Structural and biophysical studies of single-stranded DNA binding proteins and dnaB helicases, proteins involved in DNA replication and repair

Vinu Johnson

The University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation

http://utdr.utoledo.edu/theses-dissertations/1289
A Dissertation

Entitled

*Structural and Biophysical Studies of Single-Stranded DNA Binding Proteins and dnaB Helicases, Proteins Involved in DNA Replication and Repair*

By

*Vinu Johnson*

Submitted as partial fulfillment of the requirements for

the *Doctor of Philosophy in Chemistry*

Advisor: Timothy C. Mueser, Ph.D.

Graduate School

The University of Toledo

*December 2007*
The processes of DNA replication, repair and recombination have been studied for many years using model systems such as *E. coli*. One key aspect of DNA replication is the role of proteins at the replication fork. It is important to understand the interactions of these proteins with the other proteins involved in the process or with a DNA substrate.

In PART I of this dissertation, studies of the Archaeal *Aeropyrum pernix* and *Archaeoglobus fulgidus* single-stranded DNA binding protein are presented. Archaea are more closely related to eukaryotes than the commonly studied *E. coli* model system. The SSB proteins have an involvement in Okazaki fragment processing. The FEN-1 enzyme removes DNA flaps that are 3 to 5 nucleotides long. If the length of the flap is more than
5 to 7 nucleotides, two other proteins are then involved in the removal of these flaps, the SSB protein that binds to the flap, and dna2 needed to shorten the flap.

Using PCR and molecular cloning techniques and protein chemistry, milligram quantities of the purified *Ape* SSB protein were obtained. Preliminary attempts to crystallize the full length protein failed, most likely due to highly flexible C-terminal region. To overcome this problem, a truncated version of the protein was prepared. The truncated protein crystallized in two different crystallization conditions. One produced beautiful large 1mm hexagonal crystals that appeared twinned on diffraction. The other crystal form, rod shaped small crystals, diffracted to almost 1.6 Å. The structure of the truncated protein was solved using molecular replacement with an already known SSB protein from the species *Sulfolobus sulfataricus*. Biophysical studies, Dynamic Light Scattering (DLS) and Fluorescence Anisotropy, have been carried out to characterize the *Ape* SSB protein.

In PART II, work done on DNA helicases (dnaB) from different organisms is presented. Expression and purification protocols for *Vibrio cholerae* dnaB were successfully established and crystal screens set up. Few hits were obtained in the screens but expansion of some of the conditions did not yield any crystals. Cloning, expression and purification protocols were established for *Yersinia pestis* dnaB and *Helicobacter pylori* dnaB.
Acknowledgements

First and foremost I would like to thank my advisor Dr. Mueser for all the help and advice. The time I spent during this work has helped me in understanding and learning science in a much more effective way. I would also like to thank my committee members, Dr. Funk, Dr. Huang, and Dr. Ashburner for their help and assistance. I also wish to thank the University of Toledo, Department of Chemistry and their staff especially Charlene. I would also like to extend my gratitude to Dr. Leif Hanson for his help. Special thanks also go to all the members of the Mueser lab – past and present for their help and support in my work, especially Pooja who insisted on me joining the lab and Juliette for her support and help especially when I was bogged down and ready to quit. I would also like to thank Dr. David Dignam of UT Health Science campus for his assistance and advice.

I would also like to thank my friends and family, especially my parents and my brother, for all their support and encouragement. Special thanks also go to my “American family” – Bill & Ella Werth who not only provided me a place to stay but also supported and encouraged me at every step during the course of my Ph.D. I would also like to thank a few people who made a difference in my life in Toledo – Dr. Jeff Wirebaugh, Bob and Cindy Hoenig.
# Table of Contents

ABSTRACT ............................................................................................................................... III

ACKNOWLEDGEMENTS .............................................................................................................. V

TABLE OF CONTENTS .............................................................................................................. VI

LIST OF TABLES ....................................................................................................................... XI

LIST OF FIGURES ..................................................................................................................... XIII

LIST OF ABBREVIATIONS ........................................................................................................ XXIV

PART I: Studies of Single-Stranded DNA Binding Protein ........................................... 1

Chapter 1: Introduction ........................................................................................................... 2

1.1. T4 gp32 ................................................................................................................................. 4

1.2. E. coli SSB ........................................................................................................................... 8

1.3. Replication Protein A (RPA) ............................................................................................ 12

1.4. Archaeal SSB .................................................................................................................... 16

1.4.1. *Aeropyrum pernix* ........................................................................................................ 19

1.4.2. *Archaeoglobus fulgidus* ............................................................................................ 21

Chapter 2: Materials and Methods .................................................................................. 23

2.1. Molecular Biology ............................................................................................................ 23

2.1.1. Polymerase Chain reaction (PCR) ............................................................................. 23

2.1.2. Gel Purification ............................................................................................................. 26

2.1.3. Blue-white cloning using pDrive vector ................................................................. 27

2.1.4. Preparation of Competent Cloning hosts – XL10 ................................................... 28

2.1.5. Transformation and isolation of Blue-white colonies .......................................... 29

2.1.6. Miniprep to isolate plasmid ....................................................................................... 30

2.1.7. Use of Restriction Enzymes ...................................................................................... 31

2.1.8. Ligation of Restriction products and Transformation into Cloning host ............ 32
2.1.9. Transformation into XL10 cloning host ................................................................. 33
2.1.10. Transformation of cloned Expression vector into Expression Host .................... 34
2.1.11. Quick Change for truncated protein ................................................................. 34

2.2. **Protein Expression, Purification and Solubility** ...................................................... 36
   2.2.1. Small Scale Protein Expression and Solubility Study ........................................... 36
   2.2.2. Large Scale Protein Expression Study ............................................................... 38
   2.2.3. Cell Lysis .......................................................................................................... 39
   2.2.4. Protein Purification ............................................................................................ 40
   2.2.5. Solubility Profile ............................................................................................... 43
   2.2.6. Protein concentration ....................................................................................... 45

2.3. **Protein Crystallography** ....................................................................................... 49
   2.3.1. Preliminary Crystal Screening ............................................................................ 49
   2.3.2. Expansion and Optimization of Preliminary Crystal Hits .................................... 51
   2.3.3. Crystal manipulation for X-ray diffraction ......................................................... 52
   2.3.4. X-ray Diffraction Screening and Data Collection ............................................... 53
   2.3.5. Data Processing ............................................................................................... 54
   2.3.6. Structure Determination ................................................................................... 55
   2.3.7. Model Building and Refinement ................................................................. 56

2.4. **Analysis of Quaternary State and DNA Binding Affinity** .................................... 57
   2.4.1. Dynamic Light Scattering .................................................................................. 57
   2.4.2. Size Exclusion Chromatography ....................................................................... 58
   2.4.3. Fluorescence Anisotropy .................................................................................. 59
   2.4.4. Analytical Ultra-Centrifugation (AUC) .............................................................. 63
      2.4.4.1. Sample preparation ....................................................................................... 63
      2.4.4.2. Sedimentation Velocity (SV) ....................................................................... 63
      2.4.4.3. Sedimentation Equilibrium (SE) ................................................................. 64

Chapter 3: **Aeropyrum pernix SSB (Ape SSB)** ................................................................. 66
3.1. **Ape SSB full length** ............................................................................................. 66
   3.1.1. Cloning ............................................................................................................. 66
   3.1.2. Protein Expression study .................................................................................. 67
   3.1.3. Solubility Profile ............................................................................................. 67
   3.1.4. Protein Purification ......................................................................................... 69
   3.1.5. Crystal Screening ......................................................................................... 75
   3.1.6. Analysis of Quaternary State and DNA Binding Affinity ................................ 75
3.1.6.1. Dynamic Light Scattering ................................................................. 75
3.1.6.2. Size Exclusion Chromatography (SEC) ........................................ 79
3.1.6.3. Fluorescence Anisotropy ............................................................... 79

3.2. Ape SSBΔC (truncated) ...................................................................... 85
  3.2.1. Cloning (Quick Change) ............................................................ 85
  3.2.2. Protein Expression study ......................................................... 86
  3.2.3. Protein Solubility Study ......................................................... 87
  3.2.4. Solubility Profile ................................................................. 88
  3.2.5. Protein Purification ............................................................... 90
  3.2.6. Crystal Screening ................................................................. 96
  3.2.7. Optimization of crystal hits ............................................. 97
  3.2.8. Preliminary X-ray diffraction screening ..................... 99
  3.2.9. Data Collection ................................................................. 105
  3.2.10. Data Processing .............................................................. 106
  3.2.10. Molecular Replacement .................................................... 109
  3.2.11. Refinement and Model Building ........................................ 111
  3.2.12. Analysis of Quaternary State and DNA Binding Affinity ...... 120
    3.2.12.1 Dynamic Light Scattering .............................................. 120
    3.2.12.2. Size Exclusion Chromatography (SEC) ....................... 124
    3.2.12.3. Fluorescence Anisotropy ........................................... 124
    3.2.12.4. Analytical Ultracentrifugation ..................................... 132

Chapter 4: Archaeoglobus fulgidus SSB (Afu SSB) ............................ 134
  4.1. Cloning .................................................................................. 135
    4.1.1. PCR .............................................................................. 135
    4.1.2. Blue-white cloning using pDrive cloning vector .............. 136
    4.1.3. Restriction .................................................................... 137
    4.1.4. Ligation and transformation into cloning host ................ 138
  4.2. Protein Expression study ......................................................... 139
    4.2.1. Small scale expression study ....................................... 139
    4.2.2. Solubility study .......................................................... 139
    4.2.3. Solubility Profile ........................................................ 140
  4.3. Protein Purification ............................................................... 142
  4.4. Crystal Screening ................................................................. 147
    4.4.1. Study to determine protein concentration for crystal screening 147
4.5. Analysis of Quaternary State and DNA binding Affinity............................... 148
  4.5.1. Dynamic Light Scattering ................................................................. 148
  4.5.2. Fluorescence Anisotropy ................................................................. 154

Chapter 5: Discussion ....................................................................................... 157

PART 2: DNA HELICASE ............................................................................. 165

Chapter 6: Introduction .................................................................................. 166
  6.1. Classification of Helicases ................................................................. 166
    6.1.1. SuperFamily 1 (SF1) ................................................................. 168
    6.1.2. SuperFamily 2 (SF2) ................................................................. 168
    6.1.3. SuperFamily 3 (SF3) ................................................................. 170
    6.1.4. SuperFamilies SF4, SF5 and SF6 ................................................. 173
  6.2. Disorders associated with Helicases ....................................................... 174
  6.3. Helicases in this study .......................................................................... 175
    6.3.1. 
      *Vibrio cholerae* (Vc) dnaB ......................................................... 175
    6.3.2. 
      *Yersinia pestis* (Yp) dnaB .......................................................... 177
    6.3.3. 
      *Helicobacter pylori* (Hp) dnaB ................................................. 179

Chapter 7: Materials and Methods ................................................................. 182
  7.1. Molecular Biology .............................................................................. 182
    7.1.1. Polymerase Chain Reaction (PCR) ........................................... 182
    7.1.2. Insertion into cloning vector using Gateway® Entry vector ..... 183
    7.1.3. Switching from Entry to Destination vector ......................... 184
  7.2. Protein Expression, Purification and Solubility .................................... 187
    7.2.1. Protein Expression, Solubility and Cell Lysis .......................... 187
    7.2.2. Solubility Profile (modified protocol) ...................................... 187
    7.2.3. Protein Purification (modified protocol) ................................. 187

Chapter 8: 
  *Vibrio cholerae* (Vc) dnaB ................................................................... 189
  8.1. Molecular Biology .............................................................................. 189
  8.2. Protein Expression Study ................................................................... 190
    8.2.1. Small Scale Expression study ............................................... 190
    8.2.2. Solubility Study ...................................................................... 191
List of Tables

Table 1.1: Protein characteristics of the full length Ape SSB obtained using ProtParam program (ExPASy Proteomics server) ............................................................. 20

Table 1.2: Protein characteristics of Ape SSBΔC obtained using ProtParam program (ExPASy Proteomics server) ................................................................. 21

Table 1.3: Afu SSB gene and protein characteristics determined using Protparam program (ExPASy Proteomics server) ............................................................ 22

Table 2.1: Components of a typical PCR ......................................................................... 24

Table 2.2: Outline of a PCR program .............................................................................. 25

Table 2.3: Components of pDrive cloning reaction ........................................................... 28

Table 2.4: Components of restriction enzyme reaction .................................................. 31

Table 2.5: Components of a ligation reaction .................................................................. 32

Table 2.6: Quick change® reaction components. .............................................................. 35

Table 2.7: Program for Quick change® reaction ............................................................... 35

Table 2.8: List of different salts and buffers used in the solubility screen ...................... 44

Table 2.9: List of different crystal screens used in the study ............................................. 50

Table 3.1: Summary of results of fluorescence anisotropy experiment for Ape SSB ...... 84

Table 3.2: Summary of results for calculation of Mathew’s coefficient for different space groups analyzed .............................................................................. 107

Table 3.3: Summary of results of SCALA for different space groups analyzed ............. 108

Table 3.4: Determining success of PHASER ................................................................... 110

Table 3.5: Result of molecular replacement using PHASER ........................................... 111
Table 3.6: Summary of Refinement steps for determining *Ape* SSBD\(\Delta\)C crystal structure.  
........................................................................................................................................... 115

Table 3.7: Data Collection, Structure Determination and Refinement of *Ape* SSBD\(\Delta\)C . 117

Table 3.8: Summary of results from fluorescence anisotropy experiment for *Ape* SSBD\(\Delta\)C  
........................................................................................................................................... 131

Table 4.1: Summary of *Afu* SSB DLS experiment................................................................. 149

Table 6.1: Protein characteristics of the full length *Vc* dnaB.............................................. 177

Table 6.2: Protein characteristics of the full length *Yp* dnaB.............................................. 179

Table 6.3: Protein characteristics of the full length *Hp* dnaB .............................................. 181

Table 7.1: Components of a TOPO cloning reaction............................................................. 183

Table 7.2: Gateway\(^\text{®}\) LR Clonase reaction components ............................................. 185

Table 8.1: Table showing solution state of *Vc* dnaB.......................................................... 198

Table 8.2: Summary of *Vc* dnaB DLS experiment.............................................................. 205
List of Figures

Figure 1.1: Cartoon showing involvement of RPA (eukaryotic SSB protein) in Okazaki fragment processing. ................................................................. 3

Figure 1.2: Cartoon showing cooperative binding of gp32 protein to ssDNA ............... 5

Figure 1.3: Plot showing DNA synthesis in presence and absence of gp32 protein ...... 6

Figure 1.4: Crystal structure of the gp32 core (PDB: 1GPC) ................................. 7

Figure 1.5: Side view of the crystal structure of the ssDNA binding domain of E. coli SSB (PDB: 1SRU) showing the homotetramer. .............................. 10

Figure 1.6: Two different binding modes (a) [SSB]65 (b) [SSB]35 for E. coli SSB showing model of ssDNA bound in all the four subunits in the former and only in two in the latter. ................................................................. 11

Figure 1.7: Cartoon showing FRET studies by Lohman and co-workers indicating binding modes of the E. coli SSB. .................................................. 11

Figure 1.8: Domain architecture of RPA heterotrimer. ................................. 12

Figure 1.9: Front view of crystal structure of hsRPA70 with both DNA binding domains A and B (PDB: 1FGU). ................................................................. 14

Figure 1.10: Front view of crystal structure of DBD-A and DBD-B of hsRPA70 subunit bound to a single-stranded DNA (PDB: 1JMC). ................................. 14

Figure 1.11: Close up view of DBD-A and DBD-B of hsRPA70 subunit with bound single-stranded DNA (PDB: 1JMC). ....................................................... 15

Figure 1.12: Cartoon showing the divergence of the three main families from the Last Common Unicellular Ancestor (LUCA). .......................................... 16
Figure 1.13: Schematic representation of the domains of the single-stranded DNA binding proteins in all three domains of life................................................. 17

Figure 1.14: Crystallographic dimer of Sso single-stranded DNA binding protein showing the characteristic OB fold.......................................................... 18

Figure 1.15: (a) Amino acid sequence of Ape SSB obtained from the NCBI website.
(b) Sequence alignment of full length Sso SSB against Ape SSB .......... 19

Figure 1.16: Amino acid sequence of Ape SSB............................................................... 20

Figure 1.17: Amino acid sequence of Afu SSB obtained from the NCBI website. ....... 22

Figure 2.1: Flow chart showing steps of the cell lysis protocol. ......................... 40

Figure 2.2: Flow chart of a typical HPLC purification scheme used in this study....... 43

Figure 2.3: Summary of Ape SSB lysis and purification protocol.......................... 46

Figure 2.4: Summary of Ape SSBΔC lysis and purification protocol .................... 47

Figure 2.5: Summary of Afu SSB lysis and purification protocol ......................... 48

Figure 2.6: Flow chart of different steps involved in preliminary processing of X-ray diffraction data................................................................. 55

Figure 2.7: Calibration curve for Superdex-75 using proteins with known molecular weight and elution time......................................................... 59

Figure 2.8: Cartoon of graphs based on Lohman and Bujalowski’s analysis to determine dissociation constant,......................................................... 62

Figure 2.9: Sedimentation velocity and equilibrium cell design. ......................... 65

Figure 3.1: Gene sequence for Ape SSB................................................................. 66

Figure 3.2: SDS-PAGE of Ape SSB expression study using BL21(DE3) RP and RIL . 67

Figure 3.3: Solubility profile of Ape SSB................................................................. 68
Figure 3.4: SDS-PAGE gel of *Ape* SSB solubility study shows soluble protein...... 69

Figure 3.5: Chromatogram of SP-Sepharose run for *Ape* SSB purification..............72

Figure 3.6: SDS – PAGE of samples from SP-Sepharose run for *Ape* SSB .............72

Figure 3.7: Chromatogram of POROS-HS run for *Ape* SSB purification.................73

Figure 3.8: SDS – PAGE of samples from POROS-HS run.................................73

Figure 3.9: Chromatogram of Superdex-75 run for *Ape* SSB purification.............74

Figure 3.10: SDS – PAGE of samples from Superdex-75 run................................74

Figure 3.11: *Ape* SSB DLS at 25 °C................................................................. 76

Figure 3.12: *Ape* SSB DLS at 4 °C................................................................. 77

Figure 3.13: Chromatogram of Superdex-75 analytical run for *Ape* SSB.............78

Figure 3.14: SDS – PAGE of sample from Superdex-75 analytical run..................78

Figure 3.15: Graph of anisotropy versus *Ape* SSB concentration at four different 9mer ssDNA concentrations ................................................................. 80

Figure 3.16: Graph of *Ape* SSB concentration versus ssDNA concentration........ 80

Figure 3.17: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.................................................. 81

Figure 3.18: Final graph of binding density versus free protein concentration fitted to Hill equation......................................................................................... 81

Figure 3.19: Graph of anisotropy versus *Ape* SSB concentration at four different 15mer ssDNA concentrations ................................................................. 82

Figure 3.20: Graph of *Ape* SSB concentration versus ssDNA concentration........ 82

Figure 3.21: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.................................................. 83
Figure 3.22: Final graph of binding density versus free protein concentration fitted to
Hill equation .................................................................................................................. 83

Figure 3.23: Quick change® site directed mutagenesis ............................................. 85

Figure 3.24: Gene sequence for Ape SSBΔC ................................................................. 85

Figure 3.25: SDS-PAGE gel showing Ape SSBΔC expression study ......................... 86

Figure 3.26: SDS-PAGE gel showing Ape SSBΔC solubility study ......................... 87

Figure 3.27: Solubility profile of Ape SSBΔC ................................................................. 89

Figure 3.28: Chromatogram of Q-Sepharose run for Ape SSBΔC purification .......... 92

Figure 3.29: SDS – PAGE of samples from Q-Sepharose run .................................. 92

Figure 3.30: Chromatogram of modified step gradient Q-Sepharose run ............... 93

Figure 3.31: SDS – PAGE of samples from modified step gradient Q-Sepharose run 93

Figure 3.32: Chromatogram of POROS-HQ run for Ape SSBΔC purification ........... 94

Figure 3.33: SDS – PAGE of samples from POROS-HQ run ..................................... 94

Figure 3.34: Chromatogram of Superdex-75 run for Ape SSBΔC purification .......... 95

Figure 3.35: SDS – PAGE of samples from Superdex-75 run ................................... 95

Figure 3.36: Crystal hits obtained for Ape SSBΔC ...................................................... 96

Figure 3.37: (a) to (e) shows crystals obtained on expanding crystal screen condition #5. ........................................................................................................................................... 97

Figure 3.38: Crystals seen in wells A2 and A3 of the Costar™ tray aperpa-c–15 .... 98

Figure 3.39: Crystals picked from wells A2 and A3 respectively of the Costar™ tray aperpa-c–15 and used for cryo freezing and diffraction study. .................. 99

Figure 3.40: Diffraction image of a hexagonal shaped crystal seen in Figure 3.37. .... 100

Figure 3.41: Glutaraldehyde (~ 6.7%) cross-linking of Ape SSBΔC crystal ............. 102
Figure 3.42: Glutaraldehyde (~ 4.3%) cross-linking of Ape SSB∆C crystal ............... 103
Figure 3.43: Diffraction images recorded on R-AXIS IV detector ..................... 104
Figure 3.44: Diffraction images recorded on ADSC Quantum 4 detector at beamline 14-BM-C at APS, Argonne ................................................................. 105
Figure 3.45: First half of the sequence alignment between Ape SSB and Sso SSB .... 112
Figure 3.46: Model of Ape SSB∆C crystal structure visualized using Pymol .......... 118
Figure 3.47: Crystal packing of Ape SSB∆C seen using COOT ................................ 119
Figure 3.48: Ape SSB∆C DLS at 25 °C ................................................................. 121
Figure 3.49: Chromatogram of Superdex-75 analytical run for Ape SSB∆C ........ 122
Figure 3.50: SDS – PAGE of samples from Superdex-75 analytical run ............... 122
Figure 3.51: Chromatogram of Superdex-75 analytical run for Ape SSB∆C ........ 123
Figure 3.52: SDS – PAGE of samples from Superdex-75 analytical run ............... 123
Figure 3.53: Graph of anisotropy versus Ape SSB∆C concentration at four different 9mer ssDNA concentrations ................................................................. 125
Figure 3.54: Graph of Ape SSB∆C concentration versus ssDNA concentration plotted as described in section 2.4.3 ................................................................. 125
Figure 3.55: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density ........................................ 126
Figure 3.56: Final graph of binding density versus free protein concentration fitted to Hill equation ................................................................. 126
Figure 3.57: Graph of anisotropy versus Ape SSB∆C concentration at four different 9mer ssDNA concentrations in presence of higher salt concentration ........ 127
Figure 3.58: Graph of Ape SSBΔC concentration versus ssDNA concentration plotted as described in section 2.4.3................................................................. 127

Figure 3.59: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density. ........................................... 128

Figure 3.60: Final graph of binding density versus free protein concentration fitted to Hill equation........................................................................................................... 128

Figure 3.61: Graph of anisotropy versus Ape SSBΔC concentration at four different 15mer ssDNA concentrations in presence of higher salt concentration. 129

Figure 3.62: Graph of Ape SSBΔC concentration versus ssDNA concentration plotted as described in section 2.4.3................................................................. 129

Figure 3.63: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density. ........................................... 130

Figure 3.64: Final graph of binding density versus free protein concentration fitted to Hill equation........................................................................................................... 130

Figure 3.65: Data analysis using software DCDT+ for Sedimentation Velocity experiment........................................................................................................... 132

Figure 3.66: Data analysis using software SEDNTERP for Sedimentation Equilibrium experiment........................................................................................................... 133

Figure 4.1: Gene sequence for Afu SSB .................................................................. 135

Figure 4.2: (a) Forward and reverse primers for PCR. (b) 1% agarose gel stained with SYBER GOLD™ dye, showing Afu SSB gene PCR product....................... 135

Figure 4.3: Agarose gel comparing Afu SSB inserted into pDrive with pDrive only from three different white colonies................................................................. 136
Figure 4.4: Agarose gel showing restriction reaction products of pET21a and *Afu* SSB along with control reactions................................................................. 137

Figure 4.5: Agarose gel showing five colonies of *Afu* SSB in pET21a.................... 138

Figure 4.6: SDS-PAGE showing over expression of *Afu* SSB in BL21(DE3) RIL .... 139

Figure 4.7: SDS-PAGE showing solubility of *Afu* SSB in low salt, high salt and bug buster after heating at 85 °C for 15 minutes. ............................................... 140

Figure 4.8: Solubility profile of *Afu* SSB ................................................................. 141

Figure 4.9: SDS-PAGE gel showing *Afu* SSB protein before and after heating the lysate. ........................................................................................................ 142

Figure 4.10: Chromatogram of Q-Sepharose run for *Afu* SSB purification ............ 144

Figure 4.11: SDS – PAGE of samples from Q-Sepharose run............................... 144

Figure 4.12: Chromatogram of POROS-HQ run for *Afu* SSB purification.......... 145

Figure 4.13: SDS – PAGE of samples from POROS-HQ run for *Afu* SSB.......... 145

Figure 4.14: Chromatogram of Superdex-200 run for *Afu* SSB purification.......... 146

Figure 4.15: SDS – PAGE of samples from Superdex-200 run for *Afu* SSB........ 146

Figure 4.16: Graph showing the maximum solubility of *Afu* SSB in presence of three different salt concentrations................................................................. 147

Figure 4.17: *Afu* SSB DLS experiment at 25 °C .................................................. 150

Figure 4.18: *Afu* SSB DLS experiment at 4 °C ..................................................... 151

Figure 4.19: *Afu* SSB DLS experiment at 25 °C in presence of 15mer ssDNA....... 152

Figure 4.20: *Afu* SSB DLS experiment at 4 °C in presence of 15mer ssDNA....... 153

Figure 4.21: Graph of anisotropy versus *Afu* SSB concentration at three different 15mer ssDNA concentrations ............................................................. 154
**Figure 4.22:** Graph of *Afu* SSB concentration versus ssDNA concentration plotted as described in section 2.4.3................................................................. 155

**Figure 4.23:** Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.................................................. 155

**Figure 4.24:** Final graph of binding density versus free protein concentration fitted to Hill equation........................................................................................................ 156

**Figure 5.1:** OB fold domain of *E. coli* Aspartyl tRNA synthetase .......................... 158

**Figure 5.2:** *Ape* SSB△C monomer (chain A) structure exhibiting typical OB fold architecture........................................................................................................ 159

**Figure 5.3:** Cartoon of two *Ape* SSB△C monomers with ssDNA from *hsRPA70* PDB superposed in the binding cleft of the monomers........................................ 160

**Figure 5.4:** Superposition of two *Ape* SSB△C monomers (shown in red) on to the structure of *hsRPA70* DNA binding domains without DNA...................... 161

**Figure 5.5:** Superposition of two *Ape* SSB△C monomers (shown in red) on to the structure of *hsRPA70* DNA binding domains with a ssDNA............... 161

**Figure 6.1:** Inchworm model for translocation of SF1 enzymes .............................. 167

**Figure 6.2:** Proposed mechanism for the unwinding of dsDNA of NS3 helicase from *Hepatitis C* virus ....................................................................................... 169

**Figure 6.3:** *Bovine papilloma virus E1 (BPV E1)* helicase X-ray crystallographic structure................................................................................................................ 171

**Figure 6.4:** Cartoon showing the coordinated escort mechanism for *E1* hexameric helicase........................................................................................................ 172

**Figure 6.5:** Amino acid sequence of *Vc dnaB* .................................................... 176
Figure 6.6: (a) Amino acid sequence of Yp dnaB obtained from the NCBI website
                  (b) Sequence alignment of Yp dnaB against Taq dnaB ...................... 178

Figure 6.7: (a) Amino acid sequence of Hp dnaB obtained from the NCBI website
                  (b) Sequence alignment of Hp dnaB against Taq dnaB ...................... 180

Figure 7.1: Flow chart showing the different steps involved in Gateway® cloning .... 186

Figure 8.1: Gene sequence for Vc dnaB ................................................................. 189

Figure 8.2: SDS-PAGE gel of Vc dnaB expression study ........................................ 190

Figure 8.3: SDS-PAGE showing solubility of Vc dnaB ........................................... 191

Figure 8.4: Solubility profile of Vc dnaB ................................................................. 193

Figure 8.5: Chromatogram of Q-Sepharose run for Vc dnaB purification .......... 195

Figure 8.6: SDS – PAGE gel of samples from Q-Sepharose run for Vc dnaB ...... 195

Figure 8.7: Chromatogram of POROS-HQ run for Vc dnaB ................................. 196

Figure 8.8: SDS – PAGE of samples from POROS-HQ run for Vc dnaB ............ 196

Figure 8.9: Picture of an SDS-PAGE gel showing samples without and with increasing concentration of TCEP ................................................................. 197

Figure 8.10: Crystal hits obtained for Vc dnaB ....................................................... 199

Figure 8.11: DLS of Vc dnaB in 10 mM TAPS pH 8.5 and 1 mM β-ME .............. 201

Figure 8.12: DLS of Vc dnaB in 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% glycerol
                  ........................................................................................................ 202

Figure 8.13: DLS of Vc dnaB in 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% glycerol in presence of 2 mM ADP and 10 mM MgCl₂ ........................................ 203

Figure 8.14: DLS of Vc dnaB in 10 mM HEPES pH 7.5 ......................................... 204

Figure 9.1: Yp dnaB gene sequence .................................................................. 207
Figure 9.2: Forward and reverse PCR primers for \textit{Yp} dnaB .................................................. 207

Figure 9.2: Picture of an agarose gel showing samples of the four PCR reactions of \textit{Yp} dnaB ........................................................................................................ 208

Figure 9.3: Picture of an agarose gel showing insertion of \textit{Yp} dnaB into pENTR-D entry vector........................................................................................................ 209

Figure 9.4: Picture of an agarose gel showing \textit{Yp} dnaB in pDEST-17 destination vector ................................................................................................................. 210

Figure 9.5: Picture of an agarose gel showing \textit{Yp} dnaB in pET45-DEST and pKM596-H ................................................................................................................. 211

Figure 9.6: SDS-PAGE gels of \textit{Yp} dnaB expression study using destination vector pDEST 17 in different cell lines ................................................................. 212

Figure 9.7: SDS-PAGE gel of \textit{Yp} dnaB expression study using destination vector pET45-DEST ................................................................................................................. 213

Figure 9.8: SDS-PAGE gel of \textit{Yp} dnaB-MBP expression study using the destination vector pKM596-H ................................................................................................................. 213

Figure 9.9: SDS-PAGE gel of \textit{Yp} dnaB solubility study showing insoluble protein in low salt, high salt and bugbuster................................................................. 214

Figure 9.10: SDS-PAGE gel of \textit{Yp} dnaB-MBP solubility study showing higher solubility in high salt compared to low salt ........................................................................... 215

Figure 9.11: Chromatogram of Amylose column run for \textit{Yp} dnaB purification......... 216

Figure 9.12: SDS – PAGE gel of samples from Amylose column run ..................... 216

Figure 9.13: SDS – PAGE gel showing reaction products of TEV protease reaction... 217

Figure 9.14: Solubility profile of \textit{Yp} dnaB-MBP .................................................. 219
Figure 10.1: *Hp* dnaB gene sequence ................................................................. 221

Figure 10.2: (a) Forward and reverse PCR primers for *Hp* dnaB. (b) Picture of agarose gels showing two samples of polymerase chain reactions of *Hp* dnaB. 222

Figure 10.3: Picture of agarose gel showing insertion of *Hp* dnaB PCR product. 223

Figure 10.4: Picture of agarose gels showing four samples of topoisomerase reactions for insertion of *Hp* dnaB gene into pET101 ................................................. 224

Figure 10.5: SDS-PAGE gel of *Hp* dnaB expression study using *Hp* dnaB gene in pET101 ................................................................. 225

Figure 10.6: SDS-PAGE gel of *Hp* dnaB-MBP expression study using *Hp* dnaB gene in pKM596-H ................................................................. 226

Figure 10.7: SDS-PAGE gel of (a) *Hp* dnaB solubility study showing insoluble protein and (b) partially soluble *Hp* dnaB-MBP ........................................ 227

Figure 10.8: Chromatogram of Talon column run for *Hp* dnaB purification........... 229

Figure 10.9: SDS – PAGE gel of samples from Talon column run for *Hp* dnaB-MBP. 229

Figure 11.1: Crystal structure of *Tag* dnaB (PDB: 2Q6T) .................................... 231
List of Abbreviations

ACS .................. American Chemical Society
ADSC ............... Area Detector Systems Corporation
AEBSF ............. 4-(2-Aminoethyl) benzenesulfonyl fluoride
Afu .................. Archaeoglobus fulgidus
Ape .................. Aeropyrum pernix
APS ................. Advanced Photon Source
β-ME ............... β-mercaptoethanol
BSA .................. Bovine Serum Albumin
Bis-Tris ............. 2,2-Bis(hydroxymethyl)-2,2’,2”-nitrilotriethanol
CAPS ............... N-Cyclohexyl-3-aminopropanesulfonic acid
CCD .................. Charge-Coupled Device
CCP4 ............... Collaborative Computational Project 1994
CNS ................. Crystallographic and NMR System
DBD ................. DNA Binding Domain
DLS ................. Dynamic Light Scattering
DNA ................. Deoxyribonucleic acid
DTT .................. Dithiothreitol
E. coli .............. Escherichia coli
EDTA ............... Ethylene Diamine Tetra Acetic acid
FEN-1 ............... Flap endonuclease-1
FRET ............... Förster Resonance Energy Transfer
GOI .................. Gene of Interest
HEPES ............... 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC ............... High-Performance Liquid Chromatography
IPTG ............... Isopropyl-β-D-thiogalactopyranoside
ITC .................. Isothermal Titration Calorimetry
LB .................. Luria Broth
MES ............... 4-morpholineethanesulfonic acid monohydrate
MPD ............... 2-methyl-2,4-pentanediol
MW ............... Molecular Weight
MWCO ............... Molecular Weight Cut-Off
NER .................. Nucleotide Excision Repair
OMCC ............... Ohio Macromolecular Crystallography Consortium
OD600 ............... Optical Density at wavelength of 600 nm
PCR .................. Polymerase Chain Reaction
PDB ............... Protein Data Bank
PEG ............... Polyethylene glycol
PEI ............... Polyethylenimine
Pfu .................. Pyrococcus furiosus
pI .................. Isoelectric point
PIPEC .................. Piperazinebis(ethanesulfonic) acid
PMSF ............... Phenylmethylsulphonyl fluoride
R_h .................. Hydrodynamic radius of gyration
RMSD ............... Root Mean Square Deviation
RNA .................. Ribonucleic acid
RNase H…………… Ribonuclease H
RPA……………… Replication Protein-A
SDS-PAGE …….. Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SEC……………… Size Exclusion Chromatography
S.O.B…………… Super Optimal Broth
S.O.C…………… Super Optimal Catabolite repression broth
SSB……………… Single-Stranded DNA Binding protein
TAE……………… Tris-Acetate-EDTA
TAPS…………….. N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
Taq……………… Thermus aquaticus
TEV……………… Tobacco Etch Virus
T_m ……………… Melting temperature
Tris……………… Tris(hydroxymethyl)aminomethane
Xgal……………… 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside
PART I: Studies of Single-Stranded DNA Binding Protein
Chapter 1: Introduction

An important component of life, whether single cell or multicellular, bacteria or human, is the DNA. Deoxyribonucleic acid, DNA, is the cell’s blueprint and is responsible not only for the long term storage of information but also for the very existence of the cell since it carries the information to synthesize proteins vital for the integrity and proper functioning of various cellular processes. Several proteins in turn are involved in maintaining the integrity of DNA.

