

Crystallographic studies of bacterial single-stranded
DNA-binding proteins

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem

Rat der Biologisch-Pharmazeutischen Fakultät

der Friedrich-Schiller-Universität Jena

von Master of Science Santosh Panjekar

geboren am 24.03.1973 in Banka (Indien)

Date of defense

17 Dec. 2001

Gutachter

1. Rolf Hilgenfeld, Friedrich-Schiller-University, Jena
2. Siegmund Reißmann, Friedrich-Schiller-University, Jena
3. Claus Urbanke, Medizinische Hochschule Hannover

Dedicated to my grandfather

Abstract

The architecture of bacterial single-stranded DNA binding proteins has been investigated by elucidating the structures of SSBs from *Brucella abortus*, *Proteus mirabilis* and *Serratia marsecens* by X-ray crystallographic methods. The structure of the SSB from *Escherichia coli* has been refined to a high resolution.

Two cryo-techniques: (i) repeated-annealing using dried paraffin oil and (ii) cryo-cooling using PanjellyTM were developed to overcome the problematic bottlenecks arising from poorly diffracting crystals and premature crystal decay arising from radiation damage.

The essential structure of SSBs across all species consists of five β -strands and one α -helix interspersed by four loops. It is similar to the already described Oligonucleotide/Oligosaccharide Binding (OB) fold. Out of the four loops, two loops involving residues 40 to 50 and 85 to 100 (*E. coli* numbering) are always involved in DNA-binding. The core regions comprising the entire regular secondary structure is almost identical between the different SSBs; the loop regions exhibit variable conformation. These loops are often involved in crystal packing giving rise to different combinations of loop-loop interactions. Different SSBs show identical head-to-head tetramer formation similar to the situation in human mitochondrial SSB. The principal role of the conserved His-55 residue in tetramerisation, as earlier demonstrated by biochemical experiments could be rationalised. The conserved Tyr-78 residue at the dimer-dimer interface may be also essential for tetramer formation. The conserved salt bridge between Lys-7 and Glu-80 at the same interface indicates its dominant role in stabilising the quaternary structure. Three inter hydrogen-bonded water molecules proximal to the ND1 atom of the conserved His-55 could be located in the well-resolved crystal structures; these water molecules may be functionally important. The DNA-binding aromatic residues (Trp-40, 54, 88, Phe-60) are sequentially, structurally and conformationally conserved corroborating their specific role in binding ssDNAs. Three-dimensional structural comparison of the homotetrameric SSBs along with other proteins containing the common OB-fold revealed that the fold of the monomeric SSB (Gene 32 protein) is more similar to the homotetrameric form than to the dimeric (Gene V protein) or heterotrimeric form (human replication protein 70).

Content	Page
1 Introduction	1
1.1 Historical perspective.....	1
1.2 Single-stranded DNA-binding protein (SSB) classes.....	2
1.2.1 Monomeric class.....	3
1.2.2 Dimeric class	5
1.2.3 Heterotrimeric class	6
1.2.4 Homotetrameric class	7
1.2.4.1 Human mitochondrial SSB.....	8
1.2.4.2 <i>E. coli</i> SSB (<i>EcoSSB</i>).....	9
1.2.4.2.1 Function of <i>EcoSSB</i>	9
1.2.4.2.1.1 Replication.....	9
1.2.4.2.1.2 Recombination.....	11
1.2.4.2.1.3 Repair.....	13
1.2.4.2.1.4 The C-terminal domain.....	14
1.2.4.2.2 Biochemical and biophysical properties of <i>EcoSSB</i> ...	15
1.2.4.2.3 Crystallisation of <i>EcoSSB</i>	15
1.2.4.2.4 Structure of <i>EcoSSB</i>	15
1.2.4.2.5 Interaction between <i>EcoSSB</i> and DNA.....	16
1.2.4.2.5.1 Multiple mode.....	16
1.2.4.2.5.2 Co-operativity.....	17
1.2.4.2.5.3 DNA-binding domain.....	17
1.2.4.2.5.4 Structure of the <i>EcoSSB</i> /ssDNA complex	18
1.2.4.3 Sequence comparison of bacterial SSBs.....	19
1.3 Trends in cryocrystallography.....	20
1.4 Aim of the work.....	21
2 Materials and Methods	22
2.1 Materials.....	22
2.1.1 Chemicals.....	22
2.1.2 Buffers and solutions.....	22
2.1.3 Protein and DNA samples.....	23
2.1.4 Instruments.....	23
2.2 Methods.....	24

Index

Content	Page
2.2.1 Purity determination.....	24
2.2.2 Crystallisation.....	24
2.2.3 Characterisation of protein crystals.....	24
2.2.3.1 Light microscope.....	24
2.2.3.2 Gel analysis.....	25
2.2.3.3 Staining with Izit™.....	25
2.2.4 Structure determination.....	25
2.2.4.1 Diffraction theory.....	25
2.2.4.2 X-ray diffraction data collection	28
2.2.4.2.1 X-ray sources.....	28
2.2.4.2.1.1 Rotating anode tubes.....	28
2.2.4.2.1.2 Synchrotron.....	28
2.2.4.2.2 Monochromators.....	29
2.2.4.2.3 Detectors.....	29
2.2.4.2.3.1 Imaging plate.....	29
2.2.4.2.3.2 CCD detector.....	30
2.2.4.2.4 Preparation of crystals.....	30
2.2.4.2.5 Preparation of dried paraffin oil and Panjelly™.....	31
2.2.4.2.6 Crystal storage and retrieval.....	31
2.2.4.2.7 Strategy.....	32
2.2.4.3 Data processing.....	33
2.2.4.3.1 Autoindexing.....	33
2.2.4.3.2 Intensity integration.....	33
2.2.4.3.3 Scaling	34
2.2.4.3.4 Quality of data.....	34
2.2.4.3.5 Calculation of structure factor amplitudes.....	35
2.2.4.4 Method of phase determination.....	35
2.2.4.4.1 Multiple isomorphous replacement (MIR).....	36
2.2.4.4.2 Multiple-wavelength anomalous dispersion (MAD).....	37
2.2.4.4.3 Molecular replacement (MR).....	38
2.2.4.5 Electron density and model building.....	41
2.2.4.5.1 (F _o -F _c)-map.....	41
2.2.4.5.2 (2F _o -F _c)-map.....	42

Index

Content	Page
2.2.4.5.3 Omit map.....	42
2.2.4.5.4 SigmaA weighted maps.....	42
2.2.4.6 Refinement of protein structures.....	42
2.2.4.6.1 Refinement theory.....	43
2.2.4.6.2 The free R-factor (R_{free}).....	44
2.2.4.6.3 The program CNS.....	45
2.2.4.6.4 The program REFMAC.....	46
2.2.4.6.5 Introduction of water molecules in the structure.....	46
2.2.4.7 Validation of the model.....	47
2.2.4.8 Structural analysis	47
3 Results and Discussion.....	48
3.1 Cryo-cooling techniques.....	48
3.1.1 Repeated-annealing using dried paraffin oil.....	48
3.1.2 Flash-cooling using Panjelly TM and single-step annealing with paraffin oil.....	52
3.2 Proteolysis of SSBs.....	55
3.3 Crystal structure of <i>EcoSSB</i>	56
3.3.1 Crystallisation of selenomethionine (SeMet) <i>EcoSSB</i>	56
3.3.2 MAD and native data collection.....	57
3.3.3 Refinement of the Webster- <i>EcoSSB</i> structure to high resolution.....	57
3.3.4 Quality of <i>EcoSSB</i> model.....	58
3.3.5 Comparison with the Webster-, Raghunathan- and Matsumoto- <i>EcoSSB</i> model.....	61
3.3.6 Overall structure of <i>EcoSSB</i>	62
3.3.7 Crystal packing of the <i>EcoSSB</i> structure.....	67
3.4 Crystal structure of <i>BabSSB</i>	69
3.4.1 Crystallisation of <i>BabSSB</i>	69
3.4.2 Native data collection.....	69
3.4.3 Structure determination of <i>BabSSB</i>	70
3.4.4 Refinement and model building of <i>BabSSB</i> structure.....	71
3.4.5 Quality of <i>BabSSB</i> model.....	72
3.4.6 Comparison between the <i>BabSSB</i> and <i>EcoSSB</i> model.....	73
3.4.7 Crystal packing of the <i>BabSSB</i> structure	74

Index

Content	Page
3.4.8 Model of <i>Bab</i> SSB-ssDNA complex.....	76
3.5 Crystal structure of <i>Pmi</i> SSB.....	78
3.5.1 Crystallisation of <i>Pmi</i> SSB-(dT) ₂	78
3.5.2 Native data collection.....	78
3.5.3 Structure determination of <i>Pmi</i> SSB.....	79
3.5.4 Refinement and model building of <i>Pmi</i> SSB structure.....	80
3.5.5 Quality of <i>Pmi</i> SSB model.....	80
3.5.6 Comparison between the <i>Pmi</i> SSB and <i>Eco</i> SSB model.....	81
3.5.7 The dithymidene (dT) ₂ electron density.....	83
3.5.8 Crystal packing of the <i>Pmi</i> SSB structure.....	84
3.6 Crystal structure of <i>Sma</i> SSB.....	86
3.6.1 Crystallisation of <i>Sma</i> SSB.....	86
3.6.2 Native data collection.....	86
3.6.3 Structure determination of <i>Sma</i> SSB.....	87
3.6.4 Refinement and model building of <i>Sma</i> SSB structure.....	89
3.6.5 Quality of <i>Sma</i> SSB model.....	89
3.6.6 Comparison between the <i>Sma</i> SSB and <i>Eco</i> SSB model.....	90
3.6.7 Crystal packing of the <i>Sma</i> SSB structure	93
3.7 Crystallisation of <i>Eco</i> SSB-(dT) ₃₅	96
3.8 Overall comparison between SSBs.....	98
3.8.1 Common axis.....	98
3.8.2 Comparison of SSB monomers.....	98
3.8.2.1 N-terminus.....	99
3.8.2.2 C-terminus.....	99
3.8.2.3 Helix.....	99
3.8.2.4 Protein core.....	99
3.8.3 Interface.....	100
3.8.3.1 Monomer-monomer interface.....	100
3.8.3.2 Dimer-dimer interface.....	101
3.8.4 Comparison of tetramers.....	105
3.8.5 Conservation of water molecules in SSB structures.....	106
3.8.6 Comparison of crystal packing of SSBs.....	107
3.9 Global comparison of OB-fold containing proteins.....	112

Index

	Content	Page
4	Summary	115
5	References	116
6	Appendix	128

List of figures

Figure		Page
1	Description of the OB-fold.....	2
2	Ribbon diagram of core gp32.....	3
3	Ribbon diagram of AdSSB.....	4
4	Ribbon diagram of gene V protein.....	5
5	Ribbon diagram of the RPA70/ssDNA complex.....	6
6	Ribbon diagram of RPA32/RPA14.....	7
7	Ribbon diagram of human mitochondrial SSB.....	8
8	Schematic representation of a replication fork.....	10
9	Kinetic steps of DNA strand exchange.....	12
10	Ribbon diagram of the <i>Eco</i> SSB/ssDNA complex.....	18
11	Sequence comparison of prokaryotic SSBs	19
12	Schematic diagram of a hanging drop setup.....	24
13	Bragg's Law.....	26
14	Line diagram representation of Ewald construction.....	27
15	Schematic representation of crystal storage/retrieval.....	32
16	Harker construction for protein phase determination.....	37
17	Energy dependence of the anomalous scattering factors $\Delta f'$ and $\Delta f''$	38
18	Diffraction pattern of a <i>Bab</i> SSB crystal.....	54
19	SDS-PAGE of SSB solutions and dissolved SSB crystals.....	56
20	SDS-PAGE of <i>Bab</i> SSB solution and dissolved <i>Bab</i> SSB crystals.....	56
21	SDS-PAGE of <i>Sma</i> SSB solution and dissolved <i>Sma</i> SSB crystals... ..	56
22	Monoclinic crystal of the <i>Eco</i> SSB.....	57
23	Ramachandran plot for the <i>Eco</i> SSB.....	61
24	Ribbon diagram of the <i>Eco</i> SSB B-monomer.....	62
25	Ribbon diagram of the <i>Eco</i> SSB CD-dimer.....	63
26	Schematic diagram of the hydrogen-bonding network at the monomer-monomer interface of <i>Eco</i> SSB.....	64
27	Ribbon diagram of two types of the <i>Eco</i> SSB tetramer.....	66
28	Crystal packing of the <i>Eco</i> SSB.....	68
29	Tetragonal crystal of <i>Bab</i> SSB.....	69
30	Ramachandran plot for <i>Bab</i> SSB.....	72
31	Superposed C_{α} trace of the <i>Bab</i> SSB and <i>Eco</i> SSB tetramers.....	74
32	Crystal packing of the <i>Bab</i> SSB	76
33	Theoretical model of the <i>Bab</i> SSB-ssDNA complex.....	77
34	Orthorhombic crystal of <i>Pmi</i> SSB-(dT) ₂	78
35	Ramachandran plot for <i>Pmi</i> SSB.....	81
36	Ribbon representation of the octamer of <i>Pmi</i> SSB.....	82
37	Superposition of the C_{α} atoms of the <i>Eco</i> SSB-ABCD, the <i>Pmi</i> SSB-ABCD and the <i>Pmi</i> SSB-EFGH tetramers.....	83
38	Ball-and-stick model of Trp-88 and Lys-87 from G-monomer and symmetry mate of F-monomer with (F _o -F _c)-map at 1.5 σ for the nucleotides.....	84
39	Crystal packing of the <i>Pmi</i> SSB.....	85
40	Tetragonal crystals of <i>Sma</i> SSB.....	86
41	Ramachandran plot for <i>Sma</i> SSB.....	90
42	C_{α} trace of non-crystallographic and crystallographic tetramers of <i>Sma</i> SSB.....	91
43	R.m.s. deviations for the C_{α} atoms in the <i>Sma</i> SSB monomers.....	92
44	Superposition of the C_{α} traces of the non-crystallographic <i>Eco</i> SSB,	

Index

Figure		Page
	<i>Sma</i> SSB and crystallographic <i>Sma</i> SSB	93
45	Crystal packing of the <i>Sma</i> SSB.....	94
46	Crystal packing of crystallographic and non-crystallographic <i>Sma</i> SSB tetramers along the z-axis.....	95
47	Superposition of 21 monomers of SSBs.....	98
48	Superposition of the C α atom traces from the SSBs core.....	100
49	Partial view of the tetramer interface of SSBs embedding residues (e.g. Tyr-78) from β -strand IV.....	101
50	The electrostatic surface potential of the residues constituting the interfacial region.....	102
51	Diagram showing residues with different side chain conformations..	104
52	Superposition of SSB-tetramers.....	105
53	Diagram showing hydrogen-bonding interactions between various loop regions in bacterial species.....	108
54	The electrostatic surface potential of the residues of SSB tetramers..	111
55	Dendrogram of the OB-fold containing proteins.....	113
A1	Washing procedure of protein crystal in dried paraffin oil.....	130
A2	B-factor plot of <i>Eco</i> SSB, <i>Bab</i> SSB and <i>Hsmt</i> SSB monomers.....	153
A3	B-factor plot of <i>Pmi</i> SSB monomers.....	154
A4	B-factor plot of <i>Sma</i> SSB monomers.....	155

List of tables

Table		Page
1	Structural classes of SSBs.....	2
2	Statistics on <i>Eco</i> SSB structures.....	16
3	Chemical items and their manufacturers.....	22
4	Buffer, protein storage and crystallisation solutions and oil.....	22
5	Instruments and their manufacturers.....	23
6	Rotation range for SSB crystals in different crystal classes.....	33
7	MAD data statistics.....	59
8	Native data statistics.....	59
9	Refinement statistics of SSBs.....	60
10	Hydrogen-bonding distances between various donor-acceptors atoms as shown in Figure 26.....	65
11	The molecular replacement procedure for <i>Bab</i> SSB.....	70
12	Salt bridges in the <i>Bab</i> SSB structure.....	74
13	The molecular replacement procedure for <i>Pmi</i> SSB.....	79
14	The molecular replacement procedure for <i>Sma</i> SSB.....	88
15	SSB crystal forms.....	98
16	Nomenclature of loops.....	107
17	Hydrophobic moment of SSBs.....	110
A1	Protein crystals used to study the effects of repeated-annealing in dried paraffin oil.....	128
A2	Comparison of diffraction properties of crystals of seven different proteins at 293 K, after flash-cooling and after repeated-annealing in dried paraffin oil....	129
A3	Comparison of diffraction properties of crystals of nine different proteins after flash-cooling using dried paraffin oil and in Panjelly™.....	129
A4	R.m.s. deviations between monomers of SSBs.....	131
A5	R.m.s. deviation between SSB-tetramers.....	132
A6	Conserved water molecules, their B-factors and r.m.s. deviations.....	133
A7	Dimer-dimer interfaces of SSB-tetramers.....	137
A8	Buried surface area of individual residues in each monomer upon tetramer formation.....	137
A9	Distance between the side chain oxygen atoms from the residues mutually pointing towards each other from two monomers at the dimer-dimer interface (shown in Figure 49).....	140
A10	Monomer-monomer interface of SSB-dimers and their extent of burial.....	141
A11	Buried surface area of individual residues in each monomer upon dimer formation.....	141
A12	Hydrophobic residues which belong to the core of SSBs.....	145
A13	Symmetry operators.....	146
A14	Intermolecular hydrogen-bonding contacts of <i>Eco</i> SSB.....	147
A15	Intermolecular hydrogen-bonding contacts of <i>Bab</i> SSB.....	148
A16	Intermolecular hydrogen-bonding contacts of <i>Pmi</i> SSB.....	149
A17	Intermolecular hydrogen-bonding contacts of <i>Sma</i> SSB.....	151
A18	Intermolecular hydrogen-bonding contacts of <i>Hsmt</i> SSB.....	152
A19	Comparison of OB-fold containing proteins.....	156

Abbreviations	
DNA	Deoxyribonucleic acid
ssDNA	Single-stranded DNA
SSB	Single-stranded DNA-binding protein
<i>ssb</i>	SSB gene
<i>ssb-1</i>	Thermosensible SSB-1 gene from <i>E. coli</i>
<i>ssb-113</i>	Thermosensible SSB-113 gene from <i>E. coli</i>
SSB-1	Thermosensible SSB from <i>E. coli ssb-1</i> mutants
SSB-113	Thermosensible SSB from <i>E. coli ssb-113</i> mutants
oriC	Starting point of chromosome replication
OB-fold	oligonucleotide/oligosaccharide - binding fold
<i>EcoSSB</i>	SSB from <i>Escherichia coli</i>
<i>PmiSSB</i>	SSB from <i>Proteus mirabilis</i>
<i>BabSSB</i>	SSB from <i>Brucella abortus</i>
<i>SmaSSB</i>	SSB from <i>Serratia marcescens</i>
SDS	Sodiumdodecylsulfat
HEPES	2-[4-(2-Hydroxyethyl)-1-piperazinyl] ethanesulphonic acid
M	Mole(s)/litre
MES	2-Morpholinoethanesulphonic acid
ml	Millilitre
mM	Millimolar
PAGE	Polyacrylamide gel electrophoresis
PEG 400	Polyethylene glycol 400
PEG 4000	Polyethylene glycol 4000
Cacodylate	Cacodylic acid sodium salt trihydrate
DTT	1,4-Dithiotheritol
Tris	Trishydroxymethylaminomethane
EDTA	Ethylene dinitrilotetraacetic acid disodium salt dihydrate
B-factor	Isotropic temperature factor
l	Litre
µl	Microlitre
v/v	Volume/volume
w/v	Weight/volume
V_m	Matthews coefficient

1 Introduction

Single-stranded DNA binding proteins (SSBs) constitute a class of polypeptides that preferentially bind single-stranded DNA (ssDNA) with high affinity and without apparent nucleotide sequence specificity. Most organisms, from viruses to vertebrates, encode their own SSBs. SSBs play a vital role in the cell by participating in the biological processes of DNA replication, recombination and repair (Alberts *et al.*, 1970; Wobbe, 1987; Meyer *et al.*, 1990). They prevent the formation of secondary structures of ssDNA and/or facilitate the actions of enzymes involved in these biological processes, such as DNA polymerase. SSBs are essential for the cell, deletion of the *ssb* gene may cause cell death (Porter *et al.*, 1990).

1.1 Historical perspective

The studies of DNA-binding proteins date back to the development of DNA-cellulose affinity chromatography in the laboratory of Bruce Alberts in the late 1960s (Alberts *et al.*, 1968, 1970, 1971). This led to the discovery of the gene 32 protein of bacteriophage T4 (Alberts *et al.*, 1970), which binds to DNA-cellulose very tightly. 2 M NaCl are required to elute the protein from this column. Gene 32 protein also binds strongly, selectively and co-operatively to ssDNA (Alberts *et al.*, 1970; Deluis *et al.*, 1972). Shortly thereafter, Alberts and others, discovered an analogous protein in infected *E. coli* cells (Sigal *et al.*, 1972). They named this protein the *E. coli* DNA-unwinding protein because of its ability to unwind (i.e. destabilise) a DNA double helix (Sigal *et al.*, 1972).

The SSBs have been referred to by a number of different names since their discovery in 1970, such as DNA-unwinding protein (Molineux *et al.*, 1974; Sigal *et al.*, 1972), DNA-binding protein (Mackay *et al.*, 1976; Molineux *et al.*, 1975), DNA-binding protein I (Geider *et al.*, 1978), DNA-melting protein, DNA-extending protein, and helix-destabilising protein (Alberts *et al.*, 1977). None of these names appropriately conveys the basic features of these proteins, for example, neither gene 32 nor *E. coli* SSB protein can unwind native duplex DNA (dsDNA) by itself. The name SSB is used most commonly, but this term is still not specific enough and tells too little about the functions of these proteins. RNA polymerase, RecA protein and lactate dehydrogenase also bind to ssDNA to various extents, but they are not the type of proteins which will be discussed here. The name SSB was chosen to reflect the fact that SSBs also bind to RNA (Molineux *et al.*, 1975; Ruyechan *et al.*, 1976; Weiner *et al.*, 1975) but with far less affinity than to DNA.

1.2 Single-stranded DNA-binding protein (SSB) classes

According to the above viewpoint, SSBs can be regarded as one big family of proteins, which share common functional and mechanistic features. To date, the known SSB proteins can be classified into at least four different classes: the monomeric class, the dimeric class, the heterotrimeric class, and the homotetrameric class (Table 1). These classes have little in common except their ability to bind to ssDNA (Suck, 1997). Structural information is now available for at least one member of each of the four classes of SSB (Table 1). There is no sequence homology between the proteins of any two classes. However, within the same class, sequence identities between SSBs are significant.

