INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
PROTEIN ENGINEERING: 
PRELIMINARY CHARACTERIZATION AND FURTHER 
DEVELOPMENT OF A NEW RESTRICTION ENDONUCLEASE, 
SPLASE 

DISSERENATION

Presented in Partial Fulfillment of the Requirements 
for the Degree Doctor of Philosophy in the Graduate School of 
The Ohio State University 

By 

Christopher J. Schaeffer, B. S. 

*****

The Ohio State University 
1998 

Dissertation Committee:
Ming-Daw Tsai, Advisor 
James A. Cowan 
Charles Russell Hille 
Pravin Kaumaya 

Approved by 

[Signature]

Advisor 
The Ohio State Biochemistry Program
ABSTRACT

The field of protein engineering has grown significantly over the past decade. The solution of many new protein structures combined with extensive functional analysis has lead to significant advances in our knowledge of how enzymes are constructed to perform their unique functions with near perfect efficiency. The next step in this evolutionary process is the use of this knowledge to construct proteins and enzymes with new functions.

The goal of this project was to generate a new enzyme which would perform two functions. We proposed that if we could target a non-specific nuclease to the LTR of HIV-1 then this might be useful as an anti-HIV gene therapy. This enzyme might also function as a new rare-cutter restriction enzyme. Specifically, one which would be capable of cutting at any DNA sequence desired. The specificity of each enzyme could be determined prior to gene fusion of the DNA recognition domain and the nuclease domain. The result of this design process was the construction of Splase. Splase, or sometimes known as HSplase due to the presence of a histidine tag for purification, is comprised of the zinc finger DNA recognition elements from the Sp1 transcription factor and the nuclease domain from FokI.

My research focuses on the improvement of this unique enzyme, Splase, with the hope of generating an enzyme which could fulfill the intended functions and find its way into the biotechnology marketplace. There were several obstacles to this dream. The initial problem was that the enzyme was very unstable upon purification. This would
eventually be solved by expressing the protein as a gene fusion with Glutathione-S-Transferase. The majority of my work was dedicated to achieving this result. The second problem was that the substrate was not completely turned over. Research in several other laboratories had shown that zinc fingers have very long dissociation half-lives. This dissociation constant appeared to be modulated by the off rate of the peptide from the DNA target. The protein structure revealed the presence of two types of interaction between the protein and the DNA. I have termed these affinity and specificity residues. The affinity residues are responsible for general peptide DNA interactions and may be used to modulate the overall affinity of the peptide for the DNA target thus allowing us to modulate the binding constant via the off rate. If successful this would lead to an enzyme with an improved off rate and thus an improved turnover of substrate to product.

However, mutagenesis experiments at the positions identified as possible affinity residues resulted in significant negative perturbations to the activity of the enzyme. We concluded that before this approach could be utilized in this situation we must extensively study the structural and functional properties of both the zinc finger as well as the FokI restriction enzyme before any further engineering is conducted. At the conclusion of this work the idea behind this project is very much alive, however, the experimental design toward improving the functional properties of the enzyme must take another path before Splase can become a reality.
I Would Like To Dedicate This Dissertation To My Parents, James And Nancy,
My Wife, Heather,
And In Memory Of My Uncle
Robert A. Schaeffer
For Without These People This Would Not Be Possible
ACKNOWLEDGEMENTS

I would like to thank everyone who has helped contribute to the development of this thesis.

I would like to thank my research advisor, Professor Ming-Daw Tsai, for his encouragement throughout the development and execution of this project.

I am thankful for the help Dr. Charles E. Cottrell gave me in teaching me the basic operation of the 500 MHz NMR. I would also like to acknowledge his patience when answering my questions. While I am thanking people for their help with NMR, I would also like to thank Alexander Kravchuk for teaching me the basics of operation on the 600 MHz NMR.

I would like to thank everyone in the Tsai lab for their advice during the development of this project and for their support during the low points of the project.

I would like to thank Mrs. Woodard, my seventh and eighth grade science teacher for stimulating the inquisitive scientist within me.

I would like to thank my father for the stimulating scientific discussions throughout my life which have created a curiosity with how nature works. I would also like to thank him for his guidance during this process, this was invaluable.

Finally, I would like to thank my wife, Heather, who without her patience and friendship this would not be possible.
VITA

January 5, 1969 ..................................................... Born- Rochester, New York
May 19, 1991 ......................................................... B. S. in Chemistry
John Carroll University

1992- present ......................................................... Graduate Research Associate
The Ohio State University
Columbus, Ohio

Publications


Field of Study: Biochemistry
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>General Background</td>
<td>1</td>
</tr>
<tr>
<td>Design of A New Restriction Endonuclease Designed</td>
<td>5</td>
</tr>
<tr>
<td>To Cleave The Long Terminal Repeat of HIV</td>
<td></td>
</tr>
<tr>
<td>Characterization</td>
<td>17</td>
</tr>
<tr>
<td>Evaluation</td>
<td>18</td>
</tr>
<tr>
<td>Experimental Strategy</td>
<td>21</td>
</tr>
<tr>
<td>Evaluation of Initial Situation</td>
<td>26</td>
</tr>
<tr>
<td>Evaluation of the Strategy</td>
<td>28</td>
</tr>
<tr>
<td>2. MATERIALS AND METHODS</td>
<td>39</td>
</tr>
<tr>
<td>Materials</td>
<td>39</td>
</tr>
<tr>
<td>Methods</td>
<td>40</td>
</tr>
<tr>
<td>PCR Cloning</td>
<td>40</td>
</tr>
<tr>
<td>Isolation of Genomic DNA from Flavobacterium okeanokoites(Fokl)</td>
<td>42</td>
</tr>
<tr>
<td>PCR Cloning of Fragments of Fokl Restriction</td>
<td></td>
</tr>
<tr>
<td>Enzyme</td>
<td>43</td>
</tr>
<tr>
<td>PCR Construction of pETSp1</td>
<td>45</td>
</tr>
<tr>
<td>DNA Purification</td>
<td>45</td>
</tr>
<tr>
<td>Preparation of Competent Cells</td>
<td>46</td>
</tr>
<tr>
<td>Transformation</td>
<td>47</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Site-Directed Mutagenesis</td>
<td>48</td>
</tr>
<tr>
<td>DNA Sequencing</td>
<td>49</td>
</tr>
<tr>
<td>Subcloning</td>
<td>52</td>
</tr>
<tr>
<td>Purification of GST-HSplase and GST-Splase80</td>
<td>53</td>
</tr>
<tr>
<td>Growth and Harvest</td>
<td></td>
</tr>
<tr>
<td>Preparation and Equilibration of Glutathione Agarose Resin</td>
<td>54</td>
</tr>
<tr>
<td>Cell Lysis and Glutathione Affinity Chromatography</td>
<td>54</td>
</tr>
<tr>
<td>Purification of HSplase</td>
<td>56</td>
</tr>
<tr>
<td>Growth and Harvest</td>
<td>56</td>
</tr>
<tr>
<td>Preparation and Equilibration of His-bind Resin</td>
<td>57</td>
</tr>
<tr>
<td>Cell Lysis and Ni²⁺-Affinity Chromatography</td>
<td>57</td>
</tr>
<tr>
<td>DNA Mobility Shift Assay</td>
<td>58</td>
</tr>
<tr>
<td>Assay of HSplase Activity</td>
<td>59</td>
</tr>
<tr>
<td>Cleavage Site Determination</td>
<td>61</td>
</tr>
<tr>
<td>Reactions with HSplase</td>
<td>61</td>
</tr>
<tr>
<td>Maxam and Gilbert DNA Sequencing</td>
<td>62</td>
</tr>
<tr>
<td>Purification of FokI</td>
<td>64</td>
</tr>
<tr>
<td>Growth and Harvest</td>
<td>64</td>
</tr>
<tr>
<td>Purification</td>
<td>65</td>
</tr>
<tr>
<td>Activity of FokI</td>
<td>67</td>
</tr>
<tr>
<td>Purification of Sp1 zinc finger</td>
<td>67</td>
</tr>
<tr>
<td>Growth and Harvest</td>
<td>67</td>
</tr>
<tr>
<td>Cell Lysis and Purification</td>
<td>68</td>
</tr>
<tr>
<td>NMR</td>
<td>70</td>
</tr>
<tr>
<td>Sample preparation</td>
<td>70</td>
</tr>
<tr>
<td>One-Dimensional NMR Methods</td>
<td>70</td>
</tr>
<tr>
<td>Purification of GST-HSplase</td>
<td>70</td>
</tr>
<tr>
<td>Growth and Harvest</td>
<td></td>
</tr>
<tr>
<td>3. Attempt to solve the Solubility Problem</td>
<td>72</td>
</tr>
<tr>
<td>Introduction</td>
<td>72</td>
</tr>
<tr>
<td>Creation of pGEX-Splase80</td>
<td>73</td>
</tr>
<tr>
<td>Discussion</td>
<td>76</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Sequence Specificity table for Zif268 Specificity Residues</td>
<td>37</td>
</tr>
<tr>
<td>2. Table of Mutants Constructed</td>
<td>120</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Schematic of Ubx-F&lt;sub&gt;N&lt;/sub&gt; Chimera.</td>
</tr>
<tr>
<td>2.</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Life Cycle of HIV.</td>
</tr>
<tr>
<td>3.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Schematic of Structure of Zif268 Specificity Residues.</td>
</tr>
<tr>
<td>4.</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Schematic of Specificity Matrix.</td>
</tr>
<tr>
<td>5.</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Schematic of Structure of FokI.</td>
</tr>
<tr>
<td>6.</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>The DNA Sequence of Splase.</td>
</tr>
<tr>
<td>7.</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>The Amino acid Sequence of Splase.</td>
</tr>
<tr>
<td>8.</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Digestion of Linearized DNA by Splase.</td>
</tr>
<tr>
<td>9.</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Schematic For Approach To Designing New Restriction Enzymes With New Specificities.</td>
</tr>
<tr>
<td>10.</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Schematic Showing Hypotheses A and B.</td>
</tr>
<tr>
<td>11.</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Sequence Homology of Zinc fingers.</td>
</tr>
<tr>
<td>12.</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Stereoview of Zif268 Showing Affinity Residues.</td>
</tr>
<tr>
<td>13.</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Primers Used to Amplify Nuclease Domains.</td>
</tr>
<tr>
<td>14.</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Primers Used for Mutagenesis.</td>
</tr>
<tr>
<td>15.</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Schematic of HSplase Activity Assay.</td>
</tr>
<tr>
<td>16.</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Schematic of Cleavage Site Determination Assay.</td>
</tr>
</tbody>
</table>
37. Expression of GST-HSsplase ................................................... 112
38. Purification of GST-HSsplase ................................................ 113
39. Activity Assay of GST-HSsplase ............................................ 114
40. Thrombin Cleavage of GST-HSsplase .................................. 115
41. Activity Assay of WT GST-HSsplase ................................... 122
42. Activity Assay of GST-HSsplase K52A .................................. 123
43. Activity Assay of GST-HSsplase K52D .................................. 124
44. Activity Assay of GST-HSsplase K52Q .................................. 125
45. Activity Assay of GST-HSsplase K52S .................................. 126
46. Activity Assay of GST-HSsplase K52T .................................. 127
47. Activity Assay of GST-HSsplase K80A .................................. 128
48. Activity Assay of GST-HSsplase K80D .................................. 129
49. Activity Assay of GST-HSsplase S85T .................................. 130
50. Activity Assay of GST-HSsplase S85D .................................. 131
51. Expression of Spl Zinc Finger .............................................. 135
52. Purification of Zinc Finger/Oxidation Problem ..................... 137
53. 1D-Proton NMR Spectrum of Zif268 finger 1 ....................... 139
54. 1D-Proton NMR Spectrum of Spl, Aliphatic Region .......... 141
55. 1D-Proton NMR Spectrum of Spl, Amide Region .......... 143
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Deoxyribo Nucleic Acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene Diamine Tetra-acetic acid</td>
</tr>
<tr>
<td>AT</td>
<td>Adanine and Thymine</td>
</tr>
<tr>
<td>GC</td>
<td>Guanine and Cytosine</td>
</tr>
<tr>
<td>Fok</td>
<td>Flavobacterium okanokoites</td>
</tr>
<tr>
<td>LTR</td>
<td>Long Terminal Repeat</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>AIDS</td>
<td>Aquired Immune defficiency Syndrome</td>
</tr>
<tr>
<td>AZT</td>
<td>Azidovuidine Triphosphate</td>
</tr>
<tr>
<td>ddITP</td>
<td>dideoxy inosine triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>dideoxy Cytosine triphosphate</td>
</tr>
<tr>
<td>CDC</td>
<td>Center for Disease Control</td>
</tr>
<tr>
<td>NFκB</td>
<td>Nuclear Factor Kappa B</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphoshate</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>$k_{off}$</td>
<td>Off rate</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michealis Constant</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasmon Resonance</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>dNTP's</td>
<td>Deoxy Nucleotide triphosphates</td>
</tr>
<tr>
<td>MOPS</td>
<td>4-Morpholinopropane sulphonic acid</td>
</tr>
<tr>
<td>HEPES</td>
<td>2-(N-Morpholino) ethane sulfonic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double stranded DNA</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-thio Galactose</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-Transferase</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>MWC</td>
<td>Molecular Weight Cut-off</td>
</tr>
<tr>
<td>EB</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>RT</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra Violet Radiation</td>
</tr>
<tr>
<td>Ci</td>
<td>Currie</td>
</tr>
<tr>
<td>PNK</td>
<td>Polynucleotide Kinase</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris- borate EDTA buffer</td>
</tr>
<tr>
<td>HSplase</td>
<td>Splase with a 6 his tag</td>
</tr>
<tr>
<td>GSTHSplase</td>
<td>Splase with GST and 6 his tag</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>DEAE</td>
<td>Diethyl amino ethane</td>
</tr>
<tr>
<td>HPLC</td>
<td>Highperformance Liquid Chromatography</td>
</tr>
<tr>
<td>TFA</td>
<td>Tirfluoroacetic acid</td>
</tr>
<tr>
<td>K</td>
<td>Lysine</td>
</tr>
<tr>
<td>D</td>
<td>Aspartate</td>
</tr>
<tr>
<td>T</td>
<td>Threonine</td>
</tr>
<tr>
<td>Q</td>
<td>Glutamine</td>
</tr>
<tr>
<td>N</td>
<td>Asparagine</td>
</tr>
<tr>
<td>R</td>
<td>Arginine</td>
</tr>
<tr>
<td>A</td>
<td>Alanine</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered Saline</td>
</tr>
<tr>
<td>AMP</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>LB</td>
<td>Lauria Broth</td>
</tr>
<tr>
<td>P</td>
<td>Proline</td>
</tr>
<tr>
<td>Kb</td>
<td>Killobasepairs</td>
</tr>
<tr>
<td>C-terminus</td>
<td>Carboxy terminus</td>
</tr>
<tr>
<td>N-terminus</td>
<td>Amino terminus</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenyl Methyl Sulfonyl Fluride</td>
</tr>
<tr>
<td>LC-ESI-MS</td>
<td>Liquid Chromatography-Electrospray ionization-Mass spectrometry</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix assisted Laser Desorption Ionization -Mass spectrometry</td>
</tr>
<tr>
<td>OmpT</td>
<td>Outer membrane Protein T</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>A&lt;sub&gt;NNN&lt;/sub&gt;</td>
<td>Absorbance at wavelength NNN</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (Hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>BIA</td>
<td>Biological Interaction Assay</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

General Background

The discovery of restriction and modification enzymes has revolutionized the fields of biochemistry, molecular biology, and medicine. In 1978, Hamilton Smith, Werner Arber, and Daniel Nathans were awarded the Nobel Prize in Medicine for the discovery of these systems. Since their discovery, many new restriction and modification enzymes have been identified. However, we still lack a complete library of restriction endonucleases with specificity for any desired sequence. The next revolution in this area will be driven by the design of new enzymes. One day the protein engineer will design the specificity of these enzymes. When this is successfully completed, we will no longer rely on nature to obtain our tools for DNA or RNA manipulation; we will construct them ourselves. Although, restriction endonucleases have become the main tools of DNA manipulation, they have several limitations, including a small recognition sites, restricted sequence specificity, and limited availability.

Due to these limitations, the search for a complete library of natural restriction enzymes has not yet been successful. Therefore, in order to fill the gaps left by nature and address the limitations of naturally occurring restriction endonucleases, several research groups have turned to the design of artificial nucleases. The basic approach is to engineer novel nucleolytic agents that can cleave DNA or RNA with a similar accuracy and
precision as natural restriction enzymes. Several creative strategies have been utilized to solve this problem. Novel nucleolytic agents have been constructed by fusing a chemical moiety, such as ferrous-EDTA, to a DNA or RNA recognition element (1). Other strategies involve genetically fusing a DNA recognition element to a naturally occurring nuclease (2). A modular approach has been taken in each of the examples described above. This approach allows the engineer to generate nucleases with a variety of specificities simply by altering the DNA recognition element.

The modular approach, however, is the result of the evolution of the design of artificial nucleases. Nuclease design has progressed from small chemically synthesized nucleases to very complex chimeras composed of individual protein domains.

The method of synthesis has progressed from total chemical synthesis to chemical modification and ultimately to genetic modification. The more advanced methods of synthesis include techniques such as semisynthetic protein modification, site-directed mutagenesis and gene fusion. The structure of the artificial nucleases has evolved from the small molecule to the more complex modular molecule, such as the oligonucleotide-protein conjugate and the protein chimera. Specificity has also evolved from the nonspecific chemical cleavage agents to the highly specific DNA binding domains of transcription factors. Transcription factors utilize DNA recognition motifs such as the helix-turn-helix, leucine zipper and the zinc finger. Several of these motifs have been genetically engineered to construct proteins with new specificities. By genetically combining the highly specific, yet modular DNA-binding domains of transcription factors with the nucleolytic domains of naturally occurring nucleases the most advanced modular approach to the construction of endonucleases is achieved.

Just as the discovery of natural restriction endonucleases played a significant role in the development of several areas of science, protein and enzyme design may be the next
major revolution in biochemistry, molecular biology and medicine. The development of
catalytic agents that can mimic nature in their catalytic efficiency and specificity will allow
us to increase our repertoire of synthetic reactions. These new synthetic catalytic agents
may allow us to create new enantiomerically pure drugs or precursors to drugs or allow us
to clean up our environment from toxic waste. Therefore, the design of novel catalysts is
one of the very difficult problems which biochemists and protein engineers face today.
Through extensive structural and functional studies of many proteins we are beginning to
understand how proteins and enzymes work. The next step is to use this knowledge to
create a protein or enzyme which performs a function not yet identified in nature and then
optimize the function of this protein for practical utilization in research, medicine and
other scientifically related areas.

In an attempt to create site specific nucleases from nonspecific chemical cleavage
agents, several unique strategies have been used. Initial attempts were made to increase
specificity by attaching the nonspecific nucleolytic agents to DNA intercalators and
antibiotics known to bind DNA with limited sequence specificity. At this point, sequence
specificity was limited to AT or GC rich regions. An example of this type of artificial
nuclease is the Distamicin- Fe^{2+}(EDTA) conjugate (1). Distamycin is an antibiotic that
binds B form DNA predominantly within the minor groove of AT-rich regions.
Treatment of DNA with this agent resulted in the production of single stranded nicks at
AT rich regions.

This approach did advance the development of new restriction enzymes by
enhancing the sequence specificity over previous methods. The drawbacks of this
approach are that substantial background strand scission is observed and site specific
cleavage is hard to control. At this point in the design of artificial nucleases we see the
initial steps toward site specific nucleases, however, their activity is yet uncomparable with
restriction enzymes.
This initial advancement of the field of protein engineering fostered a curiosity in the synthesis of nucleolytic agents which could be targeted to specific DNA sequences by utilizing the naturally occurring DNA recognition elements found in transcription factors and other DNA binding proteins.

DNA-binding proteins recognize sequences as large as 10 to 16 bp. This decreases the frequency of phosphodiester hydrolysis to once in $4 \times 10^9$ bp in a completely random sequence. This decrease in frequency of cleavage provides the molecular biologist with a tool to cleave genomic DNA into small fragments. These fragments can be used later for genome mapping experiments. Researchers soon realized that by altering the specificity of the DNA recognition element a nuclease could be generated for any known DNA sequence, as long as a protein could be found to recognize it. However, despite its improvement in specificity, the chemical conjugate nuclease still has one major limitation, nonspecific DNA strand scission. Therefore, compared to the naturally occurring restriction enzymes, specificity has improved however strand scission is not limited to a single phosphodiester bond.

The next logical step in the progression toward the development of a artificial nuclease is the chemical fusion of a DNA recognition element to a naturally occurring nuclease. A nuclease was designed in which an oligonucleotide was linked to a naturally occurring nonspecific nuclease, staphylococcal nuclease (2) or by linking the repressor to staphylococcal nuclease via a aliphatic linker (3). The $\lambda$-repressor-staphylococcal nuclease hybrid demonstrated increased specificity due to the 17 bp recognition sequence of $\lambda$-repressor and showed more efficient cleavage of both linear and supercoiled duplex DNA than previous cleavage agents. However the major drawbacks of this system were incomplete cleavage and limited catalytic turnover. The catalytic turnover of the nucleolytic agent was sacrificed for the high selectivity of the DNA recognition domain.
They found that the protein:DNA complex was very stable which limited the off rate of the chimeric enzyme. This resulted in little or no catalytic turnover.