During cell division, DNA also needs to divide. This is an important event whereby a new double-stranded DNA (dsDNA) is synthesized from the original dsDNA. As part of this process, the double stranded DNA at the replication fork is unwound and a single stranded region is formed. Under normal physiological conditions the existence of single-stranded DNA (ssDNA) is energetically unfavorable and needs the assistance of accessory factors to help stabilize it. The semi-conservative, semi-discontinuous process of DNA replication requires a multitude of proteins. Okazaki fragment processing requires more involvement of accessory proteins and enzymes when compared to the leading strand synthesis.\(^1\) The single-stranded DNA binding protein or SSB fulfills the role of protecting and stabilizing ssDNA. The existence of single-stranded DNA, coated with SSB provides a structure recognized by a variety of enzymes in replication, repair and recombination.
In prokaryotes, SSB’s protect genomic DNA during replication while other proteins process Okazaki fragments. In contrast, Replication Protein-A (RPA), the eukaryotic version of SSB, plays a direct role in Okazaki fragment processing. During synthesis, the replicating polymerase displaces the 5’ end of downstream fragments resulting in structures known as flap DNA. It was thought that flap DNA was processed entirely by two enzymes, Flap endonuclease-1 (FEN1) and Ribonuclease H1 (RNase H1). Later studies indicate that FEN1 could not cleave flaps longer than five to six nucleotides and the action of another nuclease, Dna2, is required for processing of long flaps. RPA is involved in switching between FEN1 and Dna2 endonuclease. RPA was shown to recruit Dna2 to the flap for processing and ultimately when the flap is cleaved to around six nucleotides, both RPA and Dna2 dissociate making way for FEN1 to bind and cleave the shortened flap.

Figure 1.1: RPA binds to 5’ flap created by the replicating polymerase. RPA binding inhibits further progression of the polymerase. RPA then recruits Dna2 endonuclease for processing the 5’ flap of the Okazaki fragment.
Single-strand binding proteins were first discovered in the laboratory of Bruce Alberts in the late 1960’s using DNA-cellulose affinity chromatography. T4 gp32 was one of the several proteins that bound tightly to the DNA-cellulose and required almost 2 M NaCl to elute from the column. Biochemical and biophysical studies have shown that SSB’s have a strong binding preference for DNA compared to RNA and specificity for ssDNA compared to duplex DNA.¹⁵ ⁶ Due to this preference, these proteins are now more commonly referred to as single-stranded DNA binding protein.

1.1. T4 gp32

T4 gp32, a 301 residue, 33.5 kDa protein, was the first SSB to be discovered. The 32 protein binds to ssDNA with an affinity 10⁴ to 10⁵ times greater than that for dsDNA, has a preference for AT rich regions, and has no sequence specificity. AT rich regions are easily melted compared to GC rich regions of a dsDNA and the energy cost involved in this process is the least.⁷ Gp32 protein also shows a high degree of cooperativity, with individual protein molecules lining up next to each other on ssDNA rather than binding randomly to a free stretch. This has been seen using electron microscopy.⁸ A free energy gain of -4 kcal is associated with one such neighbor-neighbor binding and occurs mainly because of protein–protein interaction and also because it is easier to bind to a polynucleotide extended by an attached protein.
Gp32 protein is a critical component of *T4* replication and has a structural function in maintaining the viability of the phage. Gp32 protein binds to single-stranded regions of the replication fork and promotes unwinding which is mainly accomplished by other enzymes with helicase activity. Gp32 protein bound ssDNA is protected from nuclease activity and other undesirable recombination or repair activities. Experiments show *T4* DNA polymerase preferentially stimulated by gp32 protein in the replication of single-stranded DNA template. Synthesis of a new strand by *T4* polymerase is enhanced at a nicked duplex DNA in the presence of gp32 protein as against its absence.9

Gp32 protein levels in *T4* are auto regulated with approximately 10,000 copies per cell. Experiments indicate replication and recombination rates are directly related to
the quantity of gp32 protein present. The protein first binds to the ssDNA present at a particular time and the excess then binds to the mRNA that codes for gp32 protein which results in the inhibition of gp32 protein synthesis. Auto regulation of gp32 protein synthesis is structure specific and based on recognition of a specific structure called “Pseudoknot” on the mRNA that codes for gp32 protein.\textsuperscript{10}

\textbf{Figure 1.3:} Plot showing DNA synthesis in presence and absence of gp32 protein using (a) single-stranded DNA template and (b) nicked DNA template. Figure adapted from reference \#9

The N-terminal domain of the gp32 protein consists of 21 amino acid residues and has a basic pI. Studies have shown that this domain is responsible for self-aggregation and cooperative binding to DNA.\textsuperscript{11} The C-terminal domain consisting of 22 amino acid residues is shown to be responsible for protein-protein interactions including that with the DNA polymerase and DNA helicases.\textsuperscript{9,12,13} The central core of the protein, lacking both
the N and the C terminus, contains the intrinsic DNA binding site.\textsuperscript{14, 15} The structure of the gp32 core was solved by Shamoo and co-workers in 1995 revealing structural details of this domain.\textsuperscript{15}

![Crystal structure of the gp32 core (PDB: 1GPC). α-helices are colored in red and β-sheets in yellow. This and all subsequent cartoons of protein structures made using Pymol and drawing tools (Microsoft Word).](image)

**Figure 1.4:** Crystal structure of the gp32 core (PDB: 1GPC). α-helices are colored in red and β-sheets in yellow. This and all subsequent cartoons of protein structures made using Pymol and drawing tools (Microsoft Word).

The crystal structure of this domain complexed to ssDNA shows the DNA located within a positively charged cleft. Though the oligonucleotide is disordered in the crystal, modeling of the substrate shows the side chains of three basic amino acids – lysine 110, arginine 111 and lysine 112 in contact with the backbone phosphates of the nucleic acid and thereby positioning the heterocyclic bases in contact with the hydrophobic pockets formed by clusters of aromatic side chains.\textsuperscript{15}
1.2. *E. coli* SSB

A protein homologue to the *T4* gp32 protein was discovered by Bruce Alberts and co-workers from uninfected *E. coli*. Like the *T4* gp32, this protein bound to DNA affinity columns and required high salt concentration to elute from the column. This protein was initially named as the *E. coli* DNA-unwinding protein due to its apparent ability to destabilize and unwind a DNA double helix. With the discovery of the gene for this protein and the fact that it could also bind to RNA although with very weak affinity, the protein was named single-stranded binding protein. Purification techniques involved DNA affinity columns and a brief boiling step since *E. coli* SSB was found to be stable to brief heating which denatured most other *E. coli* proteins. The heat step was later abandoned when biochemical and biophysical studies found that the binding properties were affected upon heating. Detailed studies showed that the *E. coli* SSB is a homotetramer in solution and aggregates at concentrations above 1 mg/mL and at low ionic strength. The in-vivo concentration of SSB in a *E. coli* cell is estimated to be around 0.5 to 1.0 µM.

Extensive studies have shown a variety of binding modes of the *E. coli* SSB depending on the solution conditions with two major binding modes, referred to as $[\text{SSB}]_{35}$ and $[\text{SSB}]_{65}$ where the subscripts indicate the number of nucleotides covered by the SSB tetramer. The two binding modes reflect two types of cooperativity, unlimited and limited. The unlimited nearest neighbor cooperativity, seen in $[\text{SSB}]_{35}$, occurs at low salt concentrations and high protein concentrations and involves the tetramer binding to the ssDNA using only two out of the possible four subunits and occluding ~ 35 nucleotides. This type of intertetramer positive cooperativity results in the
formation of clusters of protein along the ssDNA which can be seen using electron microscopy.26 The limited cooperative binding mode, [SSB]_{65}, occurs at high salt concentrations and at relatively lower protein concentrations. Under these conditions, the SSB tetramer binds to ssDNA using all the four subunits with around 65 nucleotides wrapped around the tetramer. Electron microscopy images have shown beaded structures corresponding to tetramers and dimers of tetramers (octamers).26 Different modes of binding may be used selectively during replication, recombination and repair.27

The *E. coli* SSB with 177 amino acids and 8.8 kDa molecular weight has a pI of 6.0. The C-terminal domain is involved in protein-protein interaction while the N terminal domain and the core domain of the protein bind to the ssDNA. The crystal structure of the DNA binding domain alone and in complex with ssDNA has been solved and provides detailed structural information for the domain arrangements, showing different possible interactions of the ssDNA with the protein.28-30 The protein has OB fold domain (Oligonucleotide / Oligosaccharide Binding domain) clearly seen in the structure (Figure 1.5).31, 32 The characteristic OB fold architecture consisting of antiparallel β-sheets forming a β-barrel is seen intertwining an α-helix packed against the bottom of the barrel. The L_{45} loop (the loop between β-sheets 4 and 5) is extended and forms a β-hairpin. Two anti-parallel β-sheets from two SSB subunits form an extended six-stranded anti-parallel β-sheet. The monomer-monomer interface has several main-chain hydrogen bonds including interactions between the two L_{45} loops in the dimer interface. The SSB tetramer is formed by two six-stranded β-sheets from the dimer and mutation studies of residues involved in the tetramer interface have confirmed the existence of the six-stranded β-sheet mediated tetramer.33
Recent biophysical studies indicate that in case of unlimited cooperativity, only two subunits of the homotetramer are available for ssDNA binding whereas in case of limited cooperativity, all the four subunits are available.\textsuperscript{34} This kind of arrangement is shown in the model in Figure 1.6. Förster Resonance Energy Transfer (FRET) studies (Figure 1.7) have also indicated the existence of the two binding modes. Depending upon the condition and the functional role, \textit{E. coli} SSB can switch from one binding mode to the other.\textsuperscript{34}
**Figure 1.6:** Two different binding modes (a) $[\text{SSB}]_{65}$ (b) $[\text{SSB}]_{35}$ for *E. coli* SSB with ssDNA bound in all the four subunits in the former and only in two in the latter. Figure adapted from reference #34 and made using Pymol along with Microsoft Word drawing tools.

**Figure 1.7:** FRET studies by Lohman and co-workers indicating binding modes of the *E. coli* SSB. (c) shows partial duplex DNA in absence of SSB (d) binding of a single SSB tetramer in $[\text{SSB}]_{65}$ binding mode and (e) binding of another SSB tetramer in $[\text{SSB}]_{35}$ binding mode. Figure based on reference #34.
1.3. Replication Protein A (RPA)

Replication Protein A (RPA) also known as Replication Factor A (RFA), is the single-stranded DNA binding protein found in Eukaryotes. RPA is a heterotrimeric protein composed of three subunits: RPA70, RPA32 and RPA14 (70, 32 and 14 indicating the molecular weight in kDa of the respective subunits) of which RPA70 is the central subunit involved in binding to DNA. Consistent with the SSB’s of the eubacteria and eukaryotes, RPA exhibits the presence of OB folds. There are in all six OB folds in the heterotrimer, with four of them in RPA70 and one each in RPA32 and RPA14. The four OB folds in RPA70 are denoted as DNA Binding Domain-A (DBD-A), DBD-B, DBD-C and DBD-F while RPA32 has DBD-D and RPA14 has DBD-E.

**Figure 1.8:** Domain architecture of RPA heterotrimer. The three subunits have six DNA binding domains characterized by the presence of OB folds. Figure based on reference #37
Biochemical analyses have shown that DBD-F, the N-terminal domain of RPA70, interacts with various proteins and binds weakly to DNA.\(^ {38}\) A 60 amino acid linker connects DBD-F with DBD-A.\(^ {39}\) The central domains of RPA70, DBD-A and DBD-B, binds DNA with high affinity and constitutes the binding core of the RPA.\(^ {40}\) Like the DBD-F, DBD-C binds weakly to DNA and has been shown to interact specifically with damaged DNA.\(^ {41}\) RPA32 with DBD-D is involved in protein-protein interactions and RPA14 with DBD-E is important for the formation of a stable heterotrimer.\(^ {6,42}\)

Studies show that binding of RPA to ssDNA is initiated with interaction of DBD-A and DBD-B with around 8 to 10 nucleotides.\(^ {43}\) While both DBD-A and DBD-B are essential for binding of the RPA to ssDNA, experiments indicate that DBD-A binds first at the 5’ side of the ssDNA followed by DBD-B which seems to be the weaker of the two binding sites.\(^ {43,44}\) This is followed by binding of DBD-C and in all a total of 13 to 22 nucleotides are occluded resulting in a more stable intermediate.\(^ {42}\) Cooperative binding of all the subunits results in 30 nucleotides being occluded by a RPA heterotrimer.\(^ {37}\)

Extensive studies of the human RPA (\(hsRPA\)) have shown its interaction with several DNA repair and tumor suppressor proteins.\(^ {45-47}\) Structural studies of \(hsRPA\) using X-ray crystallography and NMR have shown the structural characteristics of the protein and its binding interactions with ssDNA. Crystal structures of DBD-A and DBD-B in the presence and absence of DNA\(^ {32,43}\) (Figures 1.9 and 1.10), RPA32 and RPA14 complex\(^ {48}\), DBD-C/RPA32/RPA14 trimer\(^ {49}\) have been solved.
Figure 1.9: Front view of crystal structure of hsRPA70 with both DNA binding domains A and B (PDB: 1FGU). The characteristic β-barrel is seen formed by β-sheets (yellow) with α-helices shown in red.

Figure 1.10: Front view of crystal structure of DBD-A and DBD-B of hsRPA70 subunit bound to a single-stranded DNA (PDB: 1JMC).
Figure 1.11: Close up view of DBD-A and DBD-B of hsRPA70 subunit with bound single-stranded DNA showing different interactions between the two (PDB: 1JMC).

DNA binds at the upper part of the protein in the cleft formed by the β-sheets and the loops and extends through the two domains in the cleft. Superposition of the hsSSB structures with and without DNA shows the movement of the loops upon DNA binding. Several interactions can be seen in the binding site of the protein with ssDNA, consisting of a combination of polar and non-polar residues. As seen in Figure 1.11, non-polar Π stacking interactions can be seen between F238 and dC1, F269 and dC3, dC2 and dC3, R210 and dC4, dC4 and dC5, W361 and dC6, F386 and dC8 and dC7.
1.4. Archaeal SSB

Studies of archaeal proteins, including SSB’s, provide an evolutionary link between bacteria and eukaryotes (Figure 1.12) with several proteins having striking structural and functional similarity to those of eukaryotes. The main advantage of studying the archaeal proteins is that they are hyperthermophilic and hence are convenient for structural, biochemical and biophysical studies. The archaeal kingdom is further subdivided into two subkingdoms – Euryarchaeae and Crenarchaeae.

![Diagram of evolutionary tree showing divergence of three main families from the Last Common Unicellular Ancestor (LUCA).](image)

**Figure 1.12**: Cartoon showing the divergence of the three main families from the Last Common Unicellular Ancestor (LUCA).

Archaeal SSB’s have been shown to have a variety of domain arrangements (Figure 1.13) depending on whether they belong to the euryarchaeal or crenarchaeal subkingdoms with some architecture similar to the bacterial arrangement while others are similar to the eukaryotic domain architecture.
Figure 1.13: Schematic representation of the domains of the single-stranded DNA binding proteins in all three domains of life. Overview of the attributes of SSB from each domain is indicated in the bold text box. Scheme adapted from reference #52.
Crystal structures of a few SSB’s have been solved (appendix #1) including that from the crenarchaea, *Solfalobus solfataricus* (*Sso*). The structure of the *Sso* SSB without the C-terminus was solved to a resolution of 1.26 Å by Kerr and co-workers in 2003 after initial attempts to crystallize the full length *Sso* SSB failed.\(^53,54\) Although studies using Size Exclusion Chromatography (SEC) showed that *Sso* SSB could exist as a tetramer in solution, detailed experimental analysis indicated that at biologically relevant concentrations the protein existed as monomer in solution.\(^52,55\)

Figure 1.14: Crystallographic dimer of *Sso* single-stranded DNA binding protein showing the characteristic OB fold. β-sheets are shown in yellow and α-helices are shown in red (PDB: 1JMC).

Using ITC (Isothermal Titration Calorimetry), the full length *Sso* SSB showed a dissociation constant of ~ 90 nM for a 21mer oligonucleotide.\(^54\)
1.4.1. *Aeropyrum pernix*

The SSB in this study was from *Aeropyrum pernix* (*Ape*) which like *Sso* belongs to crenarchaeal subkingdom. The organism was first discovered in the mid-nineties from a coastal hydrothermal vent at Kodakara-Jima Island, Japan by Sako and co-workers. The cells are 0.8 to 1.2 µM in diameter and grow optimally at 90 °C to 95 °C at pH 7.

Sequence alignment (using the BLAST2sequence program, ExPASY proteomics webserver) of *Sso* SSB with *Ape* SSB showed a sequence similarity of 72% with a sequence identity of 45%. This information was helpful since the phase information from the *Sso* SSB structure could be used for solving and determining the structure of *Ape* SSB.

(a)

1 MDLREGLRNV SISGRVLETG EPKVMETKRG PATLSEAVVG DESGRVKVTL 50
51 WGSHAGTLKE GEAVRIEGAW TTSYRGKVQV NVGRESTIEK VDSEDVFQAE 100
101 DIPEEMPEAQ YRGFG [**Q**R**Q**PY] **R**SGGFGGFRG **G**GYQP**R**GGR **R**F 143

(b)

*Ape* SSB: 2 DLREGLRNVSISGRVLETGEPKVMETKRGPATLSEAVVGDESGRVKVTLWGSHAGTLKEG 61
1 +L+ + +V+++ RVLE E + ++TR G + + +SE + + + +G E + + + + G + + ++KR + + + + + + + P A + Q RG G + + + + + + + +
*Sso* SSB: 7 NLKNMESVNVTRLASEARQIQTKNGVRTISEAVGDTRGKVLWGLQHAGSIKEG 66
*Ape* SSB: 62 EAVRIEGAWTTSYRGKVQVNQGRESTIEKVDSEDVFQAEIDIPEEMPEA--QYRGFGQRQP 119
1 + V+I**E** ANTT+++G+G++VT N G ++ I ++ + P++ + + +E P A + Q RG G +
*Sso* SSB: 67 QVVKIENAWTTAFKQGQLNAGSKTK1AEASEDGFPESQIPENTPTAPQMRGGGRGFR 126

Figure 1.15: (a) Amino acid sequence of *Ape* SSB obtained from the NCBI website. The highlighted region is the C-terminus sequence which was not present in the *Ape* SSBAC sequence (b) Sequence alignment of full length *Sso* SSB against *Ape* SSB using BLAST2sequences program available on the NCBI website shows alignment of only the N terminus and the core region.

Table 1.1 shows some of the characteristics of the *Ape* SSB gene and protein determined using ProtParam program available on the ExPASy proteomics server. The
A full length protein with 143 amino acids has a low molecular weight and has a highly basic calculated pI.

Table 1.1: Protein characteristics of the full length *Ape* SSB obtained using ProtParam program (ExPASy Proteomics server)

<table>
<thead>
<tr>
<th>Gene length</th>
<th>432 bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td>143</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>15.7 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>9.04</td>
</tr>
</tbody>
</table>

*Ape* SSB without the C-terminus (*Ape* SSBΔC) was also studied. The gene for truncated protein was obtained by introducing a stop codon in the gene for the full length protein using site-directed mutagenesis as discussed in section 2.1.11. Figure 1.16 shows the sequence of *Ape* SSBΔC. The highly basic C-terminus composed of 28 residues is not present in this protein. Table 1.2 shows the characteristics of the protein.

```
1  MDLREGLRNVSISGRVLETGEPKMVEKTGPATLSEAVGDESGRVKVTL 50
51 WGSHAGTLKEGEAVRIEGAWTTSYRGKVQVNVGRESTIEKVDSEDVPQAE 100
101 DIPEEMPEAQYRGFG 115
```

*Figure 1.16:* Amino acid sequence of *Ape* SSBΔC obtained from the NCBI website showing 115 amino acids.
Table 1.2: Protein characteristics of *Ape* SSBΔC obtained using ProtParam program (ExPASy proteomics server)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene length</td>
<td>348 bp</td>
</tr>
<tr>
<td>Amino acids</td>
<td>115</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>12.5 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>4.87</td>
</tr>
</tbody>
</table>

1.4.2. *Archaeoglobus fulgidus*

SSB from *Archeoglobus fulgidus* which belongs to euryarchaeota was also studied. *Archeoglobus fulgidus* is a hyperthermophile that can be found in hydrothermal vents, oil deposits and hot springs. Because of its tendency to produce the noxious gas hydrogen sulfide, this archaea is considered pathogenic. It reduces the quality of oil reserves and causes corrosion of iron and steel in oil and gas processing systems.\(^{57}\) *Afu* grows at temperatures anywhere from 75 °C to 85 °C.

Search was done to identify *Afu* SSB homologues using gene mining program “BLAST” available on the NCBI website. But no such homologues with any significant sequence homology were found, implying that the structure may be different from the other SSB’s. *Afu* SSB gene and protein characteristics were determined using ProtParam program available on the ExPASy proteomics server and are summarized in Table 1.3. Compared to *Ape* SSB, *Afu* SSB is a much bigger protein and has a pI which is much lower.
Figure 1.17: Amino acid sequence of *Afu* SSB obtained from the NCBI website.

Table 1.3: *Afu* SSB gene and protein characteristics determined using Protparam program (ExPASy proteomics server)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene length</td>
<td>939 bp</td>
</tr>
<tr>
<td>Amino acids</td>
<td>312</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>35.6 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>6.34</td>
</tr>
</tbody>
</table>
Chapter 2: Materials and Methods

Chemicals for the various experiments were purchased from Sigma, Fisher and Aldrich unless otherwise indicated. All solutions were made using deionized, filtered water (0.22 µm filter) from the Milli Q academic water purification system from Millipore. Reagent and stock preparation recipes are included in the appendix.

2.1. Molecular Biology

2.1.1. Polymerase Chain reaction (PCR)

The forward and reverse primers for the Polymerase Chain Reaction (PCR) were obtained from IDT Inc. The nucleotide sequences for the primers were based on the gene sequence to be amplified. The primers were supplied by the company in lyophilized form and 250 µM stocks were prepared by adding Tris-EDTA (TE) buffer to the lyophilized primers and vortexing briefly using Vortexing Genie2® from Fisher Scientific. Working primer solution was prepared by adding 2 µL each of reverse and forward primer stock solutions to 46 µL of Tris-HCl pH 8.0 buffer. The reaction was set up in a 0.2 mL thin walled PCR tube from eppendorf as shown in Table 2.1.
Table 2.1: Components of a typical PCR

<table>
<thead>
<tr>
<th>PCR component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 X Proof Start buffer</td>
<td>5 µL</td>
</tr>
<tr>
<td>dNTP (10 mM)</td>
<td>3 µL</td>
</tr>
<tr>
<td>10 µM primer mix</td>
<td>10 µL</td>
</tr>
<tr>
<td>5x Q-Solution</td>
<td>10 µL (optional)</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>optional</td>
</tr>
<tr>
<td>Proof Start Polymerase</td>
<td>1 µL</td>
</tr>
<tr>
<td>Distilled water</td>
<td>As required to make up final volume</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>49.0 µL</strong></td>
</tr>
</tbody>
</table>

Once all the components were added and mixed, 1 µL of the template (with-in the concentration range of 100 ng to 1000 ng) was added, mixed lightly by pipetting and then vortexed briefly. The tube was centrifuged briefly at 2000 x g for less than a minute on a bench top Centrifuge 5417C from Eppendorf Scientific Inc and then placed on PCR Mastercycler Personal from Eppendorf Scientific Inc. The program for the reaction on the PCR machine was based on the annealing temperature of the primers. For stringency, the annealing temperature was set at the calculated annealing temperature of the primers instead of 2° below it as suggested in the protocol. Reactions with varying primer
concentrations or additional MgSO₄ were run parallel. Table 2.2 shows an example of a typical PCR program.

Table 2.2: Outline of a PCR program

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Activation of DNA polymerase</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>Annealing</td>
<td>Based on primer length and sequence</td>
</tr>
<tr>
<td>3</td>
<td>Extension</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 min / kb : &gt; 2 kb</td>
</tr>
<tr>
<td>4</td>
<td>Denaturation</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>Annealing</td>
<td>Based on primer length and sequence</td>
</tr>
<tr>
<td>6</td>
<td>Extension</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 min / kb : &gt; 2 kb</td>
</tr>
<tr>
<td>7</td>
<td>Go to Step 4, 22 to 30 cycles in all.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Additional Extension</td>
<td>72</td>
</tr>
<tr>
<td>9</td>
<td>End of PCR</td>
<td>4</td>
</tr>
</tbody>
</table>

After the reactions were completed, 2 µL samples per reaction were mixed with 2 to 3 µL of gel loading dye. An agarose gel was then run to check for the success and purity of the PCR products. The remaining samples were stored at -20 °C until the next step.

1% agarose gel was prepared by adding 40 mL of TAE buffer to 0.4 g of Analytical SeaKem GTG® agarose. The agarose was dissolved in the buffer by boiling
and once cooled; it was poured on to the setting tray and allowed to set for couple of hours after which it was placed in the electrophoresis chamber and TAE buffer added. Samples and base pair markers were then loaded into the appropriate lanes and the gel was run at a constant voltage of 90 V for 60 to 90 minutes. The gel was then transferred into a sandwich box and stained with working solution of SYBR® Gold nucleic acid stain from Invitrogen™, prepared by diluting 4 µL of 10,000X concentrate stain with 45 mL of TAE buffer. The container with the gel was placed on a Lab rotator (Lab-Line Instruments, Inc.) and allowed to shake very slowly for an hour. Samples were visualized using Illumatool light source (Lightools Research). Photograph of the gel was taken using a Nikon Coolpix 4600 digital camera. The setting on the camera was for close up and without flash. Multiple pictures at different zoom and resolution were taken and the best picture was used for recording the results.

2.1.2. Gel Purification

PCR product was mixed with equal or lesser volume of 50% autoclaved glycerol and depending on the thickness of the gel, around 5 to 8 µL was loaded into each lane. The gel was run for around 90 minutes at a constant voltage of 90 V and then stained using the SYBR® Gold nucleic acid stain just as described in the previous section. Once stained, the band containing the sample was cut out of the gel using a sharp sterile blade. Care was taken not to take too much of the gel and also not to cross-contaminate the lanes (if running more than two different samples). The cut piece of gel containing the sample was then placed carefully in a pre-weighed 1.5 mL eppendorf tube.
MiniElute™ Gel Extraction kit from QIAGEN Sciences was used for gel purification. First, buffer QG (~ 50% Guanidinium thiocyanate) was added at a volume three times the weight of the gel cut out (for example 300 µL to each 100 mg gel). The tube was then placed in a water bath at a temperature of 50 °C for 10 minutes or until the gel completely dissolved. Next, 1 gel volume of 100% isopropanol was added and mixed by inverting the tube gently several times. The contents of the tube were then transferred to a MiniElute® column placed in a 1.5 mL eppendorf tube and DNA was bound to the silica membrane in the column by centrifuging the tube on bench top centrifuge at 10,000 x g for a minute. The flow-through was discarded and 500 µL of buffer QG was again added to the spin column and centrifuged for another minute. 750 µL of buffer PE (containing ethanol) was added to the spin column and centrifuged for a minute. The flow-through from this was also discarded and the spin column centrifuged again for an additional minute. The spin column was placed in a fresh 1.5 mL eppendorf tube and 10 µL of buffer EB (10 mM Tris-HCl, pH 8.5) added to the center of the spin tube and then allowed to stand for a minute after which it was centrifuged for a minute. The eluent contained the purified PCR product. The purity of the PCR product was again checked on a 1% agarose gel (as described in the previous section).

2.1.3. Blue-white cloning using pDrive vector

The gel purified PCR product was inserted into pDrive cloning vector. This was done by taking the PCR product, pDrive vector and ligation master mix as shown in Table 2.3 in a 0.2 mL thin walled PCR tube.
Table 2.3: Components of pDrive cloning reaction

<table>
<thead>
<tr>
<th>Reaction component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDrive cloning vector</td>
<td>1 µL</td>
</tr>
<tr>
<td>PCR product</td>
<td>1- 4 µL</td>
</tr>
<tr>
<td>Ligation Master Mix</td>
<td>5 µL</td>
</tr>
<tr>
<td>Distilled Water</td>
<td>Final volume to 10 µL</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10 µL</strong></td>
</tr>
</tbody>
</table>

The reaction components were mixed and incubated at 37 °C for 30 minutes on the PCR machine. The tube was placed on ice for fresh transformation or stored at -20 °C for later use.

2.1.4. Preparation of Competent Cloning hosts – XL10

XL10 cells from glycerol stock at -80 °C were inoculated into 5 mL SOB media taken in a 50 mL Fisher tube. The tube was incubated overnight at 37 °C with constant shaking in a New Brunswick Scientific Innova 4000 Incubator Shaker at 200 rpm. 50 mL of fresh SOB was then inoculated with 120 µL of the overnight culture in a 250 mL autoclaved culture flask. The flask was incubated at 37 °C in the shaker at 200 rpm until OD$_{600}$ was around 0.6 and then it was transferred into an ice bath for 10 minutes. The culture was transferred into two 50 mL Fisher tubes and centrifuged at 2700 x g at 4 °C.
for 10 minutes. The supernatant was then decanted off and the pellet gently resuspended in 16 mL of transformation buffer (250 mM KCl, 10 mM PIPES; pH 6.5, 15 mM CaCl$_2$, 55 mM MnCl$_2$). The tubes were placed on ice for 10 minutes and then centrifuged at 2700 x g at 4 °C for 10 minutes. After discarding the supernatant, the pellet was again resuspended in 16 µL of transformation buffer. Cells from both tubes were pooled into one tube and DMSO added to a final concentration of 7%. The tube with the cells was then placed on ice and used immediately for transformation.

2.1.5. Transformation and isolation of Blue-white colonies

Transformation of the ligation product into the competent XL10 cells was done by adding 2 µL of ligation product into 50 µL of competent cells in a 1.5 mL eppendorf tube. After incubating on ice for 15 to 20 minutes, the cells were heat shocked by placing the eppendorf tube in a water bath at 42 °C for 30 seconds. The tube was then placed on ice for two minutes. After addition of 1 mL of SOB, the tube with the cells was incubated on a shaker at 200 rpm and 37 °C for an hour, centrifuged for a minute and 800 µL of the supernatant removed. Cells were carefully resuspended in the remaining 250 µL SOB and 50 to 100 µL plated on a freshly made LB agar plate containing Xgal and IPTG. The plate was incubated overnight at 37 °C in an incubator. Two or three white colonies were picked and inoculated into separate 50 mL fisher tubes containing 5 mL LB. The tubes were incubated on a shaker at 200 rpm for 6 to 10 hours at 37 °C.

The pDrive with the inserted gene was then isolated from the XL10 cells using miniprep kit as discussed in the following section.
2.1.6. Miniprep to isolate plasmid

Using QIAprep® spin Miniprep kit from Qiagen Sciences, the plasmid pDrive with the inserted gene was isolated. For this, the pelleted cells were resuspended in 250 µL re-suspension buffer P1 (containing RNaseA) and transferred into a 1.5 mL eppendorf tube. 250 µL of buffer P2 (containing NaOH) was added and mixed thoroughly by inverting six times. At this point care was taken to not let the lysis reaction proceed for more than five minutes. The lysis reaction was terminated by adding 350 µL of buffer N3 (containing guanidine hydrochloride and acetic acid) and mixing the tube by inverting six times which lead to the solution becoming homogenous. The tube was then centrifuged for 10 minutes at 13000 rpm on a bench-top centrifuge and the supernatant applied on to a QIAprep spin column. After centrifuging for a minute the flow-through was discarded and the column washed by adding 500 µL of buffer PB (containing guanidine hydrochloride and isopropanol) and centrifuging for a minute. After discarding the flow-through, 750 µL of buffer PE (containing ethanol) was added to the spin column and centrifuged for a minute. The flow-through was discarded and the column centrifuged for an additional minute to remove any traces of buffer PE. The spin column was transferred to a clean 1.5 mL eppendorf tube and 40 to 50 µL of buffer EB (10 mM Tris-HCl, pH 8.5) was added to the center of the membrane of the spin column. After a minute, the column was centrifuged to elute out the plasmid.

The Minipreped samples and supercoiled DNA ladder from Promega (2-10kb) were run on a 1% agarose gel to check for insertion of the gene.
2.1.7. Use of Restriction Enzymes

The next step was to restrict the Gene of Interest (GOI) from the pDrive vector using restriction enzymes. Since the forward and the reverse primers for the PCR step had two different sequence-specific restriction sites, the gene amplified and inserted into pDrive was restricted out using the specific restriction enzymes. The forward primer had \textit{nde I} restriction site and the reverse primer had the \textit{xho I} restriction site. At the same time, the expression vector, pET21a was also cut using the same two restriction enzymes. The reaction components were taken as shown in Table 2.4.

\textbf{Table 2.4: Components of restriction enzyme reaction}

<table>
<thead>
<tr>
<th>Reaction component (Tube 1)</th>
<th>Control (Tube 2)</th>
<th>Control (Tube 3)</th>
<th>Control (Tube 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDrive with GOI / pET21a</td>
<td>25 µl</td>
<td>4 µL from tube 1 (before addition of Enzyme 1 and 2)</td>
<td>4 µL from tube 1 (before addition of Enzyme 1 and 2)</td>
</tr>
<tr>
<td>Buffer (Tris-HCl pH 7.5, MgCl$_2$, KCl)</td>
<td>5 µl</td>
<td>4 µL from tube 1 (before addition of Enzyme 1 and 2)</td>
<td>4 µL from tube 1 (after addition of Enzyme 1)</td>
</tr>
<tr>
<td>BSA</td>
<td>1 µl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>13 µl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enzyme 1 (\textit{nde I})</td>
<td>3 µl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enzyme 2 (\textit{xho I})</td>
<td>3 µl</td>
<td>0.3 µL</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>50 µl</td>
<td>4 µL</td>
<td>4.3 µL</td>
</tr>
</tbody>
</table>

The reaction components were taken in 0.2 mL thin walled PCR tube and incubated at 37 °C for an hour on the PCR cycler. After an hour of incubation at 37 °C, the tubes were incubated at 65 °C for 20 minutes. The tube containing pDrive with GOI
and treated with both restriction enzymes as well as the tube containing pET21a and treated with both enzymes were incubated for another hour at 37 °C after addition of 0.3 µL of the enzyme alkaline phosphatase (to prevent ligation of the restricted DNA). All the samples including the control were then ran on a 1% agarose gel to confirm restriction. Once restriction was confirmed, the band containing the GOI and the band containing the restricted pET21a were gel purified and then re-run on an agarose gel to check for purity.

2.1.8. Ligation of Restriction products and Transformation into Cloning host

The gel purified restricted GOI and pET21a were then ligated after adding the components as shown in Table 2.5.