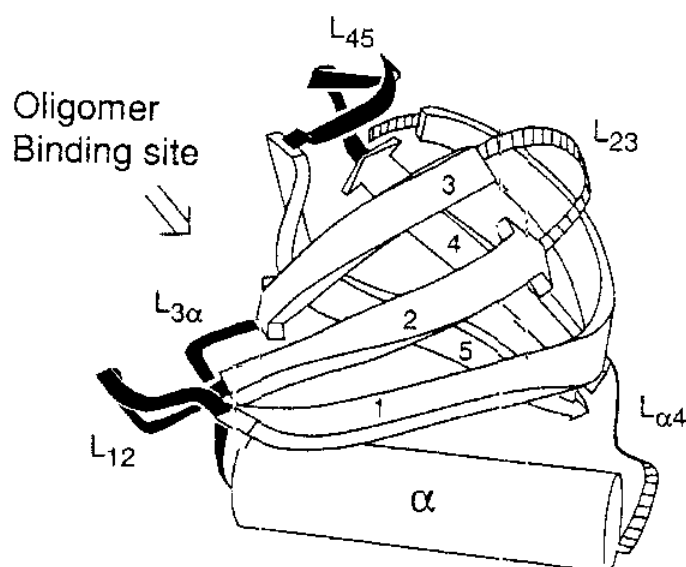


Figure 1. Description of the OB-fold. Five β -strands (numbered arrows) form a closed β -sheet (β -barrel), capped by an α -helix (cylinder). The loops are shown as ribbons, connecting the structural segments and numbered accordingly. Three variable loops (black) contribute residues to the oligomer(s) binding site. The picture is adapted from Murzin (1993).

Table 1 Structural classes of SSB

Class	SSB	PDB [‡] entry code
Monomeric	T7 gene 32	1GPC
	Adenovirus SSB	1ADT
Dimeric	Filamentous Pf3 SSB	1PSF
	Gene 5 protein	1GVP
Heterotrimeric	Human cellular SSB (RPA)	1JMC
Homotetrameric	Human mitochondrial SSB	3ULL
	<i>E. coli</i> SSB	This work
	<i>Proteus mirabilis</i> SSB	This work
	<i>Serratia marcescens</i> SSB	This work
	<i>Brucella abortus</i> SSB	This work

[‡] PDB: Protein data bank (Berman *et al.*, 2000).

At first glance, common structural features are not easily detected if one compares the three-dimensional structures of SSBs from the four different classes. Indeed, the structures are very different. If, however, one looks more closely at the individual domains involved in ssDNA-binding and compares their topology, a common fold becomes apparent. This fold is known as oligonucleotide/oligosaccharide (OB) fold (Murzin, 1993). It is characterised by a five-stranded Greek Key β -barrel capped on one side by an α -helix (Figure 1).

1.2.1 Monomeric class

This includes the bacteriophage T7 gene 32 protein (Shamoo *et al.*, 1995) and the adenovirus SSB (AdSSB) (Tucker *et al.*, 1994).

T7 gene 32

The T7 gene 32 protein (gp32) is a monomer of 33 kDa that binds ssDNA co-operatively, with a binding site size of 8 nucleotides (Giedroc *et al.*, 1987; Prigodich *et al.*, 1984).

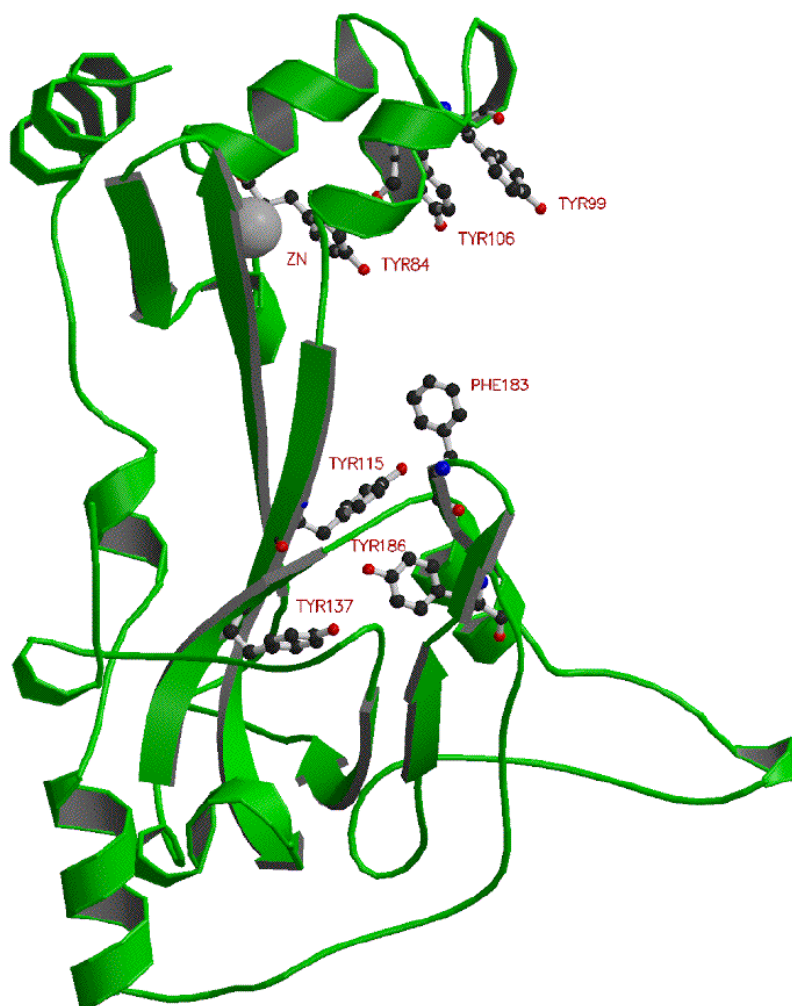


Figure 2 Ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) of core gp32 (PDB code: 1GPC) showing the structure and aromatic residues which are involved in ssDNA-binding (Shamoo *et al.*, 1995).

Introduction

The structure of the gp32 DNA-binding domain complexed to ssDNA has been determined by X-ray crystallography at 2.2 Å resolution and shown to contain a hydrophobic pocket composed of five β strands that contact the ssDNA bases and an electropositive cleft that is close to the phosphate backbone (Shamoo *et al.*, 1995). In the PDB (Berman *et al.*, 2000) file of this structure (PDB code: 1GPC), ssDNA molecules are not included because of their weak electron density. A ribbon diagram showing the complete fold of native gp32 and the residues involved in ssDNA-binding (Tyr-84, 99, 106, 115, 137, 186, Phe-183) (Shamoo *et al.*, 1988,1989) as well as the zinc ion is depicted in Figure 2.

Adenovirus SSB

The adenovirus 5 protein (Ad5) is made of 529 amino acids (Kruijer *et al.*, 1982). It has been shown by limited chymotrypsin digestion that this protein can be divided into two domains (Tsernoglou *et al.*, 1985). The N-terminal domain (1-173) contains the nuclear location signal (Morin *et al.*, 1989) and is involved in determining the host range (Klessig & Grodzicker, 1979).

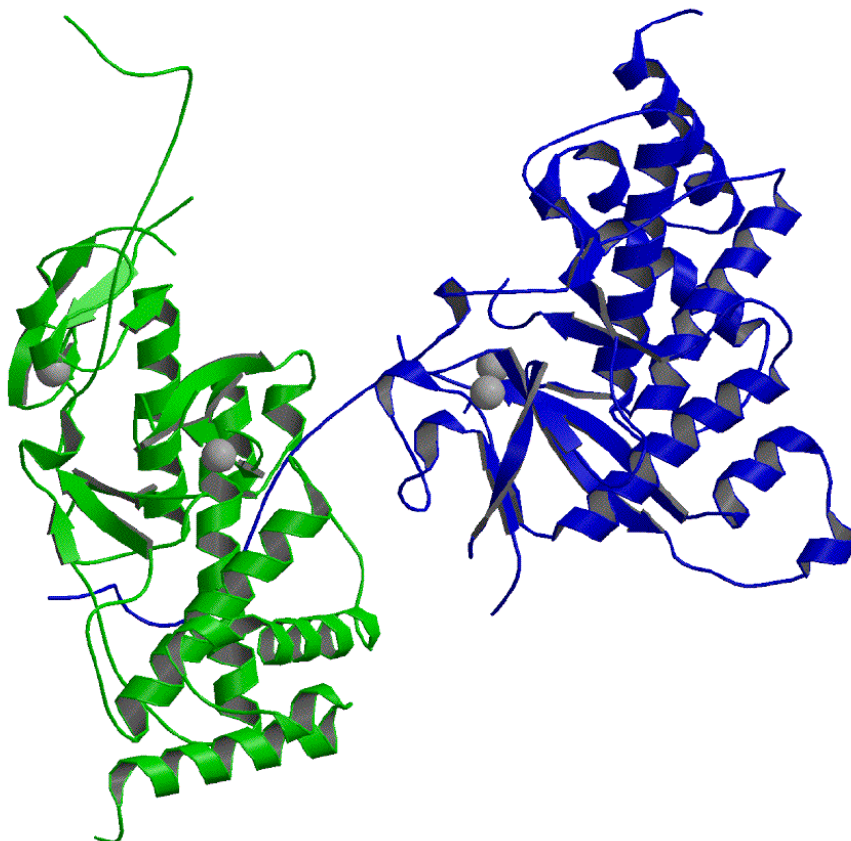


Figure 3 Ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) of AdSSB (PDB code: 1ADT). The interaction of the C-terminal hook of one molecule, in blue, with the adjacent molecule in green shows that the hook passing between two helices (Tucker *et al.*, 1994).

The C-terminal domain (174-529) contains the nucleic acid binding properties. The domain is functional in DNA replication in a manner similar to the intact molecule (Tsernoglou *et al.*,

1985). Ad5 binds to ssDNA, dsDNA, as well as to RNA in a sequence-dependent manner. The crystal structure of the nucleic acid binding domain of this protein has been solved (Tucker *et al.*, 1994). The structure (Figure 3) shows that the protein contains a 17 amino acid C-terminal extension which hooks onto a second molecule (green), thereby forming an arm. Deletion of this C-terminal arm reduces co-operativity in DNA-binding, suggesting a hook-on model for co-operativity.

1.2.2 Dimeric class

This class includes SSBs from the filamentous phages f1, fd, M13 and Pf3 (Skinner *et al.*, 1994; Brayer *et al.*, 1983; Folkers *et al.*, 1994; Folmer *et al.*, 1995). The filamentous phages f1, fd and M13 are collectively called Ff. These Ff viruses are almost identical in sequence and behaviour (Model & Russel, 1988).

The gene V protein (GVP) encoded by the filamentous bacteriophage Ff has 87 amino acids per monomer with a molecular weight of 9.7 kDa. It exists as a dimer in solution, and binds specifically to DNA in a highly co-operative manner (Fulford & Model, 1988). The complete three-dimensional structure of GVP is known; both its solution (Folkers *et al.*, 1994) and crystal (Skinner *et al.*, 1994) structures have been determined. The two aromatic residues, Tyr-26 and Phe-73 are involved in DNA binding as indicated by binding experiments of spin-labelled oligonucleotides to wild type GVP (van Duynhoven, 1992) and also from the analysis of structural data (Folkers *et al.*, 1994; Skinner *et al.*, 1994).

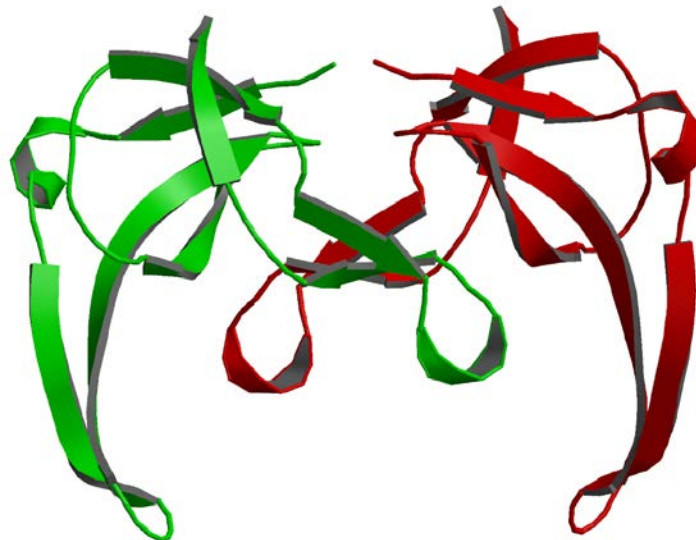


Figure 4 Ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) of the gene V protein dimer structure (PDB code: 1GVP). One monomer is shown in green and the other in red (Skinner *et al.*, 1994).

Ff GVP contains a five-stranded antiparallel β -sheet from which three β -hairpins protrude (Figure 4).

1.2.3 Hetrotrimeric class

This class is represented by the Replication protein A (RPA), the eukaryotic SSB. RPA is a heterotrimer composed of subunits of molecular weight 70, 32 and 14 kDa. The central 250 amino acids of the large subunit (RPA70) contain two tandem single-stranded DNA-binding domains (Bochkarev *et al.*, 1997; Gomes *et al.*, 1995, 1996; Lin *et al.*, 1996). The N-terminal region of RPA70 mediates interactions with a variety of cellular proteins including other replication proteins and certain transcription factors (Wold, 1997). The C-terminal region of RPA70 interacts with RPA32 and is required for assembly of the trimeric complex (Gomes *et al.*, 1995, 1996; Lin *et al.*, 1996). The middle subunit of RPA (RPA32) contains an additional single-stranded DNA binding domain that is likely to be similar in structure to those present in RPA70 (Philipova *et al.*, 1996; Bochkareva *et al.*, 1998).

A crystal structure of the two DNA-binding domains of RPA70 bound to a single-stranded octanucleotide at 2.4 Å resolution (Figure 5) has been reported (Bochkarev *et al.*, 1997). Each domain is composed of an OB fold (Murzin, 1993) and contains a channel that binds three nucleotides in the DNA chain. The DNA-binding channels of the two domains are oriented so that the ssDNA extends from one domain to the other in roughly a straight line, with two nucleotides bridging the space in between.

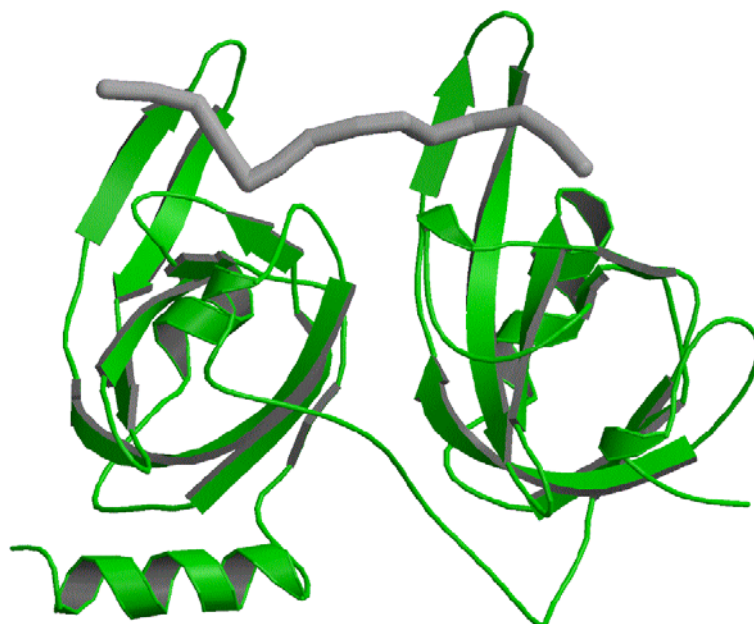


Figure 5 Ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) of the RPA70/ ssDNA complex (PDB code: 1JMC). RPA70 is shown in green and the ssDNA in grey (Bochkarev *et al.*, 1997).

Both domains interact with the DNA via hydrogen bonds to the bases and the phosphate backbone, as well as via stacking interactions between conserved aromatic residues and the bases.

Introduction

The complex of the two smaller subunits RPA32 and RPA14 has weak DNA-binding activity but the mechanism of DNA binding is unknown. The crystal structure of the proteolytic core of RPA32 and RPA14 (Figure 6), which consists of the central two-third of RPA32 (residues 45-170) and the entire RPA14 subunit, has been solved to 2.5 Å (Bochkarev *et al.*, 1999). The structure revealed that RPA14 and the central part of RPA32 are structural homologues. Each subunit contains an OB fold domain, which also resembles the DNA-binding domain in RPA70 and an N-terminal extension that interacts with the central OB-fold domain and a C-terminal helix that mediate heterodimerisation via a helix-helix interaction (Figure 6). The OB-fold of RPA32, but not that of RPA14, possesses additional similarity (i.e. conservation of aromatic residues) to the RPA70, supporting a DNA-binding role for RPA32.



Figure 6 Structure of RPA32/RPA14 (PDB code: 1QUQ). A ribbon representation (prepared using MOLSCRIPT; Kraulis, 1991) of the RPA14 (in red) and RPA32 (in green) (residues 45-170) is shown (Bochkarev *et al.*, 1999).

1.2.4 Homotetrameric class

This class includes the human mitochondrial SSB (Curth *et al.*, 1994), bacterial SSB from *E. coli* (Meyer *et al.*, 1990), *Serratia marcescens* (de Vries *et al.*, 1993), *Proteus mirabilis* (de Vries *et al.*, 1994), and *Brucella abortus* (Zhu *et al.*, 1993).

1.2.4.1 Human mitochondrial SSB (HsmtSSB)

The HsmtSSB functions as a tetramer. Each tetramer can bind 50-70 nucleotides with high affinity. It is also involved in DNA replication, repair and recombination. HsmtSSB exhibits

36% sequence identity and 50% similarity to *Eco*SSB. The crystal structure of HsmtSSB has been solved to 2.4 Å resolution (Yang *et al.*, 1997). The tetramer is D₂-symmetric and is formed by two dimers interacting head-to-head (Figure 7). The OB-fold of individual monomeric molecules has one α-helical segment and two β-pleated sheets. The front face of the barrel consists of three antiparallel strands and the back face of the barrel consists of four antiparallel strands. Between the antiparallel strands, three large β-hairpin loops are formed. Some residues in the N and C-terminal segments of the polypeptide chain (residues 1-9, 125-132) and one of the loop regions (residues 52-60) are disordered. The dimer is stabilised by two intermolecular antiparallel strands which are part of a large β-pleated sheet formed by six antiparallel strands. His 69 (which is equivalent to His 55 in *Eco*SSB) has been shown to be involved in tetramerisation (Curth *et al.*, 1991). It has also been shown that upon ssDNA binding the fluorescence of the exposed Trp-49 and Trp-68 in HsmtSSB protein, which are also conserved in *Eco*SSB, is drastically quenched (Curth *et al.*, 1994).

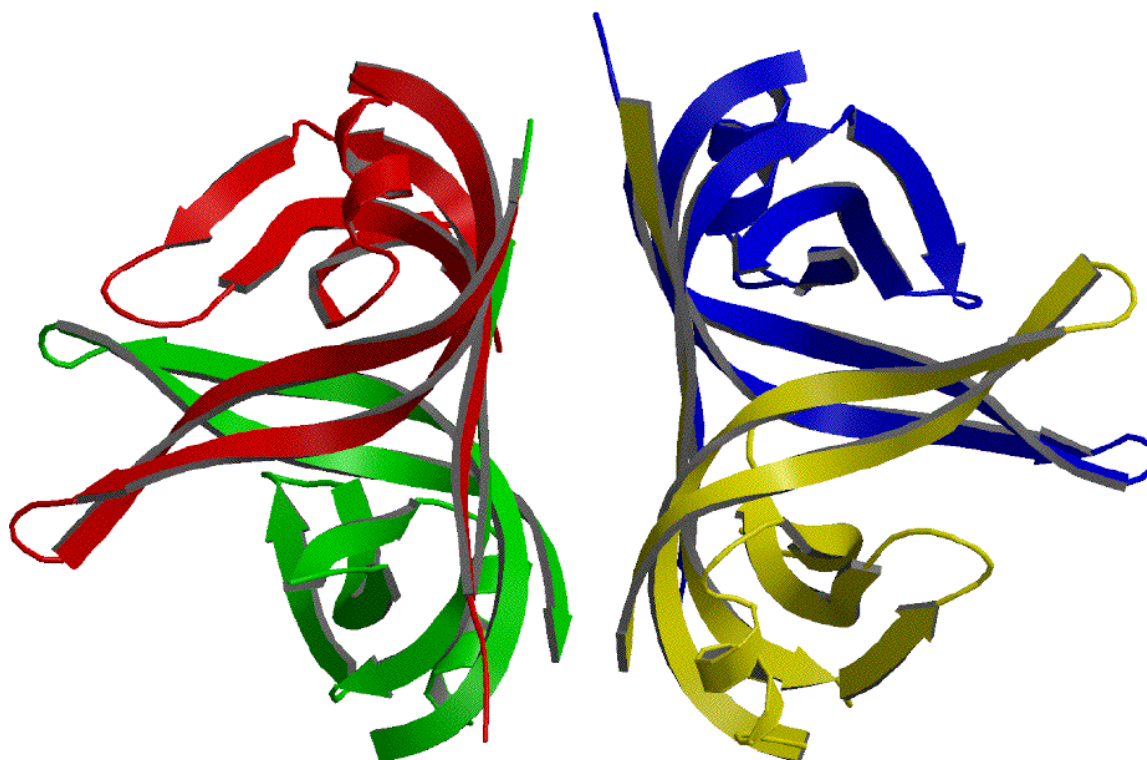


Figure 7 Ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) illustrating the position of the four HsmtSSB molecules (PDB code: 3ULL) found in two asymmetric units. Two molecules depicted in yellow and blue colour are related by a non-crystallographic pseudo two-fold axis, as are the red and green coloured molecules (Yang *et al.*, 1997).

1.2.4.2 EcoSSB

1.2.4.2.1 Function of EcoSSB

EcoSSB plays a vital role in various aspects of DNA metabolism and as a result, *ssb* mutations are generally pleiotropic (Chase *et al.*, 1986). It is essential for DNA replication (Glassberg *et al.*, 1979; Vales *et al.*, 1980; Meyer *et al.*, 1979), some repair mechanisms (Vales *et al.*, 1980; Meyer *et al.*, 1979; Whittier *et al.*, 1983) and it also plays a role in recombination (Glassberg *et al.*, 1979; Vales *et al.*, 1980; Golub *et al.*, 1983; Whittier *et al.*, 1981). Estimates of the number of SSB tetramers per *E. coli* cell range from 1000 to 2000, which is more than enough than what stoichiometric binding of SSB to transient ssDNA would require at the replication fork (Weiner *et al.*, 1975; Bobst *et al.*, 1985).