Kim and Chandrasegaran (1994) synthesized a novel chimeric endonuclease by linking the Drosophila Ultrabithorax homeodomain to the cleavage domain of FokI (4) (Figure 1). This enzyme can be expressed in E. coli and purified to homogeneity. This approach is unique because the DNA recognition domain of a naturally occurring restriction endonuclease is replaced by the DNA recognition element of a naturally occurring transcription factor. Although the fusion of the Ultrabithorax homeodomain to the FokI nuclease domain created a new sequence-specific endonuclease, non-specific cleavage by the nuclease domain of FokI was observed.

It was at this point in the design of new artificial restriction enzymes that we decided to engage in the design of an enzyme which would be able to bind and cleave the Long Terminal Repeat (LTR) of HIV. The section below describes the design strategy that was invoked, a summary of the important results and a critical evaluation of our initial progress.

**Design of A New Restriction Endonuclease Designed To Cleave The Long Terminal Repeat of HIV**

Presently, Acquired Immune Deficiency Syndrome is still classified at pandemic levels. There are 16,000 new cases of AIDS reported daily. With the use of the drug cocktail (AZT, ddITP or ddCTP and protease inhibitors) we have managed to reduce the viral load within many of the current patients. This has allowed for a reduction in the number of deaths due to AIDS. The number of new infections has also decreased due to extensive education programs in the media and in schools. However, we have just learned
A schematic of the first *FokI* nuclease/DNA binding protein fusion produced in the lab of Srinavasan Chandrasegaran at Johns Hopkins University. This figure was taken from Kim et al. (4). The tubes represent the alpha helices of the DNA binding protein, Ubx, the large circle represents the nuclease domain and the line in between represents the linker.

**Figure 1.** Schematic of Ubx-\(F_N\) Chimera
that patients who have been treated with the drug cocktail still harbor the latent form of the virus within their T-Cells (5). Therefore, we must still attempt to find ways that will allow us to completely clear the body of any remnants of this virus. The approach described here, will give new cells the ability to battle future infection we will allow the body to defeat this disease by itself thus eliminating the need for indefinite use of expensive drug therapy. However, at this point the protein we describe is not fully functional and must undergo drastic re-engineering to fulfill this goal.

The life cycle of HIV (Figure 2) is a very simple process which begins with the binding of the virion to the cell surface via the CD4:T-Cell receptor complex (CD4:TCR). The virion is then endocytosed and fuses with the vesicle membrane thus injecting the capsid into the cellular cytoplasm. This capsid is a caged structure which protects the RNA encoded genome of the virus from the degradation by cellular RNase molecules. Within this caged structure the RNA is reverse transcribed into an RNA:DNA hybrid, the RNA is removed by the RNase H domain of Reverse Transcriptase and concurrently the second DNA strand is produced. The provirus is then transported across the nuclear membrane where integrase facilitates the successful integration of the genome into the host chromosome. Viral proteins then trigger the host cell to begin transcription of the viral genes and the production of new virus begins. We have chosen to target the genome after reverse transcription and prior to translocation to the nucleus. At this point the virus is believed to uncoat, thus exposing the newly constructed DNA molecule to cellular site specific nucleases which might degrade the DNA prior to translocation.

The genome of HIV contains the promoter and enhancer regions for RNA transcription within the Long Terminal Repeat (LTR). The LTR contains binding sites for the required transcription factors such as Sp1, NFκB and AP-1. Therefore, we decided
Threatening life. Each step in HIV's life cycle is a potential target for interruption by therapeutic drugs.

This figure was taken from Cohen (84).

**Figure 2.** Life cycle of HIV.
that if we could target a nuclease to one of these positions we could possibly destroy the
ability of the genome to translocate and integrate into the host genome. We chose the Spl
zinc finger as our DNA recognition element due to the presence of three consensus
binding sites located within the LTR. This would increase the effective concentration of
substrate, thus increasing our chances for success.

The zinc finger DNA recognition motif was first identified in 1986 by Aaron Klug
as a structural element found in transcription factor IIIA [6]. This family of
zinc finger containing proteins has grown to over 200 members and consists of 1300 zinc
fingers. This group of transcription factors is presently the largest family of DNA binding
proteins known. Jeremy Berg first identified the theoretical basis of this structural motif in
1988 [7]. The important structural elements are an antiparallel β-sheet followed by an α-
helix. These secondary structural elements are held in three-dimensional space via the
tetrahedral coordination of a zinc atom. The zinc atom is coordinated by two sulfur atoms
from cysteines located near the bend of the two β-sheets and two nitrogen atoms from
histidines located in the α-helix. This structure was confirmed when the structure of a
single zinc finger, Xfin, was solved by NMR [13].

In 1991 the structure of a three-zinc-finger motif of Zif268 complexed with DNA
was published by Pavletich and Pabo [8] (Figure 3). This structure identified the critical
residues responsible for sequence-specific binding. Residues in the -1, 2, 3 and 6 positions
on the α-helix were identified as critical players in this role. Mutagenesis of residues in
these positions confirmed that these are the residues responsible for sequence specificity
[9]. Also, this structure showed that each zinc finger recognizes a single DNA triplet.
Subsequently, X-ray [10-12] and NMR [13-15] have solved many other zinc finger
structures.
A schematic of the structure of Zif268 depicting the residues which interact with the DNA bases and define the sequence specificity of the protein. This figure was taken from Pavletich and Pabo (8).

Figure 3. Schematic of Structure of Zif268 Specificity Residues
In order to identify the rules, which govern DNA sequence identification by the zinc finger amino acid residues, several groups have used phage display libraries or site-directed mutagenesis [16-24]. These studies have generated a table of sequences and the peptide sequences that bind them (Figure 4). On the basis of these “specificity rules” which have resulted we hope to further broaden our method by constructing new restriction enzymes with designed specificity.

One example of a transcription factor with a zinc finger DNA recognition element is Spl. Spl is a 600 amino acid protein which binds to the 5'-GGGGCGGGGC-3' consensus recognition sequence [25] utilizing a three-zinc finger motif. Kriwacki et al. have applied the minimalist approach to determine the smallest fragment of this protein which still retains DNA recognition with a similar affinity. They have demonstrated that the 92 amino acids, which comprise the zinc finger motif, will bind the target DNA with an affinity similar to that of the native Spl [26,27].

Now that we have the DNA recognition domain for our artificial nuclease we need a nuclease domain. Currently, there are three types of restriction enzymes. Type I restriction enzymes [28-38] cleave DNA more than 1000 bp away from the recognition site and are ATP dependant. Type II restriction enzymes (BamHI, EcoRI, etc.) [39-43] cleave DNA at the recognition site and are commonly used in molecular biology research, but do not require ATP. Type III enzymes [44-56] are also ATP dependent and recognize an asymmetric sequence and cleave DNA 25-26 base pairs (bp) downstream. These enzymes are rare and are not currently used in molecular biology.

The type IIS restriction enzymes are a subset of type II. These enzymes bind a non-palindromic DNA sequence as a monomer [57] and cleave DNA a short distance downstream of the recognition sequence. FokI is one example of a type IIS restriction
A schematic depicting the protein primary sequence of the DNA recognition alpha helix (left) and the DNA recognition sequence which is generated (right). This is a table of DNA specificity rules. This figure was taken from Choo and Klug (16).

**Figure 4.** Schematic of Specificity Matrix
endonuclease. The enzyme is isolated from Flavobacterium okanokoites [58] and is the most extensively studied type IIS restriction enzyme. The structural and functional properties of FokI are distinctly different from those of type II restriction enzymes. Type II enzymes are homodimers that cut each strand of DNA symmetrically within the DNA recognition sequence. In contrast, FokI is believed to bind DNA as a monomer [59] and cuts the two strands of DNA asymmetrically. FokI is a classic example of a class IIS restriction endonuclease. FokI is a monomer of approximately 66 kD and binds an asymmetric sequence 5'-GGATC-3'. Cleavage occurs on the DNA substrate 9 bp downstream on the upper strand and 13 bp on the lower strand (Figure 5). The work of Li, Wu and Chandrasegran (1992) showed that FokI is composed of two domains, a DNA binding domain and a nuclease domain by digesting the protein DNA complex with trypsin (60). Subsequent to digestion the protein fragments were purified and their activity assayed. They found that the 41 kD fragment bound DNA and the 28 kD fragment retained nuclease activity, which was monospecific. This supported the theory that this protein consists of two separate domains. The separation of the functional domains was instrumental to the conceptualization and subsequent construction of the new chimeric restriction enzymes Ubx-FN (4) and the zinc-finger-FokI chimeras (61,62).

Utilizing these two domains we have genetically engineered a novel chimeric endonuclease called Splase. Splase was genetically constructed from the sequence encoding the 92 amino acids derived from the zinc finger motif of SpI and fused to sequences encoding the C-terminal cleavage domain of FokI, as shown in Figures 6 and 7.
This is a schematic showing the two domain structure of the FokI enzyme. It also depicts the DNA recognition sequence and the cleavage pattern observed with the enzyme (shown by the presence of the arrows)

**Figure 5** Schematic of Structure of FokI
ATG GCT AAA AAG AAA CAG CAT ATT TGC CAC ATC CAA GGC TGT GGG
AAA GTG TAT GGC AAG ACC TCT CAC CTG CGG GCA CAC TTG CGC TGG
CAT ACA GGC GAG AGG CCA TTT ATG TGT ACC TGG TCA TAC TGT GGG
AAA CGC TTC ACA CGT TCG GAT GAG CTA CAGA GGC ACA AAC GTA CAC
ACA CAG GTG AGA AGA AAT TTG CCT GCC CTG AGT GTC CTA AGC GCT
TCA TGA GGA GTG ACC ACC TGT CAA AAC ATA TCA AGA CCC ACC AGA
ATA AGA AGG TAC CTA ATC GTG GTG TGA CTA AGC AAC TAG TCA AAA
GTG AAC TGG AGG AGA AGA AAT CTG AAC TTC GTC ATA AAT TGA AAT ATG
TGC CTC ATG AAT ATA TTG AAT TAA TTG AAA TTG CCA GAA ATT CCA CTC
AGG ATA GAA TTC TTG AAA TGA AGG TAA TGG AAT TTT TTG TGA AAG TTT
ATG GAT ATA GAG GTA AAC ATT TGG GTG GAT CAA GGA AAC CCG ACG
GAG CAA TTT ATA CTG TCG GAT CTC CTA TTG ATT ACG GTG TGA TCG TGG
ATA CTA AAG CTT ATA GCG GAG GTT ATA ATC TGC CAA TTG GCC AAG CAG
ATG AAA TGC AAC GAT ATG TCG AAG AAA ATC AAA CAC GAA ACA AAC
ATA TCA ACC CTA ATG AAT GGT GGA AAG TCT ATC CAT CTT CTG TAA CGG
AAT TTA AGT TTT TAT TTG TGA GTG GTC ACT TTA AAG GAA ACT ACA AAG
CTC AGC TTA CAC GAT TAA ATC ATA TCA ATG GTA ATG GAG CTG TCC
TTA GTG TAG AAG AGC TTT TAA TTG GTG GAG AAA TGA TTA AAG CCG
GCA CAT TAA CCT TAG AGG AAG TGA GAC GGA AAT TTA ATA ACG GCG
AGA TAA ACT TTT AAG GAT CCT GCA GGC

Figure 6. The DNA sequence of Splase.
MAKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFM
CTWSYCGKRFTSRSDSDELRHKRTHTGEKKFACPECPKRFM
RSDHLSKHITKHQNKKV\textbf{P}NRGVTQQLVKSLEEEKKSEL
HKLKVYPHEYIELIEIARNSTQDRILEMKEFMKVVGY
RGKHLGSGRKPDDAIYTVGSPIDYGVIVDKAYSGGYNL
PIGQADEMQRYVEENQTRNKHINNPENWWKVYPSSVTEFK
FLFVSHFKGNKYKAQLTRLNHTNCGAVLSEEILGGE
MIKAGTLTLEEVRRKFNNGEINF

The P in bold represents the junction point between the two domains.

\textbf{Figure 7.} The amino acid sequence of Splase.
Characterization

Splase was initially expressed at 37 °C, however this resulted in the formation of inclusion bodies. The inclusion bodies were isolated, solubilized and refolded under denaturing conditions. Splase was refolded and purified to homogeneity. This protein was shown to bind plasmid DNA however no nuclease activity was observed. At this point we concluded that we needed to find a way to express this protein in a soluble form. We broke off into two groups working on expression of this as a fusion protein with GST and the other adjusting the conditions for expression.

Expression experiments conducted at lower temperatures showed that when the temperature was lowered to 25 °C during induction, a small fraction of the protein was expressed as a soluble protein. Ni²⁺-chelating chromatography was implemented as the primary means of partially purifying this protein. Purification using this method produced a typical yield of 5-10 mg from a 2.5 L culture and the protein was between 40 and 60 percent pure.

Substrates containing between one and six consensus Sp1 binding sites were constructed by cloning a DNA fragment containing the consensus binding site into the multiple cloning site of pUC19. These substrates were labeled pUCSp1 through pUC5Sp1 (the number preceding Sp1 designates the number of consecutive Sp1 binding sites cloned into the plasmid). The activity of Splase on these substrates was carried out by initially linearizing each of the substrates with AlwNI. The substrates were subsequently digested with Splase. The results of a digestion are shown in Figure 8. Due to the absence of a consensus Sp1 binding site detectable cleavage was not observed in the lanes containing the linearized pUC19 DNA (lanes 2 and 3), while pUC5Sp1 was digested into two fragments with the expected sizes of 1.9 and 0.8 kb (lanes 4 and 5). The sizes of the two fragments are consistent with DNA cleavage near the Sp1 sites. Because the sites
are so close together, multiple products are not detected using agarose gel electrophoresis. 

**Figure 8** also depicts a more important experiment, the *BamHI*-linearized pUC-BENN-CAT carrying the HIV-1 LTR sequence with three consecutive Sp1 sites was also cut specifically by Splase. As expected, the cleavage of the *BamHI*-linearized pUC-BENN-CAT by Splase near the Sp1 sites generated two fragments of 4.2 kb and 1.7 kb (lanes 6 and 7). Substrates containing fewer than three consecutive Sp1 binding sites did not produce detectable quantities of product (data not shown). This raised several questions about the binding of Splase to these substrates.

**Evaluation**

The construction of Splase has demonstrated the feasibility of generating a new enzyme with altered specificity. There are, however, a number of problems and limitations with the current Splase. Several experimental conditions for Splase, such as expression level, purity, stability, and assay, remained to be optimized. Two of the problems, however, required fundamental solutions: the catalytic efficiency of Splase was very low and a significant degree of nonspecific digestion was observed. As a consequence of the two limitations compounded together, specific digestion can only be obtained at the initial stage of digestion. Improvement of these two limitations was the main focus of my work.

While emphasizing the limitations of the specificity and catalytic efficiency of Splase, it is important to put things in perspective. To the best of our knowledge, the chimeric enzymes designed by Kim et al. have similar problems. In fact, Splase showed our designed specificity but their zinc finger restriction enzyme displayed a somewhat altered specificity [62]. The Ubx-F$_N$ also showed nonspecific nuclease activity unless low Mg$^{2+}$ concentrations and high DNA/enzyme molar ratio were used [4]. Specificity, however, is a relative term. Even the commonly used type II restriction enzymes have weak affinity
Digestion of linearized DNA by HSplase. Lanes 1 and 8 contain molecular weight marker VII purchased from Boehringer Mannheim. Lane 2 contains linearized pUC19. Lane 3 contains linearized pUC19 digested with HSplase. Lane 4 contains linearized pUC5Sp1. Lane 5 contains linearized pUC5Sp1 digested with HSplase. Lane 6 contains linearized pUC-BENN-CAT. Lane 7 contains linearized pUC-BENN-CAT digested with HSplase. This figure was taken from Huang, et. al. (61).

Figure 8. Digestion of linearized DNA by HSplase.
sites referred to as “star activity”. For Splase, we have been able to demonstrate specific DNA phosphodiester hydrolysis under certain conditions. We believe that the low efficiency of Splase is not due to fundamental problems in its design but this is due to the tight binding of the DNA recognition domain.

In addition to the problems of specificity and catalytic efficiency, there are two intrinsic properties that have become problematic. First, Splase requires substrates with multiple binding sites and second, like natural restriction enzymes, Splase also bound and cleaved a few “weak sites” (star activity). An understanding of these intriguing problems will require detailed structure-function analysis of FokI, Sp1 and Splase.

In order to overcome some of these limitations we have considered using other nucleases in place of the nuclease domain of FokI, such as Staphylococcal nuclease, DNase I, and Endonuclease III. However, the number of nucleases that cleave double stranded DNA nonspecifically is very limited. The fact that the FokI nuclease domain is naturally used for this purpose and that the FokI crystal structure has been recently completed makes the FokI nuclease domain the best choice for this study (63).

If this project is fully successful, the possible applications in molecular biology will be numerous. For example, Splase is a rare cutter restriction enzyme due to the large 10 bp recognition sequence conveyed by the Sp1 zinc fingers. This could be very useful in generating large DNA fragments for genomic sequencing. Other designed restriction enzymes could be used to cleave DNA at any desired position. This would allow easier gene fusion, facilitate the mapping of DNA secondary structure and promoters, and facilitate the removal of introns from cloned genomic fragments.

Once our technology has been optimized so that a specific sequence can be detected by choosing the correct enzyme. If a particular disease results from a point mutation in a particular gene, that gene can be amplified and the restriction enzymes
corresponding to the three possible point mutations can be used to assay the amplified product for a particular sequence. The presence of cleavage products in one of the lanes identifies a possible mutation. This assay will be faster and easier than Restriction Fragment Length Polymorphism (RFLP) or direct gene sequencing.

In summary a new restriction endonuclease, named Splase, was constructed by genetically fusing the DNA-cleavage domain of the restriction endonuclease FokI with the zinc-finger DNA-binding domain of the transcription factor Sp1. The resulting protein was expressed in E. coli, and shown to selectively digest plasmid DNA harboring consensus Sp1 sites. Splase was also shown to selectively digest the long terminal repeat of the HIV-1 DNA at Sp1 sites [64,65]. Splase recognizes a ten base-pair DNA sequence and hydrolyzes phosphodiester bonds upstream of the binding sequence. The binding specificity of Splase makes this a rare cutter restriction enzyme which could be valuable in creating large DNA fragments for genome sequencing projects. The result opens the possibility to create other restriction enzymes by altering the binding specificity of the zinc finger recognition helix.

**Experimental Strategy**

Despite the extensive effort in the development of artificial restriction enzymes by many prominent researchers, the field is still in its infancy. The work presented here is an attempt to characterize the artificial enzyme, Splase, and begin work on the structural and functional approach toward development of a modular site-specific endonuclease. The goal of this project is to develop a universal method for the design and construction of new restriction endonucleases. In the preliminary studies, we have demonstrated the feasibility of the approach by combining the zinc finger binding domain of the transcription
factor Spl and the nuclease domain of the type IIS restriction enzyme FokI. However, the resulting endonuclease, Splase, requires substantial improvement in both the specificity and the efficiency of restriction digestion. The main thrust of this work is to perform structural and functional studies to improve the performance of splase and to generalize the approach to construct other restriction enzymes. The experimental design is guided by three hypotheses: (A) The specificity of the chimeric restriction enzyme can be achieved without a regulatory mechanism between the two domains. (B) Splase is not an optimal restriction enzyme because the nuclease domain is too efficient (resulting in nonspecific digestion) and/or the recognition domain binds DNA too tightly (resulting in a very low and rate-limiting $k_{\text{off}}$ for specific digestion). (C) The interactions between the zinc finger and DNA involve “affinity residues”, “specificity residues” and residues with combined roles. The goals of the project are as follows:

Goal 1: Reduce the Nonspecific Digestion of Splase by Impairing the Nuclease Domain. According to hypothesis B, the fast nonspecific digestion by the nuclease (without binding to the Sp1 site) is in competition with the slow specific digestion. Site-specific mutagenesis will be used to reduce the $k_{\text{cat}}$ of the nuclease and increase the $K_m$ such that nonspecific digestion by the nuclease can be minimized. The experimental design will be guided by the crystal structure of the FokI-DNA complex.

Goal 2: Increase the Rate of Specific Digestion of Splase by Mutating the Affinity Residues of the Sp1 Zinc Finger. According to hypothesis B, splase has a low catalytic efficiency because it binds DNA too tightly such that the release of DNA ($k_{\text{off}}$) is very slow and rate limiting. According to hypothesis C, this can be overcome by mutating the affinity residues of the Sp1 zinc finger to weaken its binding with DNA. The design of mutants will be based on the known structures of zinc fingers.
Goal 3: Create New Restriction Enzymes by Mutating the Specificity Residues of the Sp1 Zinc Finger. The results from Aims 1-2 will likely lead to an improved splase with good specificity and catalytic efficiency. On the basis of the zinc finger libraries created by others, we will mutate “specificity residues” (at positions -1, 2, 3, and 6) to create new restriction enzymes with designed specificity.