Table 2.5: Components of a ligation reaction

<table>
<thead>
<tr>
<th>Reaction components</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted and gel purified pET 21a</td>
<td>2 µL</td>
</tr>
<tr>
<td>Restricted and gel purified GOI</td>
<td>3 µL</td>
</tr>
<tr>
<td>Autoclaved Milli Q water</td>
<td>5 µL</td>
</tr>
<tr>
<td>2X quick ligation buffer</td>
<td>10 µL</td>
</tr>
<tr>
<td>Quick T4 DNA ligase</td>
<td>1 µL</td>
</tr>
<tr>
<td>Total volume</td>
<td>21 µL</td>
</tr>
</tbody>
</table>

The reaction components were taken in the order shown in the Table in a 0.2 mL thin walled PCR tube and mixed well by pipetting. After centrifuging briefly, the tube
was incubated at 25 °C for five minutes and then placed for another five minutes on ice before transforming into XL10 cells or stored at -20 °C for later use.

2.1.9. Transformation into XL10 cloning host

For transformation, 2 µL of ligation product was added into 50 µL of competent XL10 cells in a 1.5 mL eppendorf tube and then placed on ice for 15 to 20 minutes. The cells were heat shocked by placing the eppendorf tube in a water bath at 42 °C for 30 seconds. The tube was placed on ice for two minutes and 1 mL of room temperature SOB added and incubated on a shaker at 200 rpm and 37 °C for an hour. After centrifuging for a minute, 800 µL of the supernatant was removed and the cells resuspended in the remaining 250 µL SOB. 50 to 100 µL of the cells were plated on a freshly made LB agar plate containing ampicillin. The LB agar plate was then incubated overnight at 37 °C in an incubator. Two or three colonies were picked and inoculated into separate 50 mL fisher tubes containing 5 mL LB. The tubes were incubated at 37 °C on a shaker at 200 rpm for 6 to 10 hours.

pET21a vector with the inserted gene was then isolated from the XL10 cells using miniprep (described in section 2.1.6) and the insertion verified by running a sample on a 1% agarose gel along with a supercoiled DNA ladder and pET21a as reference.
2.1.10. Transformation of cloned Expression vector into Expression Host

Freshly made competent expression hosts or those from the supplier were used for transformation of the expression vector. The cells used included BL21(DE3)RIL, BL21(DE3)RILP, BL21(DE3)RP, BL21(DE3) GOLD and Rosetta2 pLysS. The transformation protocol used was the same as described in section 2.1.5. Successful transformation was confirmed with colonies on LB agar plate with appropriate selection antibiotics after overnight incubation at 37 °C. Two to three colonies were selected and small scale protein expression studies done as described in section 2.2.1.

2.1.11. Quick Change for truncated protein

The truncated protein, Ape SSBΔC, did not have the last 28 amino acids. For synthesizing C-terminal truncated protein, primers were obtained from IDT Inc. The primers were designed to introduce a stop codon, TAG, in the gene for the full length Ape SSB just before the C-terminus. PCR was set up using the full length Ape SSB in pET21a as the template and the gene amplified using Pfu Ultra High fidelity polymerase. All the reaction components for the Quick Change were obtained from the kit, QuickChange® II Site – Directed Mutagenesis Kit, from Stratagene. The reaction was set up as shown in Table 2.6.
Table 2.6: Quick change® reaction components.

<table>
<thead>
<tr>
<th>Reaction components</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x reaction Buffer</td>
<td>5.00 µL</td>
</tr>
<tr>
<td>(100 mM KCl, 100 mM (NH₄)₂SO₄, 200 mM Tris pH 8.8, 20 mM MgSO₄, 1% Triton® X-100, 1 mg/mL BSA)</td>
<td></td>
</tr>
<tr>
<td>dsDNA – <em>Ape</em> SSB in pet21 (9.858 ng/µL):</td>
<td>1.00 µL (10 ng)</td>
</tr>
<tr>
<td>Primer 1 apessb-cF (89.16 ng/µL)</td>
<td>1.40 µL</td>
</tr>
<tr>
<td>Primer 2 apessb-cR (86.34 ng/µL)</td>
<td>1.45 µL</td>
</tr>
<tr>
<td>dNTP</td>
<td>1.00 µL</td>
</tr>
<tr>
<td>Autoclaved water (final reaction volume-50 µL)</td>
<td>38.60 µL</td>
</tr>
<tr>
<td>Mix and then add <em>Pfu</em> Ultra HF DNA polymerase</td>
<td>1.00 µL</td>
</tr>
</tbody>
</table>

The reaction components were taken in a 0.5 mL thin walled PCR tube and mixed well. Program for the Quick change® reaction included an initial polymerase activation step. Details of the program are shown in Table 2.7.

Table 2.7: Program for Quick change® reaction

<table>
<thead>
<tr>
<th>Segment</th>
<th>Cycles</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>95 °C</td>
<td>30 sec</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>95 °C</td>
<td>30 sec</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55 °C</td>
<td>1 minute</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68 °C</td>
<td>5 minutes 30 seconds</td>
</tr>
</tbody>
</table>
Once the PCR was completed, the reaction was placed on ice for two minutes to lower the temperature to \( \leq 37 ^\circ C \). Next, 1 \( \mu L \) of \( Dpn \) I restriction enzyme (cleaves DNA that is methylated or hemi-methylated) was added to the reaction mixture, mixed gently and incubated at \( 37 ^\circ C \) for an hour to digest the parental i.e., the nonmutated methylated dsDNA. 1 \( \mu L \) of the reaction mix containing the mutated plasmid was inserted into XL1-Blue super competent cells. The transformation protocol is described in section 2.1.4. Three colonies were picked and a 5 mL overnight grow-up in LB was done. The plasmids were isolated from the cells using miniprep kit as described in section 2.1.6. and transformed into competent expression hosts (as explained in section 2.1.5.) The transformed cells were plated on LB agar plates with appropriate antibiotics and incubated overnight at \( 37 ^\circ C \).

2.2. Protein Expression, Purification and Solubility

2.2.1. Small Scale Protein Expression and Solubility Study

Three to five colonies were picked and inoculated into separate 50 mL Fisher tubes containing 5 mL sterile LB with the appropriate selection antibiotic. The tube was placed in a shaker at 200 rpm for six to eight hours at \( 37 ^\circ C \). Once the \( OD_{600} \) reached 0.6, an aliquot of 1.0 mL was taken for SDS PAGE (0 hour sample) and another aliquot of 1.0 mL was taken for preparing glycerol stock. IPTG was added to a final concentration of 1 mM and the tubes incubated for another three hours on a shaker at 200 rpm and \( 37 ^\circ C \). After three hours, aliquots were taken for SDS-PAGE and the remaining cells pelleted and stored at -20 \( ^\circ C \) for further studies. The 0 hour and 3 hour samples were
centrifuged to pellet the cells and then 25 µL of the BugBuster™ Protein Extraction reagent from Novagen was added to lyse the cells. 25 µL of SDS dye was added, mixed and the sample boiled for 10 minutes and then centrifuged for a minute at 5000 x g on the bench top centrifuge. 13 µL of supernatant was loaded on to a SDS-PAGE gel for analyzing the presence of protein of interest in the 3 hour sample as against the 0 hour sample. The presence of the protein with the correct molecular weight was confirmed by comparing the protein bands to a molecular weight ladder – Mark 12™ unstained standard (Invitrogen) run along with the samples.

Once expression of the desired protein was confirmed, a solubility study was done to check for soluble protein. The pelleted cells stored at -20 °C were thawed on ice and resuspended in a standard lysis buffer (100 mM Tris pH 8.0, 10 mM EDTA, 150 mM ammonium chloride, 5 mM DTT, 10 % Sucrose, ~ 0.1 µg of AEBSF / PMSF, 0.3 % PEI) in the ratio of 10 mL per gram of cells. The suspension was stirred at room temperature for 20 minutes after ~ 0.1 µg of lysozyme was added. Cells were lyzed open by sonicating the cell suspension on ice for two minutes using a sonicator (Branson™ Sonifier 250). Once this was done, 50 µL each of the lyzed cell suspension was added to three different 1.5 mL eppendorf tubes. To the first tube 50 µL of lysis buffer (low salt), to the second tube 50 µL of 1 M sodium chloride (high salt) and to the third tube 50 µL of BugBuster® were added. After thoroughly mixing, each tube was placed on ice for 10 to 15 minutes and then centrifuged at 12000 x g for five minutes on bench top centrifuge. The pellet in each tube was separated from the supernatant and then resuspended in 50 µL of BugBuster®. To all the tubes, 50 µL of the SDS dye was added, samples boiled for 10 minutes and the tubes centrifuged for two minutes at 6000 to 8000 x g before
loading the supernatant from the tubes on to a SDS-PAGE gel. Depending on the results of the gel and after determining what the protein favors, low salt (0.15 M) or high salt (1.00 M), a plan for large scale expression and lysis was formulated.

2.2.2. Large Scale Protein Expression Study

For large scale expression, 2-Liter flasks with one liter of autoclaved LB (20 g LB per liter) were used. Six such flasks were taken per experiment. The starter culture was prepared by inoculating three 250 mL flasks containing 100 mL LB with cells from the glycerol stock stored in -80 °C freezer. The flasks were kept in a shaker at 200 rpm and 37 °C. After an hour of incubation, the appropriate antibiotics were added to a final concentration of 1 mM and the flasks incubated overnight (~ seven hours) at 37 °C in the shaker.

Next morning, each flask containing 1 L LB with the appropriate antibiotics were inoculated with 50 mL of the overnight starter culture and incubated on the shaker at 37 °C. When the grow ups reached an OD₆₀₀ of 0.6, zero hour samples were taken for SDS-PAGE analysis from two of the flasks. IPTG was added to a final concentration of 1 mM to each flask and the flasks incubated at 37 °C in the shaker at 200 rpm. After three to four hours of expression, cells were isolated from the media using Millipore Pellicon™ -2 concentrator and Easy-Load® Masterflex® peristaltic pump. The peristaltic pump was kept at an rpm less than 50 and pressure was maintained below 15 psi. Once most of the cells had been pelleted, the final pelleting was done by spinning down the cells in a preweighed 50 mL Fisher tube using ultra centrifuge at 5000 x g for 20 minutes at 4 °C. The cell pellets were then stored at -20 °C until the next step or were used immediately.
2.2.3. Cell Lysis

When ready for lysis, the cell pellets at -20 °C were thawed on ice and then transferred to a beaker. Lysis buffer (10 mL per gram of cells) containing the appropriate concentration of ammonium chloride was added followed by addition of ~ 0.1 µg lysozyme. The beaker was placed on a Corning stir plate at room temperature and the cells with the lysis buffer stirred at medium speed for 20 minutes. The beaker was then transferred on to an ice bath for 15 minutes to cool the cell suspension. Cells were then sonicated for two to five minutes with constant stirring to dissipate the heat. During sonication, the beaker was kept in an ice bath so that the heat released during sonication did not denature the protein. Once the cells were lyzed open, an aliquot was taken for SDS-PAGE analysis and the remaining lysate centrifuged in an ultra centrifuge at 15,000 x g for 20 minutes at 4 °C. Once the cell debris had been pelleted down, the supernatant with the soluble protein of interest was decanted into a separate container and used immediately for next step. Figure 1.1 shows a flow chart of the cell lysis steps.

For short term storage, the supernatant was stored at 4 °C. For long term storage the supernatant was mixed with 80% autoclaved glycerol (preservative) to give a final concentration of 25% and then transferred into 50 mL Fisher tubes, frozen on dry ice before storing at -80 °C. Before use, the tubes with the supernatant at – 80°C was thawed at room temperature.
2.2.4. Protein Purification

There were several steps involved in protein purification depending on the solubility of the protein as well as the level of contaminants present in the supernatant. Since the proteins under study in this project were from thermophiles, the first step of purification involved heating the lysate to around 75 °C to 85 °C for 15 to 20 minutes. At this temperature most of the endogenous *E. coli* proteins were precipitated leaving the
protein of interest in the supernatant. Lysate was then centrifuged in an ultra centrifuge at 15,000 x g for 20 minutes at 20 °C.

The next step of purification involved ion-exchange column chromatography. The type of column used, anion or cation exchange, depended on the pI of the protein. Since for this study the pI was not experimentally determined, theoretical pI was determined using ProtParam program from ExPASy proteomics server.

The sample was filtered using a 0.2 micron syringe filter to remove any particulate matter which could block the column. The conductivity of the supernatant was measured using Corning 311 conductivity meter and appropriate volume of dilution buffer (buffer without any salts) added to lower the conductivity. The diluted supernatant was loaded on to the column using a BioLogic Duo Flow High performance Column chromatography unit from BioRad kept in a 4 °C cabinet.

Two different types of ion-exchange columns were used for both cation and anion exchange. The first type was the low resolution column – Q Sepharose (anion exchange) and SP Sepharose (cation exchange) from Amersham Biosciences. The protein sample was first passed through one of these low resolution columns to isolate the protein of interest. Although the level of purity of the isolated protein was not high, this column helped in removing most of the contaminants including lipids which could potentially block or ruin the more expensive high resolution column. Since the protein was eluted using Buffer-B which contained high salt, it was necessary to bring down the salt concentration by dilution or by dialysis. The dilution method was used since the high resolution column used for the next step worked better with high flow rates. The diluted, partially pure protein was further purified on a high resolution ion-exchange column like
POROS-HQ (anion exchange) or POROS–HS (cation exchange) from Applied Biosystems. The column was first equilibrated with Buffer-A and the sample loaded at a flow rate of around 15.0 mL a minute. The flow through was collected and the column washed with four to six column volumes of Buffer-A. Protein was then eluted using a linear gradient of Buffer-B against Buffer-A. This step was automated using the program created for different columns using the Biologic software provided with BIO-RAD HPLC. Fractions of eluent were collected into separate 10 mm glass test tubes using BIO-RAD model 2128 fraction collector controlled by the Biologic software. The HPLC has a UV detector which can monitor the absorption of the eluent at different wave-lengths. For the purpose of this study two different wavelengths were set (280 nM and 260 nM). Based on the absorption profile, any peak at 280 nm (absorption maxima for proteins) indicated the presence of protein. Since the plot also showed the fraction number corresponding to the UV profile, presence of the protein of interest was checked using SDS-PAGE analysis.

Pooled protein fractions were concentrated using Amicon® Ultra-15 centrifugal filters from MILLIPORE with 10,000 MWCO in a centrifuge at around 3500 x g and 4 °C. Once concentrated, the protein was further purified using a Superdex-75 or a Superdex-200 size exclusion column from Amersham Biosciences. For this, the column was equilibrated with the appropriate buffer and then a 1 mL sample loaded using a 1 mL sample loop. The protein of interest was thus isolated with acceptable purity for further studies and concentrated again at room temperature or at 4 °C as needed.
2.2.5. Solubility Profile

To increase the solubility of the protein, a solubility study developed in the Mueser lab was used. An aliquot containing approximately 5 mg protein was dialyzed against milliQ water at 4 °C using a dialysis cassette. Before injecting the sample into the cassette it was hydrated and the sample injected after removing most of the air from the cassette. Dialysis was carried out overnight so that all the protein precipitated out. The precipitated protein was removed from the dialysis cassette and the suspension aliquoted into 20, 1.5 mL eppendorf tubes. The tubes were centrifuged at 15000 x g for two minutes and the supernatant removed carefully into another eppendorf tube. To each of
the tubes containing the pellet, 15 µL of different salts / buffers from the solubility screen (shown in Table 2.8) were added and mixed thoroughly using a pipette.

<table>
<thead>
<tr>
<th></th>
<th>List of different salts and buffers used in the solubility screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ammonium Chloride</td>
</tr>
<tr>
<td>2.</td>
<td>Sodium Chloride</td>
</tr>
<tr>
<td>3.</td>
<td>Potassium Chloride</td>
</tr>
<tr>
<td>4.</td>
<td>Lithium Chloride</td>
</tr>
<tr>
<td>5.</td>
<td>Magnesium Chloride</td>
</tr>
<tr>
<td>6.</td>
<td>Calcium Chloride</td>
</tr>
<tr>
<td>13.</td>
<td>MES pH 5.8</td>
</tr>
<tr>
<td>14.</td>
<td>PIPES pH 6.5</td>
</tr>
<tr>
<td>15.</td>
<td>HEPES pH 7.5</td>
</tr>
<tr>
<td>16.</td>
<td>TAPS pH 8.5</td>
</tr>
<tr>
<td>17.</td>
<td>Water</td>
</tr>
</tbody>
</table>

The tubes were left undisturbed for 5 to 10 minutes at room temperature and then centrifuged to pellet out any undissolved protein. The amount of protein in the supernatant was then determined using Bradfords assay. For this, 5 µL of the supernatant was added to 995 µL of dilute Bio-Rad dye (5 mL of Bio-Rad Protein Assay Dye Reagent Concentrate from BIO-RAD Laboratories, Inc. diluted to 25 mL using MilliQ) and the absorbance measured at 595 nm on UV spectrophotometer. Absorbance of the protein in water was subtracted from each sample absorbance and a plot was made in Microsoft Excel of the absorbances. Based on the result, the best combination of salt and buffer was made and then checked for increased protein solubility.
2.2.6. Protein concentration

Once an appropriate combination was determined, all purification runs were done using the modified buffer. After purification, an aliquot of the protein was concentrated using Centricon® Centrifugal filter device from Millipore Corporation at 4 °C or at room temperature at 4000 x g on a table top centrifuge. The protein was concentrated at low salt (~150 to 200 mM) concentration until it started precipitating and the maximum concentration determined at that salt concentration. Once this was done, the entire protein sample was concentrated to half of the maximum protein concentration and then stored at 4 °C or room temperature for immediate use or stored at -80° after adding glycerol (preservative) to final concentration of 25% and freezing on dry ice for future use.
Figure 2.3: Summary of Ape SSB lysis and purification protocol

Ape SSB lysis using general lysis buffer (contains 25 mM Tris pH 7.5, 150 mM NH₄Cl)

Heat lysate to 75 °C for 10 to 15 minutes and collect supernatant after centrifugation

Dilute using dilution buffer to bring down conductivity
(Dilution buffer: 25 mM trisodium citrate pH 7.5, 25 mM NH₄Cl)

Equilibrate SP-Sepharose with Buffer-A at 22 °C
(Buffer-A: 25 mM trisodium citrate pH 7.5, 50 mM NH₄Cl)

Load diluted supernatant on to the SP-Sepharose column, collect flow through

Rinse column with Buffer-A and elute using gradient elution with Buffer-B
(Buffer-B: 25 mM trisodium citrate pH 7.5, 1.0 M NH₄Cl)

Check fractions for presence of protein using SDS-PAGE.

Pool fractions below the protein peak and dilute using dilution buffer

Load at a fast rate on to POROS HS 50/20 after equilibrating column with Buffer-A

Elute protein using gradient elution with Buffer-B

Check fractions for purity using SDS PAGE and pool appropriate fractions

Concentrate using Amicon Ultra-4, 5000 MWCO centrifugal filters with repeated addition of Buffer-A at 22 °C

Check for concentration using Buffer-A as blank and run Superdex 75 if needed

Freeze with glycerol or use immediately – temporary storage at RT
Figure 2.4: Summary of *Ape SSBAC* lysis and purification protocol

*Ape SSBAC* lysis using general lysis buffer (*contains 25 mM Tris pH 8.0, 150 mM NaCl*)

Heat lysate to 75 °C for 15 to 25 minutes and collect supernatant after centrifugation

Dilute using dilution buffer to bring down conductivity

(*Dilution buffer: 20 mM TAPS pH 8.5*)

Equilibrate **Q-Sepharose** with Buffer-A at 22 °C

(*Buffer-A: 20 mM TAPS pH 8.5, 50 mM NaCl*)

Load diluted supernatant on to the **Q-Sepharose** column, collect flow through

Rinse column with Buffer-A and elute using gradient elution with Buffer-B

(*Buffer-B: 20 mM TAPS pH 8.5, 750 mM NaCl*)

Check fractions for presence and purity of protein using PAGE (run POROS-HQ only if needed)

Pool fractions below the protein peak and dilute using dilution buffer

Load at a fast rate on to **POROS HQ** after equilibrating column with Buffer-A

Elute protein using gradient elution with 22 °C Buffer-B

Check fractions for purity using SDS PAGE and pool appropriate fractions

Concentrate using Amicon Ultra-4, 5000 MWCO centrifugal filters with repeated addition of Buffer-A at 22 °C

Check for concentration using Buffer-A as blank and run **Superdex 75** if needed

Concentrate protein fractions and freeze with glycerol or use immediately – temporary storage at 22 °C
**Figure 2.5: Summary of *Afu* SSB lysis and purification protocol**

*Afu* SSB lysis using general lysis buffer *(contains 25 mM Tris pH 8.0, 150 mM NaCl)*

> Heat lysate to 85 °C for 15 to 25 minutes and collect supernatant after centrifugation

> Dilute using dilution buffer to bring down conductivity
  
  *(Dilution buffer: 50 mM Tris-HCl pH 8.0)*

> Equilibrate Q-Sepharose with Buffer-A at 22 °C
  
  *(Buffer-A: 50 mM Tris-HCl pH 8.0, 40 mM NaCl, 10 mM MgCl₂)*

> Load diluted supernatant on to the Q-Sepharose column, collect flow through

> Rinse column with Buffer-A and elute using gradient elution with Buffer-B
  
  *(Buffer-B: 50 mM Tris-HCl pH 8.0, 750 mM NaCl, 10 mM MgCl₂)*

> Check fractions for presence and purity of protein using PAGE (run POROS-HQ only if needed)

> Pool fractions below the protein peak and dilute using dilution buffer

> Load at a fast rate on to POROS HQ after equilibrating column with Buffer-A

> Elute protein using gradient elution with Buffer-B

> Check fractions for purity using SDS PAGE & pool appropriate fractions

> Concentrate using Amicon Ultra-4, 10,000 MWCO centrifugal filters with repeated addition of Buffer-A at 22 °C

> Check for concentration using Buffer-A as blank & run Superdex 200 if needed

> Concentrate protein fraction and freeze with glycerol or use immediately – temporary storage at 22 °C
2.3. Protein Crystallography

2.3.1. Preliminary Crystal Screening

During the initial part of the research work, crystal screening was done manually. Two types of trays were used depending on the study, a 96 well Corning™ tray (Hampton catalog#: HR3-271) if only the protein was being tested or a three sample, 96 well round bottom Greiner™ tray (Hampton catalog#: HR3-095) if protein alone as well as with ligands were being tested. Crystal screens used for this study were obtained from Hampton Research, Emerald BioStructures or prepared in Mueser lab. Screening was done by pouring the plates row by row with 100 µL (per reservoir) of the mother liquor from the deep block using a 100 µL, 8-channel LTS pipette (from Rainin Research). Disposable pipette tips from Rainin Research were used for this. Next, using a 10 µL programmable pipette (Eppendorf Research Pro), 1 µL per well of protein solution was dispensed into the empty well. To this, 1 µL of the reservoir solution was added using a 10 µL, 8-channel LTS pipette (Rainin Research). The tray was sealed with clear tape and then centrifuged briefly at a slow speed to ensure mixing of the protein and the mother liquor. Trays were stored in a cabinet for room temperature condition or in a cold room at 4 °C. At least three observations were made – the first observation being on the second day after setting the tray, the second observation after three days and the next one after two weeks. The plates were observed using a Nikon SMZ1500 microscope and observations of each well were noted on a chart with short hand notations. Different notations are shown in appendix #14. Possible crystal hits were recorded using a Nikon CoolPix™ 990 digital camera and verified to be a protein crystal (as opposed to a salt crystal) by adding 0.3 µl of Izit Crystal Dye™ (Hampton Research) to the well. The dye...
was allowed to diffuse through the well solution for about 10 minutes and the drop visualized under the microscope. Protein crystals would turn dark blue as compared to salt crystals which would remain colorless.

### Table 2.9: List of different crystal screens used in the study

<table>
<thead>
<tr>
<th>Screen</th>
<th>Company</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal Screen I™</td>
<td>Hampton Research</td>
<td>Sparse matrix</td>
</tr>
<tr>
<td>Crystal Screen II™</td>
<td>Hampton Research</td>
<td>Sparse Matrix</td>
</tr>
<tr>
<td>Cryo I™</td>
<td>Emerald BioStructures</td>
<td>Sparse matrix</td>
</tr>
<tr>
<td>Cryo II™</td>
<td>Emerald BioStructures</td>
<td>Sparse Matrix</td>
</tr>
<tr>
<td>Wizard I™</td>
<td>Emerald BioStructures</td>
<td>Random sparse matrix</td>
</tr>
<tr>
<td>Wizard II™</td>
<td>Emerald BioStructures</td>
<td>Random sparse matrix</td>
</tr>
<tr>
<td>Index™</td>
<td>Hampton Research</td>
<td>Combination of Grid screen, Sparse-matrix and Incomplete factorial</td>
</tr>
<tr>
<td>Natrix™</td>
<td>Hampton Research</td>
<td>Sparse Matrix</td>
</tr>
<tr>
<td>Ion Screen</td>
<td>Prepared in-house</td>
<td>Ion/pH matrix (has PEG)</td>
</tr>
<tr>
<td>Modified Ion Screen</td>
<td>Prepared in-house</td>
<td>pH and Salt only</td>
</tr>
<tr>
<td>Additive screen</td>
<td>Prepared in-house</td>
<td>Precipitating agent/Additive screen</td>
</tr>
</tbody>
</table>

During the later part of this research work the Ohio Macromolecular Crystallography Consortium (OMCC) was established, with state of the art instruments being set up in the Arts and Sciences instrumentation center. The Honeybee robot (961/963) from Genomic Solutions® completely automated the process of setting up the
plates. Different programs could be used depending upon the type of plate and volume of protein desired. Although the robot could dispense as low as 0.1 µL of protein, for this work 0.5 or 1.0 µL protein was used. The robot dispensed 100 µL of reservoir solution from the deep well block into the tray reservoir and then dispensed 0.5 or 1.0 µL of it into the well. Next, it dispensed 0.5 or 1.0 µL of protein into each of the wells. If the 3-well Greiner plate was used, it could dispense three different samples using three separate dispensers. The trays were manually sealed using clear tape and stored at room temperature (20 °C) or at 4 °C.

2.3.2. Expansion and Optimization of Preliminary Crystal Hits

Once a possible crystal hit was identified from initial crystal screening, the hits were expanded to optimize the crystal growth conditions. This is an important step to obtain diffraction quality crystals.

Expansion trays using hanging-drop vapor diffusion technique were set up. Two types of plates were used, the 24-well Costar™ tray (Hampton Research) and the 24-well Nextal™ tray (Nextal). The Costar™ trays involved use of 18 mm cover slips and grease as a sealant. The Nextal™ trays on the other had a cap with O-ring seal which could be screwed on top of the well and so could be manipulated multiple times with ease as compared to the Costar™ trays. For this research work, the Nextal™ trays were predominantly used. Well solutions were prepared using the AB gradient technique which was quick and easy to set up. The type of gradient depended on the condition being optimized. Expansion could be a 4 x 6 format where four rows, each with 6 wells, could be used to test one set of crystallization condition or a 2 x 12 format were two
rows, with 12 wells could be used. The third alternative was to use the entire tray of 24 wells (1 x 24 format) for one set of crystallization condition. Regardless of which format was used, expansion involved preparation of two buffers: Buffer-A and B with Buffer-A containing the lower end of the expansion range and Buffer-B having the higher end of the expansion range. The buffers were filtered using Steriflip sterile disposable vacuum based filters from Millipore. Trays were prepared by pipetting both A and B buffers to give a final volume of 1.0 mL per well throughout the gradient, using an EDP 10 mL programmable pipette (Rainin). The trays were then placed on a rotator at a slow speed to mix the well solutions. Protein was then dispensed on to the Nextal cap (1 to 4 µL depending on the format) using a 10 µL programmable pipette (Eppendorf Research Pro). Next, 1 to 4 µL of the reservoir solution was added to the protein drop on the cover and the cap screwed tight on top of the corresponding well. Trays were left undisturbed for at least a week, at room temperature or at 4 °C, before checking for crystal growth. Once a specific range was established for crystal growth, production quality crystals were obtained by using a finer gradient. Crystals obtained using this technique could then be used for diffraction studies.

2.3.3. Crystal manipulation for X-ray diffraction

For low temperature diffraction studies, crystals had to be in cryo-protectant so as to prevent damage due to icing. Cryo-protectant used for freezing depended on the mother liquor. 2X mother liquor solution was prepared and equal volume of 2X cryoprotectant added to it. This solution was then checked for a good freeze by dipping a
loop into the solution and freeing it under a stream of nitrogen at 100 K. A good freeze was indicated by a clear glassy drop. Hazy or frosted drop indicated that the cryo concentration needed to be increased. Once an optimum combination was determined, the crystal was soaked in it for 5 seconds or more depending on the size. The crystal was then picked on a loop and frozen in liquid nitrogen. The frozen crystal was either stored in a dewar or directly mounted on to the goniometer head with a focused stream of nitrogen at 100 K. An alternate technique for freezing involved use of supercooled helium. Helium can be cooled to around to 20 K leading to faster and better freeze.

For room temperature X-ray diffraction study, the crystal was looped on a mylar loop mounted on a MitiGen™ head. Excess liquid was soaked using a paper wick and then a mylar plastic sleeve was mounted on top of the post. The upper part of the sleeve had 0.5 to 2 µL of the mother liquor. The MitiGen™ was then mounted on to the goniometer for X-ray diffraction study.

2.3.4. X-ray Diffraction Screening and Data Collection

Frozen crystals were screened using Rigaku FRE diffractometer and CCD detector at the instrumentation center at University of Toledo. Initial screening involved collecting three to four diffraction images at 45° separation with 0.5° oscillation and an exposure time of 10 to 15 seconds per frame. The quality and diffraction pattern for different crystals were then analyzed. Based on preliminary indexing, a possible space group was determined. This information of tentative space group was then used to determine how many rotational degrees of data needed to be collected. Based on this preliminary study, crystals were ranked and stored in liquid nitrogen for transportation to
Argonne National Laboratories, Advanced Photon Source (APS), Chicago, IL, USA for final data collection. Data sets were collected at BioCARS 14-BM-C beamline at APS using an ADSC Quantum 4 CCD detector. Important factors that were considered during data collection were resolution, completeness and redundancy which in turn were determined by degrees of rotation, distance and exposure time.

2.3.5. Data Processing

X-ray diffraction data sets were collected at APS and processing done at Mueser lab using different software programs run on a Linux operating system. Main steps involved in data processing were indexing, integration and scaling. The first two steps were done using the program IPMOSFLM.\textsuperscript{60} Indexing involved identifying the dimensions of the unit cell and which image peak corresponds to which position in reciprocal space. The program assigns Miller indices (h, k, l) to each reflection which then was used to determine the symmetry of the crystal, i.e., its space group. Having determined a possible symmetry, data was integrated. This converted all the indexed images containing thousands of reflections into a single file, consisting of records of the Miller index of each reflection, and intensity for each reflection. The third step in data processing involved scaling the integrated data set. This was done using the program SCALA which is part of the CCP4 suite.\textsuperscript{61, 62} Scaling involved normalizing intensities and merging unique reflections in the resolution range of the data collection. This is important, since the relative intensities of the peaks are a critical piece of information.
from which the structure is determined. Figure 2.6 summarizes the steps of data processing.

![Flow chart of different steps involved in preliminary processing of X-ray diffraction data.](image)

**Figure 2.6:** Flow chart of different steps involved in preliminary processing of X-ray diffraction data.

### 2.3.6. Structure Determination

Structural determination from the processed data depends on the availability of phase information. If the structure of a homologues protein (template) has been solved and has sequence identity of around 30 to 40% to the protein under study, then molecular replacement can be used for phase determination. There are several programs available to do this of which PHASER, which is part of the CCP4 group of programs, was used for this study. The success of molecular replacement depends not only on degree of sequence similarity but also on similarity to at least a part of the structure. During molecular replacement using PHASER, the template model was positioned into the crystal lattice of the protein under study using a rotation search followed by a translation
search. If molecular replacement failed because of inadequate sequence homology, the sequence of the template structure could be modified to poly-alanine. This retains the original structural confirmation but reduces the chances of clashes upon molecular replacement. Once PHASER provides a possible solution, the success is evaluated based on the Tz scores for the molecular replacement. Tz scores above 8 are indications of a good molecular replacement. Phase information from the model can then be combined with experimentally determined amplitude. The next step is to convert the sequence of the template to that of the protein under study. This can be done using the model building program, COOT.64

2.3.7. Model Building and Refinement

Once Molecular replacement was done, a preliminary model of the protein was obtained using the initial phase information. At this stage the model was highly biased towards the template search model. To remove this biasness, multiple rounds of model building and refinement were done to obtain better set of phases which agreed better with the diffraction data. Several parameters were used during this process to judge the correlation between the diffraction data and the model, which at the end should have been maximized. One of the parameters was the R-factor which is defined by the following equation:

\[ R = \frac{\sum_{\text{all reflections}} |F_o - F_c|}{\sum_{\text{all reflections}} |F_o|} \]
where $F_o$ is the observed structure factor and $F_c$ is the calculated structure factor. Another parameter was the $R_{\text{free}}$, which was based on a set of reflections ($\sim10\%$) that were not included in the structure refinement. A good model building and refinement should yield $R_{\text{free}}$ which is approximately equal to the resolution of the data set in angstroms divided by 10.

The program used for model building was COOT. The refinement programs used were CNS and REFMAC. Multiple rounds of model building and refinement were done until the $R$ values did not reduce any further. At that point, water molecules were added using the program ARP-water, which is also a part of the CCP4 suite. After multiple rounds of refinements, structure validation was done.

### 2.4. Analysis of Quaternary State and DNA Binding Affinity

#### 2.4.1. Dynamic Light Scattering (DLS)

DLS experiments were used for determine the homogeneity of the purified protein sample. Protein samples at concentrations of $\sim 1 \text{ mg/mL}$ were filtered using a centrifugal single-use Ultrafree®-MC centrifugal filter (Millipore, $0.1 \text{ µm}$ pore size). Ultra-filtered MilliQ water was obtained using Protein Solutions microfilter system with a Whatman Anodisc 13 filter ($0.02 \text{ µm}$ pore size). DLS experiments were performed on DynaPro-801 DLS instrument (Protein Solutions) and data acquired using Dynamics™ software (version 6.7.4.) at room temperature and at 4 °C in presence and absence of various additives. To check for background, ultra filtered milliQ was taken in a specialized DLS cuvette: 16.45F-Q-3 quartz spectrophotometer cell, sub-micro Fluorimeter ($45 \text{ µL}$, 3 mm,
Starna Cells) and placed in the DLS cuvette holder and allowed to equilibrate to the desired temperature for at least five minutes. Once it was made sure that the background count for water was in the acceptable range of 40,000 counts per second, the actual experiment was started. If the counts for water were higher than 40,000, the cuvette was cleaned again using a special cuvette cleaning solution from Starna Cells Inc. When ready, 15 µL of sample was loaded into the cuvette using a gel loading tip. Once the cuvette was equilibrated to the desired temperature for at least a minute, data collection was started and data analyzed for sample homogeneity. A poly-dispersity of 15% or less was considered as acceptable for crystallization. Other parameters monitored include Distribution, Baseline (0.997 to 1.005) and Sum of Squares (SOS).