1.2.4.2.1.1 Replication

SSB is required for DNA replication. *In vitro* DNA replication systems were established in 1970s by using ssDNA phage templates (ϕ X174, M13, fd, f1, G4, α 3, ST-1 and ϕ K). From the beginning, SSB has always been a standard component of assays (Benz *et al.*, 1980; Geider *et al.*, 1974; Schekman *et al.*, 1975; Wickner *et al.*, 1974). For a long time, there was no direct evidence that SSB was essential *in vivo*. Even under certain conditions, replication *in vitro* could proceed in the absence of SSB. The essential role of SSB in DNA replication was established after the discovery of an *EcoSSB* mutation (Meyer *et al.*, 1979). Strains carrying a mutated *EcoSSB* gene (*ssb-1*, *ssb-113*) were temperature sensitive with respect to DNA synthesis *in vivo*. Their replications were terminated in 1 or 2 minutes at 42.5° C. More direct evidence was then accumulated, for instance that the deletion of the *ssb* gene in *E. coli* is lethal and genomic *ssb⁻* strains have to be complemented with a plasmid encoded SSB (Porter *et al.*, 1990).

There are four basic components of a replication fork (Figure 8): (i) the DNA polymerase is required for nascent strand synthesis, (ii) DNA helicase is involved in unwinding the parental duplex DNA, (iii) primase is needed to initiate Okazaki fragment synthesis, and (iv) SSB is required to coat exposed template ssDNA.

The origin of *E. coli* chromosomal replication (*oriC*) has been cloned on plasmid and used as a substrate to determine the events involved in initiation (Friedberg *et al.*, 1985; Fuller *et al.*, 1981). Initiation first requires the binding of DnaA (Fuller *et al.*, 1984). DnaC complexes with DnaB and delivers it to the initiation complex (Funnell *et al.*, 1987), with the subsequent release of DnaC (Wahle *et al.*, 1989). The initiation complex is stabilised by SSB. The template is unwound by the helicase DnaB, DNA primase and DNA gyrase in the

presence of SSB (Baker *et al.*, 1986, 1987) in preparation for the priming event (van der Ende *et al.*, 1985).

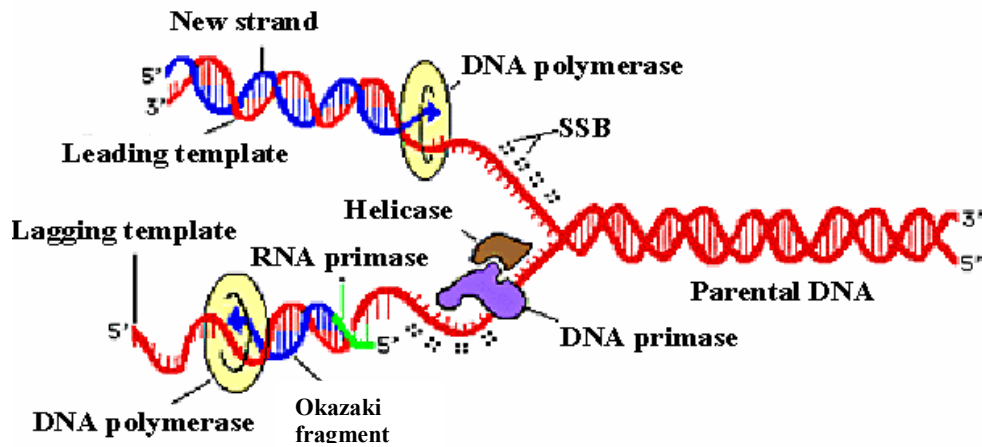


Figure 8 Schematic representation of a replication fork adapted from the website (www.accessexcellence.org/AB/GG/collaboration.html).

The **Elongation** step of DNA replication is followed by the unwinding of the parental DNA. The helicase DnaB, unwinds the parental (duplex) DNA by translocating along the lagging strand template in the 5' to 3' direction (Figure 8). The separated DNA strands behind advancing helicase are prevented from the re-annealing by the binding of SSB (LeBowitz & McMacken 1986). A fundamental asymmetry is introduced to the enzymatic requirement for DNA synthesis at the fork because of the antiparallel nature of the template strands and the fact that DNA polymerases synthesize DNA in the 5' to 3'-direction. Only one strand, the leading strand can be made in a continuous fashion. The replication fork appears to require highly processive DNA polymerase for the leading strand. Therefore, the DNA polymerase requires SSB for its high processivity and may be capable of synthesizing the leading strand with a single binding event if SSB is present (McHenry, 1988).

The other strand, the lagging strand is synthesized discontinuously, leading to generation of ssDNA in small pieces of roughly 2 kb in length (Okazaki fragments). At the lagging strand, DNA polymerase cannot initiate DNA synthesis *de novo*. The ssDNA is coated by SSB, primase acts upon this strand to form the RNA-primed sites that are then utilized by the lagging-strand polymerase for Okazaki fragment synthesis (Kelman *et al.*, 1998). In *E. coli*, new Okazaki fragment synthesis must be initiated once every 1 or 2 sec. To ensure that the primase has already access to the template when and where it is needed, its association with the replication fork is mediated by the DNA helicase acting on the lagging strand template to unwind the parental duplex in the 5'- to 3'-direction. This mobile complex of helicase and primase has been termed a primosome (Kornberg *et al.*, 1992; Marians, 1992). In *E. coli*, the

components of the primosome at the replication fork are still in question. It is clear that the primosome must include DnaB, the replication fork helicase (LeBowitz & McMacken, 1986), DnaG, the primase (Bouche *et al.*, 1975) and that it requires DnaC for function. It is also considered that the primosome contains four other proteins, DnaT, PriA, PriB, and PriC (Allen & Kornberg, 1993). The essential prerequisite for the formation of the primosomes is the coverage of ssDNA with SSB.

In summary, SSB plays a central and varied role in DNA replication: (i) it helps to organise and stabilize replication origins, (ii) it enhances helix destabilization by helicases, (iii) it prevents re-annealing of the single strands and protects against single-strand nuclease digestion, (iv) it ensures the specificity of priming, (v) it is required for primosome assembly, (vi) it enhances the fidelity of the DNA polymerase, (vii) it enhances the processivity of the polymerase by destabilising secondary structure that could cause polymerase pausing and dissociation, and (viii) it may promote binding of the polymerase to the template.

1.2.4.2.1.2 Recombination

DNA strand exchange catalysed by RecA occurs by a number of kinetically distinct phases that can be subdivided into at least three experimentally distinguishable steps: (i) presynapsis, (ii) synapsis and (iii) branch migration (Figure 9).

Presynapsis

The ordered assembly of the RecA protein homologue on ssDNA to produce a nucleoprotein complex that is the active species in DNA strand exchange is termed presynapsis. This process occurs by either of the two pathways: first, SSB binds to ssDNA (Figure 9, step 1A); this is followed by the binding of the RecA and the displacement of SSB to form the presynaptic filament (Figure 9, steps 1B and 1C, left panel). In the second pathway, RecA binds to the ssDNA substrate, forming an incomplete nucleoprotein filament due to limitations imposed by DNA secondary structure (Figure 9, step 1A, right panel). SSB then binds and removes the secondary structure and is then displaced by binding of further RecA molecules (Figure 9, steps 1B and C). The net result of either of these two pathways is the formation of a complete nucleoprotein filament that involves direct contact by the primary DNA-binding site of the strand exchange protein and results in the generation of a contiguous secondary DNA-binding site that is essential for homologous pairing and DNA strand exchange (Mazin *et al.*, 1996; 1997). SSB affects the binding of RecA protein to ssDNA, greatly stimulating the DNA strand exchange process (Kowalczykowski *et al.*, 1987).

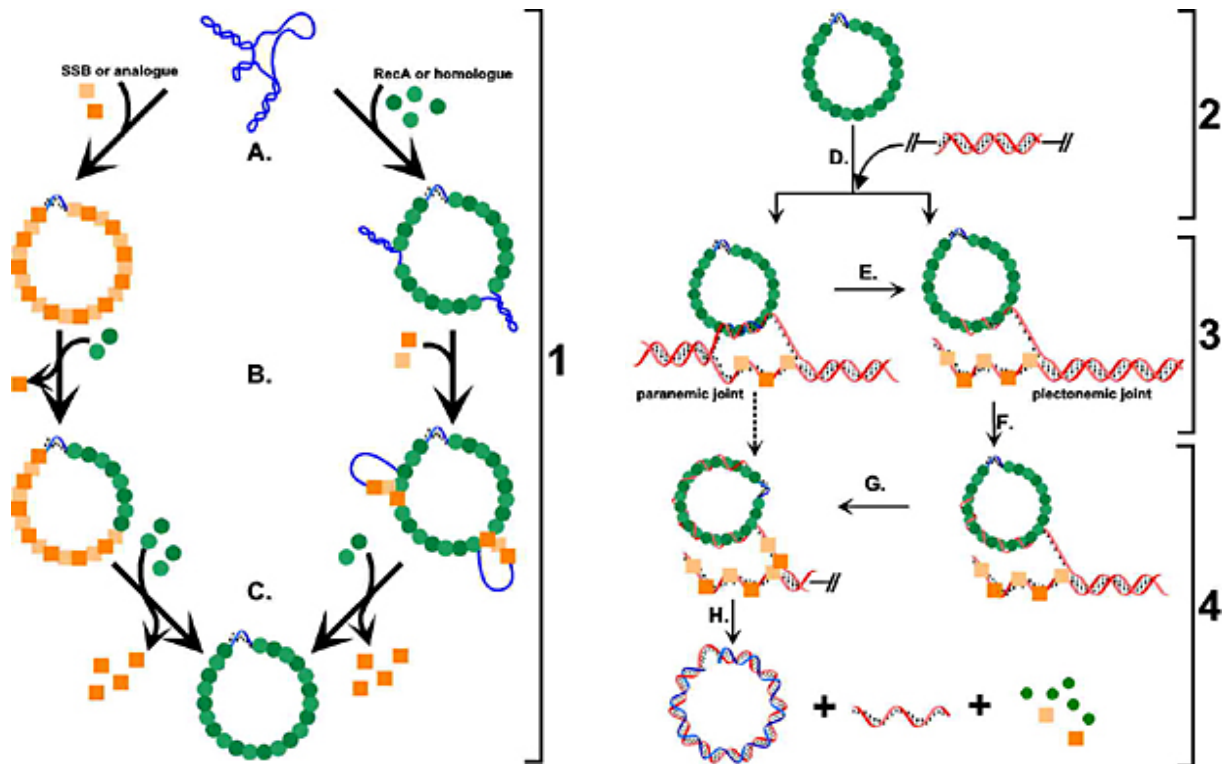


Figure 9 Kinetic steps of DNA strand exchange. The reaction between circular ssDNA and linear dsDNA is shown. The product is a nicked circle and a displaced single strand of DNA. A similar reaction scheme may apply to all of the DNA strand exchange proteins discussed but reaction polarity is not indicated; the illustration is taken from Kowalczykowski (1987). The steps shown are: 1) presynapsis; 2) and 3) synapsis; and 4) DNA heteroduplex extension. The green spheres represent the strand exchange protein, and the orange squares represent SSB.

The role of SSB in presynapsis is to dissolve secondary structure in ssDNA, which is inhibitory to the formation of the saturated presynaptic complex (Kowalczykowski *et al.*, 1987). This is consistent with the role of the protein as a helix-destabilising protein and the observation that other helix destabilising proteins can substitute for SSB in DNA strand exchange *in vitro* (Muniyappa *et al.*, 1984).

Synapsis

Once the presynaptic filament has assembled on ssDNA, synapsis occurs. In this stage of the reaction, a dsDNA molecule must be bound to the filament, homologous to the ssDNA within the filament located within the dsDNA, and a plectonemic heteroduplex joint formed. The efficiency of joint molecule formation is affected by SSB. If SSB is omitted, the displaced ssDNA is used by RecA protein to form a second joint molecule with another dsDNA molecule. This results in the formation of complex, homology-dependent network of joint molecules. In the presence of SSB, RecA protein does not form these networks, since SSB prevents re-invasion events by binding to the ssDNA displaced from the joint molecules. In addition, the yield of joint molecules is greater (Lavery *et al.*, 1992). However, under reaction

conditions where RecA protein is better able to compete with SSB (i.e. in the presence of the volume excluding agents polyethylene glycol or polyvinyl alcohol, or when dATP is used as cofactor), networks are readily formed (Lavery *et al.*, 1990, 1992).

Branch migration

Once the plectonemic joint has formed, the branch migration phase of DNA strand exchange commences. During this phase, the nascent, heteroduplex joint is extended until complete exchange of single strands of DNA occurs, resulting in a nicked, double-stranded circle. Though kinetically distinct, branch migration may not be a mechanistically separate step, but rather may represent a continuation of plectonemic joint molecule formation (Riddles *et al.*, 1985; Kowalczykowski, 1991).

In summary, SSB acts both at the pre- and post-synaptic steps of DNA strand exchange.

1.2.4.2.1.3 Repair

SSB plays a role in the SOS response, methyl-directed mismatch repair and the recombinational repair. In UV irradiated *recA* cells, SSB plays also a role in photorepair with the enzyme photolyase (Lerš *et al.*, 1989). There is not much known about its involvement in the other repair processes (Meyer *et al.*, 1990).

The **SOS response** to DNA damage in *E. coli* is mediated through the *recA*-*lexA* regulon (Little *et al.*, 1982; Walker *et al.*, 1985). Induction of the SOS regulon leads to cleavage of a variety of lysogenic phage repressors including λ , P22, 434, and ϕ 80 (Lin *et al.*, 1989). Cleavage of these repressors, is stimulated by an active form of RecA protein as an allosteric effector (West, 1988). For RecA to promote this cleavage, two cofactors are required: (i) a nucleoside triphosphate, and (ii) either an ssDNA or UV-irradiated DNA (Craig *et al.*, 1980; 1981). In this activation process SSB plays an important role. *In vitro* SSB stimulates RecA-mediated (Resnick & Sussman, 1982; Weinstock & McEntee 1981) or RecA441-mediated (Moreau *et al.*, 1984) cleavage of the λ repressor. By using the temperature sensitive SSB-113 mutant, cleavage of λ repressor at low SSB-113 concentrations can be stimulated, but becomes inhibitory at higher level (Resnick & Sussman, 1982). Cohen *et al.* (1983) observed that SSB-113 competes more efficiently than SSB with RecA for binding ssDNA. This is supported by the observation of Chase *et al.* (1984) that SSB-113 has higher affinity for ssDNA than SSB. The UV sensitivity of *ssb* mutants has been attributed in part to the failure of inducing normal levels of RecA (Baluch *et al.*, 1980; Johnson, 1984), which also supports the notion that SSB is involved in the activation of RecA.

In vitro methyl-directed mismatch repair has an absolute requirement for SSB (Lahue *et al.*, 1989; Lu *et al.*, 1986). Three key enzymes in mismatch repair interact functionally with SSB: (i) DNA helicase II (the unwinding of the parental DNA by helicase driven by SSB), (ii) exonuclease I (exonucleolytic excision of the displaced, error containing strand by exonuclease I is stimulated by SSB), and (iii) DNA polymerase III (repair resynthesis of the strand over a distance of several thousand nucleotides is effected by DNA polymerase III, requiring SSB bound to the template strand).

Recombination repair

When the presence of a DNA lesion, such as a pyrimidine dimer prevents reading of the template strand by DNA polymerase III, the polymerase may dissociate and reinitiate replication at the next Okazaki fragments, leaving a single-strand gap (Roberts *et al.*, 1982; Shwartz *et al.*, 1988). This may be repaired through a recombinational event. Recombination repair is error free. *Eco*SSB is required for recombination as discussed in section 1.2.4.2.1.2. However, it should be recalled that the *ssb* mutants are defective in recombination (Ennis *et al.*, 1987; Glassberg *et al.*, 1979). This may simply reflect a decreased efficiency of mutant SSB interacting with RecA (McEntee, 1980). Therefore, the UV sensitivity of *ssb* mutants may be attributed to three factors involved in recombinational repair: (i) the inability of the mutants to induce sufficient levels of RecA protein for enhanced recombination, (ii) a decreased efficiency of RecA-SSB interaction during the recombination process itself, leading to an accumulation of ssDNA, and (iii) the inability of mutant SSB to adequately protect against DNA degradation.

1.2.4.2.1.4 The C-terminal domain

Using immobilised SSB, Perrino *et al.* (1988) demonstrated a direct interaction between SSB and several other proteins ranging in size from 14 to 160 kDa. Furthermore, there is preliminary evidence for direct interaction of SSB with DNA polymerase (Pol) II (Molineux *et al.*, 1974), I (Molineux *et al.*, 1975) and priB (Low *et al.*, 1982). Recently, it has been shown by Kelman *et al.* (1998) that the χ -subunit of Pol III is the major contact between holoenzyme and SSB. *Eco*SSB interacts with exonuclease I via its all four C-termini (Genschel *et al.*, 2000).

Interaction between SSB and replication proteins extends beyond the *E. coli* system. A physical interaction of two well-characterised SSBs (gp32 and T7 gene 2.5 protein) with their respective DNA polymerase subunits has been demonstrated (Cha & Alberts, 1989; Kim *et al.*, 1992). The gp32 protein has also been shown to interact with T4 primase (Cha & Alberts,

1989). The T7 gene 2.5 protein also interacts with the T7 gene 4 protein which contains both helicase and primase activity (Kim & Richardson, 1994). In the eukaryotic system, the human RPA interacts with DNA polymerase α (Dornreiter *et al.*, 1992; Kowalczykowski, 2000).

1.2.4.2.2 Biochemical and biophysical properties of *EcoSSB*

EcoSSB consists of 177 amino acid (molecular weight 18.843 kDa) with an isoelectric point of 6.0 (Sancar *et al.*, 1981; Chase *et al.*, 1984). The protein forms a homotetramer which is very stable over a wide range of solution conditions, even at a protein concentration of 0.03 μM at 25°C (Weiner *et al.*, 1975; Williams *et al.*, 1983, 1984). The tetramer has a sedimentation coefficient of 4.4 to 4.9 S (Williams *et al.*, 1983) and its diffusion coefficient has been estimated as $5.6 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1}$, corresponding to a Stokes radius of 38 Å (Weiner *et al.*, 1975). The ratio of frictional coefficient of tetramer to that of equivalent mass was estimated to be around 1.36, indicating that the tetramer is either hydrated or nonspherical or some combination of these (Williams *et al.*, 1983).

1.2.4.2.3 Crystallisation of *EcoSSB*

EcoSSB has been the subject of extensive crystallisation efforts. The earlier reports describing the crystallisation of the full length of the protein were by Ollis *et al.* (1983) and Monzingo *et al.* (1983). The asymmetric unit in these crystals contained two native and two proteolytically degraded subunits with D_2 symmetry. It has been shown that the crystal grew only after spontaneous proteolytic cleavage of *EcoSSB* had occurred (Hilgenfeld *et al.*, 1984).

Initial X-ray crystallographic studies showed that *EcoSSB* crystallized in three forms. Form I crystals were of monoclinic (Ollis *et al.*, 1983; Monzingo *et al.*, 1983; Hilgenfeld *et al.*, 1984), Form II of the hexagonal space group $P6_422$ (Ollis *et al.*, 1983) and Form III of the space group $I222$ (Thorn *et al.*, 1994).

1.2.4.2.4 Structure of *EcoSSB*

The folded structure of the homotetrameric SSB from *E. coli* was solved by X-ray crystallography (Webster *et al.*, 1997; Raghunathan *et al.*, 1997) in space group C2. In both structures, the tetramer is comprised of two dimers related by a 2-fold non-crystallographic symmetry axis. Similar to HsmtSSB, individual monomers of *EcoSSB* are composed of a small α -helical segment and two β -pleated sheets that form a β -barrel whose front and back faces consist of three and four antiparallel strands, respectively. Raghunathan *et al.* (1997) and Webster *et al.* (1997) reported the *EcoSSB* structure to 2.9 Å and to 2.5 Å resolution, respectively, and recently, Matsumoto *et al.* (2000) reported the structure to 2.2 Å resolution. The Ramachandran plot (Ramachandran & Sasisekharen, 1968) of the Raghunathan-*EcoSSB* and the Matsumoto-*EcoSSB* structure show 76.6% and 74.2% residues, respectively in the

most favourable regions, whereas the Webster-*EcoSSB* structure shows only 46% residues in the most favourable regions (Table 2). In fact, a good quality model would be expected to have over 90% in the most favoured regions (Laskowski *et al.*, 1993).

Table 2 Statistics on *EcoSSB* structures

	Webster <i>et al.</i> (1997)	Raghunathan <i>et al.</i> (1997)	Matsumoto <i>et al.</i> (2000)
Resolution (Å)	2.5	2.9	2.2
Overall completeness (%)	91.9	95.0	82.7
R-factor (%)	25.5	23.0	24.7
R-free (%)	38.0	29.5	31.2
Residues distributions in Ramachandran plot †			
Most favourable (%)	46.0	76.6	74.2
Allowed (%)	44.8	17.2	18.5
Generously allowed (%)	6.6	6.2	5.3
Disallowed (%)	2.6	0	2.1

† Ramachandran & Sasisekharan, 1968; Laskowski *et al.*, 1993.

1.2.4.2.5 Interaction between *EcoSSB* and DNA

1.2.4.2.5.1 Multiple mode

One of the primary physical quantities that is needed to characterise a protein-DNA interaction is the site size, n , which is the number of nucleotides occluded upon binding the protein. For a particular protein-nucleic acid complex, the value of the site size is most often found (or assumed) to be independent of solution parameters. However, this is not the case for the *EcoSSB*/ssDNA interaction, which can form several different and distinct binding modes, characterised by different site sizes (Lohman *et al.*, 1985; Bujalowski *et al.*, 1986, 1988) referred to as $(SSB)_n$, that differ in the number of nucleotides (n) occluded by each bound tetramer. The site size of the *EcoSSB*/ssDNA complex varies between 35 ± 2 and 65 ± 5 nucleotides per tetramer depending upon the salt concentration, temperature, pH and the concentration of protein and nucleic acids (Lohman *et al.*, 1985; Bujalowski *et al.*, 1986, 1988). The higher site size binding modes are favoured at higher salt concentration (Lohman *et al.*, 1985; Bujalowski *et al.*, 1986). The effects of salt concentration are complex, but generally reflect a requirement for the uptake of cations in progressing from the lower to the higher site size binding modes (Lohman *et al.*, 1985; Bujalowski *et al.*, 1988). However, the site size transitions are sensitive to both the cation and anion type and valence, reflecting a complex set of ion binding interactions to the SSB, the ssDNA, and the SSB-ssDNA complexes (Bujalowski *et al.*, 1988). The presence of divalent cations shifts the binding mode

transition so that they occur at much lower salt concentrations. The (SSB)₃₅ to (SSB)₅₆ and (SSB)₅₆ to (SSB)₆₅ transitions are affected differently, reflecting the different nature of each transition (Bujalowski *et al.*, 1986, 1988). The binding mode transitions are also differentially affected by anion type (Bujalowski *et al.*, 1988). At 37°C, a fourth binding mode with a site size of 42±2 nucleotides per tetramer is also observed for the SSB-poly(dT) interaction (Bujalowski *et al.*, 1986). This mode is also observed with poly(U) and poly(A) at 25°C. A progressive compaction of the SSB-poly(dT) complex occurs throughout the transition from the (SSB)₃₅ to the (SSB)₆₅ as seen by sedimentation analysis (Bujalowski *et al.*, 1988).