With the objective of transforming the zinc finger motif into a restriction enzyme, we have created a zinc finger endonuclease referred to as splase. This enzyme utilizes the DNA binding specificity of the three-zinc-finger motif of the transcription factor Sp1 and the nuclease domain of FokI developed by Chandrasegaran. The approach is illustrated in Figure 9.

Evaluation of Initial Situation

The purification consisted of a single Nickel-affinity column followed by gel filtration. This resulted in a yield of 5-10 mg of enzyme from 2.5 L of culture. The purified protein was ca. 50 percent pure as estimated by SDS PAGE. There were four impurities which were visible by silver staining. Optimization of the purification will likely lead to an effective purification scheme in which splase can be purified to homogeneity.

Kinetic studies of Splase to date have been hampered by the thermal stability of the enzyme. At 37 °C the enzyme exhibits a rapid loss of activity in the first 15 minutes of the reaction. This has been lengthened to 30 minutes through the addition of glycerol and BSA to the reaction buffer. FokI is also unstable, but it is stable under similar conditions for up to eight hours. Several possibilities could lead to the problem just described. First, we may have purified a trace amount of protease that is digesting splase, resulting in a loss of activity. This will be resolved by further purification and by addition of protease inhibitors. Second, we have recently learned through long discussions with John Caradonna that significant
Figure 9. Schematic of approach for designing new restriction endonucleases with new specificities.
oxidation of the zinc coordinating cysteine residues occurs during purification of zinc fingers. This results in loss of binding activity which may also lead to loss of activity. This problem can be overcome by using anaerobic conditions for the storage and assay of splase.

We have not yet been able to determine the $k_{\text{cat}}$ and $K_m$ of splase quantitatively due to the stability problem mentioned above. Once the stability problem is resolved the standard radioactive method will be used to determine $k_{\text{cat}}$ and $K_m$.

Our work so far has included a six-histidine tag at the N-terminus of splase (i.e., the zinc finger domain) in order to facilitate the purification of splase. After working with this construct for several months we noticed a paper in *Analytical Biochemistry* by Zorbas [66] which describes a possible problem of using the histidine tag on DNA binding proteins. The paper shows evidence for an effect of the tag on the interaction of the protein with the target sequence. Therefore, we will express the zinc finger domain with and without the histidine tag and see if there are any effects on the $K_d$ and of the protein with the target sequence. If there are no observable effects we will continue to use the histidine tag to facilitate purification. However, if an effect is demonstrated then the tag will be removed and a purification scheme developed to circumvent this problem. The improvement in the level of expression will certainly make it easier to purify the splase without a His-tag.

**Evaluation of the Strategy**

Common intuition may suggest that splase does not function optimally because it does not fully mimic the structural relationship between the recognition domain and the nuclease domain of *FokI*. Whether this is true or not, it is not directly related to hypothesis A. Explicitly, hypothesis A means that, regardless of what occurs in *FokI*, we can construct new restriction enzymes by combining two independent domains, without establishing a regulatory
mechanism (e.g., a conformational change induced by binding of DNA) between the two domains.

Experimental evidence already exists which supports hypothesis A: Chandrasegaran [67,68] has shown that varying the linker length between the FokI DNA binding domain and the nuclease domain led to a change in the cleavage site while maintaining similar activity to the wild type. Our design of splase is also based on this hypothesis, since the zinc finger motif used in splase is very different from the structure of the recognition domain of FokI and is unlikely to exert the same regulatory mechanism. In our view the limitation of splase is not due to invalidity of hypothesis A; it is due to problems explained in hypothesis B.

Hypothesis B again appears to be counter-intuitive. If splase is not an optimal enzyme, one would think that the recognition domain and/or the nuclease domain should be improved. However, we think (hypothesis B) that the problem of nonspecific digestion by splase is caused by the fact that the nuclease domain is too effective, and the problem of low turnover (of specific digestion) is caused by overefficient binding by the recognition domain. We use a scheme in Figure 10 to illustrate these effects. The first half of hypothesis B has already been well demonstrated by Schultz [69] in a different system. They constructed a chimeric DNase utilizing a DNA oligomer as a DNA recognition element and Staphylococcal nuclease as the nuclease domain. The hybrid nuclease, however, behaved similarly to the Staphylococcal nuclease itself. They then reasoned that the nuclease was too efficient and thus constructed a nuclease mutant, Y113A, with \( V_{\text{max}} \) reduced by 100-fold. Upon fusing the Y113A mutant to the DNA oligomer, the new construct exhibited site-specific cleavage upstream of the recognition sequence. This is exactly what we want to do with splase. However, we have chosen not to use Staphylococcal nuclease due to the preference for AT rich regions. It is believed that FokI nuclease domain has no substrate base composition preference.
Figure 10. Schematic describing hypotheses A and B.
The second half of hypothesis B is based on a simple physical principle that \( K_d = \frac{k_{off}}{k_{on}} \). Since \( k_{on} \) is less likely to change, if \( k_{off} \) increases, \( K_d \) increases (i.e., binding affinity decreases). This relationship has already been demonstrated for zinc fingers. Barbas [24] determined the \( K_d \), \( k_{on} \), and \( k_{off} \) of a variety of DNA recognition sequences on the Zif268 zinc finger; the results showed that the sequences with higher \( K_d \) also gave higher \( k_{off} \). What we want to do in Goal 2, however, is to modify the zinc finger rather than the DNA sequence. The key question in our approach is whether we can modify the zinc finger to increase \( K_d \) without changing the specificity.

Since the recognition domain and the nuclease domain function independently (hypothesis A), the catalysis by splase can proceed in two pathways. In pathway A the nuclease domain binds DNA and cuts DNA nonspecifically, without involving the recognition domain. In pathway B the recognition domain binds DNA first, then the nuclease domain binds and cuts the DNA site-specifically. The current problem of splase is that pathway A is too fast and pathway B is too slow. Thus our strategy of improving splase is to slow down pathway a (Goal 1) and enhance pathway b (Goal 2).

The kinetic constants of splase have not yet been determined quantitatively. For the purpose of discussion, we can assign hypothetical rate constants for pathway a and pathway b according to existing knowledge. On the basis of Staphlococcal nuclease [70], the nuclease domain of \( FokI \) can be assumed a \( K_m \) of \( 10^{-4} \) to \( 10^{-5} \) M and a \( k_{cat} \) of \( 10^3 \) s\(^{-1} \). The zinc finger domain of Sp1, when expressed separately, binds target DNA with a \( K_d \) of ca. \( 10^{-8} \) M [26]. Assuming a \( k_{on} \) of \( 10^4 \) s\(^{-1}\)M\(^{-1} \) for DNA binding to the zinc finger (as determined in ref. 24 for a different zinc finger), the \( k_{off} \) will be \( 10^{-4} \) s\(^{-1} \), which limits the rate of pathway b relative to pathway a and accounts for the problems of splase. Naturally, type II enzymes overcome this problem by cutting at the recognition site. They destroy their recognition sequence and release the DNA.
One other possibility is that if we simply reduce the $k_{\text{cat}}$ of the nuclease domain pathway a will be reduced while pathway b will be unchanged ($k_{\text{off}}$ is still rate-limiting). We can further impair the nuclease by increasing its $K_m$. This will make the nuclease less efficient in binding the DNA substrate and thus further slow down pathway a. This should not affect pathway b either since the "local concentration of DNA" is very high for the nuclease domain once DNA binds to the zinc finger domain. On the basis of our experience with various enzymes, the goal of reducing $k_{\text{cat}}$ and increasing $K_m$ can be achieved using site-directed mutagenesis (possibly by multiple mutations). These operations should greatly reduce the nonspecific digestion by splase.

If the binding of zinc finger to the target DNA sequence can be "weakened" from $K_d = 10^{-8}$ M to $10^{-6}$ M by mutating certain "affinity residues", the $k_{\text{off}}$ will increase to $10^{-2}$ s\(^{-1}\). This should lead to an approximately 100-fold increase in the rate of pathway b and thus the rate of splase. The recognition domain should still compete effectively with the nuclease domain for DNA binding since the $K_m$ of the nuclease domain has also been raised by a similar magnitude. In addition, this modification can also further enhance the specificity of splase since specificity is the result of competition between the two pathways.

Without even considering the structures of zinc finger-DNA complexes, we believe that the interactions between zinc finger and DNA involve "affinity residues", "specificity residues", and residues with combined roles. An interesting example involves our work on Adenylate Kinase. We have clearly demonstrated that some residues, such as R132, R138, and R149, interact with the phosphate moiety of ATP, and R44 interacts with the phosphate moiety of AMP. When these residues are mutated, $k_{\text{cat}}$ is reduced dramatically, accounting for a transition-state binding energy of 3-6 kcal/mol (for each residue). However, the nucleotide specificity is not affected. These residues are "affinity residues". On the other hand, mutation of Q101 to E changes the adenosine specificity at the AMP site, resulting in
a significantly decreased activity on AMP and a significantly increased activity on GMP [71]. Thus Q101 is a "specificity residue".

Zinc fingers bind DNA but have no intrinsic catalytic activity. However, the binding interactions can be dissected analogously to adenylate kinase. The affinity residues could be analogous to the phosphate binding residues in adenylate kinase. In the case of adenylate kinase, such binding occurs at the transition state and the binding energy is used in enhancing $k_{cat}$. In the case of zinc fingers, such binding occurs at the ground state and contributes to the binding affinity. The specificity residues could be the residues interacting with the nucleotide bases, such as Q101 in adenylate kinase. As previously described, the two types of interactions clearly exist in the crystal structure of Zif268-DNA complex. A vast amount of previous studies have demonstrated that the specificity of zinc fingers can be modified by changing the specificity residues [16-24].

According to hypothesis B, splase has a low catalytic efficiency because it binds DNA too tightly such that the release of substrate ($k_{off}$) is rate limiting. The goal is to modify the zinc finger such that the Sp1 zinc finger-DNA binding is "loosened". Ideally we would like to increase the $K_d$ by ca. 100-fold, which could lead to an increase in $k_{off}$ by a similar magnitude.

Since the zinc finger domain of Sp1 has been expressed in our lab and in others, the mutagenesis will be performed on the zinc finger directly. The zinc finger mutants displaying the desired property will then be fused with the modified nuclease domain obtained from goal 1, and the resulting splase characterized for specificity and catalytic efficiency.

According to hypothesis C, we should modify only the residues involved in the binding affinity between the zinc finger and DNA, such that only the binding affinity, not the specificity, is affected. Since many structures of zinc finger-DNA complexes are already known, rational mutagenesis instead of random mutagenesis will be used.
On the basis of our past experience with adenylate kinase, we believe the residues interacting with the phosphate groups are likely to be “affinity residues”, i.e., they contribute more to binding affinity, while the residues interacting with the nucleoside bases are more likely to be “specificity residues”.

The structure of the Sp1 zinc finger is not yet available. However, John Caradonna at Yale told us that he has solved the solution structures of each of the three zinc fingers of Sp1 complexed to DNA by NMR and the solution structure of the entire Sp1 zinc finger-DNA complex will be completed soon.

Even without the structure of the Sp1 zinc finger-DNA complex, we can design experiments based on existing structures of other zinc fingers. Figure 11 shows the sequence homology between the zinc finger domain of Sp1, the well-studied Zif 268, and a zinc finger designed by Jeremy Berg based on the consensus sequence of 130 known zinc fingers [81]. Figure 12 shows the partial crystal structure of the Zif268-DNA complex [8]. The structure indicates that R14 hydrogen bonds to the 5’ phosphate of base 7, R42 hydrogen bonds to 5’ phosphate of base 5, R70 hydrogen bonds to the 5’ phosphate of base 2, R3 hydrogen bonds to the 5’ phosphate of base 8 and S45 hydrogen bonds to the 5’ phosphate of base 6. If the DNA strand which contains the most protein-DNA contacts is referred to as the primary strand, each of the contacts listed are primary strand contacts. Secondary strand protein-DNA hydrogen bonds are also observed. These residues, along the corresponding residues in Sp1, are potential “affinity residues” and are indicated in Figure 11. As shown by sequence alignment, many other residues are conserved in these three proteins and could be involved in minor interactions with the DNA phosphates or sugars. These residues may also be mutated and their effects on binding assessed.

Each mutant will be constructed on a T7-controlled phagemid system pETSp1 containing all three Sp1 zinc fingers. This construct will allow easy generation of mutants using PCR.
<table>
<thead>
<tr>
<th></th>
<th>Sp1</th>
<th>Berg</th>
<th>Zif268</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MKKQHICHIQGCGKVKGYKTSHLRRAHRLWHTGER</td>
<td>MEKPYKCPE--CGKSFSQSSNLQKHQRTHHTGK</td>
<td>MERPYACPVECDRRFSRSDELTHRIRIHTGQK</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 11.** Sequence homology of zinc fingers.
Stereoview of the structure of Zif268 bound to DNA. This figure shows the critical residues responsible for interaction with the phosphate backbone. This figure was taken from Pavlatich and Pabo (8).

Figure 11. Stereoview of Zif268 Showing Affinity Residues.
mutagenesis. Each mutant will be completely sequenced, expressed in BL21(DE3)pLysS cells and purified using a modification of the published protocol [72].

Although Sp1 contains three zinc fingers, only one finger will be mutated at a time. If the effect on $K_d$ and $k_{off}$ is observed but is too small, then the same mutations will be made at the same positions on the second and third fingers. The effect of the mutation on the $K_d$ and $k_{off}$ will be assayed using standard mobility shift experiments or using BIAcore 2000 from Pharmacia Biosensor. The BIAcore 2000 allows for very sensitive determination of binding constants by measurement of surface plasmon resonance (SPR). SPR is an optical phenomenon in which changes in the refractive index of the solution close to the surface of a gold sensor chip are detected. This change in refractive index is detected using polarized light and is related to the concentration of solute at the surface of the sensor. This experiment will involve the attachment of an oligonucleotide containing the desired binding sequence $(5'\text{-N(m)GGGGCGGGGCN(m)-3'})$ to the gold foil detector via avidin-biotin interaction. The zinc finger protein solution will then be allowed to flow through the detection cell and the formation and breakdown of the complex observed in real time. The SPR signal is plotted versus time. This plot can be used to determine the association rate constant, $k_{on}$, and the dissociation rate constant, $k_{off}$, in real time. The ratio $k_{off}/k_{on}$ gives the dissociation constant $K_d$. This method has been used successfully to determine the $k_{on}$, $k_{off}$, and $K_d$ of new zinc fingers obtained from a phage display library [24].

The modified zinc finger domain of Sp1 and the nuclease domain of Fok1 will then be combined to construct the “improved splase”. The method of construction will depend on the extent of modification. If only a single mutation is required for each domain, it will be easier to use site-directed mutagenesis protocols on the existing splase.

Weakening the binding of the zinc finger should not be difficult to achieve. The question is whether this can be done successfully without altering the specificity of DNA recognition.
In other words, can “pure affinity residues” be found? Four steps will be taken to ensure that the specificity does not change unexpectedly. (i) NMR will be used to identify zinc finger mutants whose global conformations have been perturbed. The mutants with perturbed global conformation are likely to show perturbed specificity and will thus be discarded. (ii) If the $k_{cat}$, $K_m$, and $K_d$ of splase determined are quite different from the prediction, the particular mutant is likely to have a perturbed specificity and will not be further used. (iii) I-genomic DNA (50 kb) will be used as substrate of splase. The larger DNA substrate allows for many more possibilities to detect unexpected changes in specificity. It is a simple assay but does not cover all possible specificities. (iv) Since we will be changing the residues at a single zinc finger, Mobility shift assays or the BIAcore 2000 will be used to determine the binding constants for each of the possible binding sites (64 possibilities). The sequence with the lowest $K_d$ will define the specificity.

More than one mutant at each position will have to be tried to find the best combination; the results may not be quantitative as predicted and fine-tuning may be necessary. If desired changes do not result from the outlined strategy we are prepared to construct random mutants of each finger and screen the library created using the yeast 1 hybrid system. This will yield mutants with a large range of binding constants. Mutants can then be assessed for actual binding constant using mobility shift and possible candidates identified. Initially we have chosen not to use this approach due to the increased length of time needed to screen libraries of zinc fingers and identify possible candidates.

As we optimize Splase we may end up with a highly toxic protein (to *E. coli*). There are several ways to circumvent this problem. The protein could be fused to a leader sequence which could export the protein [84]. If this does not solve the problem then we can express the protein in an eukaryotic system such as yeast or insect cells. One other long term possibility is to create a zinc finger methylase which could be co-expressed in *E. coli* and site-specifically
methylate the host DNA. It is interesting to note that the type IIS methylases are also two-
domain proteins and may also lend themselves to this type of engineering.

The results from Goal 2 will likely lead to an improved version of splase. The focus of
this goal is to demonstrate the feasibility of changing the specificity of the improved splase
by changing the "specificity residues" and to actually construct a number of useful new
restriction enzymes. Since there is a vast amount of published information about zinc fingers
with modified specificity, we will try to make use of the existing information rather than
creating our own. The strategic steps and procedures will be similar to those described in
goal 2, except that goal 2 focuses on affinity residues and goal 3 on specificity residues.

As shown in Figure 3, the crystal structure of Zif268 clearly indicated that certain residues
specifically interact with specific bases and are likely to control the specificity of recognition.
These residues have been identified as the residues in the -1, 2, 3 and 6 positions (see Figure
11) [8]. Subsequently, several groups have used phage display libraries to identify the rules
which govern the specificity of these residues in DNA sequence identification. Table 1
summarizes the residues in the -1, 2, 3 and 6 positions and the target DNA sequences they
confer, on the basis of the phage display libraries of Jim Wells [20], Aaron Klug [16,17], Carl
Pabo [23] and Carlos Barbas [24] and the site directed mutants of Jeremy Berg [18,19,21,81]
and Patrick Charnay [22].

The modified Spl zinc finger with attenuated Kd obtained from goal 2 will be further
mutated for specificity changes according to Table 1. Although Spl contains three zinc
fingers, only one will be mutated at a time. If the effect is observed but is too small, then the
same mutations will be made at two or three fingers. The zinc finger mutants will be assayed
for Kd and koff with the designed DNA sequence as described in Goal 2.

The first priority will be to make a few of these enzymes, demonstrate that the specificity
will change as expected, and demonstrate that the specificity mutations will not cause changes
### Table 1

A table of DNA sequence specificity and the specificity residues in the -1, 2, 3 and 6 positions which will generate this DNA specificity.
in binding affinity and catalytic efficiency. Once the feasibility has been successfully demonstrated, we will be able to design the new specificity on the basis of Table 1.

The studies leading to the specificity listed in Table 1 were all performed on a different zinc finger, ZiG268. Although the zinc fingers Sp1 and Zif268 are highly homologous as shown in Figure 9, there is a remote possibility that the results from one may not apply directly to the other. If this becomes a problem, there are three possible approaches: (i) We can switch our zinc finger to Zif268 starting from Goal 2; (ii) we can create our own zinc finger library for Sp1; or (iii) we can accept the new specificity, even though it is different than expected.

The work described below describes the process of characterizing our initial attempt at generating this artificial restriction enzyme, the identification of the problems we had to overcome and the preliminary attempts at solving some of these problems.
CHAPTER 2

MATERIALS AND METHODS

Materials

All restriction enzymes used for cloning were obtained from New England Biolabs. All synthetic oligonucleotides were purchased from Integrated DNA Technologies. Ni-NTA resin was purchased from Qiagen Incorporated. His-bind affinity resin was purchased from Novagen. All chromatography columns and Heparin Sulfate were purchased from BioRad Laboratories. Tris HCl and Tris free Base buffers were purchased from Applied Organics. HEPES buffer was purchased from Sigma. Thrombin was purchased from Sigma. pGEX-2T expression vector was purchased from Pharmacia Biotech. Glutathione agarose affinity resin was purchased from Sigma. pTrc99a was purchased from Pharmacia Biotech. pET 15a and pET21a were purchased from Novagen. Sephacryl S-200, S-100 and Sephadex G-100 gel filtration resins were purchased from Pharmacia Biotech. DTT was purchased from Amersham Life Sciences. Glutathione (Reduced) was purchased from Calbiotech. IPTG and Ampicillin was purchased from Bohrenger Manheim. Pfu polymerase was purchased from Stratagene. Plasmid miniprep kits were purchased from Qiagen. Gene fragment recovery kits were purchased from Qiagen or Bio101. JM105 cell line was purchased from Pharmacia Biotech. XL-1 Blue cell line was purchased from Stratagene. BL21 and BL21(DE3)pLysS were purchased from Novagen. Taq FS
Fluorescent Sequencing Kit was purchased from Applied Biosciences. Tryptone and Yeast Extract were purchased from Difco. Double distilled water was provided by the University. Double distilled demineralized water was purchased from University stores. All other general reagent grade chemicals were purchased from Fisher.