2.4.2. Size Exclusion Chromatography

Superdex-75 size exclusion chromatography in addition to being used as a final purification step was also used for determining the solution state of the protein. First, the column was calibrated using different standard proteins of known molecular weights. The standards which were used were Standard 1 (40.1 kDa), Standard 2 (35.5 kDa), Standard 3 (31.8 kDa), Standard 4 (25.9 kDa) and Standard 5 (14.4 kDa). A calibration curve (Figure 2.7) was then plotted based on the elution time for each of the standard proteins. This curve was then used to determine the apparent size of the sample protein based on its elution time.
2.4.3. Fluorescence Anisotropy

To determine the fluorescence anisotropy associated with binding of SSB to a fluorescent tagged ssDNA, a PTI Quanta Master 4 fluorometer equipped with film polarizers was used. Anisotropy was calculated using the following formula:

\[
A = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}
\]
where $I_{VV}$ and $I_{VH}$ are the fluorescence intensities of the parallel and the perpendicular components respectively of the emitted fluorescence signal. $G$ is a factor to correct for the difference in the sensitivity of the system to vertically and horizontally polarized light.\textsuperscript{67} Values for anisotropy were averaged over 10 measurements taken at 22 °C. Depending upon the fluorescent tag used, excitation and the emission wavelength were adjusted using the software interface Felix32 (from Photon technology International Inc.). In this study two different types of fluorescent tags were used, fluorescein Isothiocynate tag on the 9mer oligonucleotide (excitation $\lambda_{\text{max}} = 490$ nm, emission $\lambda_{\text{max}} = 515$ nm) and HEX- fluorescein tag on the 15mer oligonucleotide (excitation $\lambda_{\text{max}} = 535$ nm, emission $\lambda_{\text{max}} = 556$ nm).

First, the arc lamp was warmed up for 45 minutes and then 425 µL of buffer containing a standard concentration of fluorescent labeled ssDNA added to a micro fluorometer cuvette (type 28F, Starna cells). To this sample, 100 ng of Bovine Serum Albumin (BSA) was added and mixed well by pipetting using a gel loading tip. The first reading constituted the blank or the initial anisotropy reading which was automatically calculated by the software, Felix32. Subsequently, concentrated SSB of known concentration was titrated into the cuvette and readings taken after thorough mixing (care was taken to avoid bubble formation). Data acquired by the Felix32 software was processed, first using Excel (spreadsheet program of the Microsoft office suite) and then using the software ORIGIN (version 7).

The main idea of this experiment was to determine the dissociation constant ($K_d$) of the SSB-ssDNA binding. Determination of $K_d$ was complex and involved the use of analysis of Lohman and Bujalowski.\textsuperscript{68} The basis of Lohman and Bujalowski’s analysis
was to identify concentrations of ligand, $L_T$ (SSB in this study) that gave the same anisotropic signal at different concentrations of the acceptor, $M_T$ (fluorescent tagged ssDNA in this study). Main assumption was that these points reflect the same fractional occupancy (refer to graph (a) Figure 1.7). The purpose was to determine the degree of binding (moles of ligand per mole of acceptor) at the same concentration of free ligand. To obtain this, binding experiments were performed at four different ssDNA concentrations. Data was then fitted to a function that described it, Hill equation in this case.

Hill equation:

$$\theta = \frac{[L]^n}{K_d + [L]^n}$$

where $\theta$ is the fraction of ligand binding sites filled, $[L]$ is the ligand concentration, $K_d$ denotes dissociation constant and $n$ represents Hill coefficient with $n = 1$ indicating non-cooperative association, $n > 1$ indicating positive cooperativity and $n < 1$ indicating negative cooperativity. The graph obtained with this fit was of normalized anisotropy signal versus SSB concentration. To get a plot that related SSB concentration to ssDNA concentration, a second graph of SSB ($L_T$) versus ssDNA ($M_T$) was plotted based on the first graph. From this graph the values of binding density, $\sum v_i$ and free SSB ($L_f$) were obtained. This data of binding density and free SSB were then fitted to Hill equation which gave the value of the dissociation constant ($K_d$) and Hill constant ($n$) which described the cooperativity of the binding.
Figure 2.8: Graphs based on Lohman and Bujalowski’s analysis to determine dissociation constant, K_d.
(a) The 1st plot is based on the data obtained from a fluorescence anisotropy experiment.
(b) Based on the 1st plot, this graph gives a direct relationship between SSB concentration and ssDNA concentration. The slope of each line gives the binding density, $\sum v_i$, and the y-intercept gives the value of free SSB, LF.
(c) The plot, fitted to a 2nd degree polynomial, shows the relationship between anisotropic signal and the binding density, $\sum v_i$.
(d) This plot of $\sum v_i$ versus LF, fitted to Hill equation, gives the $K_d$ value.
2.4.4. Analytical Ultra-Centrifugation (AUC)

Preliminary sample preparation for analytical ultra-centrifugation experiment was done in Mueser lab. For the actual experiment, samples were sent to the AUC facility at Brown Cancer Research Institute, Louisville, Kentucky.

2.4.4.1. Sample preparation

Samples for analytical ultra-centrifugation had to be very pure so as to avoid any complications in analysis due to presence of contaminants. Purified samples were checked for purity on a SDS-PAGE gel and a size exclusion chromatography run was performed if needed.

The next step was to dialyze the protein extensively in the buffer to be used as blank in the reference cell of the analytical ultra-centrifuge. Protein and buffer solutions were filtered using 0.22 micron centrifugal filters. Concentration of the protein was verified using the filtered dialysis buffer as blank.

2.4.4.2. Sedimentation Velocity (SV)

Samples were loaded into three 2-sector cells with the reference sectors filled with dialysis buffer as shown in Figure 2.9 (a). The velocity experiment was performed at a speed of 60,000 rpm and a temperature of 20 °C. Samples were equilibrated for one hour after vacuum and temperature had been established. Data was collected at 280 nm as a function of radial position. Each cell was scanned sequentially with zero time delay
between scans until no further sedimentation was observed. Primary data were transferred to the sedimentation analysis software DCDT+ (version 1.995.2171.26857) and analyzed with a single species model. Data were transformed into g(s*) distribution and molecular weight was fit (s/D ratio).

### 2.4.4.3. Sedimentation Equilibrium (SE)

Calculation of physical parameters like partial specific volume and molecular weight were done using the software SEDNTERP (version 1.08) with the input being amino acid sequence and buffer density/viscosity. Samples were loaded into 6-channel cells with the reference sectors filled with dialysis buffer as shown in Figure 2.9 (b). Sedimentation Equilibrium (SE) experiment was performed at four different speeds of 28,000; 34,000; 40,000; and 48,000 rpm at a temperature of 20 °C. Data were collected at 280 nm as a function of radial position. Scans were collected every 15 minutes until equilibrium was attained as judged by monitoring the standard deviation between successive scans. Primary data in the form of A_{280} versus radius in centimeters were fitted to a single, ideal species model using the instrument supplied data analysis software (XL-A/XL-I data analysis software version 6.03 based on the Origin data analysis software from Origin Labs Software, Inc.). In addition, the primary data were cast into plots of ln(A_{280}) versus the square of the radial position. When the data was transformed into the form of ln(abs) versus the square of the radial position, a linear correlation was exhibited by a single non-interacting species, with a slope proportional to the molecular weight. Any deviation from a linear correlation was indicative of complexities such as association.
Figure 2.9: (a) Sedimentation velocity cell design. Sample is loaded into the upper sector and the reference solution (buffer) is loaded into the bottom sector. The sample is centrifuged at high rotational velocity, generating a boundary that moves towards the bottom of the cell.

(b) Six channel sedimentation equilibrium cell design. Three sample–reference pairs are loaded into the cell which is then centrifuged at moderate rotational velocity, resulting in equilibrium concentration gradients in each sample channel. The three samples are generally of three different concentrations of the same protein.
Chapter 3: *Aeropyrum pernix* SSB (Ape SSB)

3.1. Ape SSB full length

3.1.1. Cloning

Full length *Ape* SSB gene (Gene ID: APE 1323) was cloned using pDrive cloning technique explained in section 2.1. The gene was inserted into expression vector pET21a and then transformed into expression hosts. Preliminary cloning was done by Pooja Talaty, a previous lab member. Gene sequence obtained from NCBI web server is shown in Figure 3.1. Amino acid sequence is shown in Figure 1.15(a), section 1.4.1 and protein parameters are shown in Table 1.1 in the same section.

(a)

```
1 ATGGACCTAA GGGAGGTCT TAGGAACGTC TCGATCTCTG GCCGCGTCCT TGAGACGGGA
61 GAGCCCAAGA TGGTTGAAAC CAAGAGAGGC CCTGCAACCC TGAGCGAAGC CGTTGTAGGC
121 GATGAGAGTG GTAGGGTTAA GGTGACTCTC TGGGGAAGCC ACGCGGGAAC ACTAAAGGAG
181 GGCGAGGCCG TTAGGATAGA GGGCGCCTGG ACAACCAGCT ACAGGGGCAA AGTCCAGGTC
241 AACGTCGGCA GAGAGTCGAC CATAGAGAAG GTGGACTCCG AGGATGTACC CCAAGCGGAG
301 GGAGCCGCGG GTTTCGGCGG CTTCAGAGGA GGCGGCTACC AGCCCCGAAG AGGAGGCAGG
361 AGGAGGTTCT AG
```

*Figure 3.1:* Gene sequence for *Ape* SSB with 432 bases. Sequence obtained from NCBI web server (GI: APE 1323). For details of the protein refer to Table 1.1, section 1.4.1.
3.1.2. Protein Expression study

Preliminary protein expression study (section 2.2.1) was done using two cell lines, BL21(DE3)-RIL and BL21(DE3)-RP.

As seen from SDS-PAGE gel in Figure 3.2, soluble protein was obtained with both cell lines (two colonies each) at 37 °C with the expected molecular weight of around 15.7 kDa. Colony #1 of Ape SSB-pET21a in BL21(DE3) RP was selected and glycerol stock made of the same. This was the cell line used for subsequent protein expression studies.

3.1.3. Solubility Profile

Solubility study was done on a sample of Ape SSB protein obtained from a trial purification run. The protein was pure enough to do solubility study. The procedure for
Figure 3.3: Solubility profile of *Ape* SSB indicating trisodium citrate pH 7.5 and ammonium chloride as the two optimum components of a buffer giving the highest solubility of the protein.
this study is described in Chapter 2, section 2.2.5. Based on the results of the study (Figure 3.3), a buffer consisting of 25 mM tri-sodium citrate pH 7.5 and 150 to 250 mM ammonium chloride was considered as optimum for obtaining the protein with maximum solubility. For all subsequent lysis and purification runs, this was the buffer used.

### 3.1.4. Protein Purification

Preliminary protein purification involved heating the lysate to 75 °C for 10 to 15 minutes with occasional stirring.

![Figure 3.4: SDS-PAGE gel of Ape SSB solubility study shows soluble protein in supernatant after heating lysate to around 75 °C for 15 minutes.](image)

As seen from Figure 3.4, supernatant after heating was reasonably pure with most of the endogenous *E. coli* proteins precipitated out into the pellet. Soluble *Ape* SSB protein was seen in lysis buffer with low salt (0.15 M NH₄Cl), high salt (1.00 M NH₄Cl) and Bug buster®. Low salt buffer was selected for large scale lysis and purification.
Large scale expression and lysis with low salt buffer were done as described in sections 2.2.2 and 2.2.3 respectively. The buffer used was 25 mM trisodium citrate pH 7.5 and 150 mM ammonium chloride. After heating the lysate for 10 to 15 minutes at 75 °C, it was centrifuged at 18,000 x g for 30 minutes at room temperature. Supernatant was separated from the pellet and filtered using a syringe filter to remove any cell debris. Filtrate was diluted with dilution buffer (25 mM trisodium citrate, pH 7.5) to lower the conductivity to that of SP-Sepharose cation exchange column loading buffer, Buffer-A (25 mM trisodium citrate, pH 7.5 and 50 mM ammonium chloride) as described in section 2.2.4. Then using a gradient elution with Buffer-B (25 mM trisodium citrate, pH 7.5 and 1.0 M ammonium chloride), the protein was eluted from the column.

Figure 3.5 shows the chromatogram for SP-Sepharose purification run for _Ape SSB_ full length protein. Figure 3.6 shows the SDS-PAGE gel picture of different fractions from the SP-Sepharose run. As seen from the picture, _Ape SSB_ eluted from column in fractions 13 to 27. At this stage the protein was partially pure with few contaminants still present. Fractions from 17 to 26 containing the protein were pooled and dilution buffer added to match the conductivity to that of Buffer-A. POROS-HS cation exchange column, a high resolution cation exchange column was then used to further purify the protein as described in section 2.2.4. Chromatogram for POROS-HS run is shown in Figure 3.7 and SDS-PAGE gel of the fractions under the peak is shown in Figure 3.8. Based on the SDS-PAGE gel (Figure 3.8), fractions 18 to 31 were pooled and concentrated as described in section 2.2.6 to bring down the volume. Since the protein started precipitating at concentrations above 3 to 4 mg/mL, it was finally concentrated to ~ 2 mg / mL and then purified by running on Superdex-75 size exclusion column.
Chromatogram for this run is shown in Figure 3.9. Samples from fractions under the peak showed the presence of relatively pure protein (Figure 3.10). Fractions 15 to 21 were pooled together and concentrated again to around 2 mg/mL. This was then stored at room temperature for immediate use or mixed with glycerol (30% final concentration of glycerol), frozen on dry ice and stored at -80 °C for future use.
Figure 3.5: Chromatogram of SP-Sepharose run for *Ape* SSB purification shows the protein being eluted upon increasing salt concentration.

Figure 3.6: SDS – PAGE gel of samples from SP-Sepharose run for *Ape* SSB purification showing most of the protein in fractions under the peak indicated by the red arrow.

#1: *Ape* SSB SP-Sepharose, sample loaded
#2: *Ape* SSB SP-Sepharose, flow through (load)
#3: *Ape* SSB SP-Sepharose, fraction 13
#4: *Ape* SSB SP-Sepharose, fraction 17
#5: *Ape* SSB SP-Sepharose, fraction 22
#6: *Ape* SSB SP-Sepharose, fraction 25
#7: *Ape* SSB SP-Sepharose, fraction 27
#8: Molecular weight marker
Figure 3.7: Chromatogram of POROS-HS run for *Ape* SSB purification

Figure 3.8: SDS – PAGE gel of samples from POROS-HS run for *Ape* SSB purification showing purified protein with trace contaminants in fractions below the peak.
Figure 3.9: Chromatogram of Superdex-75 run for Ape SSB purification

Figure 3.10: SDS – PAGE gel of samples from Superdex-75 run for Ape SSB purification showing pure protein in fractions shown by the colored arrows.
3.1.5. Crystal Screening

Crystal screening was done using all the screens mentioned in Table 2.9, section 2.3.1. Screens were set up at a concentration of around 2 mg/mL using 0.5 µL + 0.5 µL format as described in section 2.3.1. Only room temperature screens were set up. No significant crystal hits were obtained. Some of the wells had precipitated protein whereas most of the wells looked clear. There was no change in observations even after 15 days of setup.

3.1.6. Analysis of Quaternary State and DNA Binding Affinity

Three different types of studies were performed on Ape SSB full length protein namely, Dynamic Light Scattering (DLS), Size Exclusion Chromatography (SEC) and Fluorescence anisotropy to determine the solution state of the protein.

3.1.6.1. Dynamic Light Scattering

Dynamic light scattering experiments were done using the method described in section 2.4.1. Ape SSB at a concentration of 1 mg/mL was used for these experiments. Two different temperatures, 25 °C and 4 °C were used to determine the property of the protein in solution. The sample at 25 °C showed a polydispersity of 8.5% with an apparent molecular weight of 31 kDa. The sample at 4 °C showed a polydispersity of 33% with an apparent molecular weight of 16 kDa. Figure 3.11 and Figure 3.12 shows the plot for DLS at the two different temperatures mentioned above. DLS software (DYNAMICS version 7.7.4) based data analysis is also shown in the Figures.
Figure 3.11: *Ape* SSB DLS at 25 °C in presence of 25 mM trisodium citrate pH 7.5 and 500 mM NaCl.
Figure 3.12: *Ape* SSB DLS at 4 °C in presence of 25 mM trisodium citrate pH 7.5 and 500 mM NaCl.
Figure 3.13: Chromatogram of Superdex-75 analytical run for *Ape* SSB

Figure 3.14: SDS – PAGE of sample from Superdex-75 analytical run for *Ape* SSB

#1: Molecular weight marker
#2: *Ape* SSB Superdex-75, fraction 16

*Ape* SSB
(15.7 kDa)
3.1.6.2. Size Exclusion Chromatography (SEC)

Superdex-75 size exclusion chromatography was run as described in section 2.4.2. A 500 µL sample of purified and concentrated sample at around 2.0 mg/mL was loaded on to SEC for determining the size of the Ape SSB in solution. Figure 3.13 shows the chromatogram for the analytical run along with Figure 3.14 for SDS-PAGE gel sample from fraction 16. Based on the analysis from a standard curve (Figure 1.7); elution time for Ape SSB matches with a protein that should have a molecular weight of approximately 25 kDa. The calculated molecular weight of an Ape SSB dimer is 30 kDa.

3.1.6.3. Fluorescence Anisotropy

Fluorescence anisotropy experiment was done as described in section 2.4.3 using two different ssDNA, a 9mer oligonucleotide and a 15mer oligonucleotide. Figure 3.15 shows the graph of anisotropy versus Ape SSB concentration at four different 9mer ssDNA concentrations namely, 0.05 µM, 0.20 µM, 1.25 µM and 2.50 µM in presence of 250 mM ammonium chloride and 25 mM trisodium citrate, pH 7.5 buffer. Based on the graph in Figure 3.15, a graph of Ape SSB concentration versus ssDNA concentration was plotted as shown in Figure 3.16 based on which the fourth graph shown in Figure 3.18 was plotted. Graph in Figure 3.17 shows the dependence of anisotropic signal on the binding density of SSB on ssDNA. Graph 3.18 fitted to Hill equation gives a calculated $K_d$ of $\sim 757 \pm 92$ nM. Graphs 3.19 to 3.22 are for Ape SSB in presence of four different 15mer ssDNA concentrations (0.30 µM, 0.65 µM, 0.80 µM and 1.35 µM). Graph in Figure 3.22 fitted to Hill equation gives a calculated $K_d$ of $\sim 356 \pm 7$ nM.
Figure 3.15: Graph of anisotropy versus Ape SSB concentration at four different 9mer ssDNA concentrations in presence of 250 mM ammonium chloride.

Figure 3.16: Graph of Ape SSB concentration versus ssDNA concentration plotted as described in section 2.4.3.
Figure 3.17: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.

Figure 3.18: Final graph of binding density versus free protein concentration fitted to Hill equation gives a $K_d$ of $757 \pm 92$ nM.
Figure 3.19: Graph of anisotropy versus Ape SSB concentration at four different 15mer ssDNA concentrations in presence of 250 mM ammonium chloride.

Figure 3.20: Graph of Ape SSB concentration versus ssDNA concentration plotted as described in section 2.4.3.
**Figure 3.21:** Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.

**Figure 3.22:** Final graph of binding density versus free protein concentration fitted to Hill equation gives a $K_d$ of $\sim 356 \pm 7$ nM.
Table 3.1 summarizes the results of the fluorescence anisotropy experiments for determining the binding affinity of Ape SSB with two different lengths of ssDNA oligomers. The two different dissociation constants obtained were based on whether the parameter ‘n’, also known as the Hill constant, was fixed or allowed to vary when fitting Hill equation to the data for plotting binding density, $\sum v_i$, versus free protein concentration (the final graph used for determining $K_d$).

Results show that there is tighter binding in presence of longer ssDNA ($K_d$ of 356 ± 7 nM with 15mer ssDNA compared to $K_d$ of 758 ± 42 nM with 9mer ssDNA).

<table>
<thead>
<tr>
<th>ssDNA length</th>
<th>Tag</th>
<th>ssDNA concentrations (µM)</th>
<th>Salt concentration (mM)</th>
<th>$\sim K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9mer fluorescein</td>
<td>0.05, 0.20, 1.25, 2.50</td>
<td>250</td>
<td>758 ± 42 at n = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1558 ± 110 at n = 0.75</td>
</tr>
<tr>
<td>2</td>
<td>15mer HEX-fluorescein</td>
<td>0.30, 0.65, 0.80, 1.35</td>
<td>250</td>
<td>356 ± 7 at n = 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>362 ± 11 at n = 1.81</td>
</tr>
</tbody>
</table>
3.2. Ape SSBΔC (truncated)

3.2.1. Cloning (Quick Change®)

Truncated SSB gene was cloned using Quick Change® technique as mentioned in section 2.1.11. By this technique, a stop codon was introduced before the beginning of the C-terminus in the gene for the full length protein. Transcription of this gene then resulted in the synthesis of mRNA that codes for Ape SSBΔC protein.

![Diagram](image)

**Figure 3.23:** Quick change® site directed mutagenesis. Nucleotide T is introduced at position 346 of the Ape SSB gene resulting in the stop codon TAG. Transcription of this mutated gene then resulted in synthesis of mRNA that codes for Ape SSBΔC protein.

```
1   ATGGACCTAA  GGGAGGGTCT  TAGGAACGTC  TCGATCTCTG  GCCCGCTCCT  TGAGACGGGA
61  GAGCCCAAGA  TGGTTGAAC  CAAGAGAGGC  CCTGCAACCC  TGAGCGAAGC  CGTTGTAGGC
121 GATGAGAGTG  GTAGGGTTAA  GGTGACTCTC  TGGGGAAGCC  ACGCGGGAAC  ACTAAAGGAG
181 GGCAGGCCG  TTAGGATAGA  GGGCGCTGG  ACAACCAGCT  ACAGGGGCAA  AGTCCAGGTC
241 AACGTCGGCA  GAGAGTCGAC  CATAGAGAAG  GTGGACTCCG  AGGATGTACC  CCAAGCGGAG
301 GACATACCCG  AGGAGATGCC  GGAGGCCCAG  TATAGGGGCT  TCGGGTAG
```

**Figure 3.24:** Gene sequence for Ape SSBΔC with 348 bases. Sequence obtained from NCBI web server (GI: APE 1323). For details of the protein refer to Table 1.2, section 1.4.1.
The mutated gene in pET21a was successfully transformed into XL1-Blue competent cells. Supercoiled pET21a plasmid with the inserted gene was subsequently obtained from grow up of XL1-Blue colony on an agar plate. This plasmid was then transformed successfully into expression hosts BL21(DE3) RP, RILP and Rosetta2 pLysS cell lines. These cell lines were chosen since the gene had eight rare codons for arginine. Following transformation, expression studies were done to confirm the expression of the truncated *Ape SSBΔC* protein instead of the full length *Ape SSB* protein.

### 3.2.2. Protein Expression study

![Figure 3.25: SDS-PAGE gel showing Ape SSBΔC expression study using BL21(DE3) RP, RILP and Rosetta2 pLysS cell lines. Expression was seen in case of all the three cell lines.](image)

As shown in Figure 3.25, expression studies indicated the expression of truncated *Ape SSBΔC* with an apparent molecular weight which was lower than that for the full
length protein (refer to Figure 3.2 for gel picture showing full length *Ape* SSB). The calculated molecular weight for truncated protein was 12.5 kDa. Although expression was seen with all the cell lines, during the initial part of the project only the BL21(DE3) RP cell line was tested and used for expressing soluble protein. During the latter half of the project, Rosetta2 pLysS cell line was available and was therefore used for expression of soluble protein. This was done because the glycerol stock of RP cell line stopped showing expression after two years of storage at -80 °C and the plasmid had to be retransformed into fresh competent expression host.

### 3.2.3. Protein Solubility Study

![SDS-PAGE gel showing Ape SSB\(\Delta\)C solubility study using BL21(DE3) RP cell line.](image)

<table>
<thead>
<tr>
<th>Mol.Wt. (kDa)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- #1: Molecular weight marker
- #2: *Ape* SSB\(\Delta\)C Pellet after lysis (low salt)
- #3: *Ape* SSB\(\Delta\)C Supernatant after lysis (low salt)
- #4: *Ape* SSB\(\Delta\)C Supernatant heated after lysis (low salt)
- #5: *Ape* SSB\(\Delta\)C Pellet after lysis (high salt)
- #6: *Ape* SSB\(\Delta\)C Supernatant after lysis (high salt)
- #7: *Ape* SSB\(\Delta\)C Supernatant heated after lysis (high salt)
- #8: *Ape* SSB\(\Delta\)C Supernatant after lysis (bug buster)
- #9: *Ape* SSB\(\Delta\)C Supernatant heated after lysis (bug buster)

*Ape* SSB\(\Delta\)C (12.4 kDa)

**Figure 3.26:** SDS-PAGE gel showing *Ape* SSB\(\Delta\)C solubility study using BL21(DE3) RP cell line.

The next step was to check the solubility of protein in low salt (0.15 M NH\(_4\)Cl) versus high salt (1.00 M NH\(_4\)Cl) buffer. Figure 3.26 shows the solubility of *Ape* SSB\(\Delta\)C
determined using the method described in section 2.2.1. As seen from lanes 3 and 6, soluble protein was obtained in the supernatant in presence of both low and high salt. Also, the protein was still present in the supernatant after heating the lysate to around 75 °C as seen in lanes 4 and 7. Soluble protein was also present in the supernatant after lysis with BugBuster™. Solubility study was also done on protein expressed from Rosetta2 pLysS and showed similar solubility.

3.2.4. Solubility Profile

As described in section 2.2.5., solubility study was done on a sample of *ApeSSBΔC* obtained from a trial purification run. The preliminary purification run using column chromatography with standard Q-Sepharose buffers yielded enough protein with reasonable purity to do a solubility study. As seen from Figure 3.27, *ApeSSBΔC* showed enhanced solubility with several salts but eventually NaCl was selected. The optimized buffer based on this study consisted of 20 mM TAPS pH 8.5 and 150 mM sodium chloride. For all subsequent lysis and purification runs, this was the buffer used.
Figure 3.27: Solubility profile of *Ape SSBAC* indicating TAPS, pH 8.5 and sodium chloride as the two optimum components of a buffer giving the highest solubility of the protein.
3.2.5. Protein Purification

Large scale expression study was done as described in section 2.2.2 and lysis done with low salt buffer as described in section 2.2.3. The Lysis buffer had 20 mM TAPS, pH 8.5 and 150 mM sodium chloride in addition to the standard lysis buffer components. Since Ape SSBΔC was present in the supernatant after heating the lysate to 75 °C for 15 to 20 minutes (lane #4, Figure 3.25), the first step in purification was to heat the lysate after sonication to 75 °C for 15 to 20 minutes and then centrifuge at 18,000 x g for 30 minutes at room temperature. Supernatant was separated from the pellet and filtered using a syringe filter to remove any cell debris. Filtrate was then diluted with dilution buffer (20 mM TAPS, pH 8.5) to lower the conductivity to match that of Q-Sepharose cation exchange column loading buffer, Buffer-A (20 mM TAPS, pH 8.5 and 50 mM NaCl) as described in section 2.2.4 and then loaded on to the column. After loading, the column was rinsed with two to three column volumes of Buffer-A and protein eluted using gradient elution with Buffer-B (20 mM TAPS, pH 8.5 and 750 mM NaCl). Figure 3.28 shows the chromatogram for Q-Sepharose purification run for Ape SSBΔC protein. Figure 3.29 shows the SDS-PAGE gel picture of the different fractions from the Q-Sepharose run. As seen from the picture, Ape SSBΔC eluted from the column in fractions 34 to 45. At this stage the protein was partially pure with few contaminants still present. As seen from the chromatogram, the resolution of separation was not good when a continuous gradient elution was used. So an alternative step gradient elution program was used to separate the contaminants from the Ape SSBΔC protein. As seen in Figure 3.30 the separation was much better using this technique. During the initial stages
of this study, the continuous gradient elution was used and so a high resolution, POROS-HQ anion exchange column had to be used to separate and further purify the protein from contaminants. For this, the fractions from Q-Sepharose containing the protein were pooled and dilution buffer added to match the conductivity to that of Buffer-A. This was then loaded on to POROS-HQ as described in section 2.2.4. Chromatogram for this run and the SDS-PAGE gel of the fractions under the peak are shown in Figure 3.32 and Figure 3.33 respectively. During the later half of this project, the step gradient elution with Q-Sepharose run was the only anion exchange column purification done to purify the protein.

The protein was then concentrated as described in section 2.2.6 to bring down the volume and salt concentration. *Ape* SSBΔC was concentrated to almost 60 mg/mL at ~ 100 mM NaCl and then purified further by running on Superdex-75 size exclusion column. Three different peaks were seen on the chromatogram as shown in Figure 3.34 of which pure protein was present in the 3\(^{rd}\) peak covering fractions 14 to 17 as seen from the SDS-PAGE gel shown in Figure 3.35. These fractions under the peak were then pooled together and concentrated to around 30 to 45 mg/mL and then stored at room temperature for immediate use or mixed with glycerol (30% final concentration of glycerol), frozen on dry ice and stored at -80 °C for future use.
Figure 3.28: Chromatogram of Q-Sepharose run for *Ape* SSBΔC purification

**A260**

**A280**

Figure 3.29: SDS – PAGE of samples from Q-Sepharose run for *Ape* SSBΔC purification

- **#1:** Molecular weight marker
- **#2:** *Ape* SSBΔC Q-Sepharose, fraction 31
- **#3:** *Ape* SSBΔC Q-Sepharose, fraction 33
- **#4:** *Ape* SSBΔC Q-Sepharose, fraction 35
- **#5:** *Ape* SSBΔC Q-Sepharose, fraction 38
- **#6:** *Ape* SSBΔC Q-Sepharose, fraction 40
- **#7:** *Ape* SSBΔC Q-Sepharose, fraction 43

*Ape* SSBΔC (12.4 kDa)
Figure 3.30: Chromatogram of modified step gradient Q-Sepharose run for Ape SSBΔC purification

Figure 3.31: SDS – PAGE of samples from modified step gradient Q-Sepharose run for Ape SSBΔC purification
Figure 3.32: Chromatogram of POROS-HQ run for \textit{Ape SSB}\Delta C purification

Figure 3.33: SDS – PAGE of samples from POROS-HQ run for \textit{Ape SSB}\Delta C purification

\#1: \textit{Ape SSB}\Delta C POROS-HQ, fraction 7
\#2: \textit{Ape SSB}\Delta C POROS-HQ, fraction 8
\#3: \textit{Ape SSB}\Delta C POROS-HQ, fraction 10
\#4: \textit{Ape SSB}\Delta C POROS-HQ, fraction 12
\#5: \textit{Ape SSB}\Delta C POROS-HQ, fraction 15
\#6: \textit{Ape SSB}\Delta C POROS-HQ, fraction 17
\#7: \textit{Ape SSB}\Delta C POROS-HQ, fraction 19
\#8: \textit{Ape SSB}\Delta C POROS-HQ, fraction 21
\#9: \textit{Ape SSB}\Delta C POROS-HQ, fraction 29
\#10: Molecular weight marker
Figure 3.34: Chromatogram of Superdex-75 run for *Ape* SSBΔC purification showing three peaks.

*Figure 3.35: SDS – PAGE of samples from Superdex-75 run for *Ape* SSBΔC purification*
3.2.6. Crystal Screening

Crystal screening was done using all the screens mentioned in Table 2.9, section 2.3.1. Screens were set up at a concentration of around 25 mg/mL using 0.5 µL + 0.5 µL format as described in section 2.3.1. Only room temperature screens were set up. A few good crystal hits were seen, some of which are shown in Figure 3.36.

(a)     (b)     (c)
(d)    (e)    (f)
(g)    (h)

Figure 3.36: A few of the crystal hits obtained for Ape SSBΔC (a) Index condition #3, containing 0.1 M Bis-tris, pH 5.5 and 2.0 M ammonium sulfate (b) Additive screen condition #42, containing 20% PEG 4000 and 25 mM sodium oxalate (c) Additive screen condition #60, containing 20% PEG 4000 and 10 mM sodium iodide (d) Wizard-II condition #36, containing 10% PEG3000, 0.1 M phosphate citrate pH 4.2 and 0.2 M NaCl (e) Crystal screen-II condition #5, containing 2.0 M ammonium sulfate and 5% Isopropanol (f) Crystal screen-II condition #47, containing 2.0 M ammonium sulfate and 0.1 M sodium acetate, pH 4.8 (g) Salt Rx condition #65, containing 0.1 M Bis-tris propane, pH 7.0 and 2.5 M ammonium sulfate (h) Natrix condition #2, containing 2.5 M ammonium sulfate, 0.05 M MES, pH 5.6 and 0.01 M magnesium sulfate.
3.2.7. Optimization of crystal hits

Most of the hits shown in Figure 3.36 were expanded to obtain production quality crystals as described in section 2.3.2, but only two conditions could be successfully optimized. The first condition optimized was from Crystal Screen-II, condition #5, containing 2.0 M ammonium sulfate and 5% isopropanol. Crystals ranging in size from 0.2 micron to 1.2 micron were obtained at protein concentrations ranging anywhere from 15 to 60 mg/mL. Figure 3.37 shows some of the pictures of the production quality crystals obtained from this condition.

Figure 3.37: (a) to (e) shows crystals obtained on expanding crystal screen condition #5. Pictures (c) to (e) were taken using a polarizer. Crystal shown in (a) is 1.2 micron on the long cell axis. Pictures here are not to scale and crystal sizes cannot be compared between pictures.
The second crystal condition optimized was Crystal Screen-II, condition #47 containing 2.0 M ammonium sulfate and 0.1 M sodium acetate, pH 4.8. Upon expanding this condition at approximately 30 mg/mL of protein and a gradient of 0.8 M to 2.2 M ammonium sulfate, some of the crystals obtained looked like the ones shown in picture (f) in Figure 3.36. But most of the wells were either clear (first part of the gradient) or precipitated (later half of the gradient). A few expansion trays were set up over a period of time to optimize the crystal condition. One of the trays, which was at room temperature for almost a year, showed a different crystal morphology in three of the wells at a ammonium sulfate concentration of 1.50 M to 1.59 M. Of the 24 wells in the Costar™ tray, only the first three wells showed these crystals, the rest being clear with no sign of crystals or precipitates. Crystals were small and rod shaped as shown in Figure 3.38 and varied in length from 0.1 to 0.4 micron on the long cell axis and around 0.02 micron in diameter.

\[\text{(a) ~ (b)}\]

**Figure 3.38:** Crystals seen in wells A2 and A3 of the Costar™ tray aperpa-c–15. Well A2 had 1.55 M ammonium sulfate and 0.1 M sodium acetate with a protein concentration of 31 mg/mL. Well A3 had 1.59 M ammonium sulfate and 0.1 M sodium acetate. Crystals range in size from 0.1 to 0.4 micron along the long cell axis with a diameter of approximately 0.05 micron.
Two of the crystals that were successfully separated from the cluster are shown in Figure 3.39. These crystals were approximately 0.4 micron in length with a diameter of around 0.05 microns. Crystals were transferred into a fresh 1 µL drop of mother liquor to remove any debris sticking to the surface. Two such consecutive transfers were made to clean the surface.

3.2.8. Preliminary X-ray diffraction screening

As explained in section 2.2.3, different cryo protectant solutions were checked under a stream of nitrogen at 100 K to see which one gave a good freeze. Since the crystal was grown in a high salt condition, the cryo protectants that were tried were sodium formate, sodium malonate and lithium chloride. Of these, sodium malonate gave the best freeze and was subsequently used for freezing both crystal forms. The crystals
obtained were checked for diffraction quality as explained in section 2.3.4. Four diffraction frames were collected at phi angles of -75°, -30°, 15° and 60° on the R-AXIS IV detector. Detector distance from the crystal was 200 mm with a phi oscillation width of 1° with an exposure time of four minutes. Higher exposure times were also tried and gave better resolution.