1.2.4.2.5.2 Co-operativity

Helix destabilising proteins have the general feature of co-operative binding to ssDNA and a singly contiguous protein chain has a higher affinity for nucleic acids than does an isolated protein. This is also the case for the *EcoSSB*. Two types of co-operative binding to ssDNA, referred to as “limited” and “unlimited” have been observed for *EcoSSB* (Lohman *et al.*, 1986, 1988), and these yield complexes with quite different properties *in vitro* (Lohman *et al.*, 1986; Bujalowski *et al.*, 1987; Ferrari *et al.*, 1994). The (SSB)₆₅-binding mode displays a “limited” type of cooperativity (Overman *et al.*, 1988, 1994; Bujalowski *et al.*, 1987) in which SSB clusters appear to be limited to the formation of dimers of SSB tetramers (octamers), which are seen as “beads” in electron microscopy (Chrysogelos *et al.*, 1982); long clusters of SSB tetramers do not form along ssDNA in this mode (Lohman *et al.*, 1986). In the “unlimited” co-operativity mode, which appears to be correlated with the (SSB)₃₅-binding mode (Lohman *et al.*, 1986; Ferrari *et al.*, 1994), long clusters of SSB can form (Sigal *et al.*, 1972; Ruyechan *et al.*, 1975; Lohman *et al.*, 1986), which could be observed in T4 gene 32 protein binding as well (Alberts *et al.*, 1970).

1.2.4.2.5.3 DNA-binding domain

Proteolysis studies (Williams *et al.*, 1983) indicate that the ssDNA-binding site of *EcoSSB* is contained within the first 115 N-terminal amino acids. The N-terminal “core” polypeptides, SSB*T and SSB*C, formed by trypsin cleavage after Arg115 or chymotrypsin cleavage after Trp135, respectively, also form tetramers. However, these are not as stable as wild-type tetramers, suggesting that the C-terminal region facilitates tetramerisation (Williams *et al.*, 1983). Both SSB*T and SSB*C lower the melting temperature of poly[d(A-T)] more than does intact SSB, suggesting that the C terminus modulates the ability of *EcoSSB* to melt duplex DNA.

1.2.4.2.5.4 Structure of the *Eco*SSB-ssDNA complex

Recently, the crystal structure of the chymotryptic fragment of *Eco*SSB (SSBc) corresponding to the ssDNA binding domain of the protein (residues 1-135) bound to two 35 oligodeoxycytidylates (Figure 10) has been determined to 2.8 Å resolution in C2 space group (Raghunathan *et al.*, 2000). Of the 35 nucleotides, 28 nucleotides could be resolved in the first dimer and 14 in the second (Figure 8). The PDB file (entry code 1EYG) of this structure shows 67% occupancy for each nucleotide with an overall temperature factor of around 70 Å². The Ramachandran plot of this complex structure shows that 77.4% residues are in the most favourable region again indicating the low quality of this structure.

There is not much structural difference between the free and the ssDNA-bound form of the *Eco*SSB (Webster *et al.*, 1997; Raghunathan *et al.*, 1997, 2000; Matsumoto *et al.*, 2000). In the crystals of both the apoform and the complex with ssDNA, the loop (residues from 85 – 100: Loop III) participates in crystal packing.

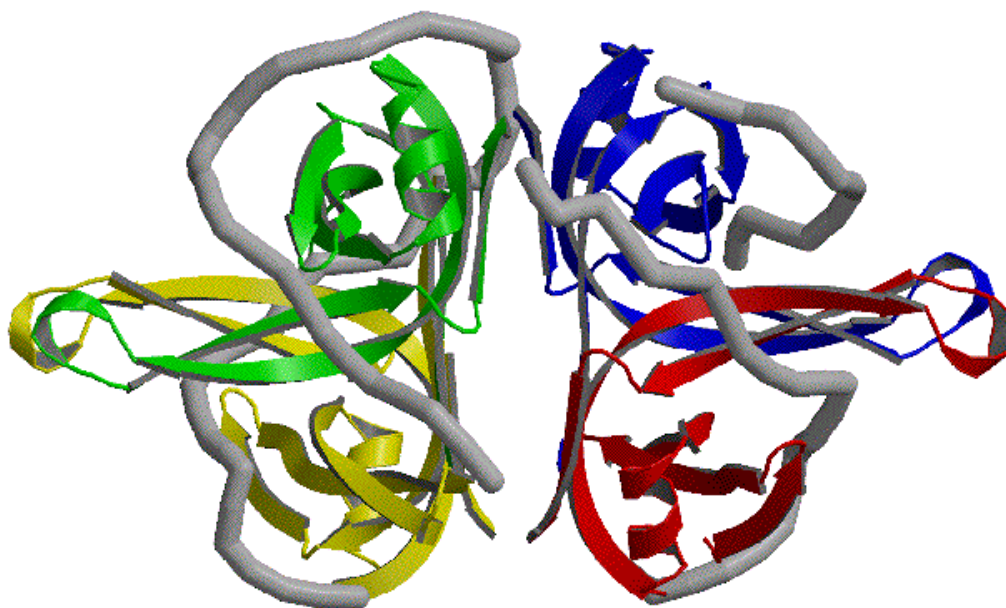


Figure 10 The *Eco*SSB/ssDNA complex (PDB code: 1EYG), a ribbon diagram (prepared using MOLSCRIPT; Kraulis, 1991) of *Eco*SSB (four subunits are in red, blue, green, and yellow) and the ssDNA in grey (Raghunathan *et al.*, 2000).

The Trp fluorescence of SSB is quenched by ~90% upon binding poly(dT) in the (SSB)₆₅ binding mode (Lohman *et al.*, 1985). In the (SSB)₃₅ mode, however, the Trp fluorescence is quenched by only ~50% due to fact that only half of the Trp residues interact with DNA in this binding mode (Bujalowski *et al.*, 1989). The role of the Lys residues and the N-terminus in ssDNA binding has also been shown by chemical modification (Chen *et al.*, 1998). It was observed that the acylation of Lys-43, 62, 73, 87, and the N-terminal NH₃-group is greatly

reduced in the presence of ssDNA. Other basic residues also interact with ssDNA, either with the bases (such as Arg-3) or with the phosphate backbone (Arg-84).

1.2.4.3 Sequence comparison of bacterial SSBs

The best alignment of the *Serratia marcescens* SSB (*Sma*SSB), *Proteus mirabilis* SSB (*Pmi*SSB) and *Brucella abortus* SSB (*Bab*SSB) sequences with *Eco*SSB is shown in Figure 11. The sequence identities between *Eco*SSB and *Sma*SSB, *Pmi*SSB and *Bab*SSB are 89%, 81% and 54%, respectively. The polypeptides of these SSBs are very similar in size to that of *Eco*SSB, ranging from 168 to 177 amino acids, and show extensive sequence similarities in the N-terminal domain (residues 1-115) and only few similarities in the C-terminal regions, despite the fact that the six out of the eight C-terminal residues are common to all four sequences.

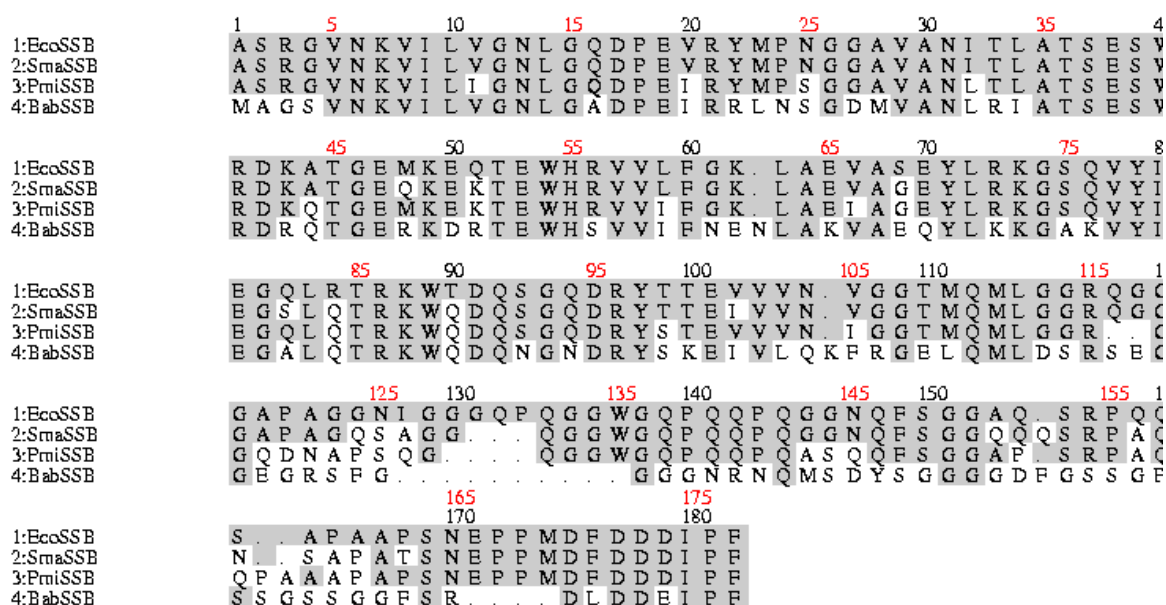


Figure 11 Sequence comparison (prepared using ALSCRIPT; Barton, 1993) of *Eco*SSB (Sancar *et al.*, 1981) with *Sma*SSB (de Vries *et al.*, 1993), *Pmi*SSB (de Vries *et al.*, 1994) and *Bab*SSB (Zhu *et al.*, 1993). Residues that are highly conserved and identical among various SSBs are shaded. The amino acid position 1 in the *Eco*-, *Sma*-, *Pmi*-, *Bab*SSB corresponds to the second position of the open reading frame in respective genes. The black numbering shows the sequential order of the residues in window. The numbering in red letter indicates the sequence position from *Eco*SSB (*Eco*SSB numbering).

Residues of known importance for *Eco*SSB function and/or DNA-binding, such as Trp-40, Trp-54, Phe-60 and Trp-88, are conserved in all of the SSBs, along with a number of other residues within the N-terminal domain. In the present study, the structures of *Eco*SSB (1.95 Å), *Sma*SSB (3.0 Å), *Pmi*SSB (2.5 Å) and *Bab*SSB (1.8 Å) have been solved and will be discussed in detail in the Results and Discussions section.

1.3 Trends in cryocrystallography

Among the biggest problems in macromolecular crystallography is the relatively weak diffraction power of the crystals and their sensitivity to ionising radiation damage. Diffraction can be dramatically improved by cryo-techniques, in which crystals are cooled to cryogenic temperature. Many factors contribute to improvements in data quality: obvious benefits are reduced thermal vibrations, enhanced signal-to-noise ratio, reduced conformational disorder, and in many cases, a higher resolution. Of primary practical importance is the decrease in X-ray damage to the crystal, permitting a complete data set to be collected from one single crystal. In addition, crystal mounting is vastly simplified over conventional capillary techniques. These combined improvements lead to enhanced contrast and sharper detail in electron density maps, facilitating model building and reducing the total time required for structure determination.

Because biological macromolecular crystals are highly hydrated with aqueous solvent contents of roughly 30-70%, to a large extent the problem of cooling involves preventing destructive ice formation. Two methods have been generally used. In one, the crystal is permeated with a diffusible cryoprotectant such as glycerol or sucrose (reviewed by Garman & Schneider, 1997; Rodgers, 1994, 1997) to retard ice nucleation. Determining the initial and optimal conventional cryoprotectant concentration is often a process of trial and error. One must find suitable conventional cryoprotectant concentrations which stabilize the crystal while at the same time combine with the crystallisation reagent to form an amorphous glass. In the second method, the liquid surrounding the crystal is replaced with immiscible hydrocarbon oil such as Paratone-N (Hope, 1988, 1990) or Paraffin oil (Riboldi-Tunnicliffe & Hilgenfeld, 1999). If the crystal is cooled rapidly enough, ice does not form and the crystal can be flash-cooled to a vitreous glass preserving the lattice order.

A further significant improvement was achieved by the introduction of crystal flash-annealing (Harp *et al.*, 1998, 1999; Yeh & Hol, 1998; Sauer & Ceska, 1997). In this procedure, the cold nitrogen gas stream is briefly diverted from the crystal before the latter is allowed to re-cool. After repeating this three times, with the crystal remaining in the cryo-loop (*in situ* annealing) Yeh and Hol (1998) were able to improve the diffraction limit of glycerol kinase from 3.6 Å to 2.8 Å. Harp *et al.* (1998) describe an annealing method where the crystal is first flash-cooled and then returned into the cryoprotectant solution which was either glycerol or MPD or Paratone-N oil.

In a recent communication, from our laboratory, Riboldi-Tunnicliffe and Hilgenfeld (1999) described flash-cooling using dried paraffin oil as a non-aqueous cryoprotectant. This

approach has been successful in nine out of ten cases in our laboratory, and many encouraging reports were received from other laboratories.

A further development of these cryo-techniques will be discussed in detail in the Results and Discussion section.

1.4 Aim of the work

The aim of this project was to elucidate the structures of bacterial ssDNA binding proteins using X-ray crystallographic methods. Development of novel techniques to assist the study was integral to this work. Deciphering the detailed architecture of this class of proteins and analyses of the factors relevant to the interactions/binding of the ssDNA was a major goal.

2 Materials and methods

2.1 Materials

2.1.1 Chemicals

Table 3 Chemical items and their manufacturers

<u>Chemical items</u>	<u>Manufacturers</u>
PEG 400, 4000	Fluka Chemie AG
EDTA	Merck
NaN ₃	Merck
Na cacodylate	Merck
Paraffin oil	Merck
Silicon fluid 200/1 cS oil	Merck
Ammonium sulphate	Merck
Sodium chloride	Merck
Magnesium sulphate	Fluka Chemie AG
Crystallisation screen	Hampton Research
Izit™	Hampton Research
β-mercaptoethanol	Merck
MES	Merck
HEPES	Merck
TRIS	Merck
Sodium acetate	Merck
Dioxane	Merck
Bicine	Merck
Dried paraffin oil	JenaBioScience
Panjelly™	JenaBioScience

2.1.2 Buffers and solutions

All buffers and solutions were prepared using deionised water at 20°C (Millipore water purification system).

Table 4 Buffer, protein storage, crystallisation solutions and oil

Oil mixture	50% paraffin and 50% silicon fluid 200/1 cS
Buffer T	Tris pH 7.5
Protein storage solution A	0.5 M NaCl, 3 mM NaN ₃ , 20 mM Buffer T
Protein storage solution B	1 M NaCl, 1 mM Na ₂ EDTA, 60% (v/v) glycerol 20 mM potassium phosphate, pH 7.5

Crystallisation solution A	4% (v/v) PEG-400, 40 mM sodium cacodylate, 10 mM β -mercaptoethanol, pH 6.5
Crystallisation solution B	12.5 mM sodium acetate, 30% (v/v) dioxane, 100 mM sodium cacodylate, pH 6.5
Crystallisation solution C	3% (w/v) PEG 4000, 5mM magnesium sulphate, 50 mM MES pH 6.5
Crystallisation solution D	1.8 M Ammonium sulphate, 25 mM magnesium sulphate, 50 mM Tris, pH 8.5

2.1.3 Protein and DNA samples

Purified *Eco*SSB, *Bab*SSB, *Pmi*SSB and *Sma*SSB were kindly provided by the group of PD Dr. C. Urbanke (Medizinische Hochschule, Hannover, Germany) and by the group of Prof. Dr. W. Wackernagel (Oldenburg University, Oldenburg, Germany).

The oligonucleotides ((dT)_N; with N = 2, 4, 8, 16, 35) were a generous gift of Dr. Birsch Herschfeld (Department of Virology; University of Jena). The DNA was suspended in Protein storage solution A, and dialysed extensively before use. Oligonucleotide concentrations were determined spectrophotometrically in Buffer T and 0.5 M NaCl, using the following extinction coefficient (per nucleotide): $\epsilon_{260} = 8.6 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ (Urbanke & Schaper 1990).

2.1.4. Instruments

Table 5 Instruments and their manufacturers

<u>Instrument</u>	<u>Manufacturers</u>
Rotating anode generator FR591	Enraf Nonius
Detector, MAR345 and MAR30 cm	Mar Research
Cryostat	Oxford Cryosystems
Light microscope	Zeiss
Water purification system	Millipore
pH meter	Schott
Analytical balance	Sartorius
Centrifuge	Heräus Instruments
Spectrophotometer (SLM 8000C)	Zeiss
Gel electrophoresis system	Pharmacia
Indy workstation	SGI
O2 graphics workstation	SGI
Indigo2 graphic workstation	SGI

2.2 Methods

2.2.1 Purity determination

Protein purity is a critical factor in crystallisation experiments: proteins used for crystallisation should be as pure as possible and completely homogeneous (McPherson, 1982). The purity and homogeneity of SSB was judged visually from SDS polyacrylamide gels stained with Coomassie Blue (Lämmli, 1970).

2.2.2 Crystallisation

The crystallisation of the protein is a prerequisite for the entire crystallographic work. After the purification of DNA and proteins, suitable crystals may be obtained by carefully searching for suitable crystallisation condition, which include pH, buffer conditions, precipitants, etc.

The most common setup to grow protein crystals is the hanging drop technique which is used in this work. The technique is based on vapour diffusion of water. A few microlitres of protein solution are mixed with about equal amount of reservoir solution containing the precipitants. A drop of this mixture is put on a siliconised microscope glass slide which covers the depression in a tray. The depression is partly filled with the required precipitant solution (reservoir solution approximately 0.5 ml). The chamber is sealed by applying grease to the circumference of the depression, before the glass slide is put into place (Figure 12). As the protein/precipitant mixture in the drop is less concentrated than the reservoir solution, water evaporates from the drop and diffuses into the reservoir. As a result the concentration of both protein and precipitant in the drop slowly increases, and crystals may form.

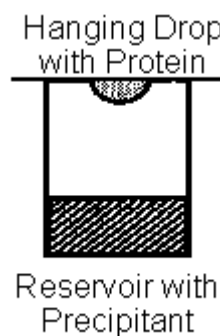


Figure 12 Schematic diagram of a hanging drop setup.

2.2.3 Characterisation of protein crystals

There are a number of ways to distinguish protein from salt crystals.

2.2.3.1 Light microscope

A light microscope is used to detect protein crystals in the small crystallisation drop. Sometimes it is difficult to recognise the crystals, especially when they are very small. In this case, a light microscope with polarising optics is required to observe birefringence. The

following path is a typical setup. Light passes from the light source through the first polarising lens then onto the sample and then to the second polarising filter. On many typical polarisation setups, the second polarising filter can be rotated while the specimen is stationary. Rotating the polarising optics will result in one seeing light, dark, light, dark, as the filter is rotated. But if a crystal with birefringent properties (all crystals except cubic ones) is positioned in between the two polarising filters, one will observe changing colours as the polarising filter is rotated. Specially, when the polarising filters are aligned such that the field is dark, a birefringent object (crystal) will glow with colour. This also depends on the size of the crystal.

Birefringence is one way we can differentiate amorphous precipitate from microcrystals in a drop when viewed under a microscope. Precipitate does not have birefringent properties while all crystals except cubic ones do have birefringent properties.

2.2.3.2 Gel analysis

The second method to verify that crystals are protein crystals is to analyse them by the SDS gel electrophoresis (Lämmli, 1970). A number of crystals are harvested and washed thoroughly in the crystallisation solution. Then they are dissolved in water and put on a 12.5% SDS gel. The gel is stained with Coomassie Blue to probe for protein. If it shows the same band as the protein solution used for crystallisation, the crystals are confirmed to be protein crystals.

2.2.3.3 Staining with IZIT™

An alternative method is to stain the crystal with IZIT™. If the crystal is a protein crystal, it turns blue otherwise it stays opaque. This can simply be done by placing 1 µl of IZIT™ in the sample drop and waiting for 24 hours. IZIT™ is a small molecule dye which fills the solvent channels in protein crystals and binds with the protein molecules. With the appropriate dilution, IZIT™ will in fact leave a clear drop with blue crystals. Salt crystals do not possess these large solvent channels, thus IZIT™ molecules cannot enter the crystal, leaving a clear crystal and a blue drop. IZIT™ works especially well for small micro-crystals or questionable precipitate.

2.2.4 Structure determination

2.2.4.1 Diffraction theory

W. L. Bragg showed that the angles at which diffracted beams emerge from a crystal can be computed by treating diffraction as if X-rays were reflected by sets of parallel planes in a crystal. This is why each spot in the diffraction pattern is called a *reflection*.

The most readily apparent sets of planes in a crystalline lattice are those determined by the faces of the unit cells (the dimension of the unit cell is designated by six numbers: the length of the edges a , b , and c ; and three unique angles α , β , and γ). These and all other regularly spaced planes that can be drawn through lattice points can be thought as sources of diffraction and can be designated by a set of three numbers called *lattice indices*. Three indices h, k, l identify a particular set of equivalent, parallel planes.

W. L. Bragg showed that a set of parallel planes with index h, k, l and interplaner spacing $d_{h,k,l}$ produces a diffracted beam when X-rays of wavelength λ impinge upon the planes at an angle θ and are reflected at the same angle, only if θ meets the condition (Figure 13)

$$2d_{h,k,l} \sin\theta = n\lambda \dots\dots\dots <1>$$

where n is an integer. The geometric construction in Figure 14 demonstrates the conditions necessary for producing a strong diffraction ray.

The path difference ($2d_{h,k,l} \sin\theta$) between two rays reflected from the successive planes is equal to an integral number of wavelengths of the impinging X-rays. The rays reflected from successive planes emerge from the crystal in phase with each other, interfering constructively to produce a strong diffracted beam.

