating millions of mutant proteins in parallel. Towards this goal, we established protocols for the *in vitro* transcription and translation of functional I-PpoI in a water-in-oil emulsion such that each emulsion droplet contained just one copy of an I-PpoI gene.

We believe we have adequately explored the potential of non-native nucleotides and the mutant dimeric enzyme as concerns the use of I-PpoI in strand displacement amplification reactions. While we were able to solve many of the technical difficulties associated with the protein engineering project and identified amino acid residues and positions in the recognition sequence that are good candidates for mutation, further work will be required to identify the desirable mutants. The platform is established but we have not yet validated it with a positive control. In the long run, the solutions that did work (non-native nucleotides, tethered dimeric form of enzyme) will never be as robust as an engineered version of I-PpoI with the desired properties. It is worth spending time, money, and effort to continue to develop a mutant nicking I-PpoI because it is likely to have many applications in genomic engineering and DNA amplification via SDA.