Methods

PCR Cloning

In order to construct pGEX-Splase several modifications needed to be made. Initially, pGEX-2T only contains three restriction sites within the multiple cloning site EcoRI, BamHI and SmaI. SmaI is a blunt cutter and provides significant difficulty during ligation and the Splase gene contains a single EcoRI site, this left BamHI as the only enzyme site. Therefore since the H-Splase gene is cloned into pTrc99a using the NcoI and BamHI this was used as template the upstream primer used, consisted of the following base sequence: 5'-GGC CGG ATC CAT GGC TAA AAA GAA ACA GCA-3'. This changes the upstream restriction site from NcoI to BamHI while leaving the NcoI site present. It also removes the six histidine tag which would no longer be needed for purification. The downstream primer was the primer originally used to clone the FokI nuclease domain. The fragment corresponding to the Splase gene was amplified in a 100 µl total volume containing 10 mM Tris-HCl, pH 8.8, 10 mM KCl, 1 mM MgCl₂, 2 µM of each primer, 10 µM deoxyribonucleotide triphosphates (dNTPs), 0.002% Tween 20 (v/v) and 1 unit of Vent polymerase. The reaction was hot started using AmpliWax beads (Perkin Elmer). The PCR was carried out in a Perkin-Elmer (Norwalk, CT, USA) DNA Thermal Cycler using the following cycling parameters: An initial denaturation of 30 seconds at 96 °C was followed by 25 cycles of denaturation at 95 °C for 30 seconds,
annealing at 60 °C for 15 seconds and extension at 72 °C for 1 minute. The final cycle was followed by a 2 minute extension period at 72 °C to complete the started extension products. The PCR products were initially analyzed for uniformity of product using Agarose gel electrophoresis, 5 μl of each reaction was loaded into a 1% agarose gel and the products separated under an electric field generated by a current of 100 mA and 150 V on a Fisher power supply. Each reaction containing a product consisting of a single band was extracted with phenol/chloroform (1:1) and the DNA was precipitated with Ethanol and dissolved in 20 μl TE (10mM Tris-HCl pH 8.0, 1 mM EDTA).

Ligations were prepared by digesting 25% of the product (5 μl) with BamHI under the conditions suggested by the manufacturer (New England Biolabs). One microgram of the pGEX-2T vector was prepared similarly and the digested products were separated on a 1% agarose gel in TAE (0.4 N. Tris-HCl, 13 mM NaOAc, 0.2 mM EDTA pH 8.0). The products were visualized by soaking the gel in 5 μg/ml ethidium bromide for 5 minutes. The DNA band of approximately 920 bp was excised from the agarose gel using a razor blade and purified using geneclean (Bio101). The purified fragments were eluted into 10 μl of ddH₂O. The ligation was set up using 250 ng of vector and 50 ng of insert, 2 μl of T4 ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 25 μg/ml BSA), and 1 μl of T4 ligase, total volume of 10 μl. The reaction was placed in a 16 °C water bath and incubated for 16 hours.

The ligation mixture was then transformed into XL-1 blue competent cells. After incubation at 37 °C for 16 hours colonies were picked and inoculated to 5 ml of LB with 100 μg/ml ampicillin. These cultures were grown for 14 hours and then their plasmid DNA purified using the standard mini-prep (Maniatis). Following purification of the plasmid DNA, each clone was screened using restriction digestion. Each plasmid was digested in a 20 μl total volume containing 5 μl of purified plasmid DNA (approximately
500ng), 2 μl of BamHI buffer (10 mM Tris-HCl pH 7.9, 150 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 100 μg/ml BSA) and 1 μl of BamHI (10 units). The reactions were placed at 37 °C for 1 hour and then loaded onto a 1% agarose gel to separate the products. The products were visualized by soaking the gel in 5 μg/ml ethidium bromide for 5 minutes and visualizing on a UV light box. The gel was destained by soaking in ddH₂O for 30 minutes. A picture of the result was taken as a record of the experiment and placed in a notebook for record keeping. Clones containing the correct size insert were kept for sequencing and expression experiments.

Isolation of genomic DNA from Flavobacterium okeanokoites (Fok)

Genomic DNA isolation was performed using the large scale protocol described in Current Protocols in Molecular Biology. Flavobacterium okeanokoites was obtained from the American tissue type culture center as a lyophilized powder. The powder was reconstituted in LB and streaked onto LB glucose plates containing 2% glucose. Flavobacterium okeanokoites was grown at 25 °C for 48 hours with shaking in 500 ml of Fok media containing 10 g of tryptone, 5 g of Yeast extract, 2 g of NaCl and 4.4 g of K₂HPO₄ per liter. The cells were harvested by centrifugation in a Beckman high speed centrifuge. The pellet was resuspended in 200 ml of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). 15 ml of 10% SDS and 3 ml of Protease K (20 mg/ml) were added and the sample was incubated at 37 °C for 1 hour. Promptly following incubation, the cell lysate was mixed thoroughly first with 50 ml of 5 M NaCl and 40 ml of CTAB/NaCl solution (10% CTAB in 0.7 M NaCl). The sample was then incubated at 65 °C for 10 minutes. The DNA comes out of solution as a stringy white solid which is collected by using a glass rod hook. The DNA was removed from the solution and placed into a 1.5 ml
microcentrifuge tube. The DNA was resuspended in 1 ml of ddH₂O. The DNA was precipitated with ethanol and washed with 70% ethanol. The final pellet was resuspended in 1 ml of TE and placed at -20 °C for storage.

**PCR Cloning of Fragments of FokI Restriction Enzyme**

The DNA fragments corresponding to the C-terminal 853 bp, 885 bp, 999 bp and 1008 bp were amplified using the upstream primers shown in Figure 13 and the downstream primer used above. PCR reactions were set up containing 300 ng of Fok genomic DNA, 10 mM Tris-HCl pH 8.8, 10 mM KCl, 10 mM MgSO₄, 2 μM each primer and 10 μM dNTPs, 0.002% Tween 20 and 1 unit of Vent polymerase. The reactions were carried out in a Perkin-Elmer DNA thermal cycler using the following parameters; an initial denaturation at 95 °C for 30 seconds, followed by 25 cycles of denaturation at 95 °C for 30 seconds, annealing at 60 °C for 15 seconds and extension at 72 °C for 3 minutes. The products were analyzed using 1% agarose gel electrophoresis for the presence of a single product of the correct size. Reactions containing the correct product were extracted with phenol:chloroform and precipitated with ethanol. The pellet was washed with 70% ethanol and redissolved in 20 μl ddH₂O. Five microliters of the product was then digested with *Kpn*I and *Hind*III. The products were separated on 1% Agarose gel electrophoresis and visualized using ethidium bromide staining and UV detection. The correct size band was excised from the gel and purified using geneclean (Bio101). The fragment was cloned into pTrcHSplase and then the upstream restriction site changed to BamHI for cloning into pGEX-2T. The resulting plasmids were screened for inserts of the
FokI Nuclease 80  
5'-GGC CGG TAC CTG CTA CAA ACC TAA CCG ATA-3'

FokI Nuclease 92  
5'-GGC CGG TAC CAC GTG TAC CTA AAA AGT ATA-3'

FokI Nuclease 130  
5'-GGC GCC TAC CAT TTA TCA TTC CTA CTT TGG-3'

FokI Nuclease 133  
5'-GGC CGG TAC CAA AAA AAG AAT TTA TCA TTC-3'

Figure 13. Primers Used to Amplify Nuclease Domains
correct size and then sequenced to determine if the correct fragment had been cloned and if the correct orientation with respect to the GST gene had been maintained.

**PCR construction of pETSp1**

Plasmid DNA was transformed into XL-1 blue cells and purified using standard miniprep procedures. The upstream primer 5'-GGC CGG CCA TGG CTA AAA AGA AAC AGC A-3' was used to remove the histidine tag prior to insertion into the pET21 vector. The downstream primer 5'-CGC CGG ATC CTT ACT TCT TAT TCT GGT GGG TCT T-3" was used to change the downstream restriction site from KpnI to BamHI and insert a stop codon after lysine 92. This product was amplified using the previously described cycling parameters. Digested with NcoI and BamHI and cloned into pET21d. Once cloned into pET21d the vector became known as pETSp1 and corresponds to the C-terminal 92 amino acids which make up the three Sp1 zinc fingers.

**DNA Purification**

The procedure for DNA purification was adapted from the alkaline lysis method described by Maniatis et al. A single colony of E. coli of XL-1 blue harboring the plasmid of interest was inoculated to 5 ml of LB with 200 µg/ml ampicillin and grown 14 hours at 37 °C. The cells were transferred as 1.5 ml aliquots to a single microcentrifuge tube and centrifuged for 6 minutes at top speed in an eppendorf microcentrifuge. The supernatant was discarded. To each tube was added 100 µl of ice-cold solution I (25 mM Tris-HCl and 10 mM EDTA pH 8.0). The cells were resuspended in solution I by gentle vortexing and kept at room temperature for 5 minutes. Two hundred microliters of freshly prepared
solution II (1% SDS and 0.2 N NaOH) was quickly added, and the lysed cells were stored on ice for 5 minutes. Following incubation on ice for 5 minutes, the cellular protein was denatured by the addition of 150 µl of ice cold solution III (5 M acetate, 3 M K+). The mixture is placed on ice for 10 minutes and then centrifuged for 10 minutes at high speed in a microcentrifuge to remove cellular debris. Following centrifugation, the supernatant is transferred to a new microcentrifuge tube and extracted with 200 µl phenol and chloroform. The contents were vortexed to mix and then centrifuged to collect the aqueous layer. This layer was then transferred to another microcentrifuge tube and 800 µl of ice cold ethanol was added to precipitate the DNA. The contents were incubated at -20 °C for 30 minutes and centrifuged at top speed for 15 minutes. The supernatant was discarded and the pellet washed with 100 µl of 70% ethanol and dried in a speed vac. The pellet was redissolved in 20 µl of ddH2O. The RNA was removed by adding DNase free RNase (1.0 µl of 10 µg/ml RNase per tube). DNA concentrations were determined by measuring the A260, or by estimating the concentration by running a sample on a 0.7 % agarose gel and comparing the DNA with a standard of known concentration.

Preparation of Competent Cells

One milliliter of an overnight culture of E. coli is added aseptically to 200 ml of SOB media (4g of tryptone, 1g yeast extract, 0.12 g NaCl, 0.10 g KCl in 200 ml ddH2O) after adding 2 ml of 1 M MgSO4, 2 ml of 1 M MgCl2. The culture is incubated at 37 °C with gentle shaking until the cell density reached an A600 of 0.3. The cell culture is removed from shaking and transferred to a sterile centrifuge tube (400 ml) and stored on ice for 10 minutes. The cells were collected by centrifugation for 5 minutes in a Beckman J2-21 centrifuge (JA -10 rotor) at 6000 rpm (6370 xg) at 4 °C. The supernatent was
removed and the cells were resuspended in 35 ml of ice cold TFB1 (30 mM KOAc, 100 mM KCl, 10 mM CaCl₂·H₂O, 50 mM MnCl₂·4H₂O, and 12% (v/v) glycerol/0.2 L, pH 5.8 sterilized by filtration and stored at 4 °C) with gentle swirling on ice. The suspension was allowed to stand on ice for an additional 10 minutes. The cells were collected by centrifugation at 6000 rpm for 5 minutes. The supernatant was discarded and the cells were resuspended in 8 ml of TFBII (10 mM MOPS, 10 mM KCl, 75 mM CaCl₂·H₂O, 12% (v/v) glycerol/0.2 L, pH 5.8 and sterilized by filtration). Two hundred microliter aliquots were transferred to pre-chilled 1.5 ml microcentrifuge tubes and immediately frozen in liquid nitrogen. The competent cells were stored at -70 °C and remained viable for up to 6 months.

Transformation

Competent cells were thawed on ice for 15 minutes at which time DNA was added by swirling the cell suspension. The cells were then incubated on ice for 45 minutes and then heat shocked at 42 °C for 90 seconds followed by storage on ice for 5 minutes. Following this incubation on ice 800 µl of LB was added and the solution was incubated at 37 °C for 1 hour. Selection for bacteria which contained the plasmid of interest was conducted by spreading between 50 µl and 200 µl of the transformation mixture out on LB agar plates containing the required antibiotic. The plates were incubated at 37 °C for 14-16 hours.
Site-Directed Mutagenesis

Site directed mutagenesis was conducted using the Quickchange Site-Directed Mutagenesis Kit obtained from Stratagene. The Quickchange system involves the use of Pfu polymerase, a high fidelity polymerase which lacks strand displacement activity. The procedure uses supercoiled double stranded DNA and two mutagenic primers. The DNA template is isolated from a strain which methylates DNA during replication. The reaction takes place in a thermocycler as a polymerase chain reaction. The reaction begins by melting the template allowing the mutagenic primers to anneal in a second step. Once annealed the primers can be extended to completely replicate a single strand of the entire plasmid. In subsequent rounds of PCR the original vector as well as the new mutagenic vectors are used as template thus amplifying the mutagenic vectors. Upon completion of the PCR cycling the resulting mutagenic vector contains staggered nicks and is composed of un-methylated DNA. The reaction mixture is then subjected to restriction digestion using \textit{DpnI}. \textit{DpnI} site specifically recognizes DNA which is methylated or hemimethylated at 5'-G\textsuperscript{m6}ATC-3' sites within the plasmid. Five microliters of the restriction digest is transformed into XL-1 Blue competent cells. The entire transformation mixture is microfuged at high speed for 2 minutes. The supernatant is removed and replaced with 50 μl fresh LB. The pellet is resuspended and plated out on LB plates containing 100 μg/ml ampicillin. The plates are incubated at 37°C for 16 hours. Colonies were isolated and grown overnight (12 hours) in 5 ml of LB containing 100 μg/ml ampicillin. Plasmid DNA was isolated and either subjected to restriction digestion using a unique enzyme or sequenced to identify positive mutants. The mutagenic primers were designed by generating a primer which contains the mutation in the center of the oligo. The melting temperature of the oligos used was between 68°C.
and 80 °C. The manual recommends that the annealing temperature be 10 degrees above the extension temperature of 68 °C. The reaction mixtures were prepared by adding 5 μl 10x Reaction buffer (100 mM KCl, 60 mM (NH₄)₂SO₄, 200 mM Tris-HCl (pH 8.0), 20 mM MgCl₂, 1% Triton X-100, 100 μg/ml BSA), 50 ng of template DNA, 125 ng of each primer, 1 μl of 10mM dNTP mix (2.5 mM each NTP) ddH₂O to 50 μl, 1 μl of Pfu polymerase. Each reaction was overlaid with mineral oil and placed in a thermal cycler (Perkin Elmer 480) for 16 cycles of denaturation at 95 °C for 30 seconds, annealing at 55 °C for 1 minute and extension at 68 °C for 12 minutes. After 16 cycles the reactions were allowed to cool to 4 °C. The reaction mixture was removed from the PCR tube and placed into a fresh eppendorf, to which was added 1 μl of DpnI, and placed into a water bath at 37 °C for 1 hour. The mixture was then removed and a 5 μl aliquot was used to transform XL-1 blue competent cells. The primers used for mutagenesis are shown in Figure 14.

DNA Sequencing

Sequencing was conducted using ABI Prism Dye Terminator Cycle Sequencing Ready Reaction Kit containing AmpliTaq DNA polymerase, FS. This method is a modification of the Sanger dideoxy termination sequencing method, however radioactive nuclides are not required. The reaction is conducted by adding an extendable primer to the template in the presence of dNTP’s as well as dye labeled ddNTP’s. Each base is labelled with a separate dye which has a unique emission spectrum. The reactions are conducted as a PCR reaction on a standard thermocycler. The reactions are initiated after an initial denaturation and annealing of the primers. The reaction proceeds, incorporating NTP’s until a ddNTP is incorporated. This reaction terminates the polymerization and
<table>
<thead>
<tr>
<th>Mutant</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>K52A</td>
<td>Sense</td>
<td>5'-TACTGTGGGGCAGCCTTCACAGC-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CGTGTGAAGCGTGGGCCAAGTA-3'</td>
<td></td>
</tr>
<tr>
<td>K52Q</td>
<td>Sense</td>
<td>5'-TACTGTGGCAACGCTTCACAGC-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CGTGTGAAGCGTGGGCCAAGTA-3'</td>
<td></td>
</tr>
<tr>
<td>K52S</td>
<td>Sense</td>
<td>5'-TACTGTGGGCAGCTTCACACAGC-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CGTGTGAAGCGTGGGCCAAGTA-3'</td>
<td></td>
</tr>
<tr>
<td>K52T</td>
<td>Sense</td>
<td>5'-TACTGTGGACACGCTTCACAGC-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CGTGTGAAGCGTGGGCCAAGTA-3'</td>
<td></td>
</tr>
<tr>
<td>K52D</td>
<td>Sense</td>
<td>5'-TCATACTGTGGGGGATCGCTTCACAGC-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-AGCTGTGAAGCGTGATCCCACAGTATGA-3'</td>
<td></td>
</tr>
<tr>
<td>S85A</td>
<td>Sense</td>
<td>5'-AAACGCGTTGACTGAGGGACCCACCTG-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CAGGCTTGTTGCGCCCTGCTGATAAGCGTCTT-3'</td>
<td></td>
</tr>
<tr>
<td>S85D</td>
<td>Sense</td>
<td>5'-CGCTTTGCTGAGGATTGACCACCTGTCA-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-TGACAGTTGCTGCTGCTGCTGATGACAG-3'</td>
<td></td>
</tr>
<tr>
<td>S85T</td>
<td>Sense</td>
<td>5'-CGCTTTGCTGAGGACCGACCACCTG-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CAGGTGTTGCTGCTGCTGATGACAG-3'</td>
<td></td>
</tr>
<tr>
<td>K80A</td>
<td>Sense</td>
<td>5'-CCTGAGTTGTCCGCTGCCTTCACATGAGG-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>5'-CCTCAGCTGACGCGCGACACTGAGG-3'</td>
<td></td>
</tr>
<tr>
<td>K80D</td>
<td>Sense</td>
<td>5'-CCTGAGTTGTCCGCGCCGACACTGAGG-3'</td>
</tr>
<tr>
<td>Anti-sense</td>
<td>3'-CCTGAGTTGTCCGCGCCGACACTGAGG-3'</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 14.** Primers Used for Mutagenesis
incorporates a unique dye which corresponds to the identity of the terminal nucleotide. The sequencing reactions are prepared by adding 500 ng of template dsDNA in a total volume of 1.5 to 2.5 μl to a 500 μl PCR tube. Subsequently 3.2 pmol of the respective sequencing primer is added to the template followed by 8.0 μl of the Terminator Ready Reaction Mix (A-Dye Terminator, C-Dye Terminator, G-Dye Terminator, T-Dye Terminator, dITP, dATP, dCTP, dTTP, Tris-HCl (pH 9.0), MgCl₂, thermal stable pyrophosphatase and AmpliTaq DNA polymerase, FS. The total volume of the mixture is brought up to a total of 20 μl, a drop of mineral oil is used to seal the mixture to prevent evaporation and the reaction is placed in a Perkin Elmer 480 DNA Thermal Cycler. The reactions are subjected to an initial thermal ramp to 96 °C for 30 seconds for template denaturation, the reactions are then cooled to 50 °C for primer annealing and are held at this temperature for 15 seconds. Finally, the reactions are heated to 60 °C for 4 minutes to enzymatically extend each annealed primer/template pair. This cycle is repeated 25 times and finally cooled to 4 °C for storage. Each Reaction is transferred to a new 1.5 ml Eppendorf tube containing 2.0 μl 3 M NaOAc pH 5.2 and 50 μl 95% ethanol to precipitate the terminated products. The Solution is then vortexed to mix the added reagents. The tubes are placed in a microcentrifuge at maximum speed (14 krpm) for 30 minutes. Upon removal from the microcentrifuge the supernatant is removed and the residual liquid is removed by careful aspiration attempting to leave the precipitated products undisturbed. The pellet is then rinsed with 250 μl of 70% ethanol to remove precipitated salts which may interfere with the product separation on an polyacrylamide gel. The supernatent is removed and the residual liquid is removed by aspiration. The tube containing the pellet is then placed in a Speed-vac on low heat with spinning for 5 minutes to dry the pellet. The tube was removed andWrapped in tinfoil and sent to The Ohio State University Biological Instrumentation Center for Analysis on an ABI 373 DNA
Sequencing instrument. The chromatograms were obtained and the sequence verified against the proposed sequence.