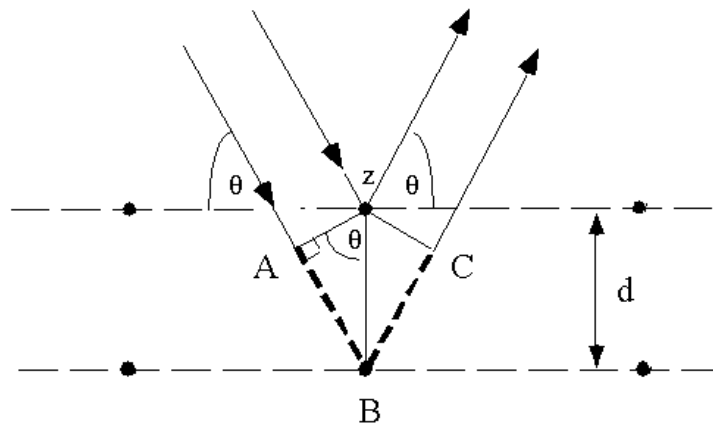


Figure 13 Bragg's Law (adapted from Blundell & Johnson, 1976).

For any other angle of incidence the path difference does not equal an integral multiple of λ the waves emerging from successive planes are out of phase, so they interfere destructively, and no beam can be observed at that angle.

Reciprocal lattice and Ewald sphere

Consider a vector perpendicular to the lattice planes with magnitude equal to $1/d_{h,k,l}$. These vectors $d_{h,k,l}^*$ drawn from a common origin constitute the reciprocal lattice, a point lattice.

If we refer the Bragg's law (equation <1>)

$$\sin\theta_{h,k,l} = \frac{n\lambda}{2} \cdot \frac{1}{d_{h,k,l}}$$

It can be seen that $\sin\theta_{h,k,l}$ is inversely proportional to $d_{h,k,l}$, the interplanar spacing. Since $\sin\theta_{h,k,l}$ is a measure of the deviation of the diffracted beam from the direct beam, structures with a large $d_{h,k,l}$ will exhibit a narrow diffraction pattern and structures with smaller $d_{h,k,l}$ will show the opposite effect. Therefore, the most important property of the reciprocal lattice is that it allows a simple visualisation of Bragg's law.

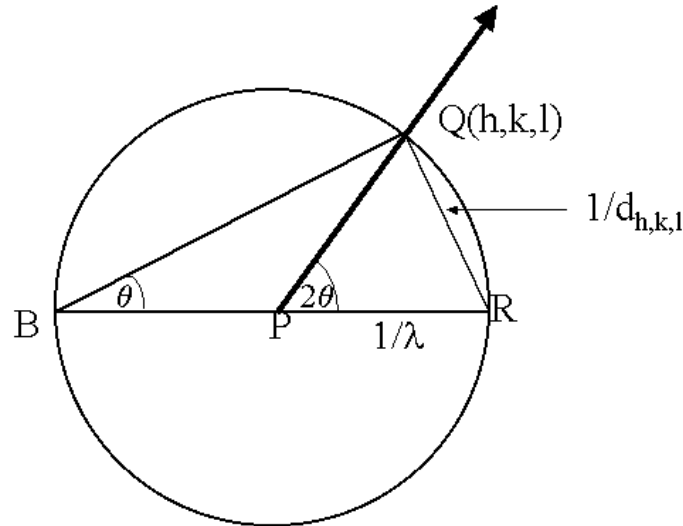


Figure 14 Line diagram representation of the Ewald construction.

The graphical representation (Figure 14) of Bragg's law in reciprocal space was proposed by P. Ewald in 1921, and is known as Ewald construction.

It is assumed that the X-ray beam is passing through P along the line BR. A circle is drawn of radius $1/\lambda$ having its centre P. The circle represents the wavelength of the X-ray in reciprocal space. (If the wavelength is λ in real space, it is $1/\lambda$ in reciprocal space). The crystal centred at P can be physically oriented so that the required reciprocal lattice point can be made to intersect with the surface of the Ewald sphere. It is shown that Q (whose indices are h,k,l) is in

$$\sin\theta = \frac{QR}{BR} = \frac{QR}{2/\lambda}$$

contact with the circle, and the lines RQ and RB are drawn. Because the triangle RBQ is inscribed in a semicircle, it is a right angle triangle

Rearranging the equation gives

$$2 \frac{1}{QR} \sin\theta = \lambda$$

Only when Q is a reciprocal lattice point, the length of the line QR is $1/d_{h,k,l}$ where h, k, and l are the indices of the set of planes represented by Q.

$$2 d_{h,k,l} \sin\theta = \lambda$$

which is Bragg's law with $n = 1$.

We can conclude that whenever the crystal is rotated so that a reciprocal lattice point intersects the sphere with radius $1/\lambda$, Bragg's law is satisfied and a reflection is observed.

Friedel pairs

The intensity of a diffracted beam is proportional to the square of its amplitude [$I(h,k,l)$ proportional to $|F(h,k,l)|^2$], the intensities $I(h,k,l)$ and $I(-h,-k,-l)$ are also equal. The reflections (h,k,l) and $(-h,-k,-l)$ are called Friedel or Bijvoet pairs. Their equal intensities give rise to a centre of symmetry in the diffraction pattern, even though a centre is not present in the crystal structure. The $I(h,k,l) = I(-h,-k,-l)$ equality is usually assumed to be true in crystal structure determination. It depends however on the condition that anomalous scattering is absent.

2.2.4.2 X-ray diffraction data collection

The main pieces of hardware needed for the collection of X-ray diffraction data are X-ray source, monochromators and detector.

2.2.4.2.1 X-ray sources

X-rays are electromagnetic radiation with wavelengths of $10^{-7} - 10^{-10}$ m. They can be produced either using a generator with sealed tube and fixed target or a more powerful system with a rotating anode, the latter being preferred for protein X-ray crystallography. The most commonly used radiation source has a wavelength of 1.5418 Å. This is the characteristic K_{α} wavelength emitted by copper. In our laboratory, we use a rotating anode. Modern technology has made it possible to generate intense X-rays also from circular particle accelerators. These synchrotrons are large, expensive facilities that are mainly used by high-energy physicists. The X-ray radiation (which is so useful to the protein crystallographer) is produced as a side effect.

2.2.4.2.1.1 Rotating anode tubes

In a rotating anode a cathode emits electrons, because the system is under vacuum and it is heated. The cathode is at a high negative potential with respect to a rotating cylinder (made of copper). With a rotating anode small source widths (0.1-0.2 mm) with high brilliance (number of photons/sec/mrad²/mm²/0.1% relative bandwidth) are possible. The advantage over the sealed tube is the higher radiation intensity, but disadvantage is that it requires continuous pumping to keep the vacuum at the required level.

2.2.4.2.1.2 Synchrotron

Synchrotrons are devices for circulating electrically charged particle (electron or positron) at nearly the speed of light. The particles are injected into the storage ring directly from a linear accelerator or through a booster synchrotron. The ring has diameter of 10 to a few hundred meters. The extremely high intensity X-ray radiation from a synchrotron is of great value for

collecting data from weakly diffracting crystals. The beam is strong and highly parallel, causing smaller but more brilliant spots on the detector. Therefore, with synchrotron radiation the resolution is somewhat better and data collection faster than using a rotating anode. Another advantage of synchrotron radiation is its tunability. Therefore, for multiple anomalous data collection, synchrotron radiation is absolutely necessary.

2.2.4.2.2 Monochromators

In the diffraction methods, monochromatic X-rays are used; a narrow wavelength band must be selected from the spectrum supplied by the source. The radiation from a tube with a copper anode contains the K_{α} doublet (1.5418 Å), K_{β} radiation and Brems-strahlung. The K_{β} radiation can be attenuated with a nickel filter but not eliminated. Much cleaner radiation can be obtained with a monochromator. For X-ray radiation from a tube the monochromator, usually is a crystal of graphite. For synchrotron radiation the preferable monochromators are made of germanium or silicon because they select a wavelength band two orders of magnitude narrower ($\delta\lambda/\lambda = 10^{-4}$ - 10^{-5}). Monochromators for synchrotron radiation are of the single or double type. Single type monochromators can be either flat or bent. The advantage of the bent monochromators is that they focus the divergent beam from the synchrotron, preferably onto specimen. Single type monochromators have a disadvantage, if they are tuned to another wavelength, the scanning angle of the monochromator changes, which is not the case with double monochromator, because the direction of beam is independent of wavelength.

2.2.4.2.3 Detectors

The detector is equally important in X-ray data collection. It records the position and the intensity of the diffracted beams. Nowadays most widely used detectors are imaging plates and charge-coupled devices (CCD).

2.2.4.2.3.1 Imaging plate

In our laboratory, we use imaging plates (18, 30 and 345 cm, from Mar Research Instruments). Image plates are made by depositing a thin layer of an inorganic storage phosphor on a flat base. X-ray photons excite electrons in the material to higher energy levels. Part of this energy is emitted as normal fluorescent light in the visible wavelength region. However, appreciable amount energy is retained in the material by electrons trapped in colour centres; it is dissipated only slowly over a period of several days. This stored energy is released on illumination with light. In practical applications a red laser is used for scanning the plate and blue light is emitted. The red light is filtered away and the blue light is measured with a photo-multiplier. Assuming certain error the light emitted is proportional to the number

of photons to which that particular position of the plate was exposed. The pixel size depends mainly on the reading system and can be between 100x100 and 200x200 μm^2 .

2.2.4.2.3.2 CCD detector

These detectors have been used as area detectors in crystallographic application since their development in the late 1980s. The driving force for the development of these detectors was the desire to have a detector, which would allow rapid read-out of the diffraction image combined with large reciprocal-space coverage. The CCD detector has become the detector of choice for synchrotron applications, and a number of commercial equipment manufactures offer CCD detectors in a wide array of configurations which have suitable software for collecting diffraction data. In an effort to increase coverage of reciprocal space, detector makers have designed cameras using either arrays consisting of multiple copies of CCD chips or employing large single-module CCD chips. These chips have X-ray-sensitive phosphor surface at the front to convert incident X-ray into a burst of visible-light photons and the visible-light is converted into electrons by the photoelectric effect. After completion of the X-ray exposure, a shutter is closed and the CCD is read out serially, one pixel at a time. As mentioned above that the X-ray is converted into charges, these charges are shifted quantitatively onto readout circuit processed by on-chip and then processed by an amplifying analog processing circuit and an analog-to-digital converter.

CCD detectors are significantly faster data collection instruments than imaging plates. The most significant differences in speed occur at exposure times less than 5 minutes or so, and it is this speed advantage which has made CCD-based detectors the best suited for synchrotron sources, where exposure times are commonly found on the time scale of 10 seconds or less.

2.2.4.2.4 Preparation of crystals

To collect X-ray diffraction data from a protein crystal at room temperature, a crystal is mounted in thin-walled capillaries of borosilicate glass or quartz. This is the traditional method in macromolecular crystallography. Very often data collection from protein crystal at room temperature suffers from radiation damage; in this case, radiation damage prohibits the collection of a complete data set from a single crystal. Radiation damage to the crystal can be greatly reduced at cryo temperature (100K), which allows collecting even multiple data sets from one single crystal. One of the great benefits of cryocrystallography is the possibility of the long-term storage of crystals at cryogenic temperature, enabling screening for diffraction quality prior to data collection and/or additional data collection at a later date. Not only is this

advantageous 'in house', but it also allows synchrotron beam time to be used far more efficiently.

In this work, prior to flash-cooling the crystal in cryo-stream, the crystal was transferred with a rayon loop (Teng, 1990) from its crystallisation drop into Panjelly™ and subsequently annealed using dried paraffin oil.

2.2.4.2.5 Preparation of dried paraffin oil and Panjelly™

The paraffin oil (Highly Liquid Paraffin Oil from Merck) was dried in a rotary vacuum centrifuge at 333K for 60 minutes. Afterwards, it was allowed to cool before being stored at the same temperature as the crystals (Riboldi-Tunncliffe & Hilgenfeld, 1999).

Panjelly™ is Indian ghee, which is clarified butter. It is obtained by the traditional way by purifying fresh cow milk, separating moisture and solids, until a pure golden hued transparent liquid containing pure milk fat is obtained. The milk is warmed to around 100°C and cooled to room temperature for three to four hours and then the upper layer of the milk is collected. The upper layer is the cream of the milk which is then centrifuged for 15 min. The centrifuged solid part is warmed to 45°C. The fat is then centrifuged again at 30°C to extract the liquid for the use of the cryo purpose. This liquid is free of any foreign materials. It can be stored at room temperature without using refrigeration for several years. It is warmed 2 or 3°C above room temperature for the use as non-aqueous cryoprotectant.

2.2.4.2.6 Crystal storage and retrieval

One of the major advantages of cryocrystallography is the potential for storing and/or transporting crystals once they have been flash-cooled. The steps involved in storing a crystal that is in a cryo-stream are as follows. The holding magnet is mounted on an adapted goniometer head equipped with an attached asymmetric arc and a movable platform (Figure 15). This arrangement enables the crystal to remain in the same position in the cold stream when rotated to point downwards. A small plastic container filled with liquid nitrogen, a "cryovial" with an approximately 2 ml capacity is then brought up around the crystal and the whole top hat removed into the vial. The vial is then immediately transferred into liquid nitrogen to refill. The vials are placed on aluminium canes that accommodate up to five samples and have small tabs that hold the loop assemblies in the vials. The base of the loop assembly is designed to fit loosely in the vials with a small shank that prevents the loop from touching the sides. Notches cut in the top rims of the vials allow them to be filled more quickly with liquid nitrogen. The vial placed on an aluminium cane can be then stored in a standard liquid nitrogen dewar. The crystal stored in liquid nitrogen dewar can be retrieved for data collection in a reverse way as described above.

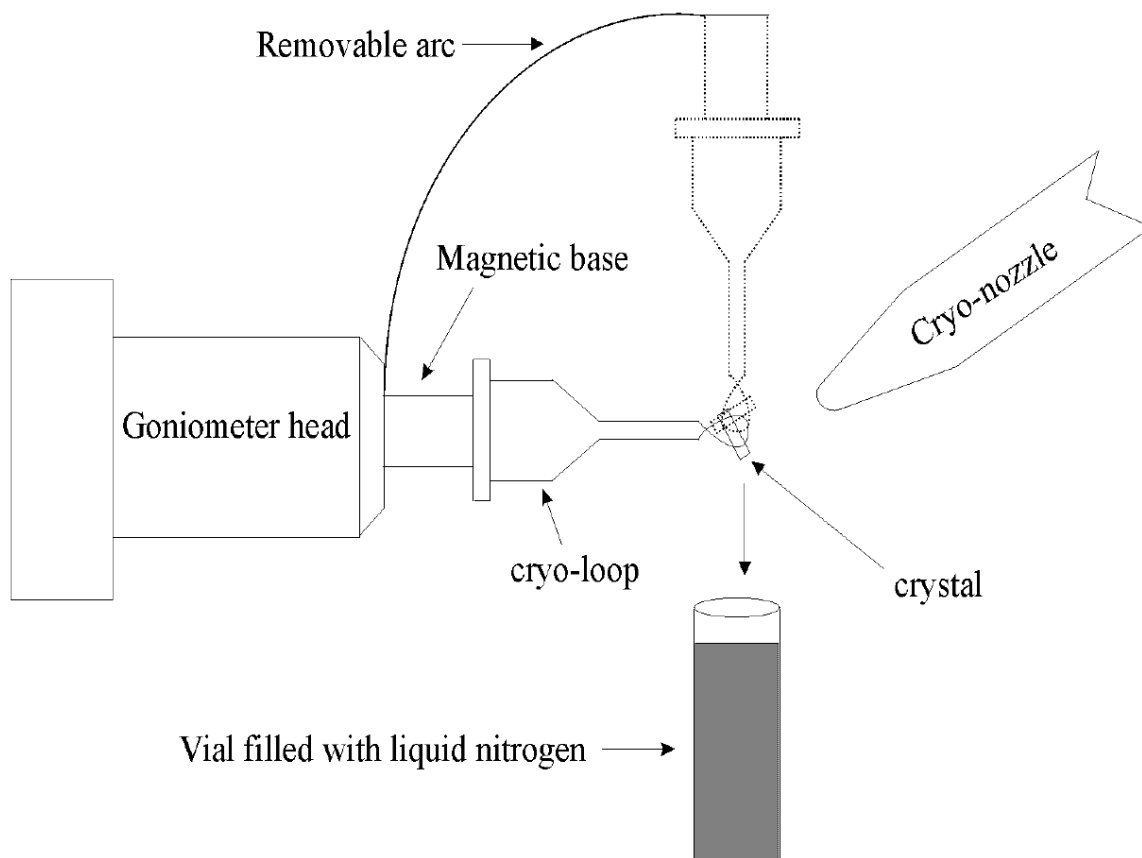


Figure 15 Schematic representation of crystal storage/retrieval.

2.2.4.2.7 Strategy

The essence of data collection strategy is to collect every unique reflection at least once. If the crystal has no symmetry (triclinic) there is no way to achieve 100% completeness from a single rotation pass during data collection, no matter what the crystal orientation is. If the crystal has symmetry higher than triclinic, it is possible to record reflections which are symmetry-equivalent to those in the blind region if the unique axis itself does not lie in it. Skewing the symmetry axis by at least θ_{\max} (maximum diffraction angle) from the spindle direction ensure that there will be no loss of completeness owing to the blind region.

Table 6 lists the required rotation range for the four crystal classes (SSB crystals) in three different point groups in various typical orientations.

If a monoclinic crystal is mounted along an axis perpendicular to its two-fold axis (the spindle axis direction lies in the a, c-plane), the minimum required asymmetric part is 90° . When a tetragonal crystal is mounted along an axis to its c-plane, the required rotation range is 45° at least, and if it is rotated along to the perpendicular to its two-fold axis (the spindle axis direction in the a, b-plane), the required rotation range is at least 90° . Orthorhombic crystals

mounted around any of its two fold axes (the spindle axis direction in the a, b-plane or a, c-plane or b, c-plane), the minimum required rotation range is 90°.

Table 6 Rotation range for SSB crystals in the different crystal class

Space group	Symmetry number†	Rotation required for native data‡
C2	4	180° (b); 90° (ac)
P4 ₃ 2 ₁ 2	96	45° (c); 90° (ab)
P4 ₂ 2 ₁ 2	94	45° (c); 90° (ab)
P2 ₁ 2 ₁ 2	18	90° (ab or ac or bc)

† Hahn, T. 1995

‡ Dauter, Z.1999

In general, the required rotation range depends upon the crystal orientation. It is not easy to orient the crystal in cryo-loop in desired orientation when X-ray data need to be collected at cryo-temperature. Generally, a crystal is oriented in the cryo-loop around an arbitrary axis not in the symmetry plane, therefore more than the required minimum rotation is needed to be collected. In such a case, the crystal orientation is determined from first diffraction image and a STRATEGY program (Ravelli *et al.*, 1997) is used to determine the minimum rotation range to acquire maximum completeness. Such a scheme is especially useful at the synchrotron for maximising the use of the available beam time.

2.2.4.3 Data Processing

After data collection, the next step is to process the X-ray diffraction images. The X-ray diffraction data processing proceeds through indexing, pre-refinement of camera parameters and crystal orientation, intensity integration, post-refinement and scaling.

2.2.4.3.1 Autoindexing

This requires a peak-picking procedure, followed by an analysis of the position of the peaks to determine the unit-cell dimensions, Bravais lattice and crystal orientation. The program DENZO (Otwinowski & Minor, 1997) has two indexing methods: automatic and interactive. The automatic method is applicable in most cases and it is fast and simple. The first step in the automatic method is a peak search, which chooses the spots to be used by the autoindexing subroutine. This program also has the ability to do the pre-refinement of the camera parameters (crystal-to-detector distance, scanning direction relative to oscillation direction, detector tilt away from being normal to X-ray beam) and crystal orientation.

2.2.4.3.2 Intensity integration

There are two distinct procedures available for determining the integrated intensities: summation, integration and profile fitting. Summation integration involves simply adding the pixel values for all pixels lying within the area of a spot and then subtracting the estimated

background contribution to the same pixels. Profile fitting assumes that the actual spot shape or profile is known (in two or three dimensions), and the intensity is derived by finding the scale factor which, when applied to the known profile, gives the best fit to the observed spot profile.

The program DENZO estimates the background value from an average detector signal in the neighbourhood of a specific reflection. In order to calculate the diffraction intensity, the background pixel values are fitted to a plane and then the background value is determined from the plane parameters and is subtracted from the reflection profile.

2.2.4.3.3 Scaling

The integrated intensities from any data collection experiment are not all on the same scale, because of various systematic differences in the collection procedure. It is the task of the data reduction protocol to place all observations on a common scale, to detect and reject outliers (reflections for which the measurement has erroneous), and produce a list of I (intensity) and $\sigma(I)$ (standard deviation of intensity).

In this work, scaling and merging of different data sets as well as the global refinement of crystal parameters is performed by SCALEPACK (Otwinowski & Minor, 1997). During the diffraction data reduction, this program determines the relative scale factors between measurements and refines the crystal parameters using the entire data set. It allows for separate refinement of the orientation of each images, but with the same unit cell value for the whole data set.

2.2.4.3.4 Quality of data

The quality of the X-ray data is assessed by four different criteria. One of them is the symmetry (R_{sym}) or merging R-factor (equation <2>) that arises from the averaging of multiple measurements of reflections of the same (h,k,l) and of symmetry-related reflections. The second quantity is the ratio of the recorded intensity and its standard deviation $I/\sigma(I)$ and the third one is the redundancy of the data, i.e. how often a given reflection and/or one of its symmetry-related reflections have been observed. The fourth quantity is the completeness of the data set overall and in the highest resolution bin.

Weiss & Hilgenfeld (1997), Weiss (2001) and Diederichs & Karplus (1997) noticed that R_{merge} is not a good quality indicator of the X-ray data. This is because R_{merge} is inherently dependent on the redundancy of the data. The more often a given reflection is observed the higher R_{merge} will be, even though by simple statistical reasoning the average value of the measurements become more accurate.

Two other R-factors proposed by Weiss & Hilgenfeld (1997) and Weiss (2001), that should be better suited to describe the quality of diffraction data, are: the so-called redundancy-independent merging R-factor ($R_{r.i.m.}$) and the precision-indicating merging R-factor ($R_{p.i.m.}$). $R_{r.i.m.}$ contains the redundancy N or the multiplicity of the observed reflection and is basically the conventional R_{merge} made independent of redundancy. $R_{p.i.m.}$ also contains the redundancy N and indicates how precisely the average measurement has been measured.