Figure 3.40: Diffraction image of a hexagonal shaped crystal seen in Figure 3.37. (a) to (d) are images at phi angles of -75°, -30°, 15° and 60° respectively at a detector distance of 200 mm, 1° oscillation with an exposure time of 4 minutes per frame.
Several crystals from the first condition were screened, but in almost all cases the
diffraction pattern seemed twined with two patterns overlapping each other as seen in
Figure 3.40. Several attempts were made unsuccessfully to get a single crystal diffraction
pattern. There were many possible reasons, one of which being that this crystal condition
resulted in twin crystals which can be seen in picture (c) in Figure 3.37. Another reason
could be that the crystal was disordered with lot of free lattice space.

To verify the possibility of twining, attempts were made to separate the two
crystals using micro tools but they were too delicate resulting in shattering of the crystal.
Some of the bigger pieces of the shattered crystal were checked for diffraction, without
any success. To reduce crystal disorder, glutaraldehyde cross-linking was tried, where in
crystals were soaked in mother-liquor containing different concentrations of
 glutaraldehyde to strengthen the crystals. Observations taken at different intervals of time
showed the stability of the crystal in a particular concentration of glutaraldehyde. Once
an optimum concentration of glutaraldehyde was determined (4.3%) in which the crystal
was stable, diffraction screening was done. Unfortunately this step also did not yield a
good diffraction profile with the overlapping phenomena still seen in the pattern. In spite
of this, it was remarkable that though the crystal shrunk, it did not destroy the lattice and
diffraction was still obtained. Figures 3.41 and 3.42 show time lapse pictures of crystals
soaked in two of the several concentrations of glutaraldehyde that were tried; with the
former Figure showing an example of too high a concentration of the desiccating / cross
linking agent and the latter showing the minimum concentration of glutaraldehyde that
did not result in the crystal shrinking.
Figure 3.41: Glutaraldehyde cross-linking of *Ape* SSBΔC crystal. The final concentration of glutaraldehyde was ~ 6.7% (b) to (h) shows the physical state of the crystal over a period of 10 minutes in the desiccating condition. Crystal has a dimension of ~ 0.7 x 0.7 x 0.4 mm³. Images taken at 10 x magnification.
(a) Time: NA
(b) Time: 0 min 15 secs
(c) Time: 0 min 50 secs
(d) Time: 2 min 0 secs
(e) Time: 4 min 20 secs

(f) Time: 6 min, 50 secs
(g) Time: 8 min 20 secs
(h) Time: 9 min 20 secs
(i) Time: 10 min 20 secs
(j) Time: 11 min 50 secs

(k) Time: 12 min 40 secs
(l) Time: 14 min 20 secs
(m) Time: 16 min 50 secs
(n) Time: 17 min 50 secs
(o) Time: 26 min 0 secs

Figure 3.42: Glutaraldehyde cross-linking of Ape SSBΔC crystal. The final concentration of glutaraldehyde was ~ 4.3% (b) to (o) shows the physical state of the crystal over a period of 26 minutes in the desiccating condition. Crystal has a dimension of ~ 0.7 x 0.7 x 0.4 mm³ Images taken at 10 x magnification.
Three crystals (rod shaped) from the second crystallization condition were frozen in 60% sodium malonate and screened for X-diffraction. As seen from Figure 3.43, diffraction was reasonably good with no apparent overlapping diffraction pattern seen as in the first crystal form. The maximum resolution during preliminary screening was around 2.5 Å. Preliminary analysis using HKL2000 indicated the space group P3. Two of these crystals were stored for data collection at synchrotron at APS as discussed in section 2.3.4.

Figure 3.43: Diffraction images recorded on R-AXIS IV detector of a rod shaped crystal seen in Figure 3.39(a). (a)-(d) are images at phi angles of -75°, -30°, 15° and 60° respectively at a detector distance of 200 mm, 1° oscillation with an exposure time of 3 minutes per frame.
3.2.9. Data Collection

Data was collected on beamline 14-BM-C, at the synchrotron at APS, Argonne as described in section 2.3.4 using ADSC Quantum 4 detector. 400 frames at an oscillation width of 0.5° per frame were collected giving a total of 200° of diffraction images. Exposure time was 15 seconds per frame at a detector distance of 250 mm. The entire data set was collected with crystal under a stream of nitrogen at 100 K. Figure 3.44 shows couple of diffraction images from the data set collected. Data set collected at Argonne was subsequently processed in Mueser lab.

(a)  (b)  (c)  (d)

Figure 3.44: Diffraction images recorded on ADSC Quantum 4 detector at beamline 14-BM-C at APS, Argonne of the rod shaped crystal seen in Figure 3.32. (a)-(d) are images at phi angles of 0°, 45°, 90° and 135° respectively at a detector distance of 250 mm, 1° oscillation with an exposure time of 15 seconds per frame.
3.2.10. Data Processing

The steps involved in data processing are discussed in section 2.3.5. Initial data processing involved indexing using the software IPMOSFLM run on a Linux operating system. The first step was to read an image into the software using the “read image” command on the IPMOSFLM Graphic User Interface (GUI). The command “find spots” was then used to pick spots on the diffraction image. Images 1,180 and 200 were read in and spots picked. On an average around 375 spots were picked per image. Indexing was then done using the function “autoindexing” on the GUI. Initial autoindexing resulted in unit cell being predicted as P3 (solution 10). The estimated mosaicity was determined to be 0.40. Spot prediction was then done to see if the predicted spots overlapped with the actual spots on the images collected, which in this case was quite close with the majority of the spot predictions being right on target. After setting the resolution limit to 1.9 Å, the cell was then refined using the command “refine cell” for two segments starting at 0° and 90° with eight images in each segment. Mosaicity after cell refinement dropped to 0.36. Next, the beam stop region was assigned to remove that region from data processing. The final step was then to integrate the indexed images to obtain an output MTZ file, ape_ssbP3.mtz.

A program called pointless was then used to check for the correct space group. Pointless indicated two possible space group solutions, P3_{1}21 (space group #152) or P3_{2}21 (space group #154). Based on this result, the output MTZ file for P3 was converted to P3_{1}21 and P3_{2}21 using the program “Sort/Reindex MTZ files” which could be accessed from Data reduction in the CCP4 GUI. Two different output files were thus obtained, ape_ssbP3_{1}21.mtz and ape_ssbP3_{2}21.mtz. The approximate dimension of the
unit cell obtained was: a=84.19, b=84.19, c=50.94, \( \alpha = 90^\circ \), \( \beta = 90^\circ \), \( \gamma = 120^\circ \). In addition to the space groups mentioned above, other possible space groups were also analyzed as shown in Table 3.2.

Once the output files were obtained, the number of Ape SSB\( \Delta \)C molecules in the asymmetric unit were determined using the program “cell content analysis” which was accessed from “Coordinate Utilities” in the CCP4 GUI. Calculation was based on determining Mathew’s coefficient which in turn indicated the possible number of molecules in the asymmetric unit and also the solvent content. Results of these calculations are shown in Table 3.2.

<table>
<thead>
<tr>
<th>Space group</th>
<th># of molecules (Total # of amino acids)</th>
<th>Mathew’s coefficient</th>
<th>% solvent</th>
<th>Probability (at highest resolution)</th>
<th>Probability (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P3 (143)</td>
<td>3 (339 aa)</td>
<td>2.81</td>
<td>56.2%</td>
<td>0.31</td>
</tr>
<tr>
<td>2</td>
<td>P3 (143)</td>
<td>4 (452 aa)</td>
<td>2.10</td>
<td>41.6%</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>P3(_2) (152)</td>
<td>2 (113 aa)</td>
<td>2.11</td>
<td>41.7%</td>
<td>0.99</td>
</tr>
<tr>
<td>4</td>
<td>P3(_2) (154)</td>
<td>2 (113 aa)</td>
<td>2.11</td>
<td>41.7%</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>P3(_2) (145)</td>
<td>3 (339 aa)</td>
<td>2.78</td>
<td>55.0%</td>
<td>0.31</td>
</tr>
<tr>
<td>6</td>
<td>P321 (150)</td>
<td>2 (113 aa)</td>
<td>2.11</td>
<td>42.0%</td>
<td>0.99</td>
</tr>
<tr>
<td>7</td>
<td>P312 (149)</td>
<td>2 (113 aa)</td>
<td>2.11</td>
<td>42.0%</td>
<td>0.99</td>
</tr>
<tr>
<td>8</td>
<td>C2</td>
<td>6 (678 aa)</td>
<td>2.10</td>
<td>41.4%</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of results for calculation of Mathew’s coefficient for different space groups analyzed
Table 3.3: Summary of results of SCALA for different space groups analyzed

<table>
<thead>
<tr>
<th></th>
<th>Space group</th>
<th>Wilsons B factor</th>
<th>Rmerge</th>
<th>Total reflections</th>
<th>Unique reflections</th>
<th>Completeness %</th>
<th>Multiplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P3 (143)</td>
<td>18.2</td>
<td>8.4</td>
<td>190,139</td>
<td>31,751</td>
<td>99.7%</td>
<td>6.0</td>
</tr>
<tr>
<td>2</td>
<td>P3121 (152)</td>
<td>18.1</td>
<td>8.7</td>
<td>189,793</td>
<td>16,682</td>
<td>99.8%</td>
<td>11.4</td>
</tr>
<tr>
<td>3</td>
<td>P3221 (154)</td>
<td>18.1</td>
<td>8.7</td>
<td>190,124</td>
<td>16,704</td>
<td>99.8%</td>
<td>11.4</td>
</tr>
<tr>
<td>4</td>
<td>P32 (145)</td>
<td>18.2</td>
<td>8.4</td>
<td>190,118</td>
<td>31,733</td>
<td>99.7%</td>
<td>6.0</td>
</tr>
<tr>
<td>5</td>
<td>P321 (150)</td>
<td>18.1</td>
<td>8.7</td>
<td>190,119</td>
<td>16,703</td>
<td>99.8%</td>
<td>11.4</td>
</tr>
<tr>
<td>6</td>
<td>P312 (149)</td>
<td>17.3</td>
<td>45.6</td>
<td>190,291</td>
<td>16,378</td>
<td>99.9%</td>
<td>11.6</td>
</tr>
<tr>
<td>7</td>
<td>C2</td>
<td>19.3</td>
<td>6.7</td>
<td>161,116</td>
<td>40,829</td>
<td>98.8%</td>
<td>3.9</td>
</tr>
</tbody>
</table>
Once the number of amino acids per asymmetric unit was estimated, scaling was done using the program SCALA which could be accessed from the program group, Data Reduction, found in the CCP4 GUI. Scaling was done separately using the output MTZ file from IPMOSFLM for each of the three main space groups (P3, P3_{12}1 and P3_21) as well as for the other space groups analyzed. Parameters obtained from the scaling are shown in Table 3.3. Output of SCALA was a new MTZ file with high resolution scaled data. A low resolution MTZ file was then created from this high resolution file using the program “convert to/modify/extend MTZ” that can be found under the program group “reflection data utilities” in CCP4 interface. This low resolution MTZ file had resolution starting from 3Å and was necessary for use with molecular replacement program.

3.2.10. Molecular Replacement

Once the experimental data was processed, two MTZ files – high resolution and low resolution scaled MTZ were obtained per space group analyzed. These MTZ files had the information of the intensities but lacked the phase information. Phase information was obtained using molecular replacement with the structure of a homologous protein, Sso SSB solved by Kerr and co-workers. Sequence alignment between Ape SSB\Delta C and Sso SSB was done using the program BLAST2sequences available on the NCBI website as shown in chapter 1, Figure 1.15. Sequence alignment showed a sequence homology of 72% with a sequence identity of 45% which was considered adequate for molecular replacement. The program used for this was PHASER, which can be accessed through CCP4 GUI. The low resolution MTZ file from SCALA along with the protein data
coordinates of \( Sso \) SSBΔC were the two files used by PHASER. The resolution range was set between 3 and 15 Å. Most of the other information required was left at the default values. Once the program finished analyzing, it generated couple of files with the two main ones being a PDB and an MTZ file. The program also generated a log file which showed the translational (\( T_Z \)) and the rotational parameters (\( R_Z \)) generated during molecular replacement. As shown in Table 3.4, higher the Z-score for the translation parameter more likely is it that the molecular replacement worked successfully.

**Table 3.4: Determining success of PHASER**

<table>
<thead>
<tr>
<th>( T_Z )</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 5</td>
<td>No</td>
</tr>
<tr>
<td>5 – 6</td>
<td>Unlikely</td>
</tr>
<tr>
<td>6 – 7</td>
<td>Possibly</td>
</tr>
<tr>
<td>7 – 8</td>
<td>Probably</td>
</tr>
<tr>
<td>More than 8</td>
<td>Definitely</td>
</tr>
</tbody>
</table>

Preliminary PHASER runs using the \( Sso \) SSB as the search model failed. Analysis of the log file indicated possibility of clashes between the amino acid side chains. To correct this problem, the \( Sso \) SSB protein sequence was mutated to polyalanine using the program “Moleman”. This resulted in a PDB file with the tertiary structure of the \( Sso \) SSB retained but the amino acid sequence changed to poly-alanine. PHASER was again run using this new PDB as the template. This time the molecular
replacement worked with high Tz values. Table 3.5 shows the Tz values obtained for molecular replacement with different MTZ files corresponding to the alternative space groups analyzed.

**Table 3.5: Result of molecular replacement using PHASER.**

<table>
<thead>
<tr>
<th>Space group</th>
<th># of Ape SSBΔC molecules in asymmetric unit</th>
<th>Tz value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P32</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>P3_21</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>P3_21</td>
<td>2</td>
</tr>
</tbody>
</table>

The result from molecular replacement showed that the space group for this data set was most likely P3_21. For further data processing/refinement, the high resolution MTZ file obtained from scaling and the PDB file from molecular replacement using the P3_21 space group were the only ones used.

### 3.2.11. Refinement and Model Building

The output PDB from PHASER now had the polyalanine structure aligned roughly to the experimental density map of the *Ape* SSBΔC. Two such molecules were present in the asymmetric unit – named chain A and chain B. The first step in refinement process involved theoretically heating the molecules to a very high temperature in order to impart enough kinetic energy to the amino acid backbone and side chains for crossing the energy barrier. Upon slow cooling, the residues would fall back in to a position that would reflect a global energy minimum. CNS (version 1.1 and 1.2) developed by Brunger and coworkers was used for preliminary refinements. One of the main inputs for CNS
was a “cv” file that was made by converting the high resolution MTZ file. After the first CNS run, the R value was 43% and the R\text{Free} was 45%.

\textbf{Ape SSB} 2 DLREGLRNVSISGRVLETGEPKMVEKTGRPGATLSEAVGDESGRVKVTWGSAGTLK 61
\textbf{Sso SSB} 7 NLKPNMESVNVTVRVLDAERQIQTKNGVMTISEAIYGDGTGRVKTTLWCKHAGSIK 66

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{alignment.png}
\caption{First half of the sequence alignment between \textit{Ape} SSB and \textit{Sso} SSB. The two yellow highlighted regions indicate sequences of $\beta_2$ and $\beta_3$ strand of the \textit{Sso} SSB structure.}
\end{figure}

Before mutating the poly-alanine residues of the PDB file obtained from molecular replacement, a starting point for mutation had to be identified. For this, a region that was highly conserved between the \textit{Sso} and \textit{Ape} SSB sequences needed to be identified. As seen in Figure 3.45, one such region was the main ssDNA binding surface, $\beta_2$ and $\beta_3$ strands along with the connecting L\textsubscript{23} loop seen in \textit{Sso} structure. The corresponding residues were the same in the \textit{Ape} SSB and upon visual inspection of the density map, $2F_o-F_c$, density for prominent amino acid side chains like that of tryptophan at position number 51 and of histidine three amino acids from it at position number 54 could be clearly seen. The program used for visualizing and model building was COOT.\textsuperscript{64} Using the mutate command in COOT, the entire stretch covering amino acids 32 to 57 in the poly-alanine sequence was mutated to the respective amino acids of the \textit{Ape} SSB\textsubscript{ΔC} sequence. At the same time, since the C-terminal region of the molecule looked highly disordered, amino acids from position 93 onwards on the polyalanine chain were deleted from both A and B chains. Another round of CNS was run on this new PDB and the output maps ($2F_o-F_c$ and $F_o-F_c$) and PDB were then visualized in COOT. The R and the R\text{Free} values at this stage did not change significantly.
In the next step of building, all the remaining poly-alanine sequences were mutated to the actual Ape SSBΔC sequence using COOT. Also the initial six residues on both the A and B chains were deleted, since the density did not look good in that region at this stage. The side chains of most amino acids were then oriented more or less closely to correlate with the density. This was done using the “Fit rotamer” function under “Model/Fit/Refine” category in COOT. Once the modifications were saved and a new PDB file created to reflect these changes, another round of CNS was run. At this point the R value dropped to 39.8% whereas R_Free value remained at 45%. The next step was to renumber the amino acids from that of the Sso to that of Ape (refer Figure 3.44). A walk-through was done on the structure and the amino-acid backbone aligned to fit the density as best as possible and then another round of CNS was run on this new PDB. The new R value was 39.5% and R_Free value rose to 47%. A composite omit map was run using the output of the last CNS. Based on the composite omit maps, model building was repeated and the C-terminal amino acids added until amino acid #103 on chain A. The built chain A was copied and the copy superposed on to chain B and the coordinates for this superposed molecule were used as the new coordinates for chain B. Another round of structure building (using COOT) and refinement (CNS) were done resulting in an R value of 39.6% and R_Free of 43%. CNS using bulk solvent correction resulted in an R value of 64.8% and R_Free of 68.5%. Since such an increase in R values was not possible, the previous version of CNS, version 1.1 was tried instead of version 1.2. This resulted in the R value dropping to 34.7% and R_Free to 38.5%. CNS with bulk solvent correction gave an R value of 29.5% and R_Free of 34.1%. After another round of model building using maps from CNS with and without bulk solvent correction, refinement was done using
REFMAC. The R value dropped significantly to 25.7% and R\textsubscript{Free} dropped to 27.5%. Model building was repeated and the C-terminal amino acids added until amino acid #106 on chain A. REFMAC was again run on the new PDB which resulted in an R value of 24.2% and R\textsubscript{Free} of 27.9%.

In the next round of model building, one more amino acid was added on to the C-terminal end of chain A (Proline at position #107). On chain B, amino acids from 93 onwards were deleted since there was not enough density to model the terminal amino acids. Refinement at this point with CNS gave R = 32.89% and R\textsubscript{Free} = 33.94%. After repositioning the side chains into the density as much as possible, a round of refinement with REFMAC gave a new R value of 23.8% and R\textsubscript{Free} of 27.8%. Next, amino acid # 108 was mutated from alanine to glutamic acid on chain A and after another round of model building, refinement was done using CNS (R = 32.02%, R\textsubscript{Free} = 34.58%). After this, couple of refinement cycles (REFMAC) alternated with model building were done which resulted in an R value of 22.12% and R\textsubscript{Free} = 25.98%. Next, REFMAC with 10 cycles of refinement to find water molecules resulted in determination of 59 water molecules (R = 19.71% and R\textsubscript{Free} = 24.5%). After visually confirming the presence of water molecules, another round of REFMAC was done to find additional water molecules but none were found (R = 19.59%, R\textsubscript{Free} = 24.51%). Table 3.6 summarizes the refinement steps involved in determining the structure of Ape SSB\textDelta C.
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Refinement</th>
<th>R</th>
<th>R_{Free}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Poly-alanine PDB from PHASER</td>
<td>CNS 1.2</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>2</td>
<td>Mutated amino acids #32 to 75 on chain A and B to that of <em>Ape SSBΔC</em></td>
<td>CNS 1.2</td>
<td>43%</td>
<td>46%</td>
</tr>
<tr>
<td>3</td>
<td>Mutated amino acids from #6 to #31 and from #76 until #92 to that of <em>Ape SSBΔC</em>, model building.</td>
<td>CNS 1.2</td>
<td>39.8%</td>
<td>45.5%</td>
</tr>
<tr>
<td>4</td>
<td>Corrected amino acid numbering to match that of <em>Ape SSBΔC</em>, Model building.</td>
<td>CNS 1.2</td>
<td>39.5%</td>
<td>47%</td>
</tr>
<tr>
<td>5</td>
<td>Composite Omit</td>
<td>CNS 1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Mutated C terminal poly-alanine residues from #93 to #103, Model building.</td>
<td>CNS 1.2</td>
<td>39.6%</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNS 1.2</td>
<td>65.8%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNS 1.1</td>
<td>34.7%</td>
<td>38.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNS 1.1</td>
<td>29.5%</td>
<td>34.1%</td>
</tr>
<tr>
<td>7</td>
<td>Model building on PDB from CNS 1.1 run from step #6</td>
<td>REFMAC5</td>
<td>25.7%</td>
<td>27.5%</td>
</tr>
<tr>
<td>8</td>
<td>Mutated C terminal poly-alanine residues from position #104 to #106, Model building.</td>
<td>REFMAC5</td>
<td>24.2%</td>
<td>27.9%</td>
</tr>
<tr>
<td>9</td>
<td>Composite Omit</td>
<td>CNS 1.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.6 (continued): Summary of Refinement steps for determining *Ape* SSBΔC crystal structure.

<table>
<thead>
<tr>
<th>Step</th>
<th>Refinement</th>
<th>R</th>
<th>R&lt;sub&gt;Free&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Deleted methionine at position #1, mutated alanine at #107 to proline, superposed A on to B, deleted amino acids from # 93 onwards on chain B, Model building.</td>
<td>CNS 1.1</td>
<td>32.89%</td>
</tr>
<tr>
<td>11</td>
<td>Model building</td>
<td>REFMAC5</td>
<td>23.8%</td>
</tr>
<tr>
<td>12</td>
<td>Mutated alanine at #108 to glutamic acid, Model building</td>
<td>CNS 1.1</td>
<td>32.02%</td>
</tr>
<tr>
<td>13</td>
<td>Composite Omit</td>
<td>CNS 1.1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Model building, worked mainly on amino acids # 93 to # 96.</td>
<td>REFMAC5</td>
<td>22.88%</td>
</tr>
<tr>
<td>15</td>
<td>Composite Omit with and without bulk solvent corrections</td>
<td>CNS 1.1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Built methionine at #1 on chain A only, Model building</td>
<td>REFMAC5</td>
<td>22.61%</td>
</tr>
<tr>
<td>16</td>
<td>Model building</td>
<td>REFMAC5</td>
<td>22.12%</td>
</tr>
<tr>
<td>17</td>
<td>10 cycles of refinement to find water molecules</td>
<td>REFMAC5</td>
<td>19.71%</td>
</tr>
<tr>
<td>18</td>
<td>10 cycles of refinement to find water molecules</td>
<td>REFMAC5</td>
<td>19.59%</td>
</tr>
</tbody>
</table>
Table 3.7 shows some of the important parameters of data collection, refinement and structure building of *Ape* SSBΔC.

**Table 3.7: Data Collection, Structure Determination and Refinement of *Ape* SSBΔC**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>14-BM-C</td>
</tr>
<tr>
<td>Wavelength</td>
<td>0.9</td>
</tr>
<tr>
<td>Resolution</td>
<td>72.93 – 1.9 Å</td>
</tr>
<tr>
<td>Space group</td>
<td><em>P</em>3₂21</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 4</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td>a = 84.188, b = 84.188, c = 50.941</td>
</tr>
<tr>
<td></td>
<td>α = 90°, β = 90°, γ = 120°</td>
</tr>
<tr>
<td>Solvent content</td>
<td>41.7%</td>
</tr>
<tr>
<td>Total reflections</td>
<td>190,124</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>16,704</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>11.4</td>
</tr>
<tr>
<td>Data completeness</td>
<td>99.75 %</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>8.7 %</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>19.82 %</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>24.5%</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>1511</td>
</tr>
<tr>
<td>Number of water molecules</td>
<td>59</td>
</tr>
</tbody>
</table>
Figure 3.46: Model of Ape SSBΔC crystal structure visualized using Pymol. The crystallographic dimer is shown. β-strands are colored in yellow and α-helix in red. The two important loops L_{12} and L_{45} are seen in the Figure. The binding cleft for the ssDNA can also be seen.

Figure 3.46 shows the cartoon of the Ape SSBΔC structure. The asymmetric unit has two molecules represented by chain A and chain B. The characteristic OB fold is seen with the β-sheets forming a β-barrel. Two important loops, L_{45} and L_{12}, along with the ssDNA binding cleft are also seen in the cartoon of the structure in Figure 3.46.

Figure 3.47 shows crystal packing visualized using COOT. The unit cell with the trigonal space group, P3_21, can be seen in the Figure with the two fold and three fold symmetry molecules (in grey). The colored molecules in the unit cell indicate the two Ape SSBΔC molecules of the asymmetric unit cell.
Figure 3.47: Crystal packing of *Ape* SSBAΔC seen using COOT. The space group for this particular crystal packing is P3₂₁, The yellow box indicates a unit cell with the following dimensions: $a = 84.188$, $b = 84.188$, $c = 50.941$, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 120^\circ$. The colored molecule represents the crystallographic dimer in the asymmetric unit while the grey colored molecules indicate the symmetry molecules generated. The two-fold symmetry can be seen in the bottom left of the unit cell and the three-fold symmetry is seen at all the four corners of the unit cell.
3.2.12. Analysis of Quaternary State and DNA Binding Affinity

Three different types of studies were performed on *Ape* SSBDΔC: Dynamic Light Scattering (DLS), Size Exclusion Chromatography (SEC) and Fluorescence anisotropy to determine the solution state of the protein.

3.2.12.1 Dynamic Light Scattering

Dynamic light scattering experiment, described in section 2.4.1., was done on *Ape* SSBDΔC at a concentration of ~ 1 mg/mL. Two different temperatures, 25 °C and 4 °C, were used to determine the quaternary state of the protein in solution. The sample at 4 °C seemed to form big aggregates indicated by the DLS instrument shutting down because of intensity signal beyond the threshold of the detector. The sample at 25 °C showed a polydispersity of 19.5% with an apparent molecular weight of 22 kDa. Figure 3.48 shows the plot for DLS at 25 °C along with DLS software (DYNAMICS version 7.7.4) based data analysis.
Figure 3.48: *Ape* SSBA C DLS at 25 °C in 25 mM TAPS pH 8.5 and 200 mM NaCl.
**Figure 3.49:** Chromatogram of Superdex-75 analytical run for \(Ape\) SSB\(\Delta\)C at a protein concentration of \(~25\) mg/mL shows a single peak eluting at fraction # 15.

**Figure 3.50:** SDS – PAGE of samples from Superdex-75 analytical run for \(Ape\) SSB\(\Delta\)C.

\(Ape\) SSB\(\Delta\)C
(12.4 kDa)

#1: \(Ape\) SSB\(\Delta\)C Superdex-75 fraction 14
#2: \(Ape\) SSB\(\Delta\)C Superdex-75 fraction 15
#3: \(Ape\) SSB\(\Delta\)C Superdex-75 fraction 16
#4: Molecular weight marker
Figure 3.51: Chromatogram of Superdex-75 analytical run for *Ape SSBΔC* at a protein concentration of ~1.5 mg/mL shows the protein eluting at fraction #22.

Figure 3.52: SDS–PAGE of samples from Superdex-75 analytical run for *Ape SSBΔC*.
3.2.12.2. Size Exclusion Chromatography (SEC)

Superdex-75 size exclusion chromatography was run as described in section 2.4.2. A 500 µL sample of purified and concentrated sample (~ 25.0 mg/mL) was loaded on to SEC for determining the size of the Ape SSBΔC in solution. Figure 3.49 shows the chromatogram for the analytical run along with the digital image of a SDS-PAGE gel (Figure 3.50) showing samples under the peak. Based on the analysis from a standard curve (Figure 2.7, section 2.4.2); the elution time for Ape SSBΔC matches with that of a protein that should have a molecular weight of approximately 22 to 25 kDa. The calculated molecular weight of an Ape SSBΔC dimer is 25 kDa.

A second SEC run of the same sample at a lower concentration of approximately 1.5 mg/mL showed a different elution time (Figure 3.51). This elution time corresponds to the elution time of a protein with a molecular weight of around 12 to 15 kDa, indicating that the protein was probably in a monomeric state.

3.2.12.3. Fluorescence Anisotropy

Fluorescence anisotropy experiment was done as described in section 2.4.3. using both ssDNA sizes, 9mer oligonucleotide and 15mer oligonucleotide. Figure 3.53 shows the graph of Anisotropy versus Ape SSBΔC concentration at four different 9mer ssDNA concentrations namely, 0.2 µM, 1.0 µM, 2.5 µM and 5.0 µM in presence of 100 mM sodium chloride. Based on the graph in Figure 3.14, a graph of Ape SSBΔC concentration versus ssDNA concentration was plotted as shown in Figure 3.54. Analysis was based on the method described in section 2.4.3.
Figure 3.53: Graph of anisotropy versus Ape SSBΔC concentration at four different 9mer ssDNA concentrations in presence of 100 mM sodium chloride.

Figure 3.54: Graph of Ape SSBΔC concentration versus ssDNA concentration plotted as described in section 2.4.3
**Figure 3.55:** Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.

**Figure 3.56:** Final graph of binding density versus free protein concentration fitted to Hill equation gives a $K_d$ of 228 ± 25 nM.
Figure 3.57: Graph of anisotropy versus Ape SSBΔC concentration at four different 9mer ssDNA concentrations in presence of higher salt concentration, 250 mM sodium chloride.

Figure 3.58: Graph of Ape SSBΔC concentration versus ssDNA concentration plotted as described in section 2.4.3.
Figure 3.59: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.

$\Delta \text{A}_{\text{obs}}$ vs $\sum v_i$

Figure 3.60: Final graph of binding density versus free protein concentration fitted to Hill equation gives a $K_d$ of $876 \pm 71$ nM.
**Figure 3.61:** Graph of anisotropy versus *Ape* SSBΔC concentration at four different 15 mer ssDNA concentrations in presence of higher salt concentration, 250 mM sodium chloride.

**Figure 3.62:** Graph of *Ape* SSBΔC concentration versus ssDNA concentration plotted as described in section 2.4.3.
Figure 3.63: Graph of normalized anisotropic signal versus $\sum v_i$ showing the dependence of anisotropic signal on the binding density.

Figure 3.64: Final graph of binding density versus free protein concentration fitted to Hill equation gives a $K_d$ of $159 \pm 10$ nM.
Table 3.8: Summary of results from fluorescence anisotropy experiment for *Ape SSBΔC*

<table>
<thead>
<tr>
<th></th>
<th>Protein</th>
<th>ssDNA length</th>
<th>Tag</th>
<th>ssDNA concentrations (µM)</th>
<th>Salt concentration (mM)</th>
<th>( \sim K_d ) (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Ape SSB ΔC</em></td>
<td>9mer</td>
<td>fluorescein</td>
<td>0.2, 1.0, 2.5, 5.0</td>
<td>100</td>
<td>229 ± 25 at n = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>385 ± 116 at n = 0.65</td>
</tr>
<tr>
<td>2</td>
<td><em>Ape SSB ΔC</em></td>
<td>9mer</td>
<td>fluorescein</td>
<td>0.2, 1.0, 2.5, 5.0</td>
<td>250</td>
<td>877 ± 72 at n = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1796 ± 568 at n = 0.762</td>
</tr>
<tr>
<td>3</td>
<td><em>Ape SSB ΔC</em></td>
<td>15mer</td>
<td>HEX-fluorescein</td>
<td>0.30, 0.65, 0.80, 1.35</td>
<td>250</td>
<td>159 ± 11 at n = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129 ± 4 at n = 1.32</td>
</tr>
</tbody>
</table>

Table 3.8 summarizes the results of the fluorescence anisotropy experiments for determining binding affinity of *Ape SSBΔC* with two different lengths of ssDNA oligomers. The two different dissociation constants obtained were based on whether the parameter ‘n’, also known as the Hill constant, was fixed or allowed to vary when fitting Hill equation to the data for plotting binding density, \( \sum v_i \), versus free protein concentration (the final graph used for determining \( K_d \)).

As shown in Table 3.8, two different salt concentrations were used during the experiment to study the effect of salt concentration on binding. Results show that there is tighter binding in presence of lower salt concentration and in presence of longer ssDNA (\( K_d \) of 159 ± 11 nM with 9mer ssDNA compared to \( K_d \) of 159 ± 11 nM with 15mer ssDNA).
3.2.12.4. Analytical Ultracentrifugation

As discussed in section 2.4.4, analytical ultracentrifugation experiment and analysis were done at Brown Cancer Research Institute, Louisville, Kentucky. Based on the analysis of data from both Sedimentation Velocity (SV) and Sedimentation Equilibrium (SE) experiment, a single species was determined to exist in solution with a molecular weight of ~ 11.11 kDa (calculated molecular weight of Ape SSBΔC is 12.5 kDa).

![Figure 3.65](image-url)

**Figure 3.65**: Data analysis using software DCDT+ for Sedimentation Velocity experiment at three different concentrations, 0.2, 0.4 and 0.8 mg/mL shown in (a), (b) and (c) respectively. Sedimentation coefficient, s (Svedbergs) value is shown in the text box. The analysis fit is consistent with a monomer with a molecular weight ~ 11.0 kDa. The calculated molecular weight of Ape SSBΔC is 12.5 kDa.
Figure 3.66: Data analysis using software SEDNTERP for Sedimentation Equilibrium experiment at three different concentrations, 0.2, 0.4 and 0.8 mg/mL shown as black, red and green data points respectively in (a) to (d). The experiment was run at four different speeds: 28,000, 34,000, 40,000 and 48,000 rpm. Ln(Absorbance) versus radial position at the four different speeds are shown in graphs (a) to (d). The data from this graph was then fitted to a residual plot to obtain a best fit that would describe the solution state of the protein. Based on this fit, a single species with an average molecular weight similar to that of an Ape SSBΔC monomer was determined.
Chapter 4: Archaeoglobus fulgidus SSB (Afu SSB)

The gene sequence for *Archaeoglobus fulgidus* (Afu) SSB was obtained from the NCBI website (Gene ID: AF 0780). Based on this sequence (Figure 4.1), forward and reverse primers were designed for PCR with the forward primer having *nde I* restriction site and the reverse primer having *xho I* restriction site (Figure 4.2 a). The gene was composed of 939 base pairs which coded for 312 amino acids. Protein characteristics are shown in Table 1.3 in Chapter 1, section 1.4.1.

```
1 ATGGACAGAG TTGAGAGTTT AGCCACCAG ATTTTTGAAG ATTACAAGGA TTATGGAGTT
  61 GACAGAAGGG AGATTGTTGA GAAGTCAGA AAACCTTTTA TCGAGTTCAG ATTTCCGGAA
  121 AATGAAGCTG TTGAACCAT CAGAAATTAC ATCATCAGGG AGTACGGAGC ACCTGCAACG
  181 GTAAGGGAGG AGAGGATTAC AAAAAATTGAG GAAATAAAGG AGCGAACCAG ATGGGTACG
  241 GTGAAGGCCCA AGGTATTGCA GCTTTGGGAG AGCAGCAGCC CCTCAATAGC CCAGGTGGGG
  301 TGATTTGGTG ATGAAACCAG CTACATCAGA TTTCTCGTTG GGACGAAAGC CAAGAAGCAG
  361 CAGATTGATG AGGGTAAGAG CTACATATTC AGGAACGTTG TTGAGCCAGA CTACGGGGGA
  421 GTTTTAAGGC TTAACGTAAC GAAATAAGC GAGATAGAGG AGATTGAGGA GGATGTGAAG
  481 GTAAGGCGGC CGGAAAGGCT GAGCGGGGAG GTGGAGTTCG TTGGGGCCTG TGTAGCGATT
  541 CAGCAGAACCA GCAGACTCAT TCAGCGCTGC AGCGTGTGGG AGTGGCAACG AGTGAAAGA
  601 ATGGGGAAAGT GTCCGGAACA CGGAAAACAAG AAGCGAAGG ACGATTGAGG GATTAAGGA
  661 GGTGTGACG AGGGATACAC AACTAGGAG GTTATCATAA ACGAGGGAGG TGTGGAATCT
  721 CTGACCGGAA TTAACCTTGA GAAGGCAAAG AAAATAGCCG AGGAAAATTT GGATAGAGGT
```
Figure 4.1: Gene sequence for *Afu* SSB with 939 bases. Sequence obtained from NCBI web server (Gene ID: AF 0780). For details of the protein refer to Table 1.3, section 1.4.1.