Subcloning

Double stranded recombinant DNA was subcloned into the polylinker of the respective plasmid vector by digesting with 10 units of the respective upstream and downstream restriction enzymes (2-3 hours at 37 °C in a total volume of 20 μl). The linearized plasmid and recombinant insert were isolated by first separating the individual products out on a 1% agarose gel. The gel was then stained with ethidium bromide. The insert and vector were subsequently excised from the gel and placed into separate marked eppendorf tubes. They were each subjected to purification from the gel slice using GENE CLEAN (United States BioChemical) or Qiaquick DNA isolation kit (Qiagen). The DNA was then resuspended in 10 μl total volume and 2μl was used to estimate the DNA concentration. The insert and vector were mixed in the appropriate 3:1 ratio and ligated by the addition of 10 units of T4 DNA Ligase and 2 μl of 10x ligation buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP and 25 μg/ml BSA). Each reaction was incubated at 16 °C for 16 hours. Ten microliters was transformed into XL-1 Blue competent cells. Recombinant colonies were screened for the presence of the ligated insert using restriction digestion with the enzymes used to clone the insert. DNA containing the insert of interest was further analyzed by sequencing the entire gene to make sure there were no unwanted mutations. Clones containing the correct DNA sequence were transformed into an expression cell line and stored as a glycerol stock at -80 °C.
Purification of GST-HSsplase and GST-Splase80

Growth and Harvest

Plasmid containing the gene of interest is transformed into competent BL21 cells, plated out on LB/Amp plates containing 100 μg/ml amp and placed at 37 °C overnight. A single colony was used to inoculate a 500 ml flask containing 100 ml of LB, 100 μg/ml Amp and 2% glucose (5 ml of 40% glucose stock). The culture was placed in a New Brunswick rotary shaker at 250 rpm for 12 hours at 37 °C. Twenty-five milliliters of this overnight culture was then used to inoculate 4 x 2 L flasks containing 0.5 L of 2 x YT media, 2% glucose and 100 μg/ml amp. These flasks were placed in a rotary shaker until the A₆₀₀ reached 1.0. At this point the cultures were induced to express the protein of interest. Each culture was induced by adding 0.5 ml of 0.3 M IPTG to a final concentration of 0.3 mM. The temperature of the incubator was changed to 25 °C and the cultures were incubated for 3 hours. Subsequent to the completion of the incubation the cells were harvested by centrifugation. The contents of each flask was transferred to a 400 ml centrifuge bottle. Each of the bottles were balanced and placed into a Beckman Ultrahigh centrifuge and spun at 6000 rpm (6437 x g) in a JA-10 rotor for 10 minutes. Upon completion the cells were collected into a single centrifuge bottle and washed with 200 ml of 1x PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ (pH 7.3)). The cells were again centrifuged at 6 krpm for 10 minutes and placed at -20 °C until purification could begin.
Preparation and Equilibration of Glutathione Agarose Resin

Glutathione Agarose (G-column) resin was purchased from Sigma as a lyophilized powder (Catalog # G-4510). The resin was swelled in a 250 ml beaker, by adding 400 mg (enough to give 10 ml of swelled resin) of the lyophilized powder to 200 ml of ddH$_2$O for 2 hours. The resin is allowed to settle to the bottom of the beaker and the water is poured off. Two hundred milliliters of 1xPBS is then added to the swelled resin and this is allowed to equilibrate at 4 °C overnight. The following day the supernatent is poured off leaving approximately 20 ml of a resin/buffer slurry. The slurry is carefully poured down the side of a 24 ml BioRad econocolumn (1.0 cm x 30 cm) while the stopcock is opened. The resin is allowed to settle while being equilibrated with 1xPBS. After 1L of buffer has been run over the column the resin is ready to be used.

Once the column has been used the resin can be regenerated by first washing with 200 ml of 1 M NaCl. The column is then washed with 200 ml of ddH$_2$O, 200 ml of Borate buffer (0.1 M NaBO$_3$ (pH 8.5), 0.5 M NaCl), 200 ml ddH$_2$O, 200 ml of Acetate buffer (0.1 M NaOAc (pH 4), 0.5 M NaCl), 200 ml of ddH$_2$O and finally 200 ml of 1xPBS. The resin can be stored in 1 M NaCl at 4 °C for short periods of time and in buffer containing NaN$_3$ for longer time periods as recommended by Sigma.

Cell Lysis and Glutathione Affinity Chromatography

Fresh or Frozen cells were resuspended in 180 ml of 1xPBS buffer by stirring for 1 hour at 4 °C. Once the cell suspension was homogeneous 80 mg of Hen Egg White Lysozyme was added to the cell suspension at 4 °C. The mixture was allowed to incubate for 1 hour to achieve complete lysis. This resulted in a very viscous liquid. This viscous
liquid was then transferred to a 250 ml glass beaker which was pre-chilled on ice for one hour. The mixture was subjected to Ultrasonic waves produced from a High Intensity Ultrasonic Processor (Sonics and Materials, 300 Watt Model). The sonication consisted of four 60 second pulses with settings of 70% Duty Cycle and power equal to 8. The tip was immersed approximately ½ inch into the mixture. This resulted in a mixture which had the consistency of H$_2$O and had a darker color than prior to sonication. The mixture was then loaded into four 50 ml centrifuge tubes and balanced. The tubes were placed into a JA-20 Beckman centrifuge rotor and spun at 19 krpm (43,700 xg) for 30 minutes. Upon completion, the supernatent was removed and collected in a 250 ml Erlenmeyer flask which was pre-chilled on ice for 30 minutes. The cell lysate was then loaded onto a 10 ml G-column with a flow rate of 1 ml/min. Upon completion of the load phase of purification the column was washed with 800 ml of 1xPBS followed by 200 ml of Wash buffer (50 mM Tris-HCl (pH 8.0)). The GST fused protein was then eluted from the column with 200 ml Elution buffer (50 mM Tris-HCl (pH 8.0), 10 mM Glutathione (GSH)). Two milliliter fractions were collected using a RediFrac Fraction Collector (Pharmacia Biotech). Spectrophotometric determination of protein concentration was conducted by determining the absorbance at 280 nm. Fractions containing protein were analyzed for purity by loading 0.5 ml samples onto 12.5% SDS-PAGE gels. Samples were prepared by adding 2 µl of 5 x gel loading dye to 8 µl of each fraction. The samples were placed in a 90 °C water bath for 5 minutes to denature the protein in each sample. Each sample is then loaded into a Phast System Gel Electrophoresis System (Pharmacia Biotech.). The Phast system loads each sample automatically onto a 12.5% polyacrylamide gel, automatically initiates electrophoresis and separates proteins based on size. Fractions containing the protein of interest were identified and used in an activity assay to compare the activity with our original HSplase construct. The protein
concentration was determined using an extinction coefficient calculated at http://expasy.hcuge.ch/sprot/protparam.html using the protein parameters tool.

**Purification of HSplase**

**Growth and Harvest**

The plasmid vector, pTrcHSplase, containing the HSplase gene was transformed into an expression cell line, JM105. The transformation mixture was plated out on LB/Amp plates containing 200 μg/ml ampicillin and incubated at 37 °C for 12 hours. A 2 L flask containing 0.5 L of LB was inoculated with 200 μg/ml ampicillin and a single colony from the overnight plate. This was grown for 12 hours, removed and used to inoculate 4 L of Rich Medium (19 g Yeast Extract, 12 g Tryptone and 10 mM MgCl₂). Each 2 L flask containing 0.5 L of Rich media was inoculated by adding 10 ml of 1 M MgCl₂, 0.5 ml 200 mg/ml ampicillin and 25 ml of overnight culture. The cultures were grown at 37 °C until the A₆₀₀ reached 0.6. Once the optimal optical density was reached the cells were initially chilled on ice until the culture temperature reached 25 °C. Once this temperature was reached the cultures were induced by adding 0.5 ml of 1 M IPTG (final concentration 1 mM) to each flask. The cultures were allowed to continue to grow for another 12 hours before harvesting the cells. The cells were harvested by spinning in a Beckman Ultrahigh centrifuge at 6000 rpm (6400 xg) in a JA-10 rotor. The cell pellets were consolidated and washed in 200 ml of binding buffer (20 mM Tris-HCl (pH 7.9 at 4 °C), 0.5 M NaCl, 5 mM Imidazole). The cell suspension was then recentrifuged at 6000 rpm to collect the cell pellet. The cell pellet was redissolved in fresh binding buffer and immediately entered the purification protocol.
Preparation and Equilibration of His-bind Resin

The His-Bind resin (Novagen) comes as a 50% slurry of resin in 20% ethanol to prevent bacterial growth. Twenty milliliters of resin slurry was removed from the stock bottle and placed in a 250 ml beaker. The resin was allowed to settle and the ethanol/water mixture was poured off. The resin was rehydrated with 200 ml of ddH₂O and allowed to settle. Approximately 150 ml of liquid was then removed and a slurry of the remaining water and resin was created. This slurry was carefully poured into a 24 ml BioRad econocolumn (1 cm x 30 cm) creating a 10 ml bed volume of resin. The resin was washed with 100 ml of ddH₂O, 100 ml of Charge buffer (50 mM NiSO₄) and equilibrated with 100 ml of Binding buffer (20 mM Tris-HCl (pH 7.9), 0.5 M NaCl, 5 mM Imidazole).

Cell Lysis and Ni²⁺ Affinity Chromatography

Cells containing HSplase were resuspended in 100 ml Binding buffer and lysed by ultrasonic waves using a High Intensity Ultrasonic Processor (Sonic and Materials, 300 watt model). The cell suspension was subjected to 6 pulses of 45 seconds duration at 70% duty cycle and power setting of 8, on ice until the suspension became a dark tan color and the viscosity was significantly reduced and similar to water. The lysis mixture was loaded into 50 ml centrifuge tubes, the tubes were balanced and placed into a Beckman JA-20 rotor and spun at 19 krpm (43700 xg) for 30 minutes. The resulting cell lysate was transferred to a pre-chilled 125 ml Erlenmeyer flask and the pH was adjusted to 7.9 with 1 M NaOH or 1 M Tris-HCl. The cell lysate was loaded onto a pre-equilibrated His bind column at a moderate flow rate of 1 ml/min. Once the lysate was loaded, the resin was washed with 200 ml of Binding buffer, 200 ml of Wash buffer (20 mM Tris-HCl (pH 7.9), 0.5 M NaCl, 60 mM Imidazole, final pH adjusted to 7.9) the protein was eluted.
with 50 ml of Elution Buffer (0.5 M Imidazole, 20 mM Tris-HCl (pH 7.9), 0.5 M NaCl, final pH adjusted to 7.9) and collected as 1.5 ml fractions. Each fraction was analyzed for the presence of protein by determining the absorbance at 280 nm. Samples of the protein containing fractions were analyzed by separating the products on a 12.5% polyacrylamide gel using SDS-PAGE. Fractions containing the majority of the protein were collected and concentrated using Centriprep-3 (a 3000 MWC ultrafiltration vessel from Amicon). Once the protein had been concentrated to a total volume of approximately 5 ml and equal volume of glycerol which was pre-chilled to -20 °C to prevent denaturation of HSplase upon addition. The resulting protein solution was then aliquoted into twenty 1 ml samples and placed at -20 °C for storage.

DNA Mobility Shift Assay

A 63 mer oligonucleotide containing a single Sp1 binding site 5'-CCG GAG TCA GAA TTC GAA GAC TTG CCG CCG GGG CGG GGC TTC TGC AAT CTG CAG GCC AGC TGT-3' and its complementary strand 5'-AAT TAG AGG TGG GGT GGA GAT TGG AGA AGG GGG GGG GGG GGG GGA AGT GTT GGA ATT GTG AGT-3' were annealed and both strands were 5' end labeled with γ-32P ATP using T4 polynucleotide kinase. The radiolabeled DNA fragment was loaded onto a biospin-6 column to remove unincorporated ATP and change the buffer to binding buffer [10 mM Tris pH 8.5, 50 mM NaCl, 100 μM ZnCl2, 5 mM DTT, 8% glycerol, and 200 ng/μl poly dA/dT (Pharmacia)]. Labeled DNA (20 fmol) was incubated with increasing amounts of HSplase (0, 50, 100, and 200 fmol) at 25 °C for 20 minutes and loaded onto a 5% native PAGE gel (30:1 polyacrylamide: bisacrylamide) in 0.25 X TBE (23 mM Tris-Borate). The
same procedure was followed for the 90-mer oligonucleotide used previously in the mapping of the site of cleavage.

**Assay of HSplase Activity**

DNA substrates (pUCSp1-pUC6Sp1) which were previously constructed by cloning ds DNA oligonucleotides containing the consensus Sp1 binding site into pUC19 utilizing the *EcoRI* and *BamHI* restriction sites, thus creating pUCSp1. This vector was then used to insert multiple copies of a similar ds DNA oligonucleotide into the *XbaI* site, thus generating pUC2Sp1-pUC6Sp1. The activity assay was conducted, as shown in Figure 15, by initially linearizing the purified plasmid with *Alw*NI (a unique restriction site in pUC19 as well as all of the daughter vectors). The linearized vector was then subjected to cleavage using HSplase. The reactions contained 2 μl of 10x Reaction Buffer (100 mM Tris-HCl (pH 8.0), 500 mM NaCl, 20 mM MgCl₂ and 250 mM DTT (final pH adjusted to 8.0 at RT)), 2 μl of 10x BSA (1 mg/ml), 2 μl of 1 mM ZnCl₂, 10 nM pUC5Sp1 and approximately 10 nM HSplase. The protein concentration was estimated by running samples of known protein concentration alongside HSplase and comparing their stained intensity on a 12.5% SDS-PAGE gel. The reactions were then placed at 37 °C for 1 hour and removed. Upon removal, an additional 2 μl of 10x DNA loading dye (0.42% bromophenol blue, 0.42% xylene cyanol FF and 50% glycerol in ddH₂O) was added and the samples immediately loaded onto a 1% agarose gel. Each gel was run under standard agarose gel electrophoresis conditions in TAE buffer (0.04 M Tris-Acetate, 1 mM EDTA (pH 8.0)). The products were visualized by staining in a solution of ethidium bromide (EB, working concentration 5 μg/ml) for 5 minutes. The gels were then
Figure 15. Schematic of HSplase activity assay.
destained for 1 hour to remove nonspecifically bound EB. The results were recorded by taking a photograph upon illumination using a Photodyne UV illuminator. Expected products were DNA fragments of 1.9 kb and 0.8 kb in length.

**Cleavage Site Determination**

**Reactions with HSsplase**

Oligos containing a single Spl binding site (5'-GGGGCGGGGC-3', labeled Spl63) and oligos containing 3 consecutive Spl binding sites (labeled Spl90) were constructed and purchased from Integrated DNA Technologies. The sequences of these oligonucleotides are shown below. Spl63: 5'CCG GAG TGG GAA TTC GAA GAC TTG CCG CCG GGG CGG GGC TTC TGC AAT CTG CAG GCC AGC TGT-3' and the complementary strand 5'-AAT TAC AGC TGG CCT GCA GAT TGC AGA AGC CCC GCC CCG GCG GCA AGT CTT CGA ATT CTG ACT-3'. Spl90: 5'-AAA CGA CGG CCA GTG AAT TCG CCG GGG CGG GGC GAT CCT CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT GCA CTA GGC CGG GGC GGG GCT TCT C4
units. Each reaction was placed at 37 °C in a temperature controlled circulating water bath for 1.5 hours to ensure completion of the phosphorylation reaction. Upon completion, each reaction was loaded onto a PD-10 column, equilibrated with ddH₂O, and spun at 6,000 rpm (Marathon Centrifuge, Fisher) in a 15 ml polypropylene centrifuge tube. The PD-10 columns were used to remove unreacted nucleotides. The collected liquid (~47 µl) was placed in a fresh eppendorf tube, to which was added 10 µl of a 100 µM stock solution of each cold oligo and each reaction was placed at 90 °C for 2 minutes in a temperature controlled water bath. After denaturing the T4 PNK, the reactions were removed from the water bath and placed in a 200 ml beaker filled with 90 °C water and allowed to slowly cool to room temperature behind a beta-shield. HSplase assay reactions were set up as previously described using the labeled oligos as substrate, the labeled oligos were diluted in ddH₂O to give the correct concentration. Activity assays were heated to 55 °C for 5 minutes and quickly placed on ice. Ten microliters of each sample was then loaded onto a 12% denaturing polyacrylamide gel containing 8 M Urea in TBE buffer (90 mM Tris-Borate, 2 mM EDTA) prerun for 30 minutes at 1500 V and washed with fresh 1x TBE buffer. Maxam and Gilbert sequencing reactions of the respective oligo were run side by side with the activity assay to determine the precise location of DNA strand scission by HSplase. Upon completion, the gel was removed from the glass plates and exposed to x-ray film to visualize the result. The signal was amplified using a Kodak Biomax intensifying screen. A schematic of this assay is shown in Figure 16.

Maxam and Gilbert Sequencing

Prepared fresh piperidine formate by adding 75 µl of fresh formic acid to 5 µl of piperidine in 20 µl of ddH₂O. This must be prepared fresh before each use. Each Maxam
This is a schematic of how the cleavage site was determined. The oligonucleotides are initially strand specifically labeled with a radioisotope and annealed to the corresponding cold complementary strand. This substrate is then used in a restriction digestion with Splase. The resulting products are then separated on a denaturing polyacrylamide gel prior to detection with autoradiography. A schematic of the observed results is shown at the right.

Figure 16. Schematic of Cleavage Site Determination Assay.
and Gilbert G+A sequencing reaction was initiated by adding 10 µl of ddH₂O, 2 µl of piperidine formate and 10 µl of labeled oligo (diluted 1:10). Each tube was placed in a temperature controlled water bath at 20 °C for 1.5 hours. The reactions were placed in a speed-vac under vacuum and exposed to medium heat until dry. Each reaction was then redissolved in 100 µl of piperidine (diluted 1:10), sealed with parafilm and placed at 90 °C in a temperature controlled water bath for 30 minutes. Subsequent to this 30 minute incubation the DNA was lyophilized to remove the piperidine. Once lyophilized the DNA was redissolved in 20 µl of ddH₂O three times and lyophilized after each addition. Upon completion, the reactions were redissolved in 20 µl of sequencing loading buffer (98% deionized formamide, 10 mM EDTA (pH 8.0), 0.025 xylene cyanol FF and 0.025% bromophenol blue) and placed at −20 °C for storage, prior to loading on 12% denaturing polyacrylamide gel.

**Purification of FokI**

**Growth and Harvest of E. coli containing FokI**

MM294/λAFB cells were obtained from New England Biolabs. This strain contains the FokI methylase gene as a lysogen in the MM294 chromosome. The pAFPI plasmid was transformed into competent MM294/λAFB and plated out on L-Broth (10 g Tryptone, 5 g Yeast Extract, 10 g NaCl, 1 g MgCl₂, 1 g Glucose in 1 L of ddH₂O, adjust pH to 7.2, MgCl₂ and glucose are added after autoclaving media) plates contained 100 µg/ml ampicillin and 50 µg/ml Kanamycin. The plates are placed at 37 °C for 12 hours. After the 12 hour incubation, 200 ml of L-Broth is inoculated with 1 ml of 20% Glucose, 200 µl of 100 mg/ml ampicillin and a single colony of MM294/λAFB/pAFPI. The liquid culture
is then placed in a rotary shaker at 37 °C for 12 hours. This starter culture is then used to inoculate four 1 L cultures of L-Broth by adding 5 ml 20% glucose, 1 ml 100 mg/ml ampicillin and 10 ml of the starter culture. These cultures are then placed at 30 °C in a rotary shaker until the optical density (OD) at 600 nm reaches 1.0. Once this OD was reached the expression of FokI was induced by adding 70 mg of IPTG. The cells continued to grow at 30 °C for another 3 hours before harvesting. Cells were harvested by placing the liquid culture in 400 ml centrifuge bottles and spinning at 6000 rpm for 10 minutes. The cells were collected in one tube and washed with lysis buffer (100 mM Tris-HCl (pH 8.5), 200 mM NaCl, 1 mM EDTA and 5% Glycerol) and recentrifuged at 6000 rpm for 10 minutes. A typical yield of wet cells was around 15 g. The cell pellet was stored at −20 °C until needed for purification.

**Purification of FokI**

FokI was purified using the published protocol of Waugh and Sauer (PNAS 90, 9596). The purification protocol is described below. The cell pellet is dissolved in 150 ml of Lysis buffer until homogeneous. The cell mixture was placed in 250 ml beaker and 20 ml of 2.5 % polyethyleneimine was added and allowed to stand for 5 minutes at 4 °C. The liquid was then loaded into a 400 ml centrifuge bottle and placed into a Beckman JA-10 rotor at 8500 rpm (10000 xg) for 30 minutes. The supernatent was removed and placed in a 1 L beaker, to which 800 ml of (NH₄)₂SO₄ saturated Lysis buffer was added and allowed to incubate for 20 minutes. The resulting suspension was then loaded into 400 ml centrifuge bottles and spun at 8.5 krpm for 40 minutes. The supernatent was carefully removed and discarded. The pellets were redissolved in 75 ml of cold Low Salt buffer (10 mM Tris-HCl (pH 8.1 at 25 °C), 0.1 mM EDTA and 5% Glycerol) by gently swirling. The solution was then transferred to a 6000 MWC dialysis tubing and dialyzed against 5L
of Low Salt buffer for 12 hours. The dialysis buffer was changed 3 times to 5 L of fresh
Low Salt buffer. The resulting solution was then placed in 40 ml centrifuge tubes and
loaded into a Beckman JA-20 centrifuge rotor and spun at 10000 rpm for 40 minutes.
The supernatant was transferred to a 250 ml Erlenmeyer flask and loaded onto a 70 ml
DEAE (Pharmacia) column at 1 ml/min. The flow through is immediately loaded onto a
20 ml Biorex-70 column (BioRad). The column was washed for 12 hours with Low Salt
buffer and eluted with a 100 ml gradient of Low Salt buffer from 0 mM to 1 M KCl.
Three milliliter fractions were collected using a Pharmacia Fraction collector. Protein
containing fractions were identified by sampling fractions and determining the A_280. A
12.5% SDS-PAGE gel was used to determine the purity of each sample. Samples which
contained protein which could be separated using Sephadex S-200 resin.