The equations for the conventional R_{merge} , $R_{r.i.m.}$ and $R_{p.i.m.}$ is given below.

$$R_{merge} = \sum_{h,k,l} \sum_i |I_i(h,k,l) - \overline{I(h,k,l)}| / \sum_{h,k,l} \sum_i I_i(h,k,l) \dots \dots \dots < 2 >$$

$$R_{r.i.m.} = \sum_{h,k,l} \sqrt{\frac{N}{N-1}} \sum_i |I_i(h,k,l) - \overline{I(h,k,l)}| / \sum_{h,k,l} \sum_i I_i(h,k,l)$$

$$\equiv R_{meas} \text{ (Diederichs \& Karplus, 1997)} \dots \dots \dots < 3 >$$

$$R_{p.i.m.} = \sum_{h,k,l} \sqrt{\frac{1}{N-1}} \sum_i |I_i(h,k,l) - \overline{I(h,k,l)}| / \sum_{h,k,l} \sum_i I_i(h,k,l) \dots \dots \dots < 4 >$$

2.2.4.3.5 Calculation of structure factor amplitudes

The program TRUNCATE (French & Wilson, 1978) is used to convert a file of averaged intensities (output from SCALA or SCALEPACK2MTZ) to a file containing mean amplitude and the original intensities. If anomalous data is present then F(+), F(-), with the anomalous difference, plus I(+) and I(-) are also written out. The amplitudes are put on an approximate absolute scale using the scale factor taken from a Wilson plot (Wilson, 1949).

2.2.4.4 Method of phase determination

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} |F(h, k, l)| e^{-2\pi i(hx+ky+lz)+i\alpha(h,k,l)} \dots \dots \dots <5>$$

The determination of the three-dimensional structure of macromolecules using single crystal X-ray diffraction techniques requires the measurement of amplitudes and the calculation of phases for each diffraction point. Amplitudes $|F(h,k,l)|$ can be directly measured from diffracting crystals, phases $\alpha(h,k,l)$ have to be determined indirectly. Thus, methods were developed to calculate phases. These methods are multiple isomorphous replacement, multiple wavelength anomalous dispersion, molecular replacement and direct methods.

2.2.4.4.1 Multiple isomorphous replacement (MIR)

Isomorphous replacement requires the introduction of atoms of a high atomic number (heavy atoms), such as mercury, platinum, uranium, and so forth, into the macromolecule under study without disrupting its structure or packing in the crystal. Thus, a perfect isomorphous derivative is one in which the only change between it and the native molecule is the incorporation of one or more heavy atoms. This is commonly done by soaking crystals of native molecules in a solution containing the desired heavy atom. The binding of these atoms to functional groups of macromolecules is facilitated by the presence of large solvent channels in protein and nucleic acid crystals into which these functional groups protrude. The addition of one or two more heavy atoms to a macromolecule introduces differences in the diffraction pattern of the derivative relative to that of the native. If this addition is truly isomorphous, these differences will represent the contribution from the heavy atoms only; thus, the problem of determining atomic positions is initially reduced to locating the position of a few atoms. Once the positions of these atoms are accurately determined, they are used to calculate a set of phases for data measured from the native crystals. Although theoretically one needs only two isomorphous derivatives to determine the three-dimensional structure of biological macromolecules, in practice one needs more than two, owing to errors in data measurement and scaling and in lack of isomorphism.

The following vector diagram (Harker construction) illustrates the relationship between native and derivative structure. The objective of a phasing experiment is to derive the unknown phase of the protein reflection F_P and α_P (Figure 16a)

From the experiment, we know only the magnitudes F_P and F_{PH} which can be represented in the complex plane as a circle of radius F_P and F_{PH} , respectively (Figure 16b). If we know both the magnitude and the phase of F_H we can draw both circle offset by vector F_H and obtain two solutions for possible values of F_P (red arrows). The phase and magnitude of F_H can be calculated easily if we know the position of a heavy metal. It can be proven that the best phase we can obtain from the 2 solutions is the mean between the 2 possibilities. In real cases, F_H is much shorter than the red vector, and the phase error will be quite large. Therefore in order to eliminate the phase ambiguity we can prepare a second derivative and repeat the procedure. Provided the heavy atom is not at the same position, we can now obtain a unique solution for α_P (Figure 16c). We have now, at least in theory, an exact solution for the phase angle of F_P . The theory is based on two assumptions: (i) ideal isomorphism and (ii) exact heavy atom positions, neither of which is perfectly met, for practical and experimental reasons in the first

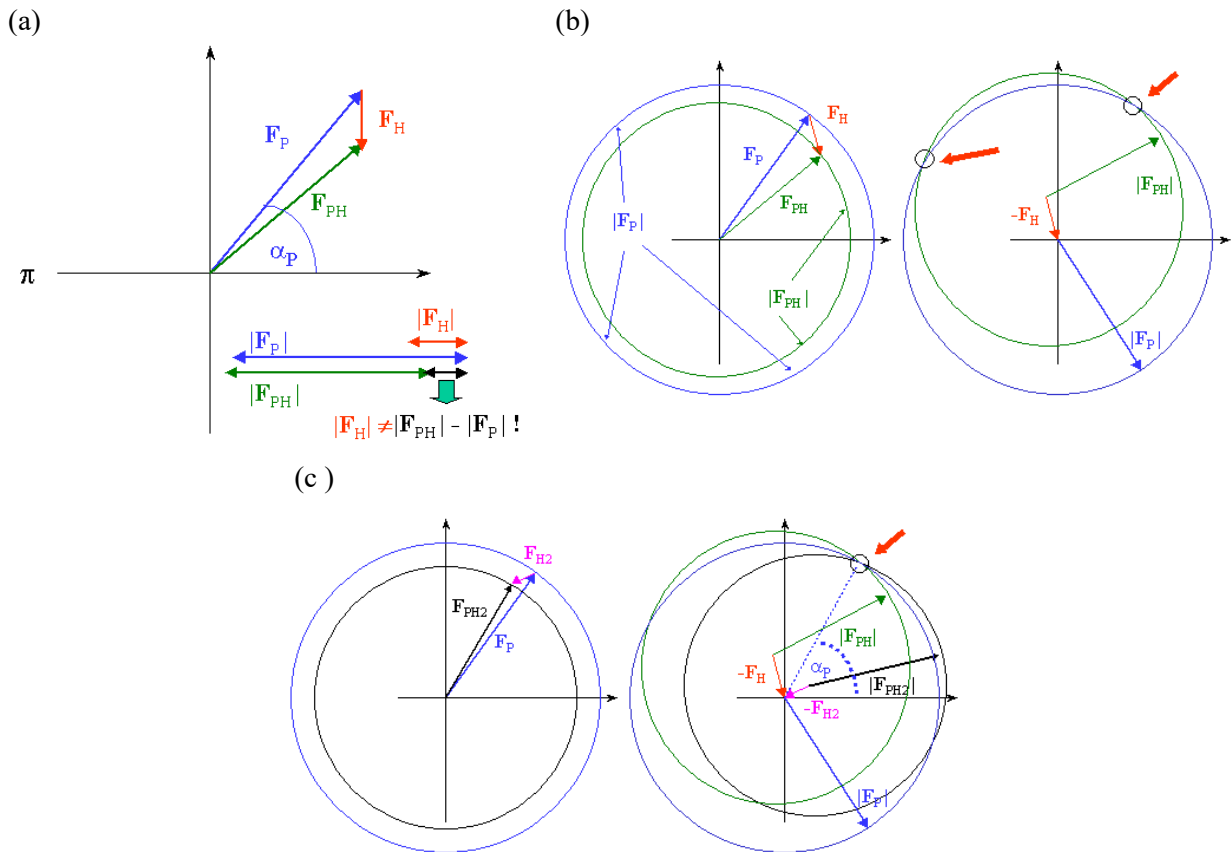


Figure 16 Harker construction for protein phase determination (a) A structure factor F_{PH} for the heavy-atom derivative (green) is the sum of the contributions from the native structure (F_P , blue) and the heavy atom (F_H , red), (b) Vectors from the origin to intersection of the two circles are showing two possibilities for the protein phase angle α_p (c) Vector from the origin to common intersection of the three circle are showing the unique protein phase angle α_p .

case and for theoretical reasons in the second. In the picture it means that the phasing circles may not intersect in exactly in one spot, and yet another derivative may be necessary to improve the quality of the phases. The method is therefore called multiple isomorphous replacement.

2.2.4.4.2 Multiple-wavelength anomalous dispersion (MAD)

If the protein contains anomalously scattering atoms, the difference in intensity between the Bijvoet pairs can be exploited for protein phase angle determination. In the multiple wavelength method the wavelength dependence of the anomalous scattering is used. The principle of this method is rather old but it was the introduction of the tuneable synchrotron radiation sources that made it a technically feasible method for protein structure determination. Hendrickson and colleagues (Hendrickson *et al.*, 1988; Krishnamurthy *et al.*, 1988) were the first to take advantage of this method and to use it for solving the protein structure. Of course, the protein should contain an element that gives a sufficiently strong anomalous signal. Therefore, the elements in the upper rows of the periodic system are not

suitable. Hendrickson showed that the presence of one Selenium (Se) atom (atomic number 34) in a protein of not more than approximately 150 amino acid residues is sufficient for a successful application of MAD (Hendrickson *et al.*, 1990); however, this depends very much on the quality of the data. With more Se atoms the size of the protein can, of course, be larger. One way to introduce Se into protein is by growing a methionine-auxotroph microorganism in a Se-methionine substrate instead of a methionine-containing substrate.

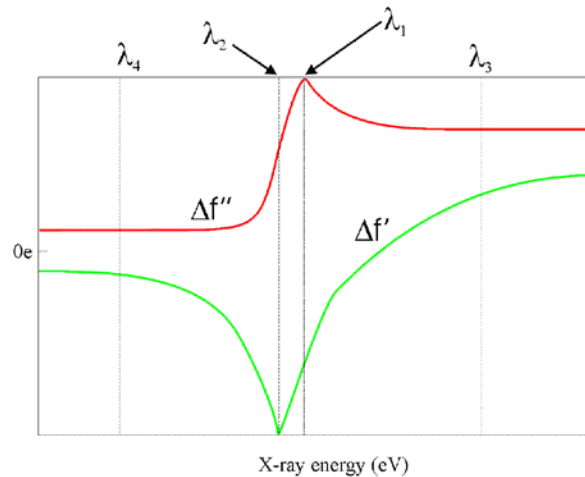


Figure 17 Schematic of experimental values for $\Delta f'$ (green) and $\Delta f''$ (red) as a function of X-ray energy.

The wavelengths have to be carefully chosen to optimise the difference in intensity between Bijvoet pairs and between the diffraction at the selected wavelengths. The anomalous scattering contributes $\Delta f'$ (the real part) and $\Delta f''$ (the imaginary part). Usually diffraction data are collected at three wavelengths (Figure 17).

- (i) First wavelength (λ_1) where $\Delta f''$ has its maximum and where the Bijvoet difference is largest (also called the “peak”).
- (ii) Second wavelength (λ_2) where $\Delta f'$ has its minimum (also called the “edge”).
- (iii) Third wavelength (λ_3 or λ_4) (remote from the edge, on the left or right) where $\Delta f'$ and $\Delta f''$ are small.

2.2.4.4.3 Molecular replacement (MR)

The MR method makes use of a known three-dimensional structure as an appropriate starting model to provide phase angles for the observed structure factor amplitudes from the unknown structure. When a whole family of homologous structure is available, it is often preferable to calculate an “average structure” for use as a starting model. Quite often, if the relative positions of the copies are not known, the presence of more than one copy of the unknown structure in the asymmetric unit complicates the search. In such a case, the phase determination can be facilitated through the use of an oligomeric search model.

Placement of the molecule in the target unit cell requires six parameters, 3 rotational and 3 translational parameters. The six-dimensional search is very time-consuming even with the fastest computers; therefore, it is only practical in limited cases (Baldwin, 1980). In 1962, Rossmann and Blow showed that the six-dimensional search could be divided into two three-dimensional searches: rotation search and translation search.

Rotation search

The basic principle of the molecular replacement method can be understood by considering the Patterson function of a protein crystal structure. Any Patterson function will contain sets of peaks representing intramolecular vectors. These are the self-vectors. There will be one set for each molecular orientation found in the crystallographic cell. They must all lie within the largest intramolecular distance 'r' from the origin of the Patterson function. On the other hand there are the intermolecular vectors which will tend on the average to be longer than the intramolecular ones. These are termed cross-vectors set. The sphere of radius 'r' must be smaller than the smallest cell dimensions; otherwise overlaps will occur with the origin of the next unit cell in Patterson space.

The next problem is to choose a criterion for the correspondence of the self-vectors at different orientations. The classical method of Rossmann and Blow (1962) is to compare the value of the product function R where

$$R(\theta_1, \theta_2, \theta_3) = \int_U P_1(u) \cdot P_2(u) du \dots\dots\dots < 6 >$$

For the Patterson P_1 and the rotated Patterson P_2 within the Patterson volume U . This will have a maximum value when the two self-vector sets are equivalently oriented. $R(\theta_1, \theta_2, \theta_3)$ is known as the rotation function, where $\theta_1, \theta_2,$ and θ_3 are Eulerian angles.

The locked rotation function

If the cross-rotation function is applied with model molecules to be oriented in an unknown crystal structure, several solutions will be found if the crystal structure has crystallographic symmetry, and also because of non-crystallographic symmetry operators. The solutions of the cross-rotation function are not independent but related through the non-crystallographic symmetry. This knowledge can be used as a constraint in the calculation of a "locked" rotation function, which is the average of n independent rotation functions and, consequently, has an improved peak-to-noise ratio (Tong & Rossmann, 1990).

PC-refinement

Once orientations of molecules are found, these may not be exact. Therefore, these rotation function peaks need to be refined before the translation search is attempted. To optimise the orientation, Brünger *et al.* (1990, 1998) have introduced Patterson correlation (PC) refinement in the programs X-plor and CNS using the Patterson correlation coefficient as target function. It consists of carrying out “Patterson refinements” of a large number of the highest peaks of a rotation function. The refinement of the orientation of the search model is performed prior to translation searches. The target function for the refinement consists of an effective Patterson energy term that is the negated standard linear correlation coefficient between the observed and calculated normalised amplitude.

Translation search

For the final solution of the MR method, the translation required to overlap one molecule (or subunit) onto another in real space must be determined, after it has been oriented in the correct way with the rotation function. The known molecule is moved through the asymmetric unit and structure factors (F_{calc}) are calculated and compared with the observed structure factor (F_{obs}) by calculating an R-factor or the correlation coefficient as a function of the molecular position.

$$R = \frac{\sum_{h,k,l} \left| |F_{obs}| - k|F_{calc}| \right|}{\sum_{h,k,l} |F_{obs}|} \dots\dots\dots < 7 >$$

The standard linear correlation coefficient C is

$$C = \frac{\sum_{h,k,l} (|F_{obs}|^2 - \overline{|F_{obs}|^2}) \times (|F_{calc}|^2 - \overline{|F_{calc}|^2})}{\left[\sum_{h,k,l} (|F_{obs}|^2 - \overline{|F_{obs}|^2})^2 \sum_{h,k,l} (|F_{calc}|^2 - \overline{|F_{calc}|^2})^2 \right]^{1/2}} \dots\dots\dots < 8 >$$

The advantage of this correlation coefficient over R-factor is that it is independent of scaling.

The program **AMoRe** (Automated Molecular Replacement) is a combination of three core programs: ROTING for the rotation function, TRAINING for the translation function, and FITING for the refinement of the results (Navaza, 1994).

ROTING is a fast rotating function, defined as the overlap of observed and calculated Patterson functions. The Patterson functions are expanded in spherical harmonics. The resulting peaks in the rotation function are tested and selected with correlation coefficient between calculated and observed structure factor amplitudes.

The output of ROTING is the input for the translation function TRAINING. The highest peaks in the translation function(s) are tested with the correlation coefficient in terms of

amplitudes because it is independent of scaling factor. On this basis, acceptable solutions are selected. MR solutions are usually followed by rigid-body refinement for the orientation and position of the search model in the target unit cell. This is done by using the program FITING. The difference with other rigid-body refinement strategies is that FITING presents the search model, as an electron density map and not as an atomic model.

In the present work, all crystal structures of SSBs have been solved by MR method using AMoRe.

2.2.4.5 Electron density and model building

Refinement (see section 2.2.4.6) is usually unable to correct very large errors in the atomic model or to correct for missing part of the structure. The atomic model needs to be corrected by visual inspection of difference electron density maps at a suitable graphics workstation. In order to improve the quality and resolution of the difference electron density map, the observed phases are often replaced or combined with calculated phases as soon as an initial atomic model has been built. These combined electron-density maps are then used to improve and to refine the atomic model. The inclusion of calculated phase information brings with it the danger of biasing the phases towards the current atomic model. This model bias can obscure the deletion of errors in atomic models if sufficient experimental phase information is unavailable.

2.2.4.5.1 (F_o-F_c)-map

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (|F_o| - |F_c|) e^{-2\pi i(hx+ky+lz-\alpha_{calc})} \dots \dots \dots < \mathbf{9} >$$

A contour map of this Fourier series is called (F_o-F_c)-map. In protein crystallography, the average of the difference map is usually set to zero and the density is assigned as positive or negative depending on the standard deviation from zero.

Positive density in the region of the map implies that the contribution of the observed intensities (F_os) to ρ are larger than the contribution of the model (F_cs), and thus the unit cell (represented by F_os) contains more density in this region than implied by the model (represented by F_cs), In other words, the map is telling us that the model should be adjusted to increase the electron density in this region by moving atoms towards the region. On the other hand, a region of negative density indicates that the model implies more electron density in the region than the unit cell actually contains. The region of negative density is telling us to move atoms away from this region.

2.2.4.5.2 (2F_o-F_c)-map

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (2|F_o| - |F_c|) e^{-2\pi i (hx+ky+lz-\alpha_{calc})} \dots\dots\dots <10>$$

This map shows, besides the electron density of the model, the difference between the actual structure and the model. The phase angles are those calculated for the model. In this map, the model influence is reduced, but not as severely as in the (F_o-F_c)-map. Unless the model contains extremely serious errors, this map is everywhere positive and the contour at carefully chosen electron density resembles the molecular surface and can be viewed as a superposition of an (2F_o-F_c)-map with (F_o-F_c)-map.

2.2.4.5.3 Omit map

An important way to overcome phase bias is the use of omit map. An omit is made by removing the residues of interest from the model for calculating the phase. In theory, this will allow the phase calculated from the rest of the model to phase the area of interest with no bias from the model left out. The method takes advantage of the Fourier transform property that every point in real space is influenced by every point in reciprocal space, and vice versa. If the rest of the model is mostly correct, then the phases calculated for this portion will be closer to the true phase and will produce a mostly correct image of the portion left out.

2.2.4.5.4 SigmaA weighted maps

The program SIGMAA (Read, 1986) calculates weighted Fourier coefficients either from the calculated phase from a (partial) model structure, or by combining phase probabilities from isomorphous phases with those from one or more (partial) structures. The calculated SigmaA weighted (2mF_o-DF_c, mF_o-DF_c) maps are less biased than the 2F_o-F_c and F_o-F_c maps.

(mF_o-DF_c)-map

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (m|F_o| - D|F_c|) e^{-2\pi i (hx+ky+lz-\alpha_{calc})} \dots\dots\dots < 11 >$$

(2mF_o-DF_c)-map

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (2m|F_o| - D|F_c|) e^{-2\pi i (hx+ky+lz-\alpha_{calc})} \dots\dots\dots < 12 >$$

where m is figure of merit and D estimated error in the (partial) structure from the coordinated error (Luzzati, 1952).

2.2.4.6 Refinement of protein structures

Approximate model co-ordinates are initially derived from electron density maps which are usually calculated either by the MAD, MIR or MR method. The atomic co-ordinates based on such model are not very accurate because the maps are of insufficient resolution to show the constituent atoms individually and in part because the phases used to calculate the electron

densities maps are not very accurate. Even for models derived from the best electron density maps the accuracy of the atomic co-ordinates is likely to be no better than about 0.5 Å (Jensen, 1985), and localised regions of models derived from average quality maps may suffer more serious errors. Such models, however, are still useful in showing the overall features of macromolecules and the pattern of chain folding; but to understand much of chemistry of these molecules and how they function, more accurate parameters are required.

The process of obtaining atomic parameters that are more accurate than those obtained from an initial model is referred to as refinement of the crystal structure. It is an iterative process of improvement of the quality of the structure model. The position x, y, and z and the atomic thermal parameters B derived from each atom, are adjusted so as to improve the agreement between the observed structure factor amplitude $|F_{obs}(h,k,l)|$ and those calculated from the structural model $|F_{calc}(h,k,l)|$.

2.2.4.6.1 Refinement theory

The refinement techniques in protein X-ray crystallography are based on the principle of least squares. The method of least squares is a very powerful method for obtaining the most reliable information possible from a set of experimental observations. The measured data set has for each reflection (h,k,l) an intensity from which the amplitude of the structure factor $|F_{obs}(h,k,l)|$ can be derived. From the preliminary model, values for the structure factors $|F_{calc}(h,k,l)|$ can be calculated and in the refinement procedure the values of $|F_{calc}(h,k,l)|$ should be brought as close as possible to $|F_{obs}(h,k,l)|$ for all reflections (h,k,l). $|F_{calc}|$ can be varied by changing the parameters of the model. For some reflections $|F_{calc}|$ will be larger than $|F_{obs}|$ and for others it just the other way around. It is assumed that the $|F_{obs}|$ values are free of systematic errors and distributed as a Gaussian error curve around their real values $|F_{real}|$, which means that the probability P of finding a value $|F_{obs}(h,k,l)|$ for the reflection (h,k,l) between $|F_{obs}(h,k,l)|$ and $|F_{obs}(h,k,l)|+d|F_{obs}(h,k,l)|$ is

$$P(h, k, l) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\left\{|F_{obs}(h, k, l)| - |F_{real}(h, k, l)|\right\}^2}{2\sigma^2}\right] d|F_{obs}(h, k, l)| \dots\dots\dots < 13 >$$

σ^2 is the variance caused by arbitrary errors in the measurement

$$\sigma^2 = \int_{-\infty}^{+\infty} \left\{ |F_{obs}(h, k, l)| - \overline{|F_{obs}(h, k, l)|} \right\}^2 P(h, k, l) d|F_{obs}(h, k, l)| \dots\dots\dots < 14 >$$

Normalisation requires

$$\int_{-\infty}^{+\infty} P(h, k, l) d|F_{obs}(h, k, l)| = 1 \dots\dots\dots < 15 >$$

With the assumption that the errors in the $|F_{obs}(h, k, l)|$ values for different reflections are independent of each other, the total probability P for finding certain set of $|F_{obs}(h, k, l)|$ is

$$\begin{aligned}
 P &= \prod_{h, k, l} P(h, k, l) \\
 &= \prod_{h, k, l} \frac{1}{\sigma(h, k, l)\sqrt{2\pi}} \exp\left[-\frac{\{|F_{obs}(h, k, l)| - |F_{real}(h, k, l)|\}^2}{2\sigma^2(h, k, l)}\right] d|F_{obs}(h, k, l)| \\
 &= \exp\left[-\sum_{h, k, l} \frac{\{|F_{obs}(h, k, l)| - |F_{real}(h, k, l)|\}^2}{2\sigma^2(h, k, l)}\right] \times \prod_{h, k, l} \frac{1}{\sigma(h, k, l)\sqrt{2\pi}} d|F_{obs}(h, k, l)| \dots\dots\dots < 16 >
 \end{aligned}$$

The problem is that the real values of the $F(h, k, l)$ s are unknown. However, it is assumed that these real values can be approximated by the calculated values. The goal is to bring the set of $|F_{calc}|$ s as close as possible to the $|F_{obs}|$ s. In the method of least squares this is defined as occurring at the maximum value of P. In other words, the optimal set of $|F_{calc}|$ s is the one that has the highest probability P. A maximum for P is obtained for a minimum of

$$\sum_{h, k, l} \frac{\{|F_{obs}(h, k, l)| - |F_{calc}(h, k, l)|\}^2}{2\sigma^2(h, k, l)}$$

This is the principle of least squares.