4.1. Cloning

4.1.1. PCR

(a) **Forward Primer:** 5’ CAT ATG GAC AGA GTT GAG AGT TTA GCC AA 3’

(b) **Reverse Primer:** 5’ CTC GAG TTA TAC CTC CTC CAG AAG GGC TTT AA 3’

Figure 4.2: (a) Forward and reverse primers for PCR. Forward primer has *nde I* restriction site and reverse primer has *xho I* restriction site (b) 1% agarose gel stained with SYBER GOLD™ dye, showing *Afu* SSB gene PCR product

Full length *Afu* SSB gene was cloned using the cloning technique described in section 2.1. Figure 4.2 (a) shows the forward and reverse primer sequences for the PCR.
Figure 4.2 (b) shows the picture of *Afu* SSB PCR product run on a 1% agarose gel along with λ-**BstE** base pair ladder.

4.1.2. **Blue-white cloning using pDrive cloning vector**

The PCR product was subsequently gel purified as explained in section 2.1.2. and then inserted into the cloning vector, pDrive as explained in section 2.1.3. Subsequently the vector with the inserted *Afu* SSB gene was transformed into XL10 cloning host. Successful insertion and transformation was indicated by white colonies seen on blue-white screening as explained in chapter 2, section 2.1.5. Three such colonies were selected and a 5.0 mL overnight grow up of the same yielded enough cells to do miniprep (refer section 2.1.6.). The cloning vector with the inserted *Afu* SSB gene isolated from the three samples in this way was then run on a 1% agarose gel to confirm the insertion. As shown in Figure 4.3, white colonies #2 and #3 showed insertion of the gene in to pDrive vector whereas colony #1 did not show presence of any plasmid.

![Agarose gel comparing Afu SSB inserted into pDrive with pDrive only from three different white colonies.](image)

**Figure 4.3**: Agarose gel comparing *Afu* SSB inserted into pDrive with pDrive only from three different white colonies.
4.1.3. Restriction

Once the *Afu* SSB gene in pDrive was isolated and verified, restriction was done to isolate the gene using restriction enzymes, *nde-I* and *xho-I*. At the same time the expression vector, pET21a, was also restricted with the same two restriction enzymes as described in section 2.1.7. A sample of the restriction reaction and controls run on an agarose gel showed successful restriction of the *Afu* SSB gene from pDrive as shown in Figure 4.4. The restricted gene and restricted pET21a were then gel purified as explained in section 2.1.2. The second band in lane #10 is seen in control reaction with *xho-I* since the pDrive vector itself has an *xho-I* restriction site.

Figure 4.4: Agarose gel showing restriction reaction products of pET21a and *Afu* SSB along with control reactions. Restricted gene is seen as the lower band in lane #8.
4.1.4. Ligation and transformation into cloning host

The gel purified restricted *Afu* SSB gene and pET21a were then ligated as explained in section 2.1.8. The ligated gene with pET21a was then transformed into the cloning host XL10 (refer to section 2.1.9). pET21a along with the inserted gene was isolated from the XL10 cells using miniprep and the insertion verified on an agarose gel as shown in Figure 4.5. The first two colonies, colony #1 and #2 did not seem to have the insertion whereas colonies 3, 4 and 5 seem to have the gene inserted into pET21a. For further studies, colony #3 was selected. Colony #4 and #5 were stored at -20 °C as backups.

![Figure 4.5: Agarose gel showing five colonies of *Afu* SSB in pET21a. Colonies 3, 4 and 5 have the inserted gene.](image-url)
4.2. Protein Expression study

4.2.1. Small scale expression study

The \textit{Afu} SSB gene in pET21a was then transformed into expression host BL21(DE3) RIL since the gene has 21 rare codons for arginine and 9 rare codons for isoleucine. A small scale expression study (refer section 2.2.1.) using two colonies showed good expression with the protein being expressed at the calculated molecular weight of 35.6 kDa as shown in Figure 4.6. Expression was done at 37 °C for three hours after induction with IPTG. As seen from Figure 4.6, \textit{Afu} SSB protein seems to have over-expressed as desired.

![Figure 4.6: SDS-PAGE showing over expression of \textit{Afu} SSB in BL21(DE3) RIL colony #1 and colony #2.](image)

4.2.2. Solubility study

The solubility of \textit{Afu} SSB in low salt (0.15 M NH\textsubscript{4}Cl) versus high salt (1.00 M NH\textsubscript{4}Cl) buffer was then examined after heating the lysate to 85 °C for 15 minutes.
Figure 4.7 shows the SDS-PAGE gel of \textit{Afu} SSB solubility study determined using the method described in section 2.2.1. As seen from lanes 3 and 4, soluble protein was obtained in the supernatant in presence of both low and high salt respectively. Although the protein appears quite pure, it may not be. Contaminants could still be present but not visible on the gel because of weak staining. This was evident when comparing Figure 4.7 with Figure 4.9. Low salt lysis approach was subsequently adopted for all further studies.

![Figure 4.7: SDS-PAGE showing solubility of \textit{Afu} SSB in low salt, high salt and bug buster after heating at 85 °C for 15 minutes.](image)

\textbf{4.2.3. Solubility Profile}

As described in section 2.2.5., solubility profile of \textit{Afu} SSB was determined on a sample obtained from a trial purification run. Based on the solubility profile seen in Figure 4.8, sodium chloride, magnesium chloride and Tris-HCl pH 8.0 were selected as the components of buffer ideal for maximum solubility of \textit{Afu} SSB. For all subsequent purification runs this was the buffer of choice.
**Figure 4.8:** Solubility profile of *Afu* SSB indicating sodium chloride, magnesium chloride and TAPS pH 8.5 as the optimum components of a buffer giving the highest solubility of the protein.
4.3. Protein Purification

Large scale expression study was done as described in section 2.2.2 and the lysis done with low salt buffer as described in section 2.2.3. The buffer used was the standard lysis buffer. The first step of purification involved heating the lysate for 10 to 15 minutes at 85 °C and then centrifuging at 18,000 x g for 30 minutes at room temperature to separate the pellet from the supernatant. As seen in Figure 4.9, the supernatant after heating was reasonably pure with most of the endogenous *E. coli* protein precipitated out into the pellet.

![Figure 4.9: SDS-PAGE gel showing *Afu* SSB protein before and after heating the lysate. Soluble protein is seen in the supernatant after heating at 85 °C for 15 minutes (lane #5). Most of the endogenous proteins have precipitated out and is seen in the pellet (lane #4).](image)

The supernatant was separated from the pellet and filtered using a syringe filter to remove any cell debris. The filtrate was diluted with dilution buffer (50 mM Tris-HCl, pH 8.0) to match the conductivity to that of Q-Sepharose cation exchange column loading.
buffer as described in section 2.2.4. Buffer-A (loading buffer) consisted of 50 mM Tris-HCl, pH 8.0, 40 mM sodium chloride and 10 mM magnesium chloride. Buffer-B (elution buffer) consisted of 50 mM Tris-HCl, pH 8.0, 1.0 M sodium chloride and 10 mM magnesium chloride. Figure 4.10 shows the chromatogram for Q-Sepharose purification run for *Afu* SSB and Figure 4.11 shows the SDS-PAGE gel picture of the different fractions from the Q-Sepharose run. As seen from Figure 4.11, *Afu* SSB eluted from the column in all the fractions below the peak. At this stage the protein was partially pure with few contaminants still present. Fractions 9 to 23 from Q-Sepharose were then pooled and dilution buffer added to match the conductivity to that of Buffer-A. POROS-HQ cation exchange column, a high resolution cation exchange column, was then used to further purify the protein as described in section 2.2.4. Chromatogram for this run is shown in Figure 4.12 and SDS-PAGE gel of the fractions under the peak is shown in Figure 4.13. Fractions 9 to 15 were then pooled and concentrated to around 10 to 15 mg/mL as described in section 2.2.6 and then purified by running on Superdex-200 size exclusion column. A single peak was obtained as shown in Figure 4.14. A sample from the fraction on the peak run on SDS-PAGE gel showed the presence of relatively pure protein (Figure 4.15). Fractions under the peak (fractions 23 to 26) were then pooled together and concentrated to around 20 mg/mL and used for further studies.
Figure 4.10: Chromatogram of Q-Sepharose run for *Afu* SSB purification shows a single prominent peak eluting out.

Figure 4.11: SDS – PAGE of samples from Q-Sepharose run for *Afu* SSB purification
Figure 4.12: Chromatogram of POROS-HQ run for Afu SSB purification

Figure 4.13: SDS – PAGE of samples from POROS-HQ run for Afu SSB purification
Figure 4.14: Chromatogram of Superdex-200 run for *Afu* SSB purification shows a single eluted peak.

Figure 4.15: SDS–PAGE sample of pooled and concentrated fractions 23 to 26 from Superdex-200 run.
4.4. Crystal Screening

4.4.1. Study to determine protein concentration for crystal screening

Once pure protein was obtained, a study was done to determine the concentration of protein that could be used for crystal screening. This was done in presence of three different salt concentrations, 100 mM, 150 mM and 200 mM NaCl in addition to 10 mM magnesium chloride and 50 mM Tris-HCl, pH 8.0. The maximum protein concentration in presence of each of these salt concentrations were then determined as shown in Figure 4.16.

![Graph showing the maximum solubility of Afu SSB in presence of three different salt concentrations.](image)

**Figure 4.16:** Graph showing the maximum solubility of *Afu* SSB in presence of three different salt concentrations.

Based on the profile seen in Figure 4.16, crystal screens were set up at 200 mM sodium chloride at 50 mg/mL of *Afu* SSB (approximately half the maximum protein
concentration). Crystal screening was done using all the screens mentioned in section 2.3.1 with and without a 15mer ssDNA oligonucleotide. Room temperature screens were set up using 0.5 µL + 0.5 µL format. No significant crystal hits were obtained. Most of the wells had precipitated protein with a few clear wells. There was no change in observations even after 15 days of setup.

4.5. Analysis of Quaternary State and DNA binding Affinity

Three different studies were performed on Afu SSB namely, Dynamic Light Scattering (DLS), Size Exclusion Chromatography (SEC) and Fluorescence anisotropy.

4.5.1. Dynamic Light Scattering

Dynamic light scattering experiment described in section 2.4.1 was done on Afu SSB at a concentration of 1 mg/mL. Different parameters like temperature, presence of reducing agent and presence of ssDNA were used to study the state of this protein in solution. Results are summarized in Table 4.1. The sample at 25 °C showed an average molecular weight of 60 kDa indicating a possibility of dimer (calculated molecular weight of a monomer is 35.6 kDa). Experiment at 4 °C showed similar molecular weight and so did the experiment in presence of reducing agent. Thus, temperature or presence of reducing agent did not seem to have any effect on dimerization. DLS of Afu SSB with ssDNA at 25 °C showed the possibility of just one monomer binding to the short 15mer ssDNA but at 4 °C, the sample had the tendency to bind a dimer. Figures 4.17 to 4.20
shows DLS plots for few of the runs along with DLS software (DYNAMICS version 7.7.4) based data analysis. Table 4.1 summarizes the different runs of DLS on *Afu SSB*.

**Table 4.1: Summary of *Afu SSB* DLS experiment**

<table>
<thead>
<tr>
<th></th>
<th><em>Afu SSB</em> 25 °C</th>
<th><em>Afu SSB</em> 4 °C</th>
<th><em>Afu SSB + DNA</em> 25 °C</th>
<th><em>Afu SSB + DNA</em> 4 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>Pd %</td>
<td>SOS</td>
<td>MW</td>
<td>Pd %</td>
</tr>
<tr>
<td>58</td>
<td>6.6</td>
<td>6.6</td>
<td>56</td>
<td>13.2</td>
</tr>
<tr>
<td>54</td>
<td>18.5</td>
<td>3.8</td>
<td>46</td>
<td>31.0</td>
</tr>
<tr>
<td>50</td>
<td>25.0</td>
<td>2.2</td>
<td>58</td>
<td>13.7</td>
</tr>
<tr>
<td>63</td>
<td>11.3</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>18.8</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>17.0</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>18.4</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>12.8</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>TCEP</strong></td>
<td></td>
<td></td>
<td><strong>TCEP</strong></td>
</tr>
<tr>
<td>59</td>
<td>16.5</td>
<td>1.0</td>
<td>57</td>
<td>19.9</td>
</tr>
<tr>
<td>67</td>
<td>12.7</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>DTT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>16.6</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.17: *Af* SSB DLS experiment at 25 °C (25 mM Tris pH 8.0, 10 mM MgCl$_2$, 200 mM NaCl).
Figure 4.18: *Afu* SSB DLS experiment at 4 °C (25 mM Tris pH 8.0, 10 mM MgCl₂, 200 mM NaCl).
Figure 4.19: *Afu* SSB DLS experiment at 25 °C in presence of 15mer ssDNA.
Figure 4.20: *Afu* SSB DLS experiment at 4 °C in presence of 15mer ssDNA.
4.5.2. Fluorescence Anisotropy

Fluorescence anisotropy experiment was done as described in section 2.4.3 using a 15mer oligonucleotide. Figure 4.21 shows the graph of Anisotropy versus Afu SSB concentration at three different 15mer ssDNA concentrations namely, 0.10 µM, 0.35 µM and 0.80 µM in presence of 250 mM sodium chloride. Based on the graph in Figure 4.21, a graph of Afu SSB concentration versus ssDNA concentration was plotted as shown in Figure 4.22. Graph in Figure 4.23 shows the dependence of anisotropic signal on the binding density of Afu SSB on ssDNA. Graph in Figure 4.24 fitted to Hill equation gives a calculated Kd of ~ 195 ± 13nM.

**Figure 4.21**: Graph of anisotropy versus Afu SSB concentration at three different 15mer ssDNA concentrations in presence of 250 mM sodium chloride.
**Figure 4.22:** Graph of *Afu* SSB concentration versus ssDNA concentration plotted as described in section 2.4.3.

**Figure 4.23:** Graph of normalized anisotropic signal versus $\Sigma v_i$ showing the dependence of anisotropic signal on the binding density.
Figure 4.24: Final graph of binding density versus free protein concentration fitted to Hill equation gives a 
$K_d$ of $195 \pm 13$ nM.
Chapter 5: Discussion

Although a lot of information is available about single-stranded DNA binding protein and its interaction with ssDNA, a comprehensive picture as to the role of different domains of the protein as well as the steps involved in the binding process are still up for speculation. As previously shown, SSB’s including RPA from eukaryotes have a major role in several important processes taking place in the cell. The SSB’s not only have an important role in protecting the single-stranded DNA, but also play a role in recruiting and interacting with several different types of proteins that are involved in processing the ssDNA. Several studies have shown elevated levels of SSBs in different types of cancers including breast cancer and colon cancer although the specific roles, if any, are still not known.\textsuperscript{50} It has also been shown that RPA mediated interaction of the tumor suppressor protein, p53, with the WRN (recQ) helicase leads to disruption of programmed cell death leading to a condition know as Werner’s syndrome.

SSB belongs to the family of OB fold proteins. In fact it has proven to be a model for these types of proteins since SSB’s from all the three branches of life – Bacteria, Archaea and Eukaryotes have shown the presence of the characteristic OB folds indicating that this type of fold is a good example of divergent evolution, accommodating a wide range of sequences and ligand binding functions. This also shows that this type of fold predates the divergence of the three kingdoms. OB folds, as previously mentioned in
the introduction, have been shown to interact with oligosaccharides and oligonucleotides. These folds were originally named and described by A.G. Murzin in the early 90’s and his original paper on OB folds was dedicated to Oleg Borisovich Ptitsyn, one of the founding fathers of protein folding whose initials coincidently were O.B.\textsuperscript{70}

In the current study SSB’s from two different archaeal organisms were chosen, one from \textit{Aeropyrum pernix} and the other from \textit{Archaeoglobus fulgidus}. The two organisms represent examples from the two subkingdoms of the archaeal kingdom with the former being an example of the crenarchaeal subkingdom and the latter being an example of the euryarchaeal subkingdom.

\textbf{Figure 5.1}: OB fold domain of \textit{E. coli} Aspartyl tRNA synthetase with characteristic OB fold domain architecture (1-106 amino acids, PDB: 1C0A). β-sheets are shown in yellow and α-helices in red.

A characteristic OB fold (Figure 5.1) is a five stranded β barrel with a common ligand binding-interface centered on β strands 2 and 3. As seen in Figure 5.1, this interface is supplemented by loops between β1 and β2 (referred to as L_{12}), β3 and α (L_{3a}),
α and β4 (Lα4), and β4 and β5 (L45). The different loops define a binding cleft that runs across the surface of the OB fold perpendicular to the axis of the β barrel. Structural studies have shown that the loops presented by the β sheets provide a recognition surface for ssDNA, allowing binding through aromatic stacking, hydrogen bonding, hydrophobic packing and polar interactions.

The *Ape* SSBΔC structure solved in this study also shows several of the characteristic features exhibited in a typical OB fold domain architecture, providing another strong example of a protein of this family possessing similar structural and functional architecture.71

![Figure 5.2: Ape SSBΔC monomer (chain A) structure exhibiting typical OB fold architecture. β-sheets are shown in yellow and α-helix in red.](image_url)

The structure of the *Ape* SSBΔC, as shown in Figure 5.2, has all the five β sheets forming the typical β-barrel with the two major loops, L12 and L45, clearly seen flanking the binding cleft. ssDNA can be visualized to pass through this binding cleft as seen in
The structure of Ape SSBΔC is very similar to that of its homolog, Sso SSB (RMSD of 1.03 Å), the structure of which has been solved by Kerr and co-workers. This shows that the members of the crenarchaeal family may have more or less the same domain and functional architecture.

Though the Ape SSBΔC and the human RPA70 (hsRPA70) do not share any significant sequence homology, they share a high degree of structural homology. This was seen when the Ape SSBΔC structure was superposed on to the structure of the hsRPA70 with and without DNA. Superposition was achieved using the SSM superposition function available in the visualization and model building software, COOT. As seen from Figure 5.4 and Figure 5.5, two monomeric structures of Ape SSBΔC superpose well with the two hsRPA70 domains (RMSD of ~2.0 Å).
Figure 5.4: Superposition of two *Ape* SS\text{BAC} monomers (shown in red) on to the structure of *hs*RPA70 DNA binding domains without DNA (shown in blue). Superposition was done using the SSM function in COOT and then displayed using Pymol. RMSD between the two structures is ~ 2.0 Å.

Figure 5.5: Superposition of two *Ape* SS\text{BAC} monomers (shown in red) on to the structure of *hs*RPA70 DNA binding domains with a ssDNA (shown in yellow). Superposition was done using the SSM function in COOT and then displayed using Pymol. RMSD between the two structures is ~ 1.9 Å.
The C-terminal region of the Ape SSBΔC is a highly flexible tail (due to the presence of a number of glycine residues) which prevents Ape SSBΔC from being concentrated and crystallized. From this study it has been shown that the protein when concentrated beyond 4 to 5 mg/mL tends to precipitate out as filaments which has also been shown by Kerr and coworkers while working with the full length Sso SSB.\textsuperscript{53} This tendency of forming long filaments by protein-protein interaction is a characteristic of this protein which enables it to coat the ssDNA effectively. Efforts to crystallize the full length Ape SSB at low concentration were futile since there was not enough protein to form crystals. The truncated protein which gave good crystal hits behaved much better with solubility increasing by 10 to 15 fold. Several crystal forms were obtained but the only one which gave a good single crystal diffraction data was that from a low pH form, which was stable even after one and half years. Because of the high salt content in the mother liquor, sodium malonate based cryo freezing was done which gave a very good freeze with no ice rings seen in the diffraction images. This freezing technique was based on earlier studies by McPherson and Petsko.\textsuperscript{72, 73}

Preliminary attempts to crystallize the protein with a 9mer ssDNA did not yield any crystals. Most of the wells were clear even after three to four months of set-up. Soaking of Ape SSBΔC crystals in mother liquor with 9mer ssDNA was not attempted and can be tried out in future to see if ssDNA can be incorporated that way without cracking the crystals.

Experiments done in this study showed that the truncated protein is possibly a monomer at low protein concentration which is consistent with the other SSB’s of the crenarchaeal family. Two experiments which show this are the SEC and the Analytical
Ultracentrifugation. DLS results were ambiguous as far as the state of the protein is concerned. At higher protein concentration, the protein has a tendency to form dimers which is an interesting observation. This dimerization is probably not seen in biologically relevant concentrations and therefore is coincidental although there is no conclusive evidence for this. Fluorescence anisotropy experiments have clearly shown that the dissociation constant ($K_d$) is in the nanomolar range indicating a relatively strong binding affinity. Results also indicate that the binding affinity is lowered with increasing salt concentration. The length of the ssDNA also affects the binding affinity with longer oligonucleotides invoking greater binding affinity which is also consistent with studies on other SSB’s. This shows that as more and more SSB’s bind to a segment of ssDNA, their binding affinity increases. The full length protein shows a slightly higher binding affinity as compared to the truncated protein when a 9mer ssDNA is used. This indicates that the C-terminal region possibly plays a role in stabilizing the binding with ssDNA. When a longer strand of ssDNA is used (15 mer), the binding affinity for truncated protein is more than the full length indicating that the C-terminal region could possibly be interacting with the adjacent SSB molecules. Another evidence for this is the fact that the full length protein forms filaments upon concentration even if ssDNA is not present, indicating protein-protein self association.

The main purpose of studying this protein is to understand the interactions that takes place when SSB’s bind to a flap DNA. Our lab is interested in understanding the events taking place during the flap DNA processing. As a first step, individual proteins are being crystallized in presence and absence of flap DNA. The structure of the apo FEN1, one of the endonucleases involved in cleaving the flap has already been solved by
a lab member (Steve Tomanicek). Eventually, the goal is to obtain structural details of the process by crystallizing the ternary complex of all the proteins involved in the process in the presence of a flap DNA.

Attempts to crystallize the protein from the euryarchaeal organism, *A. fulgidus*, failed. Although this protein did not have any solubility issues like the full length *Ape* SSB, it still did not crystallize. A crystal structure of this protein would have been interesting since no crystal structure of a homolog of the *Afu* SSB has been solved so far.
PART 2: Studies of DNA Helicase
Chapter 6: Introduction

Helicases are proteins which act like DNA zipper and unwind the dsDNA in preparation for replication. The mechanism by which these molecular motors move along DNA has been studied extensively for the past several years. Recent studies on helicases from different organisms have improved our understanding of the mechanism by which these proteins perform their action. The primary function for DNA helicases is unwinding of double-stranded DNA to form single-stranded DNA intermediates during DNA replication, repair and recombination using ATP as the energy source.

6.1. Classification of Helicases

Helicases are members of the AAA\(^+\) (ATPases Associated with various cellular Activities) proteins and can be classified into six super-families.\(^{75, 76}\) Members of these super-families have characteristic “signature” motifs which have been well defined and characterized with some of them specific for each superfamily or subfamily therein.\(^{77}\) Almost all of these enzymes have conserved residues in the core domain involved in the hydrolysis and binding of NTP (Nucleotide Tri-Phosphate) equivalent to Walker A and B motifs of many ATPases and an “arginine finger” that plays a crucial role in energy coupling.\(^{78, 79}\)
Figure 6.1: Inchworm model for translocation of SF1 enzymes shown with the example of PcrA helicase. A, B, C and D are the pockets on the enzyme involved in binding of the nucleotide bases 1 to 6 of the bound ssDNA. The critical amino acids involved in the mechanism are indicated. Figure adapted from reference #80.
6.1.1. SuperFamily 1 (SF1)

Structurally one of the best classified superfamily, SF1 helicases include two subfamilies – SF1A helicases and SF2B helicases based on the directional polarity of these enzymes. SF1A helicases translocate with 3’ to 5’ polarity whereas SF1B enzymes translocate with 5’ to 3’ polarity. Examples of the SF1A subfamily include Rep and UvrD helicases found in gram negative bacteria and PcrA helicase found in gram positive bacteria. Crystal structures of all the three have been solved and show structural features which helps in deducing the mechanism of their activity similar to the inchworm model shown in Figure 6.1.80-83

The enzymes of SF1 family have been shown to be monomers under a variety of crystallization conditions and liganded states. This along with several other biochemical assays have led to the conclusion that SF1 at the DNA fork is most likely a monomer.80, 84 There is also evidence to show that these monomers can cooperate to enhance activity.85

6.1.2. SuperFamily 2 (SF2)

One of the largest superfamilies, SF2 helicases includes NS3 helicase (Non Structural protein 3) from the Hepatitis C virus. NS3 helicase requires a 3’ single-stranded DNA overhang for unwinding. Although the functional unit of the helicase is seen to be a NS3 dimer, single molecule studies have shown that a monomer is also capable of unwinding nucleic acid.86, 87 Crystal structure of NS3 has been solved with and without a single-stranded DNA giving more insight into the mechanism of these enzymes.88 Based on the comparison between the two structures, a mechanism of action was postulated which is schematically shown in Figure 6.2.
Figure 6.2: Proposed mechanism for the unwinding of dsDNA of NS3 helicase from Hepatitis C virus by Kim and co-workers. The enzyme has three domains shown in blue, orange and green. Binding of ATP leads to movement of domain 1 and 3 in the 3' to 5' direction leading to the closure of domain 1 with 2. Release of ADP and Pi leads to the movement of domain 2 in the 3' to 5' direction. The net result of this movement is the unwinding of the dsDNA.
6.1.3. SuperFamily 3 (SF3)

The SF3 helicases are known to form hexamers or double hexamers and have a 3’ to 5’ translocation directionality. Found mainly in viruses, helicases of this superfamily have been shown to have four conserved motifs - A, B, B’ and C. The A and the B motifs resemble the typical Walker A and B motifs while motif C is specific to SF3 helicases. This class of helicases also shows the presence of “arginine finger” just after motif C. The SF3 helicases are known to be involved in origin recognition and unwinding.

The X-ray crystallographic structure of the Bovine Papilloma Virus E1 helicase solved in the laboratory of Joshua-Tor has given an insight and possible mechanism of action of SF3 class of helicases.\(^{89}\) Clearly seen from the structure (Figure 6.3) is an oligomerization domain that is responsible for the formation of the hexameric structure that encircles the ssDNA. Following the oligomerization domain is the AAA\(^+\) domain that contains the highly conserved nucleotide-binding motifs. Also seen in the structure are the six ADP molecules bound at the interface of the inter AAA\(^+\) domains. Based on the structure, a coordinated escort mechanism was proposed (Figure 6.3) wherein six consecutive nucleotides bind to DNA binding loop present in the six AAA\(^+\) domains of the hexamer. The six subunits represent six consecutive steps of a staircase. During one cycle, each subunit hydrolyzes one ATP, releases one ADP and translocates one nucleotide through the interior channel. The subunit at the bottom of the staircase releases the nucleotide, binds to a new ATP molecule and goes back to the top of the staircase to pick up the next available nucleotide (Figure 6.4). In this fashion the helicase translocates on the ssDNA unraveling the dsDNA.\(^{89, 90}\)
Figure 6.3: Bovine papilloma virus E1 (BPV E1) helicase X-ray crystallographic structure (PDB: 2GXA), shows the hexamer with a central ssDNA and six ADP molecules bound at the individual domain interfaces. Each individual subunit has an oligomerization domain and a DNA binding domain as shown in the figure. (a) View of the helicase looking down into the hexameric channel and (b) side view.
Figure 6.4: Coordinated escort mechanism for E1 hexameric helicase. The DNA binding hairpin at the bottom of the stair case releases its associated ssDNA phosphate to conclude its journey through the hexameric channel. This subunit then binds to a new ATP molecule and goes to the top of the staircase to pick up the next available ssDNA phosphate initiating its escorted journey through the channel. Figure adapted from reference # 90.
6.1.4. SuperFamilies SF4, SF5 and SF6

Helicases of this family like those of SF3 are functional hexamers and have structures that look very similar to *E1* helicase shown in Figure 6.3. Found mainly in bacteria and bacteriophages, the helicases of this family generally associate with a primase and act as the main replicative helicases. The SF4 typically exhibit five different motifs - H1, H1a, H2, H3 and H4. While H1 and H2 are similar to the Walker A and B motifs, the other motifs do not have any homologs in any other helicase families. Examples of this type of helicase family include the *E. coli* dnaB and the gene 4 protein (gp4) of the bacteriophage T7.

Rho helicase (SF5) is very similar to the SF4 helicases and has been classified into a separate superfamily because unlike the other helicases, it is responsible for the termination of transcription in bacteria. Rho binds to a specific sequence rich in cytosine, known as the rho utilization site, on the nascent RNA and then separates the DNA/RNA hybrid and eventually disengaging the polymerase. Crystal structure of the full length Rho has been solved by Berger and co-workers.91

The mini chromosome maintenance (MCM) protein complex, considered as the main eukaryotic replicative helicase belongs to the SF6 superfamily. Currently no crystal structure of the helicase domain of the MCM protein has been reported although structural information from archaeal homologues is available.
6.2. Disorders associated with Helicases

A number of human diseases have been linked to defects, including mutations, to proteins with helicase activity. Disease states such as Xeroderma Pigmentosum (XP), Cockayne syndrome, Bloom’s syndrome and Werner’s syndrome have been linked to defects in specific genes coding for DNA helicases. Xeroderma Pigmentosum is a genetic disorder resulting from a defect in Nucleotide Excision Repair (NER) of ultraviolet damaged DNA. Patients are extremely sensitive to sunlight and are highly susceptible to skin cancer. The XPD protein, mutation in which causes this disorder, has been shown to possess ATPase and helicase activity. Another well known disorder is the Bloom’s syndrome associated with a defective Bloom’s protein which is a DNA helicase. Patients suffering with Bloom’s syndrome suffer from short stature and are highly predisposed to cancer in addition to other symptoms. Another disorder associated with a defective helicase activity is the Werner syndrome characterized by premature aging in addition to other symptoms like cancer, heart disease, diabetes and cataract. Helicases are also critical for several viral genome replications and hence are the target for ongoing anti-viral therapy.

The critical role of helicases in human diseases, coupled with the importance of these proteins in the basic aspects of nucleic acid metabolism, makes this class of enzymes an attractive target for study.
6.3. Helicases in this study

The main replicative helicase, dnaB, from three different species, *Vibrio cholerae (Vc)*, *Yersinia pestis (Yp)* and *Helicobacter pylori (Hp)* were included as part of this study.

6.3.1. *Vibrio cholerae (Vc) dnaB*

*Vibrio cholerae* is a gram-negative bacterium which is responsible for causing the potentially deadly disease cholera in humans. The bacterium is rod shaped with a single flagellum. The gene for dnaB was cloned from the genomic DNA using PCR.

Sequence alignment (using the BLAST2sequence program, ExPASY proteomics web server) of *Vibrio cholerae* dnaB with *E. coli* dnaB showed a sequence similarity of 89% with a sequence identity of 79%. Using the Blastp program on the NCBI web server, a search was done to find homologues of *Vc* dnaB in the Protein Data Bank (PDB). A positive match with *Thermus aquaticus (Taq)* dnaB was determined with a sequence similarity of 66% and a sequence identity of 47%. This information indicated that the structure of *Vc* dnaB could be solved by molecular replacement using the phase information of *Taq* dnaB structure.

(a)

1 MADPRTDNRRKIPDAQVDAIKVPPHSLEAEQSVIGGLLLDNERWDTVSEHVTQDFYSR
61 PHRLIFDGVKSILEAAGKPLDLITLSEYLEQREQLEDVGGFAYLADLAKNTPSAANINAYA
121 EIVAERALVRNLIGVANEIADAGYDPQGRNAEDLLLDAESKVFAIAEARTSENEGPKNVD
181 SILERTLERIELLYKTPQDGVTVNVTGDNLKNTAGLQGSDLIIVAARPSMGKTTFAMN
241 LCENAAMEQDKPVLIFSLEMPPAEQIMMRMLASLSRVDQTKIRTGDQLDDEDWARISSTMGI
301 LMEKKNYIIDDSGLTPTEVRSPARRIAREHGGGLSLIMVLYQLMRVPALTDRNRTLEIAE
361 IRSILKALAKELNVPSVALSQRNLSERLEQRADKRPVNSDLRESGISEQADLIMFYIRDEV
421 YHPSPLKGTAEIIIGQRRNPGISVRLTFQGHYSRFNDYAGPAPFDDE

Figure 6.5: (a) Amino acid sequence of Vc dnaB obtained from the NCBI website (b) Sequence alignment of Vc dnaB against Taq dnaB using BLAST2sequences program available on the NCBI website.

Table 1.1 shows some of the characteristics of the Vc dnaB gene and protein determined using ProtParam program available on the ExPASy proteomics server. The full length protein with 468 amino acids has an acidic pI.
Table 6.1: Protein characteristics of the full length \( Vc \) dnaB obtained using ProtParam program (ExPASy Proteomics server)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene length</td>
<td>1407 bp</td>
</tr>
<tr>
<td>Amino acids</td>
<td>468</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>52.5 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>4.91</td>
</tr>
</tbody>
</table>

### 6.3.2. *Yersinia pestis* (Yp) dnaB

*Yersinia pestis* is a gram negative bacterium responsible for bubonic plague. *Yersinia pestis* is usually transmitted to humans by the bite of an infected rat flea. Complete genome sequence is available for two of the subspecies of *Yp*, strain KIM and strain CO92. The gene for dnaB was cloned from the CO92 strain using PCR.