Fractions containing protein of sufficient purity were collected and concentrated to
a total volume of 8 ml using an Amicon Centriprep 30. The protein was then loaded onto
a Sephadex S-200 column which was equilibrated with Low Salt buffer. The flow rate
was regulated at 1 ml/min using a Watson and Marlow 101U peristaltic pump. Fifteen ml
fractions were collected using a Pharmacia fraction collector. Each of the fractions were
sampled to identify the fractions containing protein. Fractions containing protein were
sampled and separated on a 12.5 % SDS-PAGE gel. Fractions containing pure protein
were collected and consolidated. These fractions were concentrated using a
Centriprep-30 to a total volume of 5 ml. Cold glycerol was added to a total concentration
of 50% and the protein solution was aliquoted into 1 ml fractions and placed at −20 °C for
storage.
Activity of *FokI*

The substrate vector, pUC19, was selected as a substrate for *FokI* due to the presence of 5 *FokI* binding (5'-GGATG-3') sites within the plasmid. Upon complete digestion this would produce fragments of 1358 bp, 616 bp, 287 bp, 244 bp and 181 bp. These fragments were separated using submarine agarose gel electrophoresis and visualized using Syber green (Molecular Probes) or ethidium bromide staining.

**Purification of Sp1 Zinc finger**

**Growth and Harvest of E. coli**

BL21(DE3) pLysS cells were transformed with pET21Sp1 containing the coding region for the c-terminal 91 amino acids (K533-K623). The addition of an n-terminal methionine brings the total amino acid number to 92 and the expected molecular weight to 11,379 Da. Transformed cells were plated out on LB/amp plates containing 200 μg/ml ampicillin and placed at 37 °C for 12 hours. After the 12 hour incubation, one liter of LB was inoculated with 1 ml of 200 mg/ml amp, and a single colony from the LB/amp plate. This flask was placed in a rotary shaker at 37 °C at 250 rpm until A<sub>600</sub> reached 0.6 (approximately 2.5 hours). After 2.5 hours 100 ml of this culture was used to inoculate 6 x 2 L flasks containing 1 L of LB and 100 μg/ml amp. Each flask was placed at 37 °C in a rotary shaker at 250 rpm until the A<sub>600</sub> reached 1.5. Once the optical density was sufficiently high the cells were induced by adding 1 ml of 0.5 M IPTG. The cells were then grown for and additional 12 hours. After growth the cells were harvested by centrifugation. The cell cultures were loaded into 400 ml Nalgene centrifuge bottles and placed in a Beckman JA-10 rotor. The cultures were subjected to speeds of 6000 rpm or
6,400 xg. The cell pellets were combined and washed with 300 ml of L-Buffer (50 mM HEPES pH 8.0, 10 mM EDTA, 50 mM NaCl, 10 mM DTT and 5 M Urea at 4 °C). The cell pellet was then placed at −20 °C until needed for purification, typical yield 40 g of wet cells.

**Cell Lysis and Purification**

Cells were removed from −20 °C freezer and resuspended in 300 ml of ice cold L-buffer at 4°C. The cell suspension was then sonicated at 70% duty cycle at a power setting of 8 until the viscosity became approximately equivalent to water. The cell lysate was then loaded into 45 ml Nalgene centrifuge tubes, balanced and loaded into a Beckman JA-20 rotor. The rotor was spun at 17 krpm for 1 hour to pellet the cellular debris. The supernatant was collected into an ice cold 500 ml Erlenmeyer flask and loaded onto a 90 ml SP-sepharose column pre-equilibrated with L-Buffer at 1 ml/min. After the protein solution was completely loaded the column was washed with Low salt buffer (50 mM HEPES pH 8.0, 10 mM EDTA 10 mM DTT, 50 mM NaCl). The column was washed until A<sub>280</sub> reached 0.01 or less. Once the A<sub>280</sub> was less than 0.01 a 1 L salt gradient was initiated. The gradient was created in a gradient maker consisting of two 500 ml beakers connected at the bottom with Tygon tubing thus allowing the contents of each beaker to mix. The first beaker contained L-buffer while the second contained L-buffer with 1 M NaCl. The protein was collected using a Phamacia fraction collector as 15 ml fractions. The protein came off the column around 400 mM NaCl. The protein containing fractions were identified by determining the A<sub>280</sub> of each fraction. Protein containing fractions were pooled and placed in 3000 MCO dialysis tubing and placed in 4 L of fresh L-buffer for 12 hours. The dialysis buffer was changed three times and left to dialyze for another 12
hours after each change. The dialyzed protein was then loaded onto another column containing 20 ml of High Performance SP sepharose resin (Pharmacia). The protein was loaded using a peristaltic pump at a flow rate of 3 ml/min. Once loaded, the column was then washed with 1 L of L-buffer and a 100 ml salt gradient was initiated. The two beakers contained 50 ml of L-buffer and 50 ml of L buffer containing 600 mM NaCl. The elution gradient was run at 3 ml/min and 3 ml fractions were collected using a Pharmacia fraction collector. Protein containing fractions were identified by determining the \(A_{280}\). Protein containing fractions were pooled and placed in a 50 ml disposable conical centrifuge tube (Fisher Scientific). The pH of the protein solution was adjusted to 9.0 and DTT was added to a final concentration of 150 mM. The protein solution was then placed in a temperature controlled water bath at 70 °C for 1 hour. The protein solution was then loaded onto a pre equilibrated 95 ml C4 column (Waters) attached to a Waters HPLC system. The protein solution was loaded on 0.75 ml at a time using a 1 ml loading loop and a 1 ml Becton and Dickinson syringe. While loading, the column was continually washed with ddH₂O (HPLC grade for Fisher Scientific) at a flow rate of 5 ml/min. Once the protein was completely loaded the column was washed with 1 L of ddH₂O. Once completely washed the protein was eluted from the column using an acetonitrile gradient from 0 to 40% acetonitrile in ddH₂O both containing 0.1% TFA. The protein eluted at approximately 24% CH₃CN and was collected in a 500 ml roundbottom flask. The protein containing solution was then immediately loaded onto a lyophilizer to remove the acetonitrile. Once lyophilized the protein was redissolved in ddH₂O which had been degassed overnight by bubbling He or Ar gas through it. This cycle was repeated three times to assure total removal of the TFA. Once lyophilized for the final time the flask was placed under vacuum and placed in a desiccator under vacuum at room temperature to reduce oxidation.
NMR

Sample Preparation

Twenty-five milligrams of Spl zinc finger protein was dissolved in 0.5 ml ddH₂O containing 0.5 % D₂O. The pH of the solution was adjusted to 8.0 at which time the solution turned a deep purple color. Upon addition of 3 equivalents of 1 M ZnCl₂ pH 2 the solution turned a light yellow. The pH was adjusted to 7.5 and placed in an NMR tube for analysis.

One Dimensional NMR Methods

NMR methods were performed on a Bruker AM 500 or DMX 600 spectrometer at 300 K. Chemical shifts were referenced to H₂O. The spectral width was 11.0 ppm, and 16 K data points were recorded using the quadrature detection mode. A 90° pulse was applied and 512 transients were obtained per sample.

Purification of GST-HSплase

Growth and Harvest of E. coli

Competent BL21 cells were transformed with pGEX-HSплase or any of the other HSплase mutants (K15A, K15Q, K15D, K52A, K52Q, K52D, K52S, K52T, K80A, K80D, S85D, S85T and QNR) and plated out on LB/amp plates containing 100 μg/ml ampicillin for selection of bacteria containing the plasmid of interest. Each plate was placed at 37 °C for 12 hours. Following this incubation a single colony was used to inoculate 100 ml of
LB containing 2% glucose and 100 μg/ml ampicillin. The liquid culture was grown for 12 hours in a rotary shaker at 37 °C. Five milliliters of this liquid culture was then used to inoculate 0.5 L of 2xYT media (total of 8 flasks containing 0.5 L) containing 2% glucose and 100 μg/ml ampicillin. Each flask was placed at 37 °C in a rotary shaker and grown until the $A_{600} = 1.0$. Once this cell density was reached, the cells were induced by adding 0.5 ml of 0.3 M IPTG. The heat was turned off and the incubator door propped open.

The cultures were allowed to grow for another 3 hours prior to harvest. The cells were harvested by transferring the liquid media to 400 ml Nalgene centrifuge bottles. These bottles were loaded into a Beckman JA-10 centrifuge rotor and spun at 6,000 rpm for 10 minutes. The supernatant was removed and the cells were combined by washing in 1 x PBS and centrifuging for another 10 min at 6,000 rpm. The cells were placed at -20 °C prior to purification of the induced protein. Purification consisted of the same procedure as previously described above for GST-Splase80.
CHAPTER 3

ATTEMPT TO SOLVE THE SOLUBILITY PROBLEM

Introduction

This project began as an attempt to create a new site-specific nuclease which would target the Sp1 binding site of the LTR of HIV-1. In our initial design phase we selected the Sp1 zinc fingers as our DNA recognition domain and the nuclease domain of FokI as our nonspecific nuclease. The fusion of the DNA coding sequences for each of these domains was very straightforward and was easily completed. Expression of the encoded protein was also completed very easily. However, as with every challenging scientific project there appear many tough problems that must be overcome. The initial problem we faced was a solubility problem. The protein was expressed, however, it aggregated within the cell as inclusion bodies. The inclusion bodies could be purified and solubilized using 6M Urea. Upon refolding, however, we did not detect any specific or nonspecific activity. Therefore, at this point I began to look into possible strategies to overcome this problem.

Several corporations were developing new strategies to address this specific problem of overcoming incorporation of the target protein into inclusion bodies by fusing the target protein to a highly soluble bacterial protein. The Glutathione-S-Transferase (GST) system from Pharmacia was among the possibilities. This is a system utilizing the highly soluble GST enzyme. The system works by genetically fusing the coding sequence of the GST enzyme to the coding sequence of the protein of interest. The coding region of the protein of interest is
fused in frame with the last codon of GST. When expressed, the highly soluble GST protein is used to complete the folding of the fusion partner. At the time I began this segment of the project this method had been attempted by many labs with a high degree of success. Therefore, I chose to use this method to attempt to express this protein as a soluble protein and retain the functional activities of each domain. Initially, we were not detecting any nonspecific nuclease activity and we began to wonder if we had obtained all of the necessary amino acid residues required for the overall activity of the nuclease domain. Therefore, as well as attempting to solve this problem by expressing the protein in a soluble form we also attempted to utilize more of the FokI restriction enzyme in an attempt to swap the structure responsible for DNA recognition with the Sp1 zinc fingers. Thus by using the predicted secondary structure as a guide, we attempted to swap structural elements within the DNA recognition domain with the Sp1 zinc fingers. Five constructs were created, Splase80, Splase92, Splase124, Splase130 and Splase133. The number following the protein name, Splase, refers to the number of residues used from the FokI endonuclease protein and begins at proline 377 and proceeds toward the n-terminus. The paragraphs that follow detail the construction, expression, purification and initial assay of these new constructs.

Construction of pGEX-Splase80

If one examines the secondary structure prediction of FokI published by Li and Chandrasegaran the trypsin cleavage site is located in between lysine 283 (K283) and glutamine 384 (Q384). This is proposed to be the beginning of a section of α-helix which is preceded by a stretch of β-sheet. These two structural elements are separated by a random sequence of 7 amino acids beginning with proline 377 (P377). This random sequence was incorporated into the original construct as a flexible region which would allow each domain to fold
independently of each other. The β-sheet segment stretches from asparagine 353 (N353) to isoleucine (I376). This structural element is preceded by a long stretch of α-helix. Glutamate 294 lies at the beginning of this stretch of α-helix. I chose this as my first site of fusion. I designed primers to amplify this region from the FokI genome, incorporating a KpnI site at the 5’ end and a BamHI site at the 3’ end. This was cloned into pTrcHSplase in place of the original nuclease domain. This vector was then used as substrate for a second round of PCR to change the upstream restriction site of Splase80 to BamHI for cloning into pGEX-2T. This amplification not only changed the upstream restriction site to BamHI but also removed the 6His tag. The histidine tag would no longer be needed for purification and I thought that maybe this might be interfering with the protein folding or protein/DNA interaction. Once created, DNA sequencing showed the presence of a K79W mutation. When the original pTrcHSplase vector was sequenced this mutation was also present. Mutating the tryptophan residue to lysine repaired both mutations. This construct expressed very well and was easily purified using Glutathione agarose resin (Sigma). The protein is greater than 90% pure after a single column as examined using SDS-PAGE and stained with silver. A representative of the purified protein can be seen in Figure 17. Therefore, at this point we had a protein which could be expressed and purified as a soluble protein.

The next logical experiment was to find out if the protein was active. Activity assays were set up in which pUC2Sp1 was used as a substrate. The substrate was first linearized with AlwNI, loaded onto a 1% agarose gel and the products separated by applying an electric field. The DNA products were stained using ethidium bromide and visualized by illuminating with ultraviolet light. The single linearized DNA fragment of 2.7 kb was excised from the gel and purified using geneclean. The purified substrate was then digested using GST-Splase80 with increasing amounts of enzyme. The enzyme exhibited nonspecific nuclease activity without specificity for the Sp1 binding site. The results of this experiment can be
This is a 12.5% SDS-PAGE gel of the purification of GST-Splase80. Lane 1, 2 and 4-6 show the first 5 fractions eluted from the Glutathione column. The protein is marked by the presence of the arrow at the left. Lane 3 contains Boehringer Mannheim low range protein molecular weight markers.

Figure 17. GST-Splase80 Purification, SDS-PAGE gel.
seen in Figure 18. This experiment showed that we could express the protein in a soluble form and maintain the structure and functional activity of the nuclease while fusing the protein to the Sp1 zinc fingers. The next logical experiment was to look at binding. Maybe, the protein could not bind the target site or maybe the nuclease domain was so active that it bound but the products were degraded by the nuclease after digestion. In any case, a band shift assay was run to determine if the protein was capable of binding the substrate under the conditions that the assay was run. The substrate, pUC2Sp1 was incubated with increasing amounts of protein. The results of this can be seen in Figures 19. This figure shows a substrate of a single band being shifted to two distinct bands. I contend that these two bands correspond to DNA bound by one GST-Splase80 molecule and one bound by two at each of the respective Sp1 binding sites. In the lanes containing pUC2Sp1 the substrate is shifted to two unique positions and the substrate contains two unique binding sites. This is further evidence to suggest that the protein is binding to the substrate. Therefore, each domain is functioning correctly, however, they do not exhibit the desired cleavage specificity. The conclusion at this point was that the cleavage domain was not oriented correctly with respect to the DNA substrate. This would explain the functionality of both domains without activity. Each of the other constructs were assembled in a similar manner with similar results, therefore this approach was abandoned due to lack of progress toward the goal.

Discussion

Although this approach was abandoned at this point during the project, we learned several things. We knew that at this point that the protein could be expressed in a soluble form. This was the first time a zinc finger had been expressed in a soluble form and with zinc bound as evidenced by the detection of specific DNA binding during the band shift assay. We also knew that the nuclease domain was also active. These experiments showed us that the
This figure shows the initial activity assay of GST-Splase80. Lanes 1-4, 6 and 7 contain 5 nM DNA substrate pUC2Sp1 and increasing amounts of enzyme. Lane 5 contains DNA Molecular Weight Marker VII from Boherenger Maniheim.

**Figure 18.** Activity Assay of GST-Splase80.
This is a mobility shift assay using supercoiled pUC2Sp1 as a substrate and the GST-Splase80 protein. Lanes 1-6 contain the DNA substrate with increasing amounts of enzyme. Lane 7 contains Molecular Weight Marker VII and Lane 8 contains a control plasmid pUC19 with the same amount of protein as lane 5. This shows that the substrate can be shifted to two distinct bands where the control plasmid is not shifted.

**Figure 19.** DNA Mobility Shift Assay of GST-Splase80.
original construct was most likely viable and at the time the decision to abandon this approach came I was beginning to put the original construct into this system. However, at this time we were able to express this protein using the pTrchSpase system in JM105 by lowering the expression temperature to 25 °C. This protein was expressed and partially purified. The protein was shown to have the desired activity while maintaining the desired specificity and the approach I have described was abandoned.

In an attempt to explain why this approach did not give the desired activity I have examined the structure of *FokI*, published July 4, 1996. Assuming that the structure of these nuclease domain remains intact upon fusion to the Sp1 zinc fingers and GST, it appears that there are reasons why this protein may not have exhibited the anticipated activity. If the *FokI* structure is used as a template and the GST-Splase80 construct is created by fusing the zinc finger to the N-terminus of the nuclease domain this would orient the nuclease domain at approximately 90 ° from the DNA substrate. This would allow for the DNA recognition activity of the zinc finger to be present as well as the nonspecific nuclease activity of the nuclease domain. However, the synergistic activity that we were looking for would be very unlikely to the spatial orientation of the nuclease domain with respect to the DNA substrate. This is a possible explanation for the results we observed during this study. Since a working construct was available, we abandoned this approach and began work on the pTrchSpase construct.
CHAPTER 4

INITIAL CHARACTERIZATION OF HSPLASE

Introduction

Initial characterization of the HSplase construct showed that the enzyme was capable of digesting a substrate containing the consensus Sp1 binding site as well the HIV-1 LTR. With the initial design phase of the project complete and a protein exhibiting the properties we desired in hand, the characterization phase was ready to begin. Through careful examination of the properties of the newly designed enzyme we noticed several things. Initially, we detected the desired activity only with substrates containing three or more binding sites as evidenced by digests containing the HIV-1 LTR (three binding sites) and pUC5Sp1 (five binding sites). The second problem we noticed was that the substrate was not completely digested to product even though an excess of enzyme was used and the incubation time was reasonably long enough for complete digestion to occur. Experiments were conducted to address these problems as well as characterize where the construct was cleaving the substrate in relation to the binding site.

The crystal structure of Zif268 shows that the zinc finger binds the DNA substrate in an antiparallel manner. By this I mean that if one looks at the DNA recognition sequence, the 5' end of the substrate would be juxtaposed to the c-terminus of the protein. Thus the first three bases of the recognition sequence would be bound by the third zinc finger and the
third triplet would be bound by the first finger. This means that the fusion of the zinc finger to the FokI nuclease domain would place the nuclease activity 5’ to the DNA recognition site.

**Determination of Substrates which are Cleavable with HSplase**

Upon initiation of this phase of development, I knew that the pUC5Sp1 substrate exhibited the best cleavage pattern under the conditions used for the assay we were using. In order to determine which substrates were able to be cleaved and examine if there was any effect of increasing effective concentration of substrate, each substrate pUCSp1 through pUC6Sp1 was digested under the same conditions. **Figure 20** shows the results of this experiment. This figure depicts the result that only substrates containing three binding sites or more can be cleaved by HSplase. I would like to draw attention to the fact that product concentration does not increase with substrate effective concentration. Therefore this does not appear to account for this phenomenon. In each case the substrate concentration is 10 pM while the enzyme concentration is in approximately 10 to 100 fold excess. This excludes the possibility that the enzyme concentration is just low enough that only this amount of substrate is turned over to product. These experiments suggested that we should examine the binding of protein to DNA using a band shift assay to see if there were any differences between the cleaved substrates and the uncleaved substrates. However, the competitive nature of the project dictated that we finish characterization prior to publication and further development. Therefore, characterization of the site of cleavage took precedence.
This figure shows the digestion of pUCSp1-pUC6Sp1 using HSplase. Lanes 1 and 14 contain DNA molecular weight marker VII. Lanes 2 and 3 contain pUCSp1 without HSplase (lane 2) and with HSplase (lane 3). The lanes are marked above for the presence of HSplase. Lanes 4 and 5 contain pUC2Sp1. Lanes 6 and 7 contain pUC3Sp1. Lanes 8 and 9 contain pUC4Sp1. Lanes 10 and 11 contain pUC5Sp1. Lanes 12 and 13 contain pUC6Sp1. The characteristic pattern is observed in lanes 7, 9 and 13. It is not observed in lane 11 however on subsequent gels this lane also depicts the characteristic fragments, of 1.9 kb and 0.8 kb, generated here.

**Figure 20.** Activity Assay of HSplase on pUCSp1-pUC6Sp1.
Characterization of the Cleavage site of HSplase

The site of cleavage is determined by comparing the length of a digested fragment against the Maxam and Gilbert sequencing products which result from the chemically digested substrate. Initially, the DNA strand of interest is labeled by phosphorylation with $\gamma^{32}$P-ATP. This transfers the radioactive phosphorus to the DNA oligonucleotide which is under examination. In this case, since substrates containing less than three binding sites did not show detectable cleavage a substrate containing three consecutive SpI consensus binding sites was used. This oligonucleotide is then annealed to the complementary strand and used as substrate in the restriction enzyme digestion assay. The assay was initially screened for the correct protein concentration using the unlabeled plasmid substrate. The same assay was conducted with the labeled substrate added as a competitive substrate. The products of these reactions were then separated under denaturing conditions on a 12% polyacrylamide gel in TBE buffer. The products were visualized by exposing the gel to x-ray film. The results of this experiment can be seen in Figures 21 and 22. This figure only depicts the results from the C-rich strand. The results for the G-rich strand were extremely light and a publication quality gel was difficult to obtain. However, on the C-rich strand one can clearly see the presence of the SpI binding site in the Maxam and Gilbert G+A sequencing reactions in lane 1 and the distinctive cleavage positions of HSplase located in lane 2. HSplase cleaves this strand very precisely 13 bp upstream of the binding site as well as 2 and 3 bp upstream. There are also three weaker cleavage sites located 10, 11, and 12 bp upstream of the SpI binding site. The G-rich strand showed very weak cleavages at positions 4 through 9 upstream of the SpI binding site. One observation at this point was that the substrate contained three equivalent SpI binding sites, however, this digestion pattern was only observed at the center site. Therefore, I concluded that maybe our problem with only observing cleavage when we
This figure shows the results from the Cleavage Determination experiments. Lane 1 shows the Maxam and Gilbert Sequencing reactions the actual sequence is presented in text alongside the figure. The lines guide the viewer to the regions of the sequence which are being read. Lane 2 shows the result of the cleavage of the same oligo with HSplase.