2.2.4.5.2 The free R-factor (R_{free})

Structure determination of macromolecules by crystallography involves fitting atomic models to the observed diffraction data. The traditional measurement of the quality of this fit is the R-value, defined as

$$R = \frac{\sum_{h, k, l} W_{h, k, l} \left| |F_{obs}(h, k, l)| - k|F_{calc}(h, k, l)| \right|}{\sum_h |F_{obs}(h, k, l)|} \dots\dots\dots < 17 >$$

Where (h, k, l) are the indices of the reciprocal lattice points of the crystal, $W_{h, k, l}$ are weights, k is a scale factor, and $|F_{obs}(h, k, l)|$ and $|F_{calc}(h, k, l)|$ are the observed and calculated structure factor amplitudes, respectively.

Despite stereochemical restraints, it is possible to overfit the diffraction data: an incorrect model can be refined to low R-values. Therefore a certain fraction of the diffraction data should be set aside for cross-validation. For cross-validation with a single test set T the free R value is defined as

$$R_{free} = \frac{\sum_{h,k,l \in T} W_{h,k,l} \|F_{obs}(h,k,l) - k|F_{calc}(h,k,l)\|}{\sum_{h \in T} W_h |F_{obs}(h,k,l)|} \dots\dots\dots <18 >$$

Where the calculated structure factors $|F_{calc}(h,k,l)|$ are obtained from a model that was constructed and refined against the working set (without the knowledge of the test set). In general, R_{free} will be higher than R since the test set has been omitted from the refinement process (Brünger *et al.*, 1998).

2.2.4.6.3 The program CNS

A new software suite Crystallography & NMR System (CNS) has been developed for macromolecular structure determination by X-ray crystallography or solution nuclear magnetic resonance (NMR) spectroscopy (Brünger *et al.*, 1998). This program allows the user to perform operations on data structures, such as structure factors, electron density maps, and atomic properties. The power of the CNS language has been demonstrated by the implementation of a comprehensive set of crystallographic procedures for phasing, density modification, and refinement. User-friendly task-oriented input files are available for nearly all aspects of macromolecular structure determination by X-ray crystallography and NMR. In this work, all crystal structures of SSBs were refined using CNS protocol.

Crystallographic refinement can be formulated as a search for the global minimum of the target function (Jack & Levitt, 1978)

$$E = E_{chem} + w_{xray} E_{xray} \dots\dots\dots <19 >$$

E_{chem} comprises empirical information about chemical interactions. It is a function of all atomic positions, describing covalent and non-covalent interactions. E_{xray} describes the difference between observed and calculated diffraction data, and w_{xray} is a weight chosen to balance the forces arising from each term.

GENERATE

This protocol generates the molecular structure by interpreting the co-ordinate file to obtain the residue sequence or by explicitly specifying the residue sequence and gives two files as output: a PDB file and an MTF file (this contains the molecular topology information which describes to covalent topology of the molecules).

RIGID

The position and orientation of molecules in the asymmetric unit is optimised using this protocol. This is especially necessary for models which have been obtained by the MR method. CNS provides the possibility of refinement of several rigid groups. Parts of the

molecule that are not specified in any “GROUp” statement remain fixed. The “constraints fix” statements have no influence on rigid-body refinement.

POSITIONAL

The positional refinement is a conventional refinement which is carried out using the conjugate gradient algorithm.

SIMULATED ANNEALING

Simulated annealing using torsion angle dynamics is used to improve the model. The use of torsion angle dynamics reduces the number of parameters being refined and hence reduces the degree of over-fitting of the data. For the initial model with relatively large errors (due to manual building or misplaced atoms) a starting temperature of 5000K is recommended, and the cooling rate is usually taken as a 50 K step. The simulated annealing refinement task files includes energy minimisation both before and after the simulated annealing.

BREFINEMENT

This statement optimises individual (restrained) isotropic B-factors. Distinctions between backbone and side-chain B-factor are made. B-factor restraints can be set-up as necessary.

2.2.4.6.4 The program REFMAC

The program REFMAC is a CCP4 (Collaborative Computational Project, Number 4, 1994) supported program (Murshudov *et al.*, 1997, 1999), which is widely used for macromolecular crystal structure refinement. This program can carry out rigid body, restrained or unrestrained refinement against X-ray data, or idealisation of a macromolecular structure. It minimises the co-ordinate parameters to satisfy either a Maximum Likelihood or Least Square residual. There are options to use different minimisation methods. It also produces a output file (in MTZ format) containing weighted coefficient for SigmaA weighted mF_o-DF_c and $2mF_o-DF_c$ maps (see section 2.2.4.5.4). Recently, fast Fourier transformations in individual atomic anisotropic refinement has been implemented in this program (Murshudov *et al.*, 1999). The anisotropic refinement not only reduces the R-factor and R_{free} but also improves the fit to the geometric target, indicating that this parameterisation is valuable for improving models derived from experimental data.

2.2.4.6.5 Introduction of water molecules in the structure

The following criterias were employed in the introduction of water molecules in the structure. They should represent density in the residual (F_o-F_c) electron density map above 4σ and have density in the ($2F_o-F_c$) map above 1σ . More importantly they should have chemically reasonable distances to potential hydrogen bonds donors/acceptors. The improvement of the model was monitored by both the conventional R-factor and the R_{free} (Brünger, 1992).

2.2.4.7 Validation of the model

The final models of all structures were validated by the program PROCHECK (Laskowski *et al.*, 1993) to check the overall quality of the model using the Ramachandran plot and other stereochemical criteria. The errors in the atomic co-ordinates of the molecular model were estimated using a Luzzati plot (Luzzati, 1952).

2.2.4.8 Structure analysis

The crystal structures of SSBs were compared using the program LSQKAB (The CCP4 Suite, 1994) or ALIGN (Cohen, 1997). Visual comparisons were made using the program O (Jones *et al.*, 1991). Further analysis of the crystal contacts were performed using the program CONTACT from the CCP4 Suite (1994). Ribbon diagram of SSBs structures were made using the programs MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1999) and RASTER3D (Merritt & Murphy, 1994). Molscript, Bobscript and Raster3D are the programs of choice to draw the three-dimensional models highlighting the secondary structures from the Protein co-ordinate files based on information in the input file and to draw a map around the desired residues from map files.

The molecular surface area buried upon dimer or tetramer formation in various structures was calculated using the algorithm of Lee and Richards (1971) as implemented in the program NACCESS (Hubbard *et al.*, 1991), taking a probe radius 1.4 Å and the following procedure: (i) the molecular surface area of the first dimer/monomer (A) was calculated, (ii) the molecular surface area of second dimer/monomer (B) was calculated and added to the first item ($C=A+B$). The sum (C) is the total molecular surface area before burial, (iii) the observed molecular surface area of whole tetramer/dimer (D) was calculated. The difference (E) between the total area to be buried (D) and the observed molecular surface area (C) is the amount of surface area buried upon tetramer/dimer formation ($E=D-C$).

Calculation of electrostatic surface potential was done using the program GRASP (Nicholls *et al.*, 1991). Protons were attached at appropriate stereochemical positions to perform all atom-based calculations. Partial atomic charges were assigned from the AMBER forcefield (Weiner *et al.*, 1984). Atomic sizes were calculated using standard radii file from the CHARMM suite (Brooks *et al.*, 1983).

3 Results and Discussion

3.1 Cryo-cooling techniques

The standard method in low-temperature crystallography of biological macromolecules is the introduction of an aqueous cryoprotectant solution into the crystal to prevent ice formation. This method has generally proven useful but requires a time- and crystal-consuming search for a suitable cryoprotectant and occasionally it fails altogether. Even flash-cooling using oil is not always successful in every case.

Although rarely mentioned in the literature, nearly every macromolecular crystallographer knows the phenomenon of well-shaped crystals of biological macromolecules diffracting X-rays only very poorly or not at all. At present, there is no accepted physico-chemical explanation for this, although it is probably safe to assume that these crystals must exhibit high internal disorder. Therefore, we developed a fast method to ‘rescue’ such crystals by a repeated-annealing method using paraffin oil. The other method is rather unconventional. It employs a non-aqueous cryoprotectant, which we call Panjelly™. Panjelly™ is Indian ghee, which is clarified butter (see section 2.2.4.2.5). This method seems to work optimally in combination with Panjelly™ prior to flash cooling and annealing in a subsequent step. More detailed description is given in the subsequent text.

3.1.1 Repeated-annealing using dried paraffin oil

Although cryoprotection by dried paraffin oil may pose less stress for protein crystals than soaking them in solutions of conventional cryoprotectants, the method does not necessarily lead to superior diffraction properties. Therefore, the protocol for repeated crystal annealing has been developed using dried paraffin oil. It is shown here that the method is particularly useful for crystals that diffract only weakly or not at all in an initial diffraction experiment. The method of repeated-annealing is different from the annealing techniques reported by Harp *et al.* (1998, 1999) and Yeh & Hol (1998). In the approach described by Harp *et al.* (1998, 1999), a crystal is first preincubated in cryoprotectant solution, then flash-cooled, and then again removed from the cryo-stream and transferred back to cryoprotectant solution. The crystal is equilibrated in the cryoprotectant solution by allowing it to remain there, at room temperature, for at least three minutes, and is then flash-cooled again. The flash-annealing method described by Yeh & Hol (1998) utilises the principle of thawing and annealing the crystal while on the loop, in a rapid cycle under the cold nitrogen stream (“*in situ* annealing”). A modification of this has been described as “annealing on the loop” (Harp *et al.*, 1999). It

involves a variable length of time for warming and does not require multiple rounds of warming and flash cooling.

We have developed an alternative protocol which consists of following steps: (i) no cryoprotectant at all is used during initial flash-cooling of the crystal; (ii) only dried paraffin oil as cryoprotectant (Riboldi-Tunncliffe & Hilgenfeld, 1999) is employed after initial flash-cooling; (iii) cooling and re-cooling of the protein crystal is performed in a repeated cycle; (hence the technique is called “repeated-annealing”); (iv) annealing is repeated at the temperature at which the crystals were grown. The first step in the protocol is highly unconventional, but it has been shown here that it is indeed crucial for ‘healing’ ill-diffracting crystals.

The procedure was applied to poorly diffracting crystals of three different proteins and a dramatic improvement of diffraction quality has been seen in all three cases. The approach has been also tried with crystals of four other proteins, which diffracted reasonably well from the outset; improvements were found minor in these cases.

The repeated-annealing procedure has been systematically studied using the available crystal systems listed in Table A1. All crystals used were grown by the hanging-drop technique, under various conditions, ranging from low to high salt concentrations, or from low to high molecular-weight polyethylene glycol. Detailed crystallisation conditions are also given in Table A1. Crystal sizes were between 0.2 and 0.6 mm in all dimensions.

The repeated-annealing procedure can be divided into five steps: (a) transfer of the crystal from the mother liquor into the cryo-stream; (b) transfer of the cooled crystal from the cryo-stream into dried oil at room temperature or 4°C; (c) washing off ice or mother liquor from the cooled crystal; (d) flash-cooling of the crystal; (e) recording the diffraction pattern. If the diffraction quality is not satisfactory, steps (b) to (e) should be repeated. In the following section, each step is explained in some detail.

(a) Transfer of the crystal from the crystallisation drop into the cryo-stream

The protein crystal should be transferred from the crystallisation drop directly into the cryo-stream with the help of a rayon loop (Teng, 1990). No dried paraffin oil or any other cryoprotectant is used at this stage. If a diffraction pattern were taken at this time, it would basically only show a strong ice ring, and perhaps some weak reflections from the protein crystal. (The importance of this step is discussed below).

(b) Transfer of the cooled crystal from the cryo-stream into dried paraffin oil

The cooled crystal is transferred directly, and as quickly as possible, from the cryo-stream to dried paraffin oil which is placed on a microscope cover slip, either at room temperature or at 4°C depending on the temperature of crystal growth.

(c) Washing off the ice or mother liquor from the cooled crystal

Once a mother liquor-containing crystal is cooled in the cryo-stream, it will form ice around the crystal. This is helpful for washing. As the crystal is brought into dried oil at room temperature or 4°C, the oil will diffuse onto the surface of the crystal between the ice and the crystal. Also, by this time, the ice will melt due to the increase in temperature. Moving the crystal in the dried oil helps to wash off the ice and/or water from its surface, and the crystal will become transparent. This is due to the refractive index of the crystal being similar to that of the oil. If the crystal edges can be recognized, aqueous mother liquor is still adherent to the surface and needs to be removed. This is an important prerequisite of successful flash-cooling or repeated-annealing of the crystal at 100 K. In addition to opaque appearance under the microscope, an indication for residual water or ice on the crystal surface is “jumping” of the crystal into the loop due to surface tension. The washing procedure is shown in Figure A1.

(d) Flash-cooling of the crystal

Once the crystal is free from mother liquor and/or ice, it should be transferred back to the cold stream as quickly as possible. It is sometimes hard to angle the mother liquor-free crystal from oil. Holding the loop directly below the center of gravity of the crystal, and moving steadily, but quite slowly, may help. Alternatively, two loops can be used to angle the crystal by fixing one loop and moving the other. One should attempt to make the oil drop adhering to the crystal as small as possible.

(e) Recording the diffraction pattern

Once the crystal is transferred to the cryo-stream, it is ready for the X-ray exposure. The diffraction pattern can be recorded using the desirable exposure time. However, the exposure time of X-ray should be constant throughout the repeated-annealing procedure to judge the improvement in diffraction pattern of crystal. If after the first or subsequent cycle of repeated annealing procedure; it is not satisfactory, then step (b) to (e) should be repeated.

Crystals of seven different proteins (or complexes between protein and nucleic acid), crystallised under widely different conditions (Table A1), have been subjected to repeated-annealing in dried paraffin oil. This approach has been compared to one-step flash-cooling in oil. Both sets of experiments were started with testing the diffraction properties of the crystals at 293 K, prior to the cooling experiments. The results are summarised in Table A2, where the diffraction limits and mosaicities after the application of the two cryo-cooling protocols are

compared with the same criteria for uncooled crystals. Of the seven crystalline systems tested, four exhibited very poor or no diffraction ($d_{\min} > 8/6 \text{ \AA}$) without cryo-cooling (at 293K).

Crystals of CPS-I, grown at 25°C, did not diffract at all at room temperature using $\text{CuK}\alpha$ radiation. They did diffract to 8.0 Å at the Elettra synchrotron source (XRD Beamline, Sincrotrone Trieste, Trieste, Italy). One-step flash-cooling using paraffin oil (Riboldi-Tunnicliffe & Hilgenfeld, 1999) did not improve the situation. However, after five rounds of repeated-annealing of the crystal, it diffracted to 6.0 Å in-house, and to 4.5 Å at the synchrotron, and a dataset was collected, with an average mosaicity of 0.7°.

Crystals of alliinase, grown at 4°C, also failed to show significant diffraction at both room temperature and after flash-cooling. A crystal was first transferred from the crystallisation drop to the cryo-stream and brought back to room temperature after 5 min, followed by washing in dried oil and again transferred to the cryo-stream. The first two rounds of repeated-annealing had little effect on diffraction. After the third round, it diffracted to 8 Å, and after the fourth round, the diffraction limit showed a dramatic improvement to 3.2 Å (using $\text{CuK}\alpha$ radiation). A complete dataset was collected, with a mosaicity of about 1.2°.

IF2-tRNA crystal (grown at 25°C), showed diffraction only to 6.0 Å at room temperature, various conventional cryoprotectants (i.e glycerol, sucrose) were tried, none of them proved to be useful for this crystal. After flash-cooling of the crystal using dried paraffin oil, it showed weak diffraction to 6.0 Å. After repeated-annealing, the crystal diffracted to 5.0 Å with smeared spots in the first round, and to 3.1 Å in the second round at the rotating anode and to 2.5 Å at the synchrotron.

Crystals of the *Brucella abortus* single-stranded DNA-binding protein (*BabSSB*) were grown at 4°C. They diffracted very weakly ($d_{\min} > 8 \text{ \AA}$) and were found to deteriorate slowly at room temperature. After applying the flash-cooling technique with dried paraffin oil as a cryoprotectant, the crystals diffracted to 2.5 Å, and a complete dataset was collected from a single crystal, with a mosaicity of about 0.4°. After data collection, repeated-annealing of the same crystal was tried. It was brought to room temperature and dipped into the paraffin oil and again transferred into the cryostream. The crystal diffracted to 2.2 Å on a rotating anode generator, but with a mosaicity of about 1.1°. Because of this, we changed the temperature at which the ice or mother liquor was washed off the crystal (another crystal), from 20 to 4°C. The crystal was brought back from the cryostream to 4°C, dipped into dried oil preincubated at this temperature, washed and transferred back into the cryo-stream. It diffracted to 2.2 Å as well, but the mosaicity was reduced to 0.3°.

An attempt to collect a full dataset at 293 K to 2.7 Å from one crystal of Des-PheB1 insulin in complex with phenol (Berchtold & Hilgenfeld, 1999), which had been grown at 278 K, failed because of severe radiation damage within 10 exposures of 15 minutes each (CuK α radiation from rotating anode generator). When flash-cooling in paraffin oil was applied, the crystals diffracted to 3.0 Å. After four rounds of repeated-annealing with washing of the crystal at room temperature, there was still no improvement. However, in the fifth round, the crystal diffracted to 2.8 Å in-house and to 2.4 Å at a synchrotron source (Beamline X11, EMBL Outstation at DESY, Hamburg), and a complete dataset was collected with mosaicity around 1.4°. When washing the crystal at 4°C instead of room temperature, the repeated-annealing procedure resulted in a diffraction limit of 2.5 Å (CuK α radiation), with a mosaicity of 1.0°.

The following experiments were carried out in order to examine the effect of repeated-annealing on better diffracting crystals. The *EcoSSB* crystal (Webster *et al.*, 1997) diffracted at room temperature to 2.8 Å resolution, using CuK α radiation. The repeated annealing technique in oil initially resulted in an improved diffraction limit of 2.7 Å, but with a high mosaicity of around 1.2°. Since this crystal had been grown at 4°C, we subsequently performed the washing step and the repeated-annealing at 4°C instead of 20°C. As a result, the crystal diffracted to 2.4 Å and sometimes even to 2.2 Å, and the mosaicity was found to be around 0.5°.

The crystal of green fluorescent protein (GFP), grown at 20°C, diffracted to 2.1 Å at room temperature, and its mosaicity was determined as 0.2°. The flash-cooled crystal diffracted to 2.4 Å, with a mosaicity of about 0.6°. After one round of repeated-annealing, the crystal diffracted to 2.2 Å with a mosaicity of about 0.4° (all experiments done using CuK α radiation).

With the exception of GFP, all protein crystals tried in this study showed improved resolution limits and mosaicities after several rounds of repeated annealing.

3.1.2 Flash-cooling using Panjelly™ and single-step annealing with paraffin oil

This method can be divided in five steps: (a) liquefying the Panjelly™, (b) transfer of the crystal from the mother liquor to Panjelly™, (c) fishing the crystal from Panjelly™ into the cryo-loop and transfer to the cryo-stream, (d) transfer of the crystal from cryo-stream to dried paraffin oil, (e) washing off Panjelly™ layer from the crystal and transfer of the crystal to the cryo-stream, and (f) recording the diffraction pattern.

(a) Liquefying the Panjelly™

It is a semisolid at room temperature. Therefore, Panjelly™ needs to be liquefied before using it for cryo-purpose. This can be done simply taking the some amount of Panjelly™ (enough to

form a drop of 200 μ l) on the microscope slide and warming it using a flame by placing a mini-burner underneath the slide. Panjelly™ will melt and become liquid. One should carefully warm and watch the Panjelly™ under the microscope such that some solid particles remain. In case one could not control the heating and the drop become completely liquefied, one can place the microscope slide on aluminium foil (for 60 to 90 seconds as the case may be) which is being indirectly cooled by liquid nitrogen. This can perhaps be done by placing an aluminium foil over an open Dewar flask containing liquid nitrogen. It is essential that Panjelly™ contains some solid particles. These solid particles are helpful in absorbing water (mother liquor) from the crystal surface. If one places the crystal in warm Panjelly™, the crystal will get a temperature shock, therefore after warming the Panjelly™, one should never forget to cool it a bit such that it is still semi liquid, and also not as much such that it starts to solidify again.

(b) Transfer of the crystal from the mother liquor to Panjelly™

The protein crystal should be transferred into the semiliquid form of Panjelly™ with the help of a cryo-loop. The semiliquid Panjelly™ contains small solid particles which easily trap the water (mother liquor) from the crystal surface and protect the crystal with a Panjelly™ layer from the open environment while it is being transferred to the cryo-stream.

(c) Fishing the crystal from Panjelly™ into cryo-loop and transfer to cryo-stream

It is easy to scoop up the crystal with a cryo-loop in Panjelly™, because Panjelly™ contains small solid particles, which reduce the surface tension between the crystal and the cryo-loop. Once the crystal is fished into the cryo-loop, it should be transferred as quickly as possible to the cryo-stream.

(d) Transfer of the crystal from the cryo-stream into dried paraffin oil

The Panjelly™ layer can be removed by bringing the cooled crystal back from the cryo-stream and washing it in the dried paraffin oil. If this step is omitted, the crystal will show good diffraction to high resolution, but a ring appears at 4 Å (Figure 18a). This ring does not obstruct the high-resolution spots during data collection, but during data processing, the resolution shell around 4 Å will be incomplete. The Panjelly™ ring is due to the Panjelly™ layer. Unlike the Panjelly™ ring, the ice ring at 3.6 Å may sometimes hinder the collection of needful reflection data beyond the ring, when flash-cooling is not properly performed by complete removable of the excess mother liquor.