Sequence alignment (using the BLAST2sequence program, ExPASy proteomics web server) of *Yp* dnaB with *E. coli* dnaB showed a sequence similarity of 92% with a sequence identity of 84%. Using the Blastp program on the NCBI web server, a search was done to find homologues of *Yp* dnaB in the Protein Data Bank (PDB). A match with *Thermus aquaticus* (*Taq*) dnaB was determined with a sequence similarity of 66% and sequence identity of 47%. This was the same as in the case of alignment of *Vc* dnaB with *Taq* dnaB. A sequence alignment between *Vc* dnaB and *Yp* dnaB was also done which showed a sequence similarity of 89% and sequence identity of 80%.
(a)

MAAKPNTKMEPRDRQMEGLKLPHPHLSDEAQSVLGLMLDNERWDNVSEVASKDFFSR
PHRIFTEMQRLLENSKPIDLITLSELQKGDLSVGGFAYLAESEKNTSAAANIGAYA
DIVRERAVVREMISVANEIADAGYDPQRSSEDDLLDAESKVFQIAERSASDEGPKSD
RILEDTVARIEQLYQRPHDGVTGVSTDLDDKCTAGLQKSDLIIVAARPSMGKTFMNN
LCENAAAMMQDKPVFLSLEMPQDIQMRMLASLSHVDQTRITGQLDDEWARISSTMGI
LMEEKRNYIDDSSQILTPETEVRSAARRIFREHGGGLSIMIDYLQLMRVPLSDNRTLIEAE
ISRSKLALKELQSVLPALCVLSQNLSEQRADKRPRVNSDLRESGSIEQDADLIMFYRDEV
YHENSDEKGIQAIIGLQQRNPISVRLKFNQWSRFNDYAAGPQYDDE

(b)

Yp dna 18 MEGLKLPPHSLEAQSVLGLMLDNERWDNVSEVASKDFFSR-FSRRPHRIFTEMQRLLENS 76
MEG +PPHSLEAQSVL +LD++ D V + S + F++ HR+I+ MQ L
Taq dnaB 1 MEGPITPPHSLEAQSVLGLGLILDSDVNDVELEGFLPSEAFYEAHRKIAAMQALRSQG 59
Yp dna 77 KPIDILITLSELQKGDLSVGGFAYLAESEKNTSAAANIGAYADIVRERAVVREMISVA 136
+P+DL+TSE L ++G L+ VGG AYL +LS+ TF+AA YA IV E+ +R +I A
Taq dnaB 60 RFVDLVLITSELFQLEEEVGTTALQLQLEATPTAYAEHYARIVAEMKRLRLITEAA 119
Yp dna 137 NEIADAGYDPQRSSEDDLLDAESKVFQIAERSASDEGPKSDVRILEDTVARIEQLYOR 196
E Y G S +++LD A K + ++A ++ + P + ++ +T IE L+Q
Taq dnaB 120 GEAMRLAYEAEAGSLDEILDTAGKtiesTKTDEARP--MRELHETFHEIEALFQN 176
Yp dna 197 PHDGVTGVSTGTDLDDKCTAGLQKSDLIIVAARPSMGKTFMNNLCENAAAMMQDKPVFLIF 256
+ V GV TGF +LD+ L L I+AARP+MGKFTFA + +NAA+ + V I+
Taq dnaB 177 KGE-VAGVRTRGKELDQLIGTLFGSLIIAIRPAMGKTAFALTIAQNAALKEGTVVG1Y 235
Yp dna 257 SLEMPGQIMMRAMLASLSHVDQTRITGQLDDEWARISSTMGIIMEKRNMYIDDSSGLT 316
SLEMP Q+ +RM+S + +D R+R GQL D D++R+ LE +YIDD+ LT
Taq dnaB 236 SLEMPAALQLTLMRCMSEARIDMNVRVLQQLTDRDSRFLVDSRSLSEA-PYYD+DTDPLT 294
Yp dna 317 PTEVRSRARRIFREHGGGLSIMIDYLQMRV---SLSDNRTLIEAEISRSKLALKEQ 373
EVR+RAR+ ++ + LI+IDYLQM P +NR EIA IRS LKALA+EL
Taq dnaB 295 LMEVRARALLVSQNQ-VGLILLIDYLQLMSGFGSSEQRGQIEAISGLKALREL 353
Yp dna 374 VFPLSQRNLQRSLEQRADKRPRVNSDLRESGSIEQDADLIMFYRDEVYHENSDEKGIQAQI 433
+F++ALSQL+R+E R +KRP+ SDLRESGSIEQDADL+MIYRDE Y+ +S++ GIA+I
Taq dnaB 354 IFIALLQLSRARVFKNRPMLSDLRESGSIEQDADLVMIYRDEYNNPHSEKAGAEI 413
Yp dna 434 ILGKQRNGPIGVSRLKFNQWSRFNDYA 461
I+GKQRNGP G+V L+F+ RF++ A
Taq dnaB 414 IVGKQRNGPTGTVELQFHASVRFNDLA 441

Figure 6.6: (a) Amino acid sequence of Yp dnaB obtained from the NCBI website (b) Sequence alignment of Yp dnaB against Taq dnaB using BLAST2sequences program available on the NCBI website shows a sequence similarity of 66% and an identity of 47%.
Table 1.1 gives an overview of the \( Yp \) dnaB gene and protein characteristics. This information was obtained using the ProtParam program available on the ExPASy proteomics server.

Table 6.2: Protein characteristics of the full length \( Yp \) dnaB obtained using ProtParam program (ExPASy proteomics server)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene length</td>
<td>1407 bp</td>
</tr>
<tr>
<td>Amino acids</td>
<td>468</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>52.5 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>5.22</td>
</tr>
</tbody>
</table>

6.3.3. *Helicobacter pylori* (Hp) dnaB

*Helicobacter pylori* is a gram negative bacterial pathogen responsible for chronic gastritis and gastrodeuodenal ulcer. \( Hp \) dnaB protein has been characterized by Soni and co-workers and has been shown to have helicase activity with 5’ to 3’ polarity. Size exclusion studies have shown the protein to be a hexamer in solution.\(^98\)

Sequence alignment (using the BLAST2sequence program, ExPASY proteomics web server) of \( Hp \) dnaB with \( E. coli \) dnaB showed a sequence similarity of only 54% with a sequence identity of 32%. Using the Blastp program on the NCBI web server, a search was done to find homologues of \( Hp \) dnaB in the Protein Data Bank (PDB). But unlike \( Vc \) dnaB and \( Yp \) dnaB, \( Hp \) dnaB has a sequence similarity of only 50% and identity of 33% with \( Taq \) dnaB.
Figure 6.7: (a) Amino acid sequence of *Hp* dnaB obtained from the NCBI website (b) Sequence alignment of *Hp* dnaB against *Taq* dnaB using BLAST2sequences program available on the NCBI website shows a sequence similarity of 50% and an identity of 33%.
Table 6.3 gives a summary of *Hp* dnaB gene and protein information including molecular weight and calculated pI obtained from the ProtParam program available on the ExPASy proteomics server.

**Table 6.3: Protein characteristics of the full length *Hp* dnaB obtained using ProtParam program (ExPASy proteomics server)**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene length</td>
<td>1461 bp</td>
</tr>
<tr>
<td>Amino acids</td>
<td>486</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>55.4 kDa</td>
</tr>
<tr>
<td>Calculated pI</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Chapter 7: Materials and Methods

Most of the materials and methods used in this project are more or less the same as described in Chapter 2. The major difference is in molecular biology steps where Gateway® technology was used in cloning the proteins instead of traditional cloning using pDrive vector (except for $Vc$ dnaB). This and other differences are discussed in this chapter.

7.1. Molecular Biology

7.1.1. Polymerase Chain Reaction (PCR)

The steps involved in PCR were similar to that discussed in chapter 2, section 2.1.1. The annealing temperature and the extension time varied depending on the length and base pair composition of the gene to be synthesized. The forward and reverse primers for the PCR were ordered from IDT Inc. The nucleotide sequences for the primers were based on the gene sequence and the forward primer included an additional “CACC” sequence required for Gateway® cloning. In some cases the forward primer also had the TEV protease cleavage site just after the “CACC” sequence.

The PCR products were checked on a 1% agarose gel and appropriate bands gel purified as described in section 2.1.2.
7.1.2. Insertion into cloning vector using Gateway® Entry vector

The purified PCR product was inserted into Gateway® entry vector, pENTR-D, using a topoisomerase based TOPO cloning reaction. Components of the reaction, as summarized in Table 7.1, were taken in a 0.5 µL thin walled PCR tube.

Table 7.1: Components of a TOPO cloning reaction

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purified PCR product</td>
<td>0.5 to 4 µL</td>
</tr>
<tr>
<td>Salt solution</td>
<td>1 µL</td>
</tr>
<tr>
<td>Sterile water</td>
<td>Add to final volume of 5 µL</td>
</tr>
<tr>
<td>TOPO vector</td>
<td>1 µL</td>
</tr>
<tr>
<td>Final volume</td>
<td>6 µL</td>
</tr>
</tbody>
</table>

Reaction components were mixed gently by pipetting and incubated at room temperature (22 °C) for 5 to 30 minutes depending on the size of the insertion. Bigger insertions required longer incubation time. The eppendorf tube was placed on ice for five minutes and the plasmid transformed into One Shot® TOP10 competent E. coli cells. For this, 2 µL of the TOPO cloning reaction product was added into a vial of One Shot® TOP10 competent E. coli cells without shaking and then placed on ice for 30 minutes. Cells were then heat shocked by placing the vial in a water bath at 42 °C for 30 seconds. The tube was then immediately placed on ice for two minutes. 250 µL of room temperature S.O.C media was added and the tube placed in a New Brunswick Scientific Innova 4000 Incubator Shaker at 200 rpm at 37 °C for an hour. 50 to 200 µL of the cells
were plated on a pre-warmed agar plate containing kanamycin. The plate was then incubated overnight in an incubator at 37 °C.

Of the several colonies seen after overnight incubation, three to five colonies were picked and inoculated into separate 50 mL Fisher tubes containing 5 mL sterile LB with kanamycin. Tubes were placed in a shaker at 200 rpm for 8 to 10 hours at 37 °C. The tubes with the culture were centrifuged using a Beckman Coulter™ TJ-25 centrifuge at 5000 x g for 10 minutes to pellet the cells. Supernatant was discarded and the cells lysed and plasmid extracted using QIAprep® spin Miniprep kit from Qiagen Sciences as described in section 2.1.6.

After all the samples were minipreped, they were run on a 1% agarose gel along with supercoiled DNA ladder from Promega (2-10 kb) as a plasmid size marker. Depending on which sample showed the correct insertion of the gene of interest, the glycerol stock for that sample was preserved and the remaining ones discarded.

7.1.3. Switching from Entry to Destination vector

Next step in Gateway® cloning was to switch the gene of interest from entry vector into destination vector. The destination vector is the expression vector which was used for expression of the gene of interest. Destination vectors used for expression studies included pDEST-17, pDEST-45 and pKM596-H. All these vectors inserted a His-tag in front of the protein of interest in addition to other features. For example, pKM596-H introduces a His-tag followed by Maltose Binding Protein (MBP) in front of the protein of interest. The vector pDEST-17 was provided by Dr. Ron Viola, University
of Toledo and pDEST-45 and pKM596-H were provided by Dr. Artem Evdokimov of P&G.

For switching from entry vector to the destination vector, a transposition reaction was carried out using Gateway® clonase reaction. Reaction components (as shown in Table 7.2) were taken in a 0.5 mL PCR tube.

**Table 7.2: Gateway® LR Clonase reaction components**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry clone (50 – 150 ng)</td>
<td>1 – 7 µL</td>
</tr>
<tr>
<td>Destination vector (150 ng / µL)</td>
<td>1 µL</td>
</tr>
<tr>
<td>TE buffer, pH 8.0</td>
<td>Final volume to 8 µL</td>
</tr>
</tbody>
</table>

The LR clonase, stored at -20 °C, was thawed on ice for two minutes and vortexed briefly before taking 2 µL and adding to the reaction tube. After mixing, the tube was incubated at 25 °C in the PCR machine for an hour. To stop the reaction, 1 µL of Proteinase K solution was added and the reaction incubated at 37 °C for 10 minutes. Reaction tube was then placed on ice for two minutes and the plasmid transformed into DH5α cloning host. Plasmids were then isolated from the cloning host as described in section 2.1.9 and then transformed into the desired expression host, competent *E. coli* cells like BL21 (DE3), using the method described in section 2.1.10.
Figure 7.1: Flow chart showing the different steps involved in Gateway® cloning. The forward primer for PCR step has a CACC sequence at the 5’ end which gives the PCR product the directionality required during insertion into entry clone. Destination vectors used included pDEST-17, pDEST-45 (incorporates His-Tag at the N-terminus of protein of interest) and pKM596-H (incorporates His-tag and Maltose Binding Protein (MBP) at the N-terminus of the protein of interest).
7.2. Protein Expression, Purification and Solubility

7.2.1. Protein Expression, Solubility and Cell Lysis

Protein expression, solubility and cell lysis methodology were the same as described in section 2.2.1, 2.2.2 and 2.2.3. The only difference was in the case of proteins expressed with Maltose Binding Protein (MBP). To avoid / minimize expression of endogenous E. coli proteins that could potentially compete with MBP for binding on an amylose affinity column, 2 g of glucose per liter of LB was added.

7.2.2. Solubility Profile (modified protocol)

Solubility screen described in section 2.2.5 was modified so as to keep the protein soluble and in solution for most part of the test. An aliquot of the purified protein containing approximately 5 mg protein was dialyzed against a minimum concentration of buffer and salt at 4 °C using dialysis cassette. After an overnight dialysis, the sample with soluble protein was removed from the dialysis cassette and used for solubility study. Remaining steps are the same as described in section 2.2.5.

7.2.3. Protein Purification (modified protocol)

Protein purification methodology was more or less similar to that described in section 2.2.4. The major difference was elimination of the heating step since none of the proteins in this project were from thermophiles. In fact, care was taken to keep the lysate as close to 4 °C as possible to avoid denaturation of the protein of interest. After cell
lysis, the lysate was centrifuged at 15,000 x g for 20 minutes at 4 °C in an ultra centrifuge. Supernatant was separated from the pellet and stored at 4 °C until the next step.

For *Vc* dnaB, purification columns used and the methodology were similar to that described in section 2.2.4. The major difference was in the case of *Yp* dnaB-MBP which had a Maltose Binding Protein (MBP) tag in front of the protein and so helped in its purification. An amylose column which could bind to MBP was used for a one-step purification of the *Yp* dnaB-MBP protein. Supernatant after lysis was filtered using a syringe filter as described in section 2.2.4 and the filtrate was loaded on to the amylose column at a slow rate. Column was rinsed with two to three column volumes of column loading buffer and the protein eluted with an elution buffer containing 10 mM maltose in addition to the other components of the column loading buffer. Elution of protein of interest was achieved using a step gradient program on the Bio-RAD HPLC system. Fractions containing the peaks were analyzed for the presence of protein of interest using SDS-PAGE gel. Fractions containing the protein were then pooled and concentrated as described in section 2.2.6.

In the case of *Hp* dnaB-MBP, Cobalt Talon column was used for purification since the protein had a his-tag in front of the protein.
Chapter 8: *Vibrio cholerae* (Vc) dnaB

8.1. Molecular Biology

*Vc* dnaB gene (Gene ID: VC 0395) was cloned and inserted into pDrive cloning vector by Dr. Roger Moore, University of Toledo using blue-white cloning technique (described in section 2.1). *Vc* dnaB gene was then restricted from pDrive and ligated into expression vector, pET28a. This expression vector, provided by Dr. Moore, was then used for transformation into different expression hosts. Figure 8.1 shows the gene sequence for *Vc* dnaB.

![Gene sequence for Vc dnaB with 1407 bases. Sequence obtained from NCBI web server (Gene ID: VC 0395). For details of the protein refer to Table 6.1, section 6.3.1.](image)

189
8.2. Protein Expression Study

8.2.1. Small Scale Expression study

\( Vc \) dnaB gene in pET28a was transformed into expression hosts BL21(DE3) RIL and Gold. The transformation protocol is described in section 2.1.5. A small scale expression study (refer section 2.2.1.) showed good expression with the protein being expressed at the calculated molecular weight of 52.04 kDa as shown in Figure 8.2. Expression was done at 37 °C after induction with IPTG. Samples were taken at 1, 2, 3, 4, 5 and 6 hours after induction. As seen from Figure 8.2, \( Vc \) dnaB seems to have over-expressed with optimum expression level seen after five hours of induction. Based on this observation, protein induction in large scale expression study was done for five hours at 37 °C.

![Figure 8.2](image)

**Figure 8.2:** SDS-PAGE gel of \( Vc \) dnaB expression study showing expression levels at different time intervals after induction of expression using IPTG.
8.2.2. Solubility Study

Once a good level of expression was shown, solubility of the expressed \( Vc \) dnaB was examined in low salt (0.15 mM) versus high salt (1.00 M) buffer. Figure 8.3 shows the SDS-PAGE gel of \( Vc \) dnaB solubility study using the method described in section 2.2.1. As seen from lanes 3 and 4, soluble protein was obtained in the supernatant in presence of both low and high salt respectively. Contaminants were still present but barely visible on the gel. Since protein was seen in the supernatant with low salt buffer, lysis in presence of low salt was subsequently adopted for all further studies.

![SDS-PAGE gel](image)

**Figure 8.3**: SDS-PAGE showing solubility of \( Vc \) dnaB in low salt, high salt and BugBuster\(^\text{TM}\). The protein is soluble in all the three cases.
8.2.3. Solubility Profile

As described in Chapter 2, section 2.2.5., solubility profile of \textit{Vc} dnaB was determined on a sample obtained from a trial purification run. Preliminary purification run using column chromatography with standard Q-Sepharose buffers yielded enough protein with reasonable purity to do a solubility study. Based on the solubility profile seen in Figure 8.4, sodium chloride, trisodium citrate and TAPS pH 8.0 were selected as the components of buffer to obtain maximum solubility of \textit{Afu} SSB. For all subsequent purification runs this was the buffer of choice.
Figure 8.4: Solubility profile of \( Vc \) dnaB indicating sodium chloride, trisodium citrate and TAPS pH 8.5 as the optimum components of a buffer giving the highest solubility of the protein.
8.3. Protein Purification

Large scale expression study was done as described in section 2.2.2 and lysis done with low salt buffer as described in section 2.2.3. The buffer used was the standard lysis buffer. After lysis the lysate was centrifuged at 18,000 x g for 30 minutes at 4 °C. Supernatant was separated from the pellet and filtered using a syringe filter to remove any cell debris. Filtrate was then diluted with dilution buffer (25 mM TAPS pH 8.5) to lower the conductivity to that of Q-Sepharose cation exchange column loading buffer as described in section 2.2.4. Buffer-A (loading buffer) consisted of 12.5 mM sodium chloride, 50 mM trisodium citrate and 50 mM TAPS pH 8.5. Buffer-B (elution buffer) consisted of 500 mM sodium chloride, 50 mM trisodium citrate and 50 mM TAPS pH 8.5. Figure 8.5 shows the chromatogram for Q-Sepharose purification run for Vc dnaB and Figure 8.6 shows the SDS-PAGE gel picture of the different fractions from the Q-Sepharose run. As seen from Figure 8.6, Vc dnaB eluted from the column in all the fractions from 25 to 50. At this stage the protein was partially pure with few contaminants still present. Fractions 25 to 50 from Q-Sepharose were then pooled and dilution buffer added to match the conductivity to that of Buffer-A. POROS-HQ, a high resolution cation exchange column, was then used to further purify the protein as described in section 2.2.4. The protein was then concentrated to around 10 to 15 mg/mL as described in section 2.2.6 to bring down the volume and salt concentration.
**Figure 8.5:** Chromatogram of Q-Sepharose run for *Vc* dnaB purification

<table>
<thead>
<tr>
<th>Mol.Wt. (kDa)</th>
<th>#1: Molecular weight marker</th>
<th>#2: <em>Vc</em> dnaB Q-Sepharose Flow through</th>
<th>#3: <em>Vc</em> dnaB Q-Sepharose fraction 16</th>
<th>#4: <em>Vc</em> dnaB Q-Sepharose fraction 23</th>
<th>#5: <em>Vc</em> dnaB Q-Sepharose fraction 30</th>
<th>#6: <em>Vc</em> dnaB Q-Sepharose fraction 37</th>
<th>#7: <em>Vc</em> dnaB Q-Sepharose fraction 44</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 8.6:** SDS – PAGE gel of samples from Q-Sepharose run for *Vc* dnaB purification

*Vc* dnaB (52.04 kDa)
Figure 8.7: Chromatogram of POROS-HQ run for \( Vc \) dnaB purification

1: \( Vc \) dnaB POROS-HQ wash after loading  
2: \( Vc \) dnaB POROS-HQ fraction 17  
3: \( Vc \) dnaB POROS-HQ fraction 31  
4: \( Vc \) dnaB POROS-HQ fraction 42  
5: Molecular weight marker

\( Vc \) dnaB (52.04 kDa)

Figure 8.8: SDS – PAGE of samples from POROS-HQ run for \( Vc \) dnaB purification
Upon concentration, the purified protein formed a dimer with a molecular weight of ~ 116.3 kDa which is seen from the SDS-PAGE gel in lane #1, Figure 8.9. Upon addition of the reducing agent TCEP, the band at 116.3 kDa disappeared with the simultaneous increase in size of the band at ~52 kDa confirming the formation of disulfide linkage at higher concentration.

![Image: SDS-PAGE gel showing samples](image)

**Figure 8.9:** Picture of an SDS-PAGE gel showing samples without and with increasing concentration of TCEP

Further analysis of the protein was done in order to obtain a monodispersed sample which could be used for crystallization trials. Three different types of reducing agents (β-ME, DTT and TCEP) were used in addition to ADP, ATP and MgCl₂. Literature had shown that proteins of this family require ATP and MgCl₂ for their activity / function. The solubility of protein in these components was determined by using 15 mg/mL per test in an eppendorf tube on ice. Reaction mixture was mixed by pipetting and the state of the solution visually observed. Table 8.1 shows the different tests done on Vc dnaB in presence of different components. From the Table it is clear that the most
optimum condition is 1 mM TCEP, 2 mM ADP, 5 to 10 mM MgCl₂ and 25 mM HEPES pH 7.5.

**Table 8.1: Table showing solution state of Vc dnaB in presence of reducing agents / ADP / ATP / MgCl₂**

<table>
<thead>
<tr>
<th></th>
<th>Vc dnaB (PIPES pH 6.5)</th>
<th>Vc dnaB (HEPES pH 7.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mM DTT</td>
<td>Soluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM TCEP</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM β-ME</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>5 mM MgCl₂</td>
<td>-</td>
<td>Soluble (faint cloudiness)</td>
</tr>
<tr>
<td>10 mM MgCl₂</td>
<td>Insoluble</td>
<td>Partially soluble on mixing</td>
</tr>
<tr>
<td>2 mM ADP</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>2 mM ATP</td>
<td>Insoluble</td>
<td>-</td>
</tr>
<tr>
<td>2 mM ADP + 10 mM MgCl₂</td>
<td>Insoluble</td>
<td>Insoluble</td>
</tr>
<tr>
<td>1 mM DTT + 2 mM ADP</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM DTT + 2 mM ADP + 5 mM MgCl₂</td>
<td>-</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM DTT + 2 mM ADP + 10 mM MgCl₂</td>
<td>Insoluble</td>
<td>Precipitate</td>
</tr>
<tr>
<td>1 mM TCEP + 2 mM ADP</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM TCEP + 2 mM ADP + 5 mM MgCl₂</td>
<td>-</td>
<td>Soluble</td>
</tr>
<tr>
<td>1 mM TCEP + 2 mM ADP + 10 mM MgCl₂</td>
<td>Insoluble</td>
<td>Soluble upon mixing</td>
</tr>
</tbody>
</table>
8.4. Crystal Screening

Crystal screening was done using all the screens mentioned in Table 2.9, section 2.3.1. Screens were set up at a concentration of 19.8 mg/mL using 0.5µL + 0.5µL format as described in section 2.3.1 in the three-well Greiner® plates.

Figure 8.10: Few of the crystal hits obtained for *Vc* dnaB. (a) to (c) were protein alone. (d) to (f) were protein in presence of 2 mM ADP. (g) to (i) were protein in presence of 2 mM ADP and 10 mM MgCl₂.

(a) Wizard-II condition #5 (20% 1,4-butanediol, 0.1 M HEPES pH 7.5, 0.2 M NaCl) (b) Cryo-I condition #25, (30% PEG 200, 0.1 M Tris pH 8.5, 0.2 M (NH₄)₂HPO₄) (c) Modified Ion screen condition #60, (0.2 M sodium citrate, 0.1 M CAPS pH 10.5) (d) Wizard-II condition #10, (1.0 M (NH₄)₂HPO₄, 0.1 M Tris pH 8.5) (e) Wizard-I condition #34, (1.0 M (NH₄)₂HPO₄, 0.1 M Imidazole pH 8.0) (f) Wizard-I condition #25, (30% PEG 400, 0.1 M Tris pH 8.5, 0.2 M MgCl₂) (g) Wizard-I condition #25, (30% PEG 400, 0.1 M Tris pH 8.5, 0.2 M MgCl₂) (h) Wizard-II condition #33 (10% 2-propanol, 0.1 M cacodylate pH 6.5, 0.2 M zinc acetate) (i) Cryo-I condition #25, (30% PEG 200, 0.1 M Tris pH 8.5, 0.2 M (NH₄)₂HPO₄).
The first well had protein alone, the second had protein in presence of 2 mM ADP and the third well had protein in presence of 2 mM ADP and 10 mM MgCl₂. Screens were set up at room temperature and at 4 °C. Few crystal hits were seen (all of them from 4 °C screens), some of which are shown in Figure 8.10. On checking the crystals with Izit® dye, none of them stained blue implying that the crystals were probably salt. But this test is known to be unreliable and the only way to confirm this is to do diffraction studies.

8.5. Dynamic Light Scattering

Dynamic light scattering was done using the method described in section 2.4.1. *Vc* dnaB at a concentration of ~1 mg/mL was used for this experiment. Different parameters like temperature, presence of reducing agent, presence of ADP, presence of MgCl₂ and different buffer conditions were used to study the state of this protein in solution. Results are summarized in Table 8.2. The sample at 4 °C in presence of 10 mM TAPS pH 8.5 buffer showed average molecular weight of ~ 300 kDa indicating a possibility of hexamer (calculated molecular weight of a monomer is 52.5 kDa). Presence of reducing agent did not seem to have any effect on oligomerization since DLS experiment in presence of β-ME showed similar molecular weight, indicating that the oligomerization probably did not involve disulfide bonds. In presence of 25 mM Tris pH 7.5, 100 mM NaCl and 1% glycerol, DLS experiment at 4 °C showed a prominent species with an apparent molecular weight of ~ 650 kDa which indicated the presence of a dodecamer. Figure 8.12 and 8.13 shows the plot for DLS with and without ADP and MgCl₂.
Figure 8.11: DLS of \( Vc \) dnaB in 10 mM TAPS pH 8.5 and 1 mM \( \beta \)-ME at 4 °C (1.0 mg/mL) shows one prominent species with apparent molecular weight of \(~ 350 \text{ kDa}\).
Figure 8.12: DLS of Vc dnaB in 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% glycerol at 4 °C shows one prominent species with apparent molecular weight of ~ 650 kDa.
Figure 8.13: DLS of Vc dnaB in 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% glycerol in presence of 2 mM ADP and 10 mM MgCl₂ at 4 °C shows a prominent species with a molecular weight of ~ 750 kDa.
Figure 8.14: DLS of $Vc$ dnaB in 10 mM HEPES pH 7.5 at 4 °C shows a prominent species with a molecular weight of ~ 1000 kDa.
Table 8.2: Summary of *Vc* dnaB DLS experiment

<table>
<thead>
<tr>
<th></th>
<th><em>Vc</em> dnaB</th>
<th><em>Vc</em> dnaB</th>
<th><em>Vc</em> dnaB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MW Pd% SOS</td>
<td>(with β-ME)</td>
<td>+ 2 mM ADP + 10 mM MgCl₂</td>
</tr>
<tr>
<td>10 mM TAPS pH 8.5</td>
<td>250 45 4.5</td>
<td>348 20.9 4.5</td>
<td>- - -</td>
</tr>
<tr>
<td>25 mM Tris pH 7.5, 100 mM NaCl, 1% glycerol</td>
<td>659 2.8 4.6</td>
<td>531 12.7 29.6</td>
<td>757 12.1 6.0</td>
</tr>
<tr>
<td>10 mM HEPES pH 8.5</td>
<td>1011 12.0 3.5</td>
<td>1034 19.1 1.6</td>
<td>- - -</td>
</tr>
</tbody>
</table>

In presence of 10 mM HEPES pH 8.5, the molecular weight obtained was approximately 10,000 kDa (calculated molecular weight of *Vc* dnaB icosatetramer is 1260 kDa).
Chapter 9: *Yersinia pestis* (*Yp*) dnaB

Gene sequence for *Yersinia pestis* (*Yp*) dnaB was obtained from the NCBI website (Gene ID: YPO 0320) as shown in Figure 9.1. Based on this sequence, forward and reverse primers were designed for PCR as shown in Figure 9.2(a).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGGCAGCGA AAAAAACAAAC CAACAAAGATG ACCGAGCCAC GAGACCCGCA GATGGGAGGG</td>
<td>Forward Primer 1</td>
</tr>
<tr>
<td>CTGAAGCTTC CGCCGCATTC GCTGGAGGCT GACGAGTCCG TGGTGGGGCGG TTTGATGCTG</td>
<td>Reverse Primer 1</td>
</tr>
<tr>
<td>GATAACGAAC GTTGGGATAA CGTATCAGAA CGTGTCGCCA GCAAGGACTT TTTCAGCCGC</td>
<td>Forward Primer 2</td>
</tr>
<tr>
<td>GCCCATCCTAC CGGAATCTAC TAAAACACCG CCAATGCGGC GGAATATCGG TGCTTATGCC</td>
<td>Reverse Primer 2</td>
</tr>
</tbody>
</table>

206
9.1. Molecular Biology

9.1.1. Polymerase Chain Reaction (PCR)

Steps involved in PCR were similar to that discussed in chapter 2, section 2.1.1. Nucleotide sequences for the primers were based on the gene sequence and the forward primer included an additional “CACC” sequence required for Gateway® cloning followed by TEV protease site. Samples were run on a 1% agarose gel to confirm the success of the reaction. The gene has 1407 base pairs and as seen in Figure 9.1, all the four reactions seem to have worked with a band at around 1407 base pairs.

**Forward Primer:**

5′ **CACC**GAGAACCTCTACTTCCAAGGAAATGGCAGCGGGAAGAA [red] **GG**GAAAAAAACCAACCAAC 3′

**Reverse Primer:**

5′ **CTACTCGTCG** TCATATTGCG GA 3′

**Figure 9.2:** (a) Forward and reverse PCR primers. The red colored sequence indicates TEV protease cleavage sequence. The sequence before that (shaded green) is required for Gateway cloning.
Figure 9.2: (b) Picture of an agarose gel showing samples of the four PCR reactions of Yp dnaB along with a 100 base pair ladder. All the four reactions seem to have worked.

The band for the gene was then gel purified using minielute as described in section 2.1.2.

9.1.2. Insertion into cloning vector using Gateway® Entry vector

After gel purification the Yp dnaB gene was inserted into the Gateway® entry vector, pENTR-D, using the topoisomerase reaction described in section 7.1.3. pENTR-D entry vector with the Yp dnaB gene was then transformed into one shot competent TOP10 (E. coli) cells as described in section 7.1.3. Transformed cells were plated on a LB agar plate with the antibiotic, kanamycin. After an overnight incubation at 37 °C, three colonies from a total of approximately 25 colonies were randomly picked and a 5.0 mL overnight grow up of the same was done using LB (with kanamycin). Supercoiled
plasmids from the three different colonies were then isolated using Miniprep as described in section 2.1.6. Samples of the plasmids were then run on an agarose gel and as shown in Figure 9.2, insertion seems to have worked since the plasmid seems to run at a level approximately equal to that of 4000 base pair plasmid.

![Figure 9.3: Picture of an agarose gel showing insertion of \( Yp \) dnaB into pENTR-D entry vector. All the three plasmids show insertion of the \( Yp \) dnaB gene.](image)

9.1.3. Switching from Entry to Destination vector

Once the insertion into entry vector was confirmed, the LR clonase (transposition) reaction was set up as described in section 7.1.5 to switch the \( Yp \) dnaB gene from entry vector to destination vector. Once this step was done and the plasmids isolated, samples were run on an agarose gel to check for the success of the transposition reaction. Three
different destination vectors were used, which included pDEST-17, pET45-DEST and pKM596-H.

As seen from Figure 9.4, \( Yp \) dnaB gene seems to have switched from pENTR-D to pDEST-17 in samples 3, 4 and 5. From Figure 9.5 it is clear that colony #2 has the gene inserted into pET45-DEST and from the same Figure it is clear that insertion of the gene was successful into pKM596-H.
Once the destination vector with the inserted gene was obtained successfully, it was transformed into different expression hosts for protein expression studies.

9.2. Protein Expression, Purification and Solubility

9.2.1. Small scale protein expression study

Small scale expression studies were done as described in section 7.2.1. Different cell lines were used for this study. For *Yp* dnaB in pDEST-17, cell lines used were BL21(DE3) GOLD, RILP, pLysS, Rosetta, Rosetta pLysS. Protein expression was not seen in any of these cell lines. Figure 9.2 shows the SDS-PAGE gels for some of the cell lines that were tried. Position at which the band for the expressed protein should have been seen is indicated in the Figure.
Figure 9.6: SDS-PAGE gels of \( Y_p \) dnaB expression study using destination vector pDEST 17 in different cell lines (a) BL21(DE3) GOLD, RILP and Rosetta (b) BL21(DE3) pLysS. None of the cell lines showed any expression of \( Y_p \) dnaB.

For \( Y_p \) dnaB in pET45-DEST, the cell line used was BL21(DE3) RILP. Protein expression was seen as shown in the SDS-PAGE gel in Figure 9.7. Expression study was done at 37 °C and 30 °C for 3 hours and 4 hours. \( Y_p \) dnaB was seen to have over expressed in all the conditions checked for.
Figure 9.7: SDS-PAGE gel of \( Yp \) dnaB expression study using destination vector pET45-DEST in BL21(DE3) RILP shows expression band at ~ 52 kDa at 30 °C and at 37 °C.

For \( Yp \) dnaB in pKM596-H, the cell line used for expression study was BL21(DE3) RILP. As seen from the SDS-PAGE gel in Figure 9.8, \( Yp \) dnaB along with MBP was expressed at a molecular weight of approximately 94 kDa (\( Yp \) dnaB : 52 kDa, His-MBP : 42 kDa).

Figure 9.8: SDS-PAGE gel of \( Yp \) dnaB-MBP expression study using the destination vector pKM596-H showing expression after induction using IPTG at around 97 kDa in two different colonies.
9.2.2. Solubility study

Solubility studies were done as described in section 2.2.1 on Yp dnaB expressed using the destination vector pET45-DEST in BL21(DE3) RILP as well as on Yp dnaB-MBP protein expressed from the BL21(DE3) RILP cell line using the pKM596-H expression vector. As seen from the SDS-PAGE gel in Figure 9.9, the protein appeared to be insoluble in low salt (0.15 M), high salt (1.00 M) and BugBuster™ when expressed alone using pET45-DEST vector. Yp dnaB-MBP protein appeared to be soluble in both low salt (0.15 M) and high salt (1.0 M) but more so in high salt (Figure 9.10).

For all further studies Yp dnaB-MBP protein in ~ 0.5 M NH₄Cl was used.

Figure 9.9: SDS-PAGE gel of Yp dnaB solubility study showing insoluble protein in low salt, high salt and bugbuster when expressed using destination vector pET45-DEST. On the left side of the molecular weight marker is the protein expressed at 30 °C and on the right is that expressed at 37 °C.
9.2.3. Large Scale Expression and Protein purification study

Large scale expression was done as described in section 7.2.1 and section 2.2.2. Cell lysis and purification were done as described in section 2.2.3 and 7.2.3 respectively. The amylose column was first equilibrated with column loading buffer (50 mM Tris-HCl pH 7.0, 600 mM NaCl, 1 mM EDTA and 1 mM DTT). After loading the lysate, the column was rinsed with two to three column volumes of column loading buffer and the protein eluted using elution buffer which consisted of column loading buffer with 10 mM maltose. As seen from the chromatogram in Figure 9.11, two peaks were obtained. Samples from the fractions were then run on a SDS-PAGE gel. Figure 9.12 shows protein was present in all the peaks but the first peak had a contaminant of approximately 66 kDa. Fractions below Peak I (fractions 6 and 7) and Peak II (fractions 9 to 12) were pooled separately and concentrated as described in section 2.2.6.
**Figure 9.11:** Chromatogram of Amylose column run for *Yp* dnaB purification.