Figure 21. Cleavage Site Determination Assay for HSplase.
Figure 22. Summary of cleavage Site Determination Assay Results.
digested substrates containing three or more binding sites was a result of the fact that at least three were required to create a high affinity binding site for the Sp1 zinc fingers. I also realized at this point that there appeared to be a strand preference for the enzyme for the C-rich strand. These results prompted me to have a band shift assay run comparing the binding to substrates containing one Sp1 binding site and three Sp1 binding sites. I also realized that we needed to know more about the FokI enzyme before proceeding any further. In order to complete these experiments I contacted a postdoctoral student in our lab, with extensive experience with bandshift experiments, and New England Biolabs, a corporation with rights to an overproducing clone of FokI. The bandshift experiments are described in the next section. Dr. Ira Shildakraut, at NEB, was kind enough to give me the production strain MM294/λAFB containing the FokI methylase and the overproducing clone pAFP, containing the coding sequence for FokI restriction endonuclease.

DNA Mobility Shift Assay

In order to address the question of weather HSplase binds DNA containing three binding sites with a different affinity than it does a substrate with a single binding site band shift assays were used. The assays were conducted using two separate substrates. The substrate which mimicked the substrate used in previous assays was a 90 mer (also used in the determination of the cleavage site just described). This 90 mer contained three consecutive binding sites each containing the 5’-GGGGCGGGGC-3’ consensus binding site. The substrate containing a single consensus binding site was a 63 mer. These two substrates were used in a mobility shift assay to determine the ability of HSplase to bind substrates with a single binding site as well as substrate with multiple binding sites. The assays were conducted by keeping the substrate constant at 20 fmol and adding an increasing amount of HSplase from
0 fmol to 200 fmol. The results of this assay are shown in Figures 23 and 24. The approximate $K_d$ for each substrate appears to be the same, therefore the affinity of HSplase for a substrate containing a single binding site versus a substrate with multiple binding sites appears to be the same.

Discussion

This result did not explain why we were unable to detect cleavage in assays with substrates with less than three binding sites. Therefore, there must be some other explanation for the observed results. There may be some steric effect in which the nuclease domain is not capable of interacting with the DNA substrate until a sufficient bend is created by the DNA sequence creating a secondary structure which places the nuclease domain in proximity to the DNA substrate. This problem could be overcome by crystallization of the protein DNA complexes of these two substrates and looking at how the nuclease domain is oriented in relation to the DNA substrate. This would then allow for further development and design to improve the activity of the enzyme. The second recommendation was to develop a collaboration with New England Biolabs where we would study the kinetics and mechanism of the $FokI$ restriction endonuclease in exchange for the $FokI$ clone and expression strain. I developed this collaboration with Dr. Ira Schildakraut, Director of Research at NEB, and obtained the $FokI$ clone. The experiments that resulted are described in the next chapter.
This figure shows a mobility shift assay using HSplase and a synthetic 63 mer oligonucleotide containing a single Spl binding site. Lane 4 contains no enzyme and enzyme is increased looking from right to left. The free probe is completely shifted to a single band labeled as bound probe.

**Figure 23.** DNA Mobility Shift Assay of HSplase and substrate With a Single Binding Site
This figure shows the same experiment as shown in figure 23 however, the substrate used was a 93 mer synthetic oligonucleotide containing three Sp1 binding sites. The substrate is not completely shifted, however, three shifted bands can be clearly seen.

**Figure 24.** DNA Mobility Shift Assay of HSplase and Substrate with Three Binding Sites
CHAPTER 5

COMPARISON OF HSPLASE WITH *FokI* RESTRICTION ENDONUCLEASE

Introduction

The initial goal of the project was to create a new restriction enzyme that would recognize the Spl binding sites of the HIV-1 LTR and would cleave double stranded DNA with this designed specificity. Having initially succeeded in accomplishing this goal we decided to improve this enzyme. In order to improve upon this enzyme we needed to know where we were in the development process in relation to the naturally occurring restriction enzymes. Splase was designed to mimic the *FokI* restriction enzyme. Therefore, I decided to compare the activity of *FokI* and Splase under the same conditions to see if we could improve upon our original design. I began talking with several people at New England Biolabs about a collaboration between our group and their corporation where they would give us the *FokI* clone to perform comparisons between our enzyme and *FokI* and we would share our research with them. They graciously accepted the offer to work on *FokI* and we received the *FokI* clone, pAFPI, and the expression strain, MM294/λAFBI, containing the *FokI* methylase. The methylase is required for expression of the restriction endonuclease. If the endonuclease were to be expressed in a cell line where the methylase were not present the host DNA would be site specifically degraded by the endonuclease. The *FokI* restriction endonuclease would be expressed and purified and then used in experiments to optimize the buffer conditions for Splase digestion,
determine if there were any problems during the assay and as a control during Splase stabilization.

**Expression and purification of FokI**

The FokI restriction endonuclease was expressed and purified using the published protocol of Waugh and Sauer. The protein is reasonably pure after the BioRex-70 column. A SDS-PAGE gel that depicts the fractions containing the FokI protein are shown in Figure 25. Fractions containing the FokI protein were pooled and dialyzed against low salt buffer. The protein was then concentrated and loaded onto a Sephadex S-200 column to separate the proteins based on molecular weight. The contamination at this point consisted of proteins which were significantly smaller than FokI. Gel filtration resulted in purification of a homogeneous protein. The results of this column are shown in Figure 26. The protein was then concentrated and stored at -70 °C.

**Initial Characterization of FokI**

The calculated molecular weight of FokI is 65,694 Da. Therefore, as a check to see if this was the correct protein and if there were any mutations I sent a sample to be analyzed by LC-ESI-MS. The results are shown in Figure 27. The mass spectrum of this protein showed that the protein had a molecular weight of 65,696 Da, a difference of two daltons. Which for a protein of this size is insignificant. The next step was to confirm that the protein function was intact. This was conducted by digesting pUC19 with FokI. The pUC19 vector contains 5 FokI binding sites. Therefore, digestion of this plasmid substrate would result in five DNA fragments as products. These fragments would be
This figure shows a representative 12.5% SDS-PAGE gel of the purification of FokI. This is the elution profile of the FokI protein off of the BioRex-70 cation exchange column. The protein is the large protein around 66 kD. Protein fractions can be seen in lanes 1-7 and 9-15. Protein molecular weight markers are located in lanes 8 and 16. The molecular weight markers are as follows from top to bottom: 97 kD, 66 kD, 38 kD, 26 kD, 20 kD and 14 kD.

**Figure 25.** Purification of FokI, SDS-PAGE gel after BioRex-70.
This figure shows a 12.5% SDS-PAGE gel representative of the elution profile of Fokl from the S-200 gel filtration column. The protein is designated by the arrow at the left. The samples of the protein can be seen in lanes 2-15. Protein molecular weight markers are located in lanes 1 and 16.

**Figure 26.** Purification of Fokl, SDS-PAGE gel after gel filtration.
This is the LC-ESI-MS spectrum obtained from the sample purified in figure 26. The protein has a calculated molecular weight of 65694 Da. The mass spectrometry shows that the protein has a molecular weight of 65696 Da. This difference is insignificant.

Figure 27. LC-ESI-MS Spectrum of FokI.
approximately 1358 bp, 616 bp, 287 bp, 244 bp and 181 bp in length and would appear on a 1% agarose gel as shown in Figure 28. The assays were set up as described previously in materials and methods. The products were separated on a 1% agarose gel. In order to examine if this enzyme retained activity during our purification an activity assay was run in which enzyme was incubated with DNA substrate, pUC19, at 37 °C for 1 hour. The reactions were removed from the water bath and the products separated using 1% agarose gel electrophoresis. The results of this are shown in Figure 29. The five fragments listed above can be clearly seen. Therefore, we had purified an active form of the FokI enzyme and we were ready to perform experiments that would begin to identify areas where Splase could be improved.

**Comparison of FokI with Splase**

One initial problem that came to mind was when mutations are made to proteins; many times they are altered such that their structural stability is weakened. Therefore, my initial focus was directed toward comparing the stability of FokI and Splase. The goal of the experiment was to determine the length of time which activity could be observed while incubating the protein at 37 °C. The experiment was conducted by placing substrate in tubes with all of the necessary reagents for DNA cleavage to occur. The protein was placed at 37 °C and the time was recorded. Samples were taken at particular intervals and placed in tubes containing the substrate. These reactions were incubated for 1 hour at 37 °C and then quenched with 0.5 M EDTA and placed at -20 °C. The products were then separated on 1% agarose gel. The results of this experiment are shown in Figure 30. The figure shows that the FokI enzyme is active for over 3 hours at 37 °C under the conditions used. At this point I talked with other people in the field and found out that FokI, under
This is a schematic of what the fragmentation pattern of the pUC19 substrate should look like when digested with *FokI*. The fragments that should be observed are 1358 kb, 616 kb, 287 kb, 244 kb, and 181 kb. Lane 1 depicts molecular weight markers and lane 2 depicts the expected DNA fragments.

**Figure 28.** Schematic of FokI activity assay results
This figure shows the actual activity assay of FokI using increasing amounts of protein from right to left. Lane 1 contains pUC19 plasmid without enzyme added. Lanes 2-7 contain pUC19 plasmid and decreasing amounts of enzyme.

**Figure 29.** Activity assay of FokI.
This figure shows a stability assay of *Fokl*. This is a time course starting in lane 3 with 1 hour. Lane 4: 2 hours, Lane 5: 3 hours, Lane 6: 4 hours, Lane 7: 5 hours, Lane 8: 6 hours.

**Figure 30.** Stability Assay of *Fokl*. 
these conditions should be active for up to 8 hours. Therefore, there was something in my technique which was causing the protein to lose activity. I attempted this same experiment with HSplase. The results are shown in Figure 31. The protein is completely inactive after the first five minute interval. I knew at this point that this problem must be overcome. I began trying new buffer systems, higher salt concentrations, higher and lower pH, adding higher concentrations of BSA and finally diluting the protein into buffer containing 5% glycerol. The last step resulted in a significant stabilization of the enzyme activity. The results of this experiment are shown in Figure 32. This figure shows that activity can be extended out to over 6 hours, a significant increase in stability. I attempted to use these new conditions with HSplase, this resulted in an increased stability out to twenty minutes. This was a huge step in stabilizing this enzyme, however, it was still not even approaching the activity of FokI. Under these conditions however I was able to use a concentration of enzyme which allowed for the complete conversion of substrate to product. The results of these experiments are shown in Figure 33.

These experiments showed me that if the enzyme concentration could be increased then complete turnover of substrate to product could be achieved. This suggested that one of a number of things were going on: (1) the protein was binding and not coming off the substrate in the time frame of our experiment, or (2) the protein was losing activity so fast that at higher concentrations of protein were able to compensate for the rapid loss in activity. At that time there was a huge controversy as to whether or not the FokI enzyme dissociated from substrate after DNA strand scission. In order to look at this phenomenon I ran an activity assay where the substrate was in 20 fold excess. Samples of the reaction were taken at 1 hour intervals and stopped by adding 0.5 M EDTA. The products were separated by agarose gel electrophoresis on a 1% agarose gel. Samples were taken for 8 hours, one sample every hour. The results of this experiment showed
This figure shows the stability assay for HSplase. Lane 1 is after 5 minutes post addition of HSplase. Lane 2 is 10 minutes post addition of HSplase. Lane 3 is 15 minutes post addition of HSplase. Lane 4 is 20 minutes post addition of HSplase.

**Figure 31.** Stability Assay of HSplase.
This figure shows an activity assay using FokI. The substrate is in 20 fold excess. Lane 1 is 2 hour post addition of FokI. Lane 2 is four hours post addition of FokI and lane 3 is 6 hours post addition of FokI. Lane 4 contains Molecular weight marker VII from Boheringer Maniheim. The enzyme is clearly still active after three hours.

**Figure 32.** Activity Assay of FokI under new conditions.
This figure shows a stability assay using the new conditions adopted from the using FokI as a model system. Lane 1 contains DNA molecular weight marker VII. Lane 2 contains pUC4Sp1 and HSplase, 5 minutes. Lane 3 contains pUC4Sp1 and HSplase, 10 minutes. Lane 4 contains pUC4Sp1 and HSplase, 15 minutes. Lane 5 contains pUC4Sp1 and HSplase, 20 minutes. Lane 6 contains pUC4Sp1 and HSplase, 25 minutes.

**Figure 33.** Stability Assay of HSplase under new Conditions.
that the substrate can be completely digested in 8 hours even though the substrate is in excess. This gives evidence for the fact that the FokI enzyme does turn over.

Therefore, if we want Splase to mimic FokI we must make sure that Splase turns over and is stable for longer than 20 minutes. However, at this point we were not able to purify HSplase such that protease contamination could be ruled out. Attempts to purify HSplase utilizing standard protease inhibitors such as PMSF did not effect the stability of HSplase. However, when I added EDTA I was able to purify a protein of approximately 36 kD which could possibly be HSplase. A gel showing the purification of this protein is shown in Figure 34. Samples of this protein were analyzed using LC-ESI-MS. The spectrum generated by this sample is shown in Figure 35. The mass spectrum, generated from the sample, identified the mass of the protein at 38596 Da. The calculated molecular weight for the protein was 35518 Da. Therefore, the protein we purified was too large to be the protein we wanted.

Additional experiments designed to purify the HSplase construct after purification using the His-bind resin were conducted. Using a protein pack gel filtration column form Bio-Rad and the HPLC I began attempting to purify the HSplase protein from the other contaminants. Samples were loaded onto the column immediately after the initial His-bind column. The chromatogram showed a several peaks, a large peak at the solvent front indicated a great deal of aggregation, however three other peaks were also observed. One corresponding to approximately 35 kD was identified by running PIPLC (35 kD) as a standard and a second peak at around 14 kD was identified using Lysozyme as a standard. The protein was allowed to stand at 4 °C overnight and a second sample was loaded onto the column. The chromatogram changed however. The peak at 35 kD was disappearing and the peak at 14 kD was increasing as well as an intermediate peak also increasing. This immediately prompted me to set up an experiment to test if the protein was being
This figure is a representative of a purification of HSplase using SP-Sepharose. The gel is a 12.5% SDS-PAGE showing the protein containing fractions identified using UV absorption to detect their presence. The protein of interest is indicated by the arrow at the left.

*Figure 34.* Purification of HSplase using SP-Sepharose Resin.
This is the LC-ESI-MS spectrum of the protein in Figure 34. The calculated molecular weight of HSplase is 35,518 Da. The molecular weight of the protein via mass spectrometry is 38,596 Da. This difference is too large to be our protein.

**Figure 35.** LC-ESI-MS of Purified HSplase.
degraded at 37 °C. The protein sample was placed in an eppendorf tube at 37 °C and samples were taken every hour, placed in a separate eppendorf tube containing SDS-PAGE loading buffer. The sample was boiled for 5 minutes and placed on ice. After four hours all the samples were collected and they were loaded onto a 12% SDS-PAGE gel. The results of this experiment are shown in Figure 36. The figure shows a dramatic decrease in the Splase band and a distinct increase in the nuclease and zinc finger band. This was conclusive evidence that the protein was being degraded by a protease. Addition of protease inhibitors indicated that EDTA was the only agent which could inhibit the degradation of the enzyme at 37 °C. A thorough search of the literature indicated that a common protease contaminant of proteins purified from E. coli was OmpT. OmpT is a zinc protease. This began to make sense with the results we were seeing and the final conclusion was that a new expression system would need to be used which lacked this protease which recognizes peptides containing a double lysine (KK). The BL21 cell line is OmpT deficient, therefore we turned to this cell line for expression. Initial attempts to express the protein using the pTrcHSplase vector were unable to produce data that supported successful expression of the protein in this cell line. Therefore, we turned to the pET system as the next alternative. The Hopkins group was able to express the protein in this system, therefore we felt that this would be a good alternative which would satisfy our goal of removing the OmpT from the expression cell line. The Hsplase gene was cloned into pET21a and expressed in BL21(DE3), BL21(DE3)pLysS, BL21(DE3)pLysE, BLR(DE3), BLR(DE3)pLysS and BLR(DE3)pLysE. Attempt to express this protein in each of these cell lines were conducted however we were unable to detect the presence of any Hsplase expressed under the conditions previously used. The expression temperature and expression times were also examined, however, no positive results were obtained. Therefore, we concluded that maybe returning to the pGEX system
This figure shows a 12.5% SDS-PAGE gel of the HSplase protein being degraded by a protease. Lane 1 is the protein directly off the column. Lane 2 contains a sample of the protein left at 4 degrees Celsius overnight. Lanes 3-6 contain a sample of the protein in lane 2 left at 37 degrees Celsius for 1 hour, 2 hours, 3 hours and 4 hours. Lane 7 contains low molecular weight protein markers from Boehringer Manheim.

**Figure 36.** Protease Digestion of HSplase.
would allow us to express the soluble protein in the BL21 cell line. The next chapter
describes this work.
CHAPTER 6

CONSTRUCTION OF GST-HSPLASE, PURIFICATION AND ASSAY OF ENZYME ACTIVITY

Introduction

In order to begin to study the properties of Splase and test our hypothesis that the zinc finger was controlling the off rate of the enzyme from substrate an enzyme which could be expressed and purified was required. Previously, we had some positive results with the pGEX system from Pharmacia, when we constructed the GST-Splase80 construct. This protein was soluble, bound DNA site specifically and retained the nuclease activity. Therefore I proposed that we return to this construct and use the HSplase gene because we have already shown that this enzyme is functional. This chapter describes the construction, expression, purification, cleavage with thrombin, and activity assay of this construct, GST-HSplase. The goal of these experiments was to create a protein which could be expressed and purified as a soluble protein and retained the designed functions which were originally outlined in our design phase.

Construction of GST-HSplase

The GST-HSplase gene was constructed using the pGEX-Splase80 and pTrcHSplase vectors. In an attempt to swap the Splase80 and HSplase genes the HSplase gene was amplified using the pTrcHSplase vector as template. The upstream restriction
site was changed from *NcoI* to *BamHI*, however the *NcoI* site was left intact. The fusion of these two sites produced a viable Thrombin site that would allow for specific cleavage yielding a protein beginning with methionine. After PCR, the fragment that resulted was ligated into the pGEX2T vector using standard techniques. The vector was sequenced and the results conclusively showed that every clone was cloned in backwards. Therefore our strategy had to be changed. I decided to use the pGEX-Splase80 gene as the vector and replace the Spl domain and extra nuclease domain with the HSp1 domain from pTrcHSplase. The HSp1 domain was removed from pTrcHSplase by restricting the plasmid with *NcoI* and *HindIII*, purifying the fragment and then ligating it into the pGEX-Splase80 vector digested with the same restriction enzymes. This resulted in a construct that contained a viable HSplase gene. The construct became known as pGEX-HSplase.

The histidine tag was retained for purification purposes once the protein was cleaved with thrombin. This would allow for the purification of the protein away from the GST and GST-HSplase contaminants.

**Optimization Of The GST-Hsplase Expression, Purification And Cleavage With Thrombin**

With a new pGEX-HSplase construct in hand, the expression, purification and Thrombin cleavage phase of this project could begin. Initial expression experiments were conducted by growing BL21 cells transformed with pGEX-HSplase in 2xYT media. These experiments resulted in very low or undetectable expression. Upon isolation of the plasmid DNA after the overnight growth the plasmid seemed to be intact and the gene did not show any signs of recombination. However, when we compared the yield of DNA from pGEX-2T to the newly constructed pGEX-HSplase the yield of the new vector was
significantly lower. This suggested that the number of cells carrying a viable construct were low. Therefore we added glucose to the growth media to further inhibit leaky expression from the pTac promoter used in the pGEX system. Expression experiments using these conditions showed a distinctive improvement. The calculated molecular weight of the GST-HS plase protein is 60,877 Da. The results of the expression experiment are shown in Figure 37. The results clearly show the presence of a new protein band which is just smaller than the 66 kD protein molecular weight marker. These expression conditions were then used to express the protein in 4 L of 2xYT media containing 2% glucose. The cells were grown overnight and then used to inoculate 8 x 0.5 L of 2xYT media containing 2% glucose and 100 μg/ml ampicillin. The cultures were grown for 3 hours or until the cell density reached A600=1.0. The cultures were induced with 0.5 ml of 0.3 M IPTG and grown for an additional 3 hours, at which time the cells were harvested and placed at -20 °C for storage prior to purification. Initial purifications showed that the protein which was expressed was functional. Figures 38 and 39 show an initial purification and an initial assay of a crudely purified GST-HS plase protein.