- **#1:** *Yp* dnaB-MBP Amylose load sample
- **#2:** *Yp* dnaB-MBP Amylose fraction 6
- **#3:** *Yp* dnaB-MBP Amylose fraction 9
- **#4:** *Yp* dnaB-MBP Amylose fraction 10
- **#5:** *Yp* dnaB Amylose wash before elution
- **#6:** Molecular weight marker

**Figure 9.12:** SDS – PAGE gel of samples from Amylose column run for *Yp* dnaB-MBP purification.
Once the purified protein was concentrated, a small scale study was done to cleave the $Yp$ dnaB from the $Yp$ dnaB-MBP protein using the TEV protease site present in the latter between $Yp$ dnaB and MBP. The concentrated $Yp$ dnaB-MBP was dialyzed in buffer that was optimum for TEV reaction (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.5 mM EDTA and 1 mM DTT) and then a 250 µL sample of protein (~ 5.0 mg/mL) was taken in an eppendorf tube and 5 µL of TEV protease (~ 2.0 mg/mL) was added, mixed and then incubated at 15 °C in the refrigerated shaker. Samples were taken after eight hours and 24 hours of incubation and run on a SDS-PAGE gel (Figure 9.13). The protein was successfully cleaved and was in a soluble form in high salt (500 mM NaCl). As soon as the salt concentration was brought down to 250 mM, the protein was no longer soluble.

![Figure 9.13](image)

**Figure 9.13**: SDS – PAGE gel showing reaction products of TEV protease reaction. The cleaved $Yp$ dnaB protein is soluble in 500 mM NaCl (lanes 2 and 3) but precipitates when salt concentration is lowered to 250 mM.
9.2.4. Solubility Profile

Solubility profile of \textit{Yp} dnaB-MBP was determined as described in section 7.2.2 on a sample of purified \textit{Yp} dnaB-MBP. Based on the solubility profile seen in Figure 9.14, trisodium citrate and MgCl\textsubscript{2} were selected as the components of buffer to obtain maximum solubility of \textit{Yp} dnaB-MBP.
Figure 9.14: Solubility profile of *Yp* dnaB-MBP shows the protein prefers magnesium chloride, sodium sulfate and trisodium citrate pH 7.5.
Chapter 10: *Helicobacter pylori* (Hp) dnaB

Gene sequence for *Helicobacter pylori* (Hp) dnaB was obtained from the NCBI website (Gene ID: JHP 1280) as shown in Figure 10.1. Based on this sequence, forward and reverse primers were designed for PCR as shown in Figure 10.1(a).

```
1 ATGGATCATT TAAAGCATTT GCAACAATTT GAAAACATTG AAAGGATCGT GCTTTCAGGC
61 ATTTGTGTAG CCAATCATAA GATGGAAGAG ATCCATAGCG TTTTAGAGCC TACGCTAGTT
121 TACTACCCGC CCCATGGCTTT GATTTTTTGA AACTCCATGA AGTGAAATGGC
181 CCATTTATATG AGAATTTTAT CCAGGCGAAA ATGGCCAAAG ACAACAAAT CAGCGAAGAC
241 GATCTAGATCG CTATTTTTGC GCGCAAGCCGC ATAGACAATA TTGAGCAGCC AGAAGTGTAT
301 ATCAAAACAC CCTCCCATTA ACAGAAGGCCT TTAACCTTTG ACAAACACC TAGAGAGCAAA
361 GCTTTAGAAA GCCGCAGAAA ATCCACGGAG ATTTAAAACG CTGAGCCCAT TACGCTAGTT
421 GCGTTTTATGA AGGCAAGGCG TTTAGGTTA TAAAGGAAGT GCTTTAACGC
481 ACAATGATAC TTATCATAGA AAAACAAAGA AAAGGAGGT TAAAGAGGAC CGCTATAGCG
541 ACTGCGCTTG CTAATGTGGA ATTATCACGG AGCGGTGTAT ACGAAGGAGA TTTAGTATAT
601 CTAGGCCAGA GAGCCGCTATG GGTAAAACGC ATGTGTATGA TAAACAGGCT CTTATACCCG
661 CTCATATGAC ATAGGGGGCT ATGCTTTTTT TTATGGAAA TGCTCCAGGA CAAACTCGGC
721 TTAAGGGCCG TATCGGACTG CACTCTTATT AACATGCAAG ATTTAGAGAG CGCAAGGTCT
781 GATGATGATC ATATGGGAAA TTTAGCAGCAA GCCTTTATGC ACTTTTGCAA AAAATAACTT
841 TTTTTTTATG ATAATGAGAA CCCTGAGGATG GATCAAAATCC GCTGGCAACT ACGAAGGAC
901 AAATCCCAAC ACAAGGAAAT GGTGATGCGT TTATATTGAT ATTGTACTCT CATGTAGGG
961 AACAAAGCCA CTAAAGGAGC GCATGAAACA ATCGCTGAAA TTGTAAGGGA GCTTAAAATC
1021 TTGAGGCAAT AATTAGAAGATT CCACTAGCAG GTGTTAGTGAC AACTCAACCG CAGCCTAGAA
1081 AAATGGGAGG ATAAACCGCG CATCTTTTACG TATTTAAAG ACACCGGAGG GATGGAACAG
1141 GAGCCTGATA TTTTTTTATT TTGGTATAGA GGCTATTATC ATACAATGAG GCTGAAGAC
```
Figure 10.1: *Hp* dnaB gene sequence obtained from the NCBI website (Gene ID: JHP 1280). For details of the protein refer to Table 6.3, section 6.3.3.

10.1. Molecular Biology

10.1.1. Polymerase Chain Reaction (PCR)

Steps involved in PCR are discussed in section 7.1.1 and section 2.1.1. Two different types of Polymerase chain reactions were done. In the first reaction, the forward primer had an additional “CACC” sequence required for Gateway® cloning. In the second type of reaction, the forward primer included the “CACC” sequence required for Gateway® cloning and a TEV protease site. PCR product from the first type of PCR (*Hp* dnaB) was used for insertion into the Gateway® expression vector, pET101 whereas PCR product of the second type of reaction (*Hp* dnaB-TEV) was used for insertion into the Gateway® entry vector, pENTR-D.

Samples of the reaction were run on a 1% agarose gel to confirm the success of the PCR. The gene has 1461 base pairs and as seen in Figure 10.2, all the reactions seem to have worked with a band at around 1500 base pairs. Band for the gene was gel purified using MiniElute as described in section 2.1.2.
Forward Primer:
\[
5' \text{CACC GAGA CCTCT TACTT CCAAGGA ATGGAT CATT TAA AG CAT TT GCAAC} 3'
\]

Reverse Primer:
\[
5' \text{TCA AGT TGT GGG CAT TTC AAA TTT A} 3'
\]

Figure 10.2: (a) Forward and reverse PCR primers. The red colored sequence indicates the TEV protease cleavage sequence. The sequence before that (shaded green) is required for Gateway cloning. The forward primer for the 2nd type of PCR does not have this TEV sequence. (b) Picture of agarose gels showing two samples of polymerase chain reactions of \textit{Hp dnaB} for use with pET101 and three samples for use with pENTR-D.

10.1.2 Insertion into cloning vector using Gateway® Entry vector

After gel purification, the \textit{Hp dnaB-TEV} gene was inserted into the Gateway® entry vector, pENTR-D, as described in section 7.1.3. The pENTR-D entry vector with the \textit{Hp dnaB} gene was transformed into one-shot competent TOP10 (\textit{E. coli}) cells as described in section 7.1.3. Transformed cells were plated on a LB agar plate with the
antibiotic, kanamycin. After an overnight incubation at 37 °C, two colonies from a total of ten colonies were randomly picked and a 5.0 mL overnight grow up of the same was done using LB (with kanamycin). Supercoiled plasmids from the two colonies were isolated using Miniprep as described in section 2.1.6. Samples of the plasmids were then run on a 1% agarose gel and as shown in Figure 10.3, the insertion seems to have worked since the plasmids seems to have run at a level approximately equal to that of 4000 base pair plasmid.

![Figure 10.3: Picture of agarose gel showing insertion of Hp dnaB PCR product (with the TEV protease site) into entry vector, pENTR-D.](image)

**Figure 10.3:** Picture of agarose gel showing insertion of *Hp* dnaB PCR product (with the TEV protease site) into entry vector, pENTR-D.

### 10.1.3. Switching from Entry to Destination vector

Once insertion into entry vector was confirmed, LR clonase (transposition) reaction was set up as described in section 7.1.4 to switch the *Hp* dnaB-TEV gene from entry vector to destination vector. Once this step was done and the plasmids isolated, a sample was run on a 1% agarose gel to check for the success of the transposition reaction. The destination vector used was pKM596-H. Unfortunately, the gel picture for
the sample run on a 1% agarose gel is not available, but the proof that the transposition worked is provided in section 10.2.1 during expression studies. Expression of the protein at the expected molecular weight is an indication that the LR clonase reaction worked.

The destination vector, pKM596-H, with the inserted *Hp* dnaB gene was then transformed into DH5α cloning host as described in section 2.1.9 to obtain supercoiled plasmid.

### 10.1.4. Insertion of *Hp* dnaB PCR product into pET101

Purified PCR product without the TEV protease cleavage sequence (PCR product type I, section 10.1.1) was then inserted into Gateway® expression vector pET101. This is a one step unidirectional topoisomerase reaction whereby once the insertion of the gene has taken place; it cannot be easily switched into any other destination vector. The reaction procedure is the same as described in section 7.1.3. Figure 10.4 shows out of three topoisomerase reactions, two worked. Use of this destination vector eliminates the step involving LR clonase (transposition reaction).

![Figure 10.4: Picture of agarose gels showing four samples of topoisomerase reactions for insertion of *Hp* dnaB gene into pET101. Samples 1 and 2 seems to have worked.](image)

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Hp</em> dnaB pET101 (~7500 bps)</td>
</tr>
<tr>
<td>2</td>
<td><em>Hp</em> dnaB pET101 #1 (duplicate)</td>
</tr>
<tr>
<td>3</td>
<td>Super coiled ladder</td>
</tr>
<tr>
<td>4</td>
<td><em>Hp</em> dnaB pET101 #2</td>
</tr>
<tr>
<td>5</td>
<td><em>Hp</em> dnaB pET101 #3</td>
</tr>
</tbody>
</table>
The expression vector pET101 with the inserted *Hp* dnaB gene was then transformed into DH5α cloning host as described in section 2.1.9 to obtain supercoiled plasmid.

10.2. Protein Expression, Purification and Solubility

10.2.1. Small scale protein expression study

Once the supercoiled destination vector with the inserted gene was successfully isolated from DH5α cells, it was transformed into BL21(DE3) Star and RILP for small scale protein expression studies as described in section 2.1.10.

![Figure 10.5: SDS-PAGE gel of *Hp* dnaB expression study using *Hp* dnaB gene in pET101 destination vector showing expression levels in two different cell lines after induction of expression using IPTG.](image)

#1: *Hp* dnaB, 3 hour expression, BL21(DE3) STAR
#2: *Hp* dnaB, 0 hour expression, BL21(DE3) STAR
#3: Molecular weight marker
#4: *Hp* dnaB, 0 hour expression, BL21(DE3) RILP
#5: *Hp* dnaB, 5 hour expression, BL21(DE3) RILP

*Hp* dnaB (55.5 kDa)
As seen from Figure 10.5 good levels of expression were seen in BL21(DE3) Star and RILP cell lines when \(Hp\) dnaB gene in pET101 destination vector was used. The calculated molecular weight of the protein is 55.5 kDa and the expressed protein has a molecular weight very close to the calculated molecular weight. Cells with the expressed protein were pelleted and used for lysis and solubility study as described in section 2.2.1 and 2.2.2.

In the same way, \(Hp\) dnaB gene in pKM596-H was transformed into BL21(DE3) RILP, STAR, T7 express, pLysS and Rosetta pLysS cell lines as described in section 2.1.10. Following successful transformation, a small scale expression study was done as described in section 2.2.1. Figure 10.6 shows an SDS-PAGE gel of samples from zero hour and three hour expression from all the above mentioned cell lines. \(Hp\) dnaB with Maltose Binding Protein (MBP) expressed in all the cell lines.

![Figure 10.6: SDS-PAGE gel of \(Hp\) dnaB-MBP expression study using \(Hp\) dnaB gene in pKM596-H destination vector showing expression levels in five different cell lines after induction of expression using 1 mM IPTG.](image)

**Figure 10.6:** SDS-PAGE gel of \(Hp\) dnaB-MBP expression study using \(Hp\) dnaB gene in pKM596-H destination vector showing expression levels in five different cell lines after induction of expression using 1 mM IPTG.
10.2.2. Solubility study

Solubility study was done as described in section 2.2.1 on *Hp* dnaB protein alone and on *Hp* dnaB with MBP protein expressed from different cell lines mentioned in section 10.2.1. As seen from the SDS-PAGE gel in Figure 10.7(a), the protein appeared to be insoluble in low salt (0.15 M NH₄Cl) and high salt (1.00 M NH₄Cl) when expressed alone but partly soluble in low salt and high salt when expressed along with MBP (Figure 10.7(b)).

![SDS-PAGE gel](image)

**Figure 10.7**: SDS-PAGE gel of (a) *Hp* dnaB solubility study showing insoluble protein and (b) partially soluble *Hp* dnaB-MBP.
10.2.3. Large Scale Expression and Protein purification study

Large scale expression of *Hp* dnaB-MBP was done as described in section 7.2.1 and section 2.2.2. Cell lysis and purification were done as described in section 2.2.3 and 7.2.3 respectively. The Cobalt-Talon column was equilibrated with column loading buffer which consisted of 50 mM Sodium Phosphate pH 7.5 and 300 mM NaCl. After loading the lysate, column was rinsed with two to three column volumes of column loading buffer and protein eluted using elution buffer which consisted of column loading buffer with 500 mM imidazole. As seen from the chromatogram in Figure 10.8, a single prominent peak was obtained. Samples from this peak along with sample loaded on to the column were then run on a SDS-PAGE gel. Figure 10.9 shows *Hp* dnaB-MBP (partially pure). Fractions below the peak (fractions 9 to 15) were pooled together and concentrated as described in section 2.2.6.
Figure 10.8: Chromatogram of Talon column run for *Hp dnaB* purification shows a single prominent peak

*Hp dnaB-MBP* (97.4 kDa)

Figure 10.9: SDS – PAGE gel of samples from Talon column run for *Hp dnaB-MBP* purification

#1: *Hp dnaB-MBP* Talon load sample  
#2: *Hp dnaB-MBP* Talon load flow through  
#3: *Hp dnaB-MBP* Talon fraction 8  
#4: *Hp dnaB-MBP* Talon fraction 10  
#5: *Hp dnaB-MBP* Talon fraction 11  
#6: *Hp dnaB-MBP* Talon fraction 13  
#7: Molecular weight marker
Chapter 11: Discussion

Obtaining structural information of helicases, especially the ones from the organisms studied in this project is important for understanding the steps involved in the helicase activity. This could provide for a potential target for structure based drug design in combating diseases like cholera and plague. The current line of treatment for many of the bacterial diseases including that for cholera and plague are antibiotics but with cases of antibiotic resistance on the rise, alternative drugs or therapy is always welcomed.

Although the main goal of this project was to obtain crystal structure of dnaB from the three species studied, none could be obtained. Of the three, Vc dnaB was the only helicase that was obtained in a soluble form. Expression and purification protocol were established for obtaining milligram quantities of Vc dnaB. Few crystal hits were obtained and preliminary attempts to expand some of the conditions did not yield any crystals. Biophysical studies showed that Vc dnaB may exist in different solution states depending on the composition of buffer. DLS experiments showed hexamer, dodecamer and icosatetramer states in solution. Yp dnaB and Hp dnaB could not be obtained in a soluble form on their own. Soluble form of the protein was obtained only when expressed with a tag, Maltose Binding Protein (MBP). Upon cleaving Yp dnaB from the MBP-Yp dnaB protein, Yp dnaB was seen to be soluble (at a NaCl concentration of ~ 500 mM). On reducing the salt concentration to 250 mM NaCl, Yp dnaB precipitated from the solution.
Figure 11.1: (a) Crystal structure of Taq dnaB (PDB: 2Q6T) monomer showing the N-terminal domain (NTD), a linker and the C-terminal domain (CTD). Part of the linker, residues 175 to 182, is missing and a portion of the CTD, residues 326 to 332, is also not shown. The structure is predominantly composed of \(\alpha\)-helices (shown in red) with few \(\beta\)-strands (shown in yellow) (b) Domain arrangement in the structure of Taq dnaB. The residues at the domain boundaries are shown.
During the later half of this study, the structure of a close homolog of $Vc$ dnaB and $Yp$ dnaB was solved by Stetiz and co-workers.\textsuperscript{100} The dnaB of \textit{Thermus aquaticus} was solved to a resolution of 2.9 Å by this group which was published in July, 2007.

As shown in the introduction, this protein has a sequence similarity of 66% and sequence identity of 47% with both $Vc$ and $Yp$ dnaB. The structure of \textit{Taq} dnaB is the first full length dnaB structure solved so far which provides an insight into the differences and similarity of this helicase with the helicases of the other superfamilies.

The structure of \textit{Taq} dnaB (Figure 12.1) shows the presence of two domains, the N-terminal domain (NTD) which is predominantly composed of $\alpha$-helices and the C-terminal domain (CTD) which is composed of $\beta$-sheets along with $\alpha$-helices. Connecting the two domains is a linker that is composed of a single $\alpha$-helix with two loops that connect the helix to the NTD and the CTD. All the five motifs that define the helicase family, H1, H1a, H2, H3 and H4 are present in the \textit{Taq} dnaB. These motifs are present in the CTD and residues from the first four motifs form the core $\beta$-sheet that provides residues involved in nucleotide binding pocket. The fifth motif, H4, is present on the $\beta$-strand which is associated with DNA binding.\textsuperscript{100}

The CTD which has the helicase activity closely resembles the helicase domain of \textit{T7} gp4 and the \textit{E1} helicase of the \textit{Bovine pappilloma virus}.\textsuperscript{89, 100} Like the \textit{T7} gp4, the helicase ring in \textit{Taq} dnaB is stabilized through interactions between the linker in one subunit with a small $\alpha$-helical pocket found in the CTD domain of the adjacent subunit. Both \textit{E1} helicase and \textit{Taq} dnaB show the characteristic hexameric ring that is involved in unwinding of the dsDNA. The major difference is in the direction of unwinding, with \textit{E1} helicase unwinding in the 3’ to 5’ direction whereas \textit{Taq} dnaB unwinds in the 5’ to 3’
direction. The translocation mechanism proposed for \textit{Taq} dnaB helicase is similar to the 
coordinated escort mechanism proposed for \textit{E1} helicase and shown in Figure 6.4 and 
described in section 6.1.3.

The structure of \textit{Vc} dnaB and \textit{Yp} dnaB should be similar to that of \textit{Taq} dnaB and 
future studies on these proteins should involve crystallization trials of these proteins in 
presence of ssDNA and nucleotides. If at all crystals are obtained of this complex and 
X-ray diffraction obtained, then the structure can be solved using molecular replacement. 
It will be interesting to see the interactions of the ssDNA with the helicase and see if it 
interacts in the same way as seen for \textit{E1} helicase. \textit{Hp} dnaB structure may or may not be 
similar. Since there is a low sequence homology with \textit{Taq} dnaB, it will be interesting to 
see whether the two dnaB’s are structurally similar in spite of the low sequence 
homology.
REFERENCES


29. Raghunathan, S.; Ricard, C. S.; Lohman, T. M.; Waksman, G., Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength X-ray diffraction on the


58. Collins, B. K.; Tomanicek, S. J.; Lyamicheva, N.; Kaiser, M. W.; Mueser, T. C., A preliminary solubility screen used to improve crystallization trials: crystallization and
preliminary X-ray structure determination of Aeropyrum pernix flap endonuclease-1.


68. Lohman, T. M.; Bujalowski, W., Thermodynamic methods for model-independent
determination of equilibrium binding isotherms for protein-DNA interactions:

69. Schramke, V.; Luciano, P.; Brevet, V.; Guillot, S.; Corda, Y.; Longhese, M. P.;
Gilson, E.; Geli, V., RPA regulates telomerase action by providing Est1p access to

70. Arcus, V., OB-fold domains: a snapshot of the evolution of sequence, structure and

71. Johnson, V.; Dignam, D.; Mueser, T. C., Structural and Biophysical studies of Single-
Stranded DNA Binding protein from *Aeropyrum Pernix*. *manuscript in preparation*.


73. Holyoak, T.; Fenn, T. D.; Wilson, M. A.; Moulin, A. G.; Ringe, D.; Petsko, G. A.,
Malonate: a versatile cryoprotectant and stabilizing solution for salt-grown
2356-8.

74. Tomanicek, S. J.; Collins, B. K.; Kaiser, M. W.; Mueser, T. C., The crenarchaeal
*Aeropyrum pernix* flap endonuclease-1 (FEN-1) DNA repair enzyme 1.4 Å
resolution. *manuscript in preparation*.

75. Erzberger, J. P.; Berger, J. M., Evolutionary relationships and structural mechanisms

76. Singleton, M. R.; Dillingham, M. S.; Wigley, D. B., Structure and mechanism of


93. Stayton, C. L.; Dabovic, B.; Gulisano, M.; Gecz, J.; Broccoli, V.; Giovanazzi, S.; Bossolasco, M.; Monaco, L.; Rastan, S.; Boncinelli, E.; et al., Cloning and


Appendices

1). Examples of Single-stranded DNA binding Proteins in the Protein Data Bank (PDB)

<table>
<thead>
<tr>
<th></th>
<th>Organism</th>
<th>Protein</th>
<th>Resolution (highest)</th>
<th>PDB #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Herpes Simplex</em> Virus-1</td>
<td>SSB</td>
<td>3.0 Å</td>
<td>1URJ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bacteriophage <em>T7</em></td>
<td>GP32 core</td>
<td>2.2 Å</td>
<td>1GPC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GP 2.5</td>
<td>1.90 Å</td>
<td>1JE5</td>
</tr>
<tr>
<td>3</td>
<td><em>Thermus aquaticus</em></td>
<td>SSB</td>
<td>1.85 Å</td>
<td>2FXQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSB</td>
<td>2.10 Å</td>
<td>2IHE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mutant SSB</td>
<td>1.9 Å</td>
<td>2IHF</td>
</tr>
<tr>
<td>4</td>
<td><em>Thermus Thermophilus</em> HB8</td>
<td>SSB</td>
<td>1.96 Å</td>
<td>2CWA</td>
</tr>
<tr>
<td>5</td>
<td><em>Thermotoga maritima</em></td>
<td>SSB</td>
<td>2.60</td>
<td>1Z9F</td>
</tr>
<tr>
<td></td>
<td><strong>Organism</strong></td>
<td>Protein</td>
<td>Resolution (Å)</td>
<td>PDB Code</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------</td>
<td>---------</td>
<td>----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>6</td>
<td><em>D. radiodurans</em></td>
<td>SSB</td>
<td>1.80</td>
<td>1SE8</td>
</tr>
<tr>
<td>7</td>
<td><em>S. solfataricus</em></td>
<td>SSB</td>
<td>1.20</td>
<td>1O7I</td>
</tr>
<tr>
<td>8</td>
<td><em>Mycobacterium</em></td>
<td>SSB</td>
<td>2.15</td>
<td>1X3E</td>
</tr>
<tr>
<td></td>
<td><em>smegmatis</em></td>
<td>SSB</td>
<td>2.70</td>
<td>1X3F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSB</td>
<td>3.00</td>
<td>1X3G</td>
</tr>
<tr>
<td>9</td>
<td><em>Mycobacterium</em></td>
<td>SSB</td>
<td>2.50</td>
<td>1UE1</td>
</tr>
<tr>
<td></td>
<td><em>tuberculosis</em></td>
<td>SSB</td>
<td>2.60</td>
<td>1UE5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSB</td>
<td>2.70</td>
<td>1UE6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSB</td>
<td>3.20</td>
<td>1UE7</td>
</tr>
<tr>
<td>10</td>
<td><em>E. coli</em></td>
<td>SSB</td>
<td>2.70</td>
<td>1QVC</td>
</tr>
<tr>
<td></td>
<td>SSB 2.90 Å</td>
<td>1KAW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSB with C terminal 3.30 Å</td>
<td>1SRU</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSB + ssDNA 3.20 Å</td>
<td>1EQQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSB + 2 ssDNAs 2.80 Å</td>
<td>1EYG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>SSB 2.7 Å</td>
<td>2DUD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Sapiens (mitochondria)</td>
<td>SSB 2.4 Å</td>
<td>3ULL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSB 2.47 Å</td>
<td>1S3O</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RPA (Trimerization core) 2.80 Å</td>
<td>1L1O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>RPA (70s subunit)</td>
<td>1FGU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Sapiens</td>
<td>RPA (70s subunit) + ssDNA 2.40 Å</td>
<td>1JMC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RPA 14 + RPA 32 2.50 Å</td>
<td>1QUQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2). **Lambda Hind III restriction:**

2 µL λ DNA  
10 µL Hind III reaction buffer  
4 µL Hind III  
1 µL BSA  
83 µL water

Total: 100 µL @840 ng / µL  
~2 hrs at 37 °C  
~20 mins at 65 °C

Then add,

20 µL: 10X gel loading solution  
40 µL: 80% glycerol  
40 µL: water

Aliquot in to 12 µL vials and store at -20 °C.

3). **100 base pair ladder:**

2 µL: ladder  
2 µL: Gel loading solution  
2 µL: 80% glycerol  
4 µL: water

<table>
<thead>
<tr>
<th>Base Pairs (ng)</th>
<th>DNA Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,517</td>
<td>45</td>
</tr>
<tr>
<td>1,200</td>
<td>35</td>
</tr>
<tr>
<td>1,000</td>
<td>95</td>
</tr>
<tr>
<td>900</td>
<td>27</td>
</tr>
<tr>
<td>800</td>
<td>24</td>
</tr>
<tr>
<td>700</td>
<td>21</td>
</tr>
<tr>
<td>600</td>
<td>18</td>
</tr>
<tr>
<td>500/517</td>
<td>97</td>
</tr>
<tr>
<td>400</td>
<td>38</td>
</tr>
<tr>
<td>300</td>
<td>29</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
</tr>
<tr>
<td>100</td>
<td>48</td>
</tr>
</tbody>
</table>
4). For ultra competent XL10 cells:

Transformation buffer: 250 mL of autoclaved water

250 mM KCl : 4.66 g  
10 mM PIPES: 0.756 g  
15 mM CaCl₂ : 0.4162 g  

Adjust pH to 6.7 using 5 N KOH  
55 mM MnCl₂ : 2.72 g  
Filter sterilize. Store in 50 ml tubes at -20 °C  

Tips:  
1) Don’t use any stir bars. Mix with an autoclaved wooden stick  
2) Make fresh KOH each time  
3) Use new filter every time.

SOB medium: 800 mL

Bacto-tryptone 2% w/v: 16.0 g  
Yeast extract 0.5% w/v: 4.0 g  
10 M NaCl : 0.47 g  
2.5 mM KCl : 0.15 g  
10 mM MgSO₄ : 0.96 g  
10 mM MgCl₂ : 1.63 g  

Check pH : 6.7 – 7.0  
Autoclave for 20 minutes.  
Store separately.
5). For expression hosts:

Transformation buffer:

60 mM CaCl₂: 500 mL

CaCl₂ : 3.33 g
PIPES : 1.51 g
100% glycerol : 75 mL

1) 400 mL of water
2) add glycerol
3) add 3.33 g of CaCl₂
4) Add 1.51 g of PIPES
5) Adjust pH: 7.0 using 6 N NaOH
6) Make final volume to 500 mL
7) Filter sterilize
8) Store in 50 mL Fisher tubes at -20°C.

6). AGAROSE ELECTROPHORESIS

TAE Buffer - 500 mL - 50 X STOCK SOLUTION

121 g : Tris
28.5 g : Glacial acetic acid
18.6 g : Disodium ethylenediamine Tetracetate

Autoclaved MQ water

Filter sterilize

Divide it in to Aliquots of 20 mL each in 50 mL tubes

Store at -20°C
Working solution 1X TAE Buffer

10 mL + 490 mL of autoclaved water = 500 mL

Gel purification gels:

0.2 to 0.4 g of Sea Kem agarose + 20 mL to 40 mL of 1X TAE buffer

Standard gels:

0.4 g of Sea Kem agarose + 40 mL of 1X TAE buffer

7). PAGE ELECTROPHORESIS

Destain solution:

300 mL Methanol
600 mL RO water
100 mL Glacial acetic acid

Gel Dry solution:

300 mL Methanol
650 mL RO water
50 mL of glycerol

8). IPTG:

*1M Stock solution*

5 g in 21 mL of autoclaved MQ water

IPTG is generally used at 1 mL / 1 L cells
9). **ANTIBIOTICS:**

All the antibiotics are used at 1 ml/L of LB

Commonly used antibiotics in the lab are-

**KANAMYCIN:**

25 mg/mL
Dissolved in autoclaved MQ water.
Store in -20°C

**AMPICILLIN:**

45 mg/mL

1) 22.57 g + 500 mL of MQ autoclaved water
2) Add 300 µL of 4 M Dibasic potassium phosphate
3) Add 200 µL of 4 M Monobasic potassium phosphate
4) Take this in a small eppendorf tube and then shake it
5) Add to the ampicillin solution
6) Neutral pH

Store in -20°C.

**CHLORENPHENICOL:**

35 mg/mL
Dissolved in 70% Ethanol

**CARBENCILLIN:**

200 mg/mL
Dissolved in 70% Ethanol
10). **PREPARATION OF ULTRA-COMPETENT XL10 E.Coli CELLS**

1) Inoculate from glycerol stock in 5 mL SOB. Grow in the shaker at 28 °C overnight
2) In the morning take 120 µL into 50 mL SOB and keep in shaker at 37°C
3) Check the OD
4) If OD₆₀₀ is ~ 0.6, keep on ice for 10 minutes.
5) Distribute it in to 2, 50 mL tubes.
6) Centrifuge @ 2700 x g for 10 minutes at 4°C. Decant the supernatant.
7) Resuspend pellet gently in 16 mL of ice cold transformation buffer.
8) Keep on ice for 10 minutes
9) Spin at 2700 x g -10 minutes at 4°C
10) Add DMSO₄ to a final concentration of 7% (75 µL)
11) Place on ice for 10 minutes
12) Use immediately for transformation

11). **PREPARATION OF ULTRA-COMPETENT E.Coli CELLS – Expression Host**

1) Inoculate from glycerol stock in 5 mL LB. Grow in the shaker at 28°C overnight
2) In the morning take 250 µl into 100 mL LB and keep in shaker at 37°C for around 1.3 hrs.
3) Check OD₆₀₀ - should be around 0.2 to 0.3.
4) Keep it on ice for 10 minutes.
5) Distribute it in to 2, 50 mL tubes.
6) Centrifuge 2500 x g - 5 minutes at 4°C.
7) Decant the supernatant.
8) Resuspend pellet gently in tube 1 - 20 mL. of ice cold transformation buffer.
9) Resuspend pellet gently in tube 2 - 20 mL. of ice cold transformation buffer.
10) Transfer the mix into 1 tube.
11) Keep on ice for 20 minutes
12) Spin at 2500 x g for 5 minutes at 4°C
13) Decant supernatant.
14) Resuspend pellet very gently in 1 mL transformation buffer.
15) Keep on ice for 2 hours
16) Use immediately for transformation and aliquot remaining into eppendorf tubes and freeze on dry ice and store in -80°C
12). **General Lysis Buffer**

Used in the ratio of 10 mL to a gram of cells

- 100 mM Tris pH 8.0
- 10 mM EDTA
- 150 mM Ammonium Chloride
- 5 mM DTT
- 10 % Sucrose
- pinch of AEBSF / PMSF
- 0.3 % PEI

13). **Antibiotic resistance**

<table>
<thead>
<tr>
<th>Vectors</th>
<th>Antibiotic resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET 101 vector</td>
<td>Ampicillin (or Carbenicillin)</td>
</tr>
<tr>
<td>pDEST-C1</td>
<td>Streptomycin</td>
</tr>
<tr>
<td>pDEST-CS</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>pET21a</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>pET28a</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>pENTR-D</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>pDrive</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>pTrcHis2 TOPO</td>
<td>Ampicillin (or Carbenicillin)</td>
</tr>
<tr>
<td>pET45-DEST</td>
<td>Ampicillin (or Carbenicillin)</td>
</tr>
<tr>
<td>pKM596</td>
<td>Ampicillin (or Carbenicillin)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cloning hosts</th>
<th>Antibiotic resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP10</td>
<td>Streptomycin</td>
</tr>
<tr>
<td>XL10</td>
<td>none</td>
</tr>
<tr>
<td>DH5α</td>
<td>none</td>
</tr>
<tr>
<td>DB3.1</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>Omnimax</td>
<td>Tetracycline</td>
</tr>
<tr>
<td>Expression hosts</td>
<td>Antibiotic resistance</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BL21</td>
<td>none</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>none</td>
</tr>
<tr>
<td>BL21(DE3) pLysS</td>
<td>Chloramphenicol (important for overnight grow up at a final concentration of 50 µg/mL)</td>
</tr>
<tr>
<td>BL21(DE3) Star</td>
<td>none</td>
</tr>
<tr>
<td>BL21(DE3)Gold</td>
<td>none</td>
</tr>
<tr>
<td>BL21(DE3) RP</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>BL21(DE3) RIL</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>BL21(DE3)RILP</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>RosettaBlue DE3</td>
<td>Chloramphenicol + Tetracycline</td>
</tr>
<tr>
<td>RosettaBlue</td>
<td>Tetracycline</td>
</tr>
<tr>
<td>Rosetta2 (DE3) pLysS</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>T7 Express LacIQ</td>
<td>Tetracycline</td>
</tr>
</tbody>
</table>
14). Examples of crystal hits interpretation

<table>
<thead>
<tr>
<th>1) Letters</th>
<th>2) Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0 Clear</td>
</tr>
<tr>
<td>O</td>
<td>1 Phase separation</td>
</tr>
<tr>
<td>DP</td>
<td>2 Surface skin</td>
</tr>
<tr>
<td>P</td>
<td>3 Flocculent precipitate</td>
</tr>
<tr>
<td>LP</td>
<td>4 Spherulites</td>
</tr>
<tr>
<td>OP</td>
<td>5 Granular precipitate</td>
</tr>
<tr>
<td>XP</td>
<td>6 Microcrystals</td>
</tr>
<tr>
<td>Sp</td>
<td>7 Needles or Urchins (1-D crystals)</td>
</tr>
<tr>
<td>F</td>
<td>8 Plates (2-D crystals)</td>
</tr>
<tr>
<td>X</td>
<td>9 Small (&lt; 0.05 mm) 3-D crystals</td>
</tr>
<tr>
<td>μX</td>
<td>10 Medium (&lt; 0.2 mm) 3-D crystals</td>
</tr>
<tr>
<td>U</td>
<td>11 Large (&gt; 0.2 mm) 3-D crystals</td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
</tr>
<tr>
<td>Pl</td>
<td></td>
</tr>
</tbody>
</table>

Crystal description modifier

<table>
<thead>
<tr>
<th>Cl</th>
<th>Cluster*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tw</td>
<td>Twin</td>
</tr>
</tbody>
</table>