At this point, while we worked on further purification, we attempted to cleave the protein into the respective domains, GST and HS plase using thrombin. The protein was initially purified using the Glutathione beads and eluted into 50 mM Tris pH 8.0 and 10 mM reduced glutathione. A sample of the fraction containing the highest concentration of GST-HS plase was used to cleave the protein with thrombin. One hundred units of Thrombin were added to the protein solution and the solution was left at room temperature and at 4 °C overnight. Very little cleavage was observed during these incubations. However, when the protein was concentrated and then exposed to thrombin the protein can be clearly shown to be cleaved into a new protein of about 35 kD, HS plase. This is shown in Figure 40. One can also clearly see the disappearance of the
This figure shows an expression experiment for GST-HSplase. Lane 1 contains a sample collected from a flask containing uninduced cells. Lanes 2-4 contain samples from flasks containing cells induced with IPTG. The presence of a new band around 66 kD becomes apparent in these lanes. This is the GST-HSplase protein and is marked by an arrow at the left side of the gel. Lane 5 contains protein molecular weight markers.

**Figure 37.** Expression of GST-HSplase
This figure shows a representative 12.5% SDS-PAGE gel for the purification of GST-HSplsae. Lane 1 contains protein molecular weight markers. Lanes 2-8 contain fractions collected from the glutathione agarose column. The presence of the protein at around 66 kD is the GST-HSplsae protein. This protein is marked by a large arrow at the left side of the figure.

**Figure 38.** Purification of GST-HSplsae.
This figure shows the activity of GST-HSsplase on pUC5Sp1 as a function of time. Lane 1, 5 minutes. Lane 2, 10 minutes. Lane 3, 15 minutes. Lane 4, 20 minutes. Lane 5, 25 minutes. Lane 6, 30 minutes. Lane 7, 35 minutes. Lane 8, overnight. The arrows indicate the positions of the expected fragments of 1.9 kb and 0.8 kb.

Figure 39. Activity Assay of GST-HSsplase.
This figure shows the cleavage of GST-HSplase by Thrombin. Lane 1 contains protein molecular weight markers. Lane 2 contains concentrated GST-HSplase. Lane 3 contains purified HSplase from inclusion bodies. Lane 4 contains GST-HSplase incubated with Thrombin for 1 hour at RT. The arrows indicate the positions of the GST-HSplase band (left) and the HSplase Band (right).

**Figure 40.** Thrombin Cleavage of GST-HSplase.
GST-HSsplase protein. This protein was assayed in the crude mixture and found to be inactive as compared with the same concentration of GST-Hsplase. Therefore we concluded that something was occurring during the digestion with thrombin to alter the structure of HSsplase thus producing an inactive form of the enzyme. Attempts to purify this enzyme using the His-bind resin also failed. The purification always resulted in the presence of GST-HSsplase, HSsplase and GST. The only explanation for this was that the HSsplase was dimerizing with a molecule of GST-HSsplase via the nuclease domain and this was dimerized with a molecule of GST, via the GST domain. GST is known to exist as a dimer, however, the Fokl domain at this point has not been shown to be a dimer. The conclusion at this point was that further purification of GST-HSsplase was required before any further experiments were conducted.

Purification was initiated by redissolving the cell pellet into 1xPBS. The cells were lysed by incubating with 80 mg of Lysozyme for 1 hour. This resulted in a very viscous lysate which was sonicated to further lyse any unbroken cells and ultrasonically shear any DNA which was released during lysis. This resulted in a water like solution of E. coli lysate which was centrifuged to remove the cellular debris left over from lysis. The soluble fraction was collected and loaded onto a pre-equilibrated Glutathione agarose column. The protein lysate was loaded at 1 ml/min and sometimes reloaded if the flow rate was too fast. The column was washed with PBS and then again with 50 mM Tris pH 8.0. The second wash was used to lower the concentration of NaCl thus removing any hydrophobically bound contaminants and changing the pH from 7.3 to 8.0. This is the pH of the elution buffer and the pH at which the assay is run. These changes resulted in significant improvement in purification and the protein was more than 90% pure at this point. Since relatively pure protein could be obtained at this point, new experiments attempting to cleave this protein with thrombin were conducted. These experiments
however failed to result in protein which could be completely cleaved and remain active. Therefore, we decided to continue with our experiments and attempt to test our hypothesis that the zinc finger was controlling the off rate of the enzyme. By mutating residues which interacted with the DNA phosphates, so called affinity residues, we could alter the off rate and increase turnover. The experiments that examine this hypothesis are described in the following chapter.
CHAPTER 7

MUTATION OF AFFINITY RESIDUES TO INCREASE TURNOVER

Introduction

The goal of this phase of the project was to address our hypothesis that the specificity of the zinc finger is defined by the overall binding constant of the protein for a particular DNA sequence. This binding constant is the ratio of the off rate and the on rate. In the case of the zinc fingers the binding constant should be modulated by the off rate of the protein for a particular DNA sequence. Therefore, if the off rate and on rate can be altered such that the ratio is not altered the overall binding constant will not change and the half-life of the complex will be significantly reduced. This will result in a shortened half-life of the protein DNA complex and hopefully an increased enzyme turnover. The approach we took was to look at the zinc finger structure and identify residues that would be good candidates for altering the affinity of the protein for DNA while leaving the specificity intact. Therefore I defined residues that were responsible for interacting with the DNA phosphates as “affinity residues” and the residues that interacted with the DNA bases as “specificity residues”. The strategy of these experiments was to mutate the affinity residues to other amino acids that might allow us to alter the binding affinity. These mutant enzymes would then be assayed for activity. The enzymes that showed a reduced binding constant and increased turnover of substrate to product would be selected as candidates for further evaluation. Mutants that
did not show any improvement or reduced activity would be discarded. The results of these experiments are shown in the following sections.

Construction of Mutants

Mutants were constructed at positions K15, K52, K80, and S85. These positions are believed to interact with the phosphate backbone via either electrostatic or hydrogen bonding interactions. This can be observed using the crystal structure of Zif268 as a model. The amino acids at these positions are different however we assumed that they would play a similar role. Mutagenesis was conducted at these positions using the Quickchange method. This method utilizes two complementary mutagenic primers which anneal to each of the template strands (which are purified from a strain which methylates DNA) and using PCR they are extended to complete the replication of the entire plasmid vector. The resulting vectors are unmethylated and thus resistant to digestion with DpnI, a restriction enzyme that recognizes methylated DNA. The reactions are then digested with DpnI to remove the unwanted template DNA and the mixture is transformed into competent cells. The cells containing plasmid are selected for using an antibiotic and then the resulting DNA is sequenced to confirm the presence of the desired mutation. The mutants that were constructed are listed in Table 2. Each mutant was constructed using quickchange, sequenced using TaqFS dyeterminator reactions, expressed, purified and assayed using the standard Splase activity assay. The substrate concentration was held constant at 10 nM (500 ng) and the enzyme concentration was varied from 50 nM to 5 μM. The concentration dependence of substrate

119
| Mutant | Expression | Purification | Activity | |
|--------|------------|--------------|----------| |
| K52Q   | Good       | Good         | Nonspecific Activity |
| K52A   | Good       | Good         | Significantly Reduced |
| K52D   | Good       | Lower        | None detected |
| K52S   | Good       | Good         | Activity restored |
| K52T   | Good       | Good         | Nonspecific Nuclease |
| K15A   | Good       | Good         | Star Activity |
| K15R   | Good       | Good         | Similar to WT |
| K15D   | Good       | Lower        | Similar Activity to WT without specificity for Sp1 sites |
| S85D   | Good       | Good         | Nonspecific Nuclease |
| S85T   | Good       | Good         | Very little detected |
| K80A   | Good       | Lower        | Slightly reduced activity with reduced star activity |
| K80D   | Good       | Lower        | Slightly higher than WT and more Star Activity |

Table 2. Table of mutants constructed, characterizations of their expression, purification and activity as compared to GST-HSplase.
digestion was examined and compared to a similar assay conducted using the WT enzyme. The results of these assays are shown in Figures 41 through 50. Figure 41 shows the activity assay conducted with WT enzyme. This assay shows that the enzyme begins to show specific digestion around 200 nM enzyme; at 300 nM enzyme significant specific digestion is observed however there is now also a significant amount of star activity. Star activity is defined as activity exhibited at other weaker sites other than the high affinity site. These can be seen as the bands that are in between the 1.9 kb and 0.8 kb fragments. This is another problem which must one day be overcome if this is to become a useful method for creating new restriction enzymes. If one looks at the K52 mutants shown in figures 42 through 46. These mutants have almost completely lost specificity except K52S. The nuclease domains of these mutants appear to be active, however, the specific DNA sequence recognition has been destroyed. The K52S mutant, figure 45, has somehow retained activity and would then be a good candidate for further study. However, the conversion of substrate to product is not enhanced. Therefore, this mutant is also not desirable. The mutations at K80 resulted in very different results. The K80A mutant resulted in a significant increase in nonspecific nuclease activity. Specific activity was observed, however, at a much higher protein concentration (approximately 700 μM enzyme). The K80D mutant retained activity, exhibited activity at a lower concentration than WT and showed increased turnover of substrate to product. However, it also showed an increased star activity. If the star activity can be minimized this would also be a good candidate for further development. The mutations at the S85 position were also either significantly reduced activity as in the case of S85T or none detected, as in the case of S85A. Without structural information on each of these mutants it becomes very difficult to determine what effect if any they are having on the overall structure of the protein or if other residues within the protein are compensating for the mutation.
Activity assay using WT GST-HSplase. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

Figure 41. Activity assay of WT GST-HSplase.
Activity assay using GST-HSplat K52A. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplat. Protein concentrations:

Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 42.** Activity assay of GST-HSplat K52A.
Activity assay using GST-HSsplase K52D. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSsplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 43.** Activity assay of GST-HSsplase K52D.
Activity assay using GST-HSplaces K52Q. Lanes 1 and 9, control, pUC5Spla cut with EcoRI. Lane 2-16 digestion of pUC5Spla using GST-HSplaces. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 44.** Activity assay of GST-HSplaces K52Q.
Activity assay using GST-HSsplase K52S. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSsplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

Figure 45. Activity assay of GST-HSsplase K52S.
Activity assay using GST-HSplase K52T. Lanes 1 and 9, control, pUC5Sp1 cut with 
EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplase. Protein concentrations:
Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. 
Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 
nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 46.** Activity assay of GST-HSplase K52T.
Activity assay using GST-HSsplase K80A. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSsplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 47.** Activity assay of GST-HSsplase K80A.
Activity assay using GST-HSplase K80D. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

Figure 48. Activity assay using GST-HSplase K80D.
Activity assay using GST-HSplease S85T. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplease. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

**Figure 49.** Activity assay of GST-HSplease S85T.
Activity assay using GST-HSplase S85D. Lanes 1 and 9, control, pUC5Sp1 cut with EcoRI. Lane 2-16 digestion of pUC5Sp1 using GST-HSplase. Protein concentrations: Lane 2, 50 nM. Lane 3, 100 nM. Lane 4, 200 nM. Lane 5, 300 nM. Lane 6, 400 nM. Lane 7, 500 nM. Lane 8, 600 nM. Lane 10, 700 nM. Lane 11, 800 nM. Lane 12, 900 nM. Lane 13, 1 μM. Lane 14, 2 μM. Lane 15, 3 μM. Lane 16, 5 μM.

Figure 50. Activity assay of GST-HSplase S85D.
Discussion

Therefore, it is my recommendation that further structural work be conducted on GST-HSpline before any further engineering is conducted. We must fully understand the orientation of the nuclease domain with respect to the DNA substrate. We must also do extensive evaluation of the mechanism of the FokI restriction enzyme. If this enzyme is functioning as a dimer we must make sure that our construct also follows this mechanism. Further work must also be conducted in the area of zinc finger DNA interactions. We must know the functional contributions of each amino acid to the overall binding affinity and specificity, before we begin attempting to alter the binding constant to change the off rate. This would be better conducted using the zinc finger system using a gel shift assay or by examination of the on and off rates using BIAcore. It is the opinion of the author that unless these detailed studies are conducted this project will go no further and no further information will be learned about protein design.
CHAPTER 8

EXPRESSION, PURIFICATION AND RECONSTITUTION OF THE Sp1 ZINC FINGER FOR STRUCTURAL AND FUNCTIONAL STUDIES

Introduction

The ultimate goal of this development phase is to engineer Splase into a more efficient enzyme. Therefore, we have taken the approach that if we are able to increase the off rate of the zinc finger for the DNA substrate without altering the specificity we will have achieved this goal. The most straightforward approach to examining this problem is to make mutations in Splase and observe their effects on the overall activity of the enzyme. At the time at which this work was undertaken, we were unable to produce a Splase protein that was stable enough to examine this hypothesis. Therefore, we decided to look at these same properties using the Sp1 zinc finger domain. Ultimately, this would need to be performed to make sure that nothing had changed structurally or functionally upon fusion to the nuclease domain. The Sp1 zinc finger had already been expressed and purified previously by Kriwacki et. al. Therefore, we could clone the gene, insert the appropriate start and stop codons, express and purify the protein under the published conditions and begin our study of the zinc finger/DNA interactions. The DNA binding would be assessed using BIAcore and the structural integrity would be examined using NMR.
The gene encoding the three zinc fingers of the Sp1 protein were cloned into the pET21d vector using PCR and utilizing the pTrcHSplase vector as template. The downstream restriction site was changed from KpnI to BamHI and a termination codon was inserted at the end of the gene prior to the BamHI restriction site. These modifications would also allow for the removal of the 6-histidine tag. This NcoI/BamHI fragment was excised from the TA cloning plasmid, pCRSp1 and cloned into the pET expression vector for expression using established protocols. The resulting vector, pETSp1 was sequenced and transformed into BL21(DE3) and BL21(DE3)pLysS for expression experiments. The vector did not transform BL21(DE3), however BL21(DE3)pLysS was transformed. This transformed strain was used to express and purify the Sp1 zinc fingers.

The Sp1 zinc fingers were expressed by initially transforming the pETSp1 vector into BL21(DE3)pLysS. The cells were plated out on LB/amp plates and grown overnight. A single colony was used to inoculate a 2 L flask containing 1 L of LB and 200 μg/ml ampicillin. This flask was placed at 37 °C for approximately 6 hours, or until the optical density at 600 nm reached 0.6. At this time, 70 ml of the culture was used to inoculate 8 x 2 L flasks containing 1 L of LB and 200 μg/ml ampicillin. These flasks were placed at 37 °C for 3 hours or until the optical density reached 1.5. At this point the flasks were induced by adding 1 ml of 1 M IPTG and the cells grown for three additional hours. The SDS-PAGE gel in Figure 51 shows the expression of a low molecular weight protein around 12 kD. The molecular weight of this protein is calculated to be 11,596 Da. The
20% SDS-PAGE gel showing the expression of the Sp1 zinc finger. Lane 1 is protein molecular weight markers. The lowest of which is 14 kD. Lane 2 contains extract from cells which did not undergo induction with IPTG. Lane 3 contains extract from cells which did undergo induction of the Sp1 zinc finger with IPTG. The protein of interest is highlighted by the arrow at the right side of the figure.

**Figure 51.** Expression of Sp1 Zinc Finger.
protein was purified using a published protocol, however significant modifications had to be made. Attempts to determine the mass using mass spectrometry (LC-ESI-MS) failed, no proteins in this mass range were observed. Initial attempts to look at protein folding by the addition of zinc using NMR also failed. Examination of the NMR sample using a 20% SDS-PAGE gel confirmed my suspicions that the protein was being oxidized. The gel in Figure 52, lanes 1-8, shows a serial dilution of this sample showing that even under the reducing conditions of an SDS-PAGE gel, oligomerization appears to be occurring.

Conversations with the leading scientists in this area indicated that we needed to be much more careful about how we purified the protein. They suggested that we reduce the protein extensively and store it in an oxygen free atmosphere. I immediately set out to put together a glove box that would satisfy this criterion. In the mean time I also began looking at ways we could satisfy this condition without using the glove box. This problem was overcome by several additions to the protocol.

Initially, as the protein comes off the HPLC column the protein is collected into a 500 ml roundbottom flask. The acetonitrile is removed using lyophilization. The rest of the water is also removed by lyophilization. The protein is then redissolved in degassed water three times. The water is degassed by bubbling He or Ar through the water for 6 hours. At the end of the lyophilization phase the protein is redissolved in degassed water, transferred to eppendorf tubes and lyophilized for storage. After lyophilization the protein is placed under vacuum at room temperature in a vacuum desiccator. The protein was stored for several months under these conditions. The results of this procedure can be observed in Figure 52, lanes 10 and 11. Therefore, one can conclude that a significant reduction in oxidation has occurred and this sample can be used for structural and functional studies of the zinc finger.
This figure shows the improvement of the new conditions upon oxidation of the Sp1 zinc finger. Lanes 1, 8 and 9 are protein molecular weight markers. Lanes 2-7 are a serial dilution of the oxidized Sp1 zinc finger. Lanes 10 and 11 are the same concentration as lanes 2 and 3 however these are dilutions of the reduced zinc finger.

**Figure 52.** Purification of Zinc Finger/Oxidation Problem.
**Zinc Titration Using NMR**

The literature was used as a source for a 1 dimensional proton NMR spectrum of the Zif268 first zinc finger. This spectrum identified four characteristic peaks in the spectrum that they assigned to the histidine and aliphatic peaks which result upon folding of the zinc finger after addition of zinc. The spectrum is shown in Figure 53 (83). The characteristic peaks are labeled with arrows. The peaks in the 0.0 to 0.5 ppm region are the aliphatic peaks and the peaks around 6.5 ppm are the histidine peaks. Therefore, since this is a single zinc finger and it has two histidines which obtain two of the coordination positions on the zinc metal and are shifted upon doing so, we should see six shifted histidine peaks in the Sp1 spectrum. Sp1 has three fingers and should have six histidines coordinating the three zinc atoms. The chemical environments of each of these histidines may be significant enough to observe all six.

The sample was reconstituted in degassed water. The protein solution, 2.5 mM (0.5 ml) was loaded into a 5 ml slidealyzer for microdialysis to allow the protein to equilibrate with the new buffer (1 mM HEPES pH 8.0, 2 mM DTT). After dialyzing overnight with bubbling with He. The sample was removed with a 1 ml syringe and placed in an eppendorf. The sample was lyophilized and redissolved in 400 μl of degassed ddH₂O. 5% D₂O was added as an isotope for signal lock during the NMR experiment and the pH was adjusted to 7.5 with dilute NaOD. The initial spectrum was taken using this sample and is shown as the top spectrum in Figures 54 and 55. This spectrum is very sharp, characteristic of an unfolded state of the protein. Upon the addition of 2.5 equivalents of ZnCl₂ the pH dropped to 6.0. The pH was adjusted back to 7.5 with dilute NaOD. The sample was returned to the NMR tube and the second spectrum, located in
This figure was taken from Parraga et. al. (83).

**Figure 53.** 1D proton NMR spectrum of Zif268 finger 1 (83).
Figure 54. Aliphatic region of the 1D-Proton NMR spectrum of the Spl zinc finger while titrating with zinc. Spectrum before adding zinc (top). Spectrum after adding 2.5 equivalents of zinc (middle). Spectrum after adding 5 equivalents of zinc.
Figure 54. 1-D Proton NMR Spectrum of Sp1, Aliphatic Region.
Figure 55. Amide region of the 1H-Proton NMR spectrum of the Sp1 zinc finger while titrating with zinc. Spectrum before adding zinc (top). Spectrum after adding 2.5 equivalents of zinc (middle). Spectrum after adding 5 equivalents of zinc.
Figure 55. 1-D Proton NMR Spectrum of Sp1, Amide Region.
Figures 54 and 55, resulted. This spectrum clearly shows the presence of six peaks around the same chemical shift as the two histidine peaks shown in the Zif268 spectrum. There is also the presence of the two aliphatic peaks located between 0 and 0.5 ppm. This is a very good representative of what the folded zinc finger proton NMR spectrum should look like. Upon further addition of zinc to 5 equivalents, the spectrum broadens and the spectrum loses the characteristic peaks. This is shown in Figures 54 and 55 as the third spectrum.

Discussion

Upon examination of the data we found that in order to determine the exact concentration of the folded protein we decided that each zinc finger would need to be titrated with zinc to determine the exact concentration needed to saturate the zinc binding sites without giving an excess. This would result in a project that would require a large amount of NMR time and was not a simple method for sample preparation. Therefore, we decided to focus our attention on the newly constructed and purified enzyme GST-HSplase. This construct would allow us to test our hypothesis without having to use an elaborate sample preparation scheme. This would tell us if our hypothesis was worth pursuing. If it was worth pursuing, then we would return to this scheme and we would work with the Sp1 zinc finger to extensively map out the overall contributions of each amino acid to the overall structure and binding affinity of the protein for DNA.

As a result of the work explained in Chapter 7, we have decided that no further engineering of HSplase should be pursued until the protein DNA interactions of Sp1 are completely understood.
REFERENCES