(e) Washing off the Panjelly™ layer from the crystal and transfer of the crystal to the cryo-stream

The Panjelly™ layer is removed by washing the crystal in dried paraffin oil with the help of a cryo-loop (Figure 18b). Sometimes, this treatment improves the diffraction and mosaicity of the crystal. To remove the Panjelly™ layer from the crystal is easier than removing water from the crystal surface because water tends to adhere to the loop due to surface tension while the Panjelly™ layer doesn't show this behaviour. The crystal should be fished with the cryo-loop and transferred to the cryo-stream again.

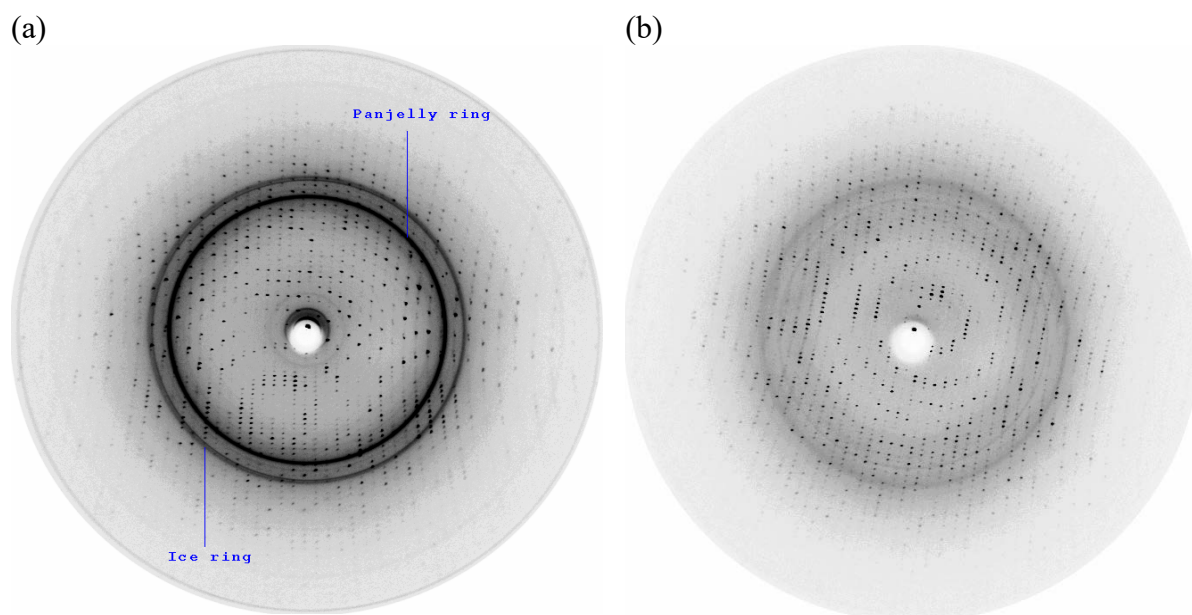


Figure 18 Diffraction pattern of *BabSSB* crystal (a) after flash-cooling using Panjelly™ and partial removal of mother liquor, (b) after treatment with dried paraffin oil.

(f) Recording the diffraction pattern

In this step, by varying the exposure times, one can check the diffraction pattern to see the maximum diffraction limits of the protein crystals.

The above-described method utilises the complete removal of water from the crystal surface than reported previously but in a more sophisticated way, which reduces any mechanical damage during the washing procedure.

The crystals of nine different proteins have been subjected to flash-cooling in Panjelly™ and subsequent single-step annealing in dried paraffin oil. This approach has been compared to flash-cooling/repeated-annealing in oil. The results are summarised in Table A3 where the diffraction limits and mosaicity are compared between the two cryo-cooling protocols.

A GFP crystal was used as a test case and to optimise the technique. The crystal was transferred into Panjelly™ and moved through it until its solid particles trapped the mother liquor from the crystal and then transferred to cryo-stream. The crystal diffracted to 1.6 Å, which was much better than the diffraction (2.2 Å) achieved by using repeated-annealing.

Afterwards, the insulin-mutant and alliinase crystals have been tried, which diffracted only on using the repeated-annealing technique as described in the section 3.1.1. Other conventional cryoprotectants had failed to improve the situation. These two crystals were flash-cooled successfully using Panjelly™ and diffracted to a higher diffraction limit (Table A3).

This approach has been applied to several other protein crystals and an improvement of diffraction properties was observed in all cases (Table A3). Therefore, for the present work, Panjelly™ was selected as the best cryoprotectant for SSB crystals.

There are three main advantages of the use of Panjelly™ over conventional cryoprotectants: (i) it improves the diffracting limits of the protein crystals, (ii) it saves time and crystals, and (iii) it is easy to use.

Due to the simplicity of this method, it is arguably that it could be tried as first choice in cryocrystallography, since it does not require the growth or soaking of crystals in solvent that could disturb the packing by diffusing into the crystal.

It is not exactly clear, why protein crystals have tendency to diffract to higher resolution limit when treated with Panjelly™ rather than any other cryoprotectant. However, it can be speculated that the Panjelly™ ring (ring at 4 Å) has some role in bringing the crystal lattice in the optimal orientation and/or removing all water from the crystal surface.

3.2 Proteolysis of SSBs

Crystallisation studies of *EcoSSB* (Ollis *et al.*, 1983) have shown that two of the four polypeptide chains in the tetramer have to be truncated at their C-termini by an unidentified proteolytic activity to obtain diffraction quality crystals. Truncated *EcoSSB* was prepared by introduction of a stop codon after the codon for amino acid number 152 by our collaborator. For crystallisation purposes, the protein solution was mixed with the intact polypeptide chains containing protein solution in a 1:1 ratio (Webster *et al.*, 1997). The gel of the *EcoSSB* protein solution and of a dissolved crystal is shown in Figure 19. The *PmiSSB*, *BabSSB* and *SmaSSB* were purified as intact polypeptide chains. The *PmiSSB* protein solution showed one band similar to intact *EcoSSB* and two bands around 14.5 kDa in the SDS gel. Similar bands were observed from the dissolved crystals of *PmiSSB*. Therefore, it can be concluded that there are three different species in the *PmiSSB* crystals, the intact polypeptide chain and polypeptide chains truncated by approximately 25 and 40 residues, respectively (Figure 19). The *BabSSB* protein solution also shows two bands similar to those found in *EcoSSB*. However, in contrast to *EcoSSB* protein crystals, the crystals of *BabSSB* show only a single 18 kDa band in the SDS gel (Figure 20). This is probably due to only one kind of species involved in the crystallisation process. The *SmaSSB* protein also showed a single band that

corresponds to molecular weight 18 kDa and no proteolysis was observed during the crystallisation process of this protein (Figure 21).

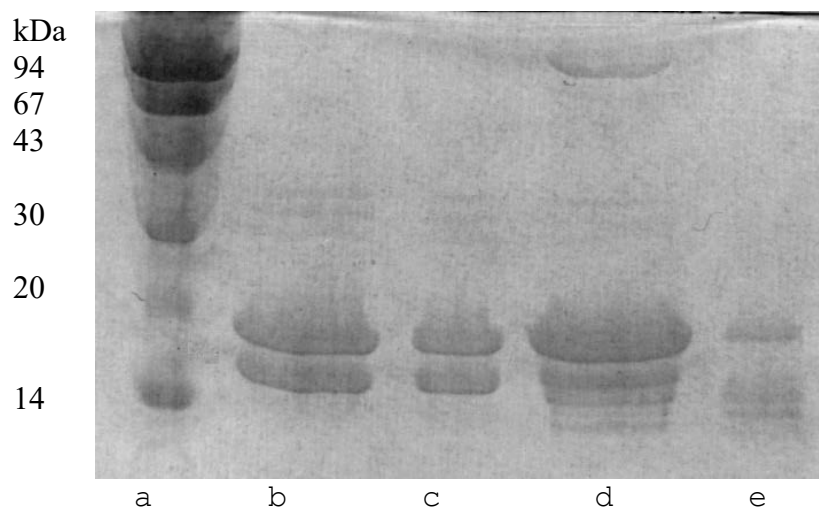


Figure 19 SDS-PAGE of SSBs (*EcoSSB*, *PmiSSB*) solution and of dissolved crystals. Protein was stained with Coomassie Blue, Lane (a) molecular weight marker (b) *EcoSSB* solution (c) dissolved crystals of *EcoSSB* (d) *PmiSSB* solution (e) dissolved crystals of *PmiSSB*.

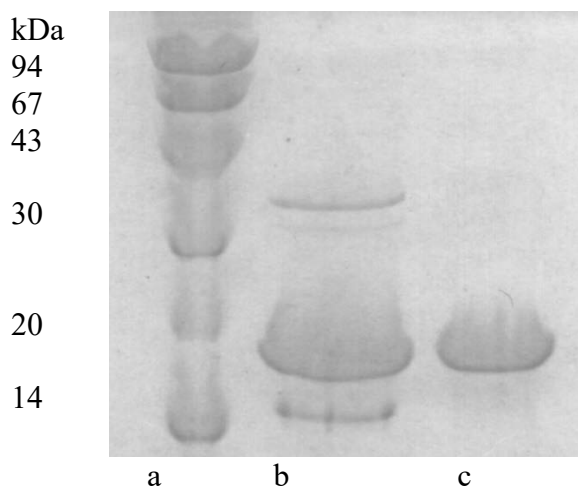


Figure 20 SDS-PAGE of *BabSSB* solution and of dissolved crystals. Protein was stained with Coomassie Blue, Lane (a) molecular weight marker (b) *BabSSB* solution (c) dissolved crystals of *BabSSB*.

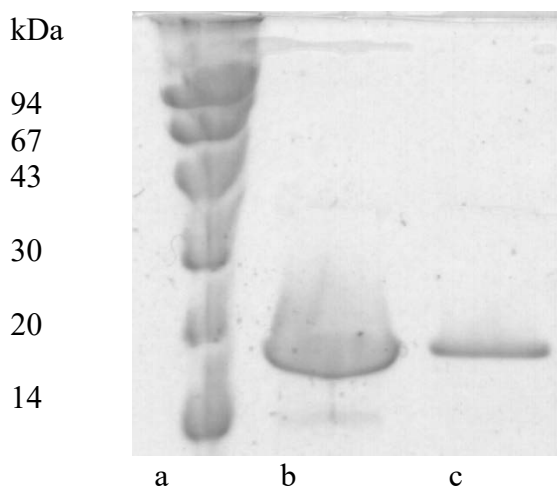


Figure 21 SDS-PAGE of *SmaSSB* solution and of dissolved crystals. Protein was stained with Coomassie Blue, Lane (a) molecular weight marker (b) *SmaSSB* solution (c) dissolved crystals of *SmaSSB*.

From the above observations, it appears that C-terminal truncation may not be a requirement for crystallisation of bacterial SSBs (i.e. *BabSSB*, *SmaSSB*).

3.3 Crystal structure of *EcoSSB*

3.3.1 Crystallisation of selenomethionine (SeMet) *EcoSSB*

In co-operation with the group of PD Dr. Claus Urbanke at the Medizinische Hochschule in Hannover, an engineered variant of *EcoSSB* has been produced, in which the four methionine

residues in the sequence have been substituted for SeMet, by using an auxotrophic expression system in which the bacteria are grown in a SeMet medium. Despite the protein yields being considerably lower than for the native system, the production was successful. The substitution of selenium was confirmed by mass spectroscopy.

Crystals of the SeMet variant of *EcoSSB* were obtained by equilibrating a 5 mg/ml solution of the protein against a reservoir containing 4% (v/v) PEG 400, 40 mM sodium cacodylate and 10 mM β -mercaptoethanol at pH 6.5 (Figure 22). The crystallisation conditions are similar to the ones of native *EcoSSB* (Webster *et al.*, 1997).

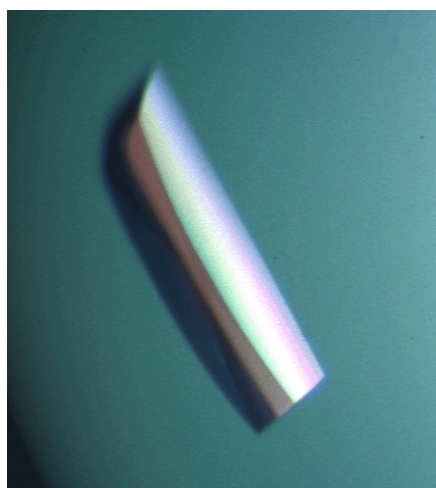


Figure 22 Monoclinic crystal of SeMet-*EcoSSB*

3.3.2 MAD and native data collection

The crystals were flash-cooled as described in section 3.1.2. The native dataset was collected at the European Molecular Biology Laboratory Outstation using the Deutsches Elektronen Synchrotron (EMBL/DESY), Hamburg beam line BW7A ($\lambda = 1.00 \text{ \AA}$). A MAD dataset from a SeMet crystal was collected at three different wavelengths, near or at the K absorption edge of selenium at the X-ray diffraction beam line ELETTRA (Trieste). Wavelengths were selected on the basis of an x-ray fluorescence spectrum of the crystal; λ_1 (0.9840 \AA), low-energy remote, λ_2 (0.9796 \AA), edge; λ_3 (0.9793 \AA), peak; λ_4 (0.9537 \AA), high-energy remote. All X-ray data were recorded at 100 K on Mar Research image plate scanner. Data processing was performed with DENZO and integrated intensities were scaled and merged by using SCALEPACK (Otwinowski & Minor 1997). Data collection statistics for the MAD datasets are summarised in Table 7 and for native dataset in Table 8.

3.3.3 Refinement of the Webster-*EcoSSB* structure to high resolution

The refinement was carried out against 95% of the measured data. The remaining 5%, which were randomly excluded from the full dataset, were used for cross-validation by calculating the free R-factor (R-free) to follow the progress of refinement (Brünger, 1992).

The same set of reference reflections was used throughout the refinement. The reference set was also excluded from the calculation of the electron density maps.

The starting model for the refinement was the room temperature structure (Webster *et al.*, 1997), which contains four molecules in the asymmetric unit. Loops (residues 23-27, 40-50, 85-100) were removed from each monomer before refinement. The volume of the unit cell decreased by 2% upon cooling to 100K and some cycles of rigid body refinement at 3.5 Å resolution were required to ensure an accurate repositioning of the four molecules in the asymmetric unit. The refinement was performed with CNS (Brünger *et al.*, 1998). In this program, bond distances and angles were restrained using the standard values suggested by Engh & Huber (1991). The refinement procedure included simulated annealing, B-factor and conjugate gradient energy minimisation against maximum likelihood targets as implemented in the program CNS (Brünger *et al.*, 1998). After each step, $3F_o-2F_c$ and F_o-F_c electron density maps were calculated and the model was visualised and rebuilt using the program O (Jones *et al.*, 1991). The rebuilding proceeded by systematically checking all the electron density peaks greater than 4σ in the F_o-F_c Fourier maps and building the missing residues which were removed during beginning of refinement; the sigma cut-off was gradually lowered during the refinement. Loops were built completely in electron density for the B- and D-monomer and partially for the A- and C-monomer.

3.3.4 Quality of *EcoSSB* model

The refined structure of *EcoSSB* (R_{free} 27.1% and R-factor 24.9%) consists of 3186 non-hydrogen atoms (monomer A: residues 1-24, 28-38, 50-89, 95-104, 106-113; monomer B: residues 2-112; monomer C: residues 2-22, 28-112; monomer D: residues 2-112) and 104 water molecules. The complete statistics of the *EcoSSB* model are shown in Table 9. The superposition of the C_α atoms (85-97 atoms) between monomers gave root mean square deviations (r.m.s.d.) varying between 0.29 and 0.70 Å (Table A4). The r.m.s.d. values from ideal bond length and bond angles for the whole tetramer are 0.006 Å and 1.29°, respectively.

The Ramachandran plot shows that 94.9% of the residues are found in the most favoured regions and 4.5% in the additionally allowed regions (Figure 23). Ala-44 of the B-monomer and the D-monomer (belong to Loop II) fall in the generously allowed and disallowed region, respectively, although these two residues are well defined in the electron density map. A Luzzati plot gives an estimated error on the co-ordinates of 0.3 Å (Luzzati, 1952). The model was also checked using an anomalous difference Fourier map.

Table 7 MAD data statistics

	Low-energy remote	Edge	Peak	High-energy remote	All Merged
Wavelength	0.9840	0.9796	0.9793	0.9537	
Completeness (%)	98.4	98.2	98.2	98.6	98.8
Mosaicity	0.59	0.59	0.58	0.59	0.58
Resolution	2.16	2.16	2.16	2.16	2.16
I/ σ (higher resolution bin)	2.21	3.36	2.76	2.22	2.36
Unique reflections	29425	29360	29367	29486	29546
Rejected reflections	264	273	438	270	171
R _{merge} (%)	4.4	4.4	4.5	4.5	5.7
R _{rim} (%)	5.3	5.3	5.4	5.4	6.0
R _{pim} (%)	2.9	3.0	3.0	3.0	1.7

Table 8 Native data statistics

	<i>Eco</i> SSB dataset1	<i>Bab</i> SSB dataset1	<i>Bab</i> SSB dataset2	<i>Pmi</i> SSB dataset1	<i>Sma</i> SSB dataset1	<i>Sma</i> SSB dataset2
X-ray source	DESY, BW7B	Rotating anode	ELETTRA	ELETTRA	Rotating anode	ESRF, BM14
Wavelength (Å)	1.000	1.54	1.000	1.000	1.54	0.9765
Space group	C2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P2 ₁ 2 ₁ 2	P4 ₂ 2 ₁ 2	P4 ₂ 2 ₁ 2
Unit cell (Å, °)	a=104.4, b=60.5, c=96.75 β =112.5°	a=b=113.3, c=52.2	a=b=113.3, c=52.2	a=146.7, b=153.2, c=61.0	a=b=111.25, c=141.76	a=b=111.25, c=141.76
Molecules/a.u.†	4	2	2	8	5	5
Resolution (Å)	100-1.9	100-2.5	100-1.78	100-2.5	100-3.5	100-2.8
Total reflections	395571	53985	103985	354872	114844	193700
Unique reflections	51359	12109	31807	48035	46617	22631
Mosaicity	0.6	0.4	0.3	0.4	0.4	0.4
Completeness (%)	96.5	98.3	95.3	98.9	97.9	99.5
I/ σ (higher resolution bin)	2.2	10.9	3.9	2.1	2.1	2.9
R _{merge} (%)	4.8	6.7	5.5	7.4	19.0	8.7
R _{rim} (%)	6.1	3.5	6.5	8.1	21.1	9.3
R _{pim} (%)	2.3	7.6	2.5	3.5	9.5	4.3

† a.u. asymmetric unit

Table 9 Refinement statistics of SSBs

Results and discussion

	<i>EcoSSB</i>	<i>BabSSB</i>	<i>PmiSSB</i>	<i>SmaSSB</i>
Resolution range (Å)	100-1.95	100-1.8	100-2.5	100-3.0
R-factor (%)	24.9	25.1	24.4	26.5
R _{free} (%)	27.1	26.2	29.3	29.8
Number of protein atoms	3186	1482	5908	3636
Number of water atoms	104	113	137	8
Main chain average B-factor (Å ²)	44.8	37.7	54.0	54.3
Side chain average B-factor (Å ²)	42.5	38.0	49.2	48.9
Water average B-factor (Å ²)	43.6	45.1	49.7	33.0
r.m.s deviation from ideal values‡				
Bond length (Å)	0.006	0.005	0.007	0.01
Bond angle (°)	1.29	1.35	1.36	1.51
Residues distribution in Ramachandran plot †				
Most favourable (%)	94.4	95.1	92.9	87.8
Allowed (%)	5.6	4.9	7.1	11.4

‡ Engh & Huber, 1991.

† Ramachandran & Sasisekharan, 1968; Laskowski *et al.*, 1993.

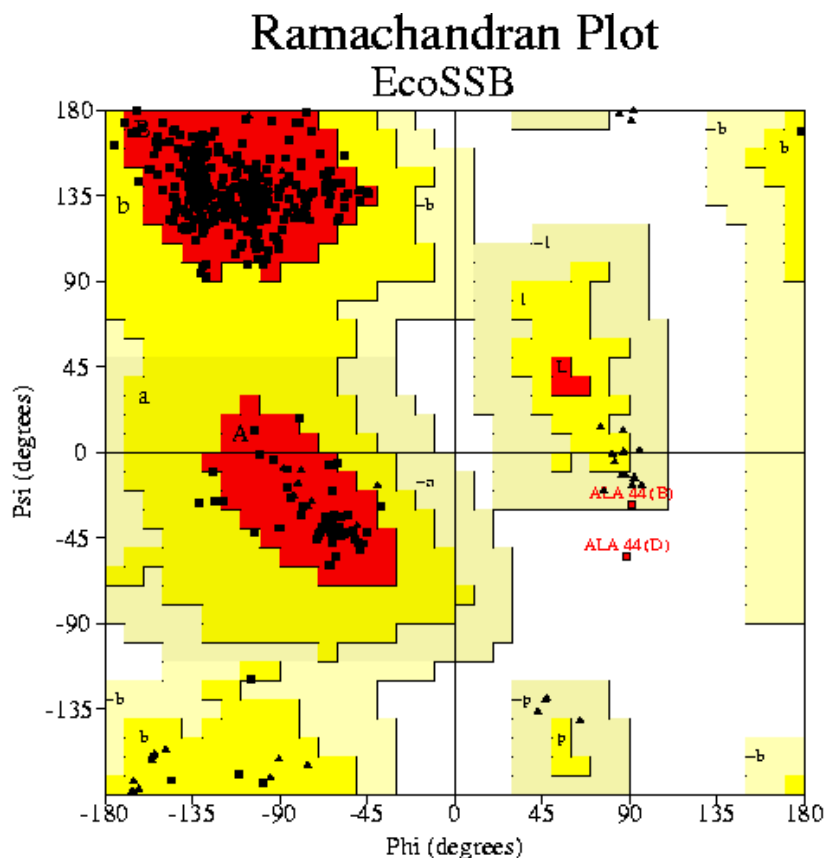


Figure 23 Ramachandran plot for *EcoSSB*, created by PROCHECK (Laskowski *et al.*, 1993). Glycine and proline residues are denoted by triangles and all other residues by squares. The different demarcated regions are labelled as A, B, L, - most favoured regions; a, b, l, p, -allowed; and -a, -b, -l, -p, -generously allowed.

The anomalous map was created using the observed anomalous differences and the phases from the model. To calculate the anomalous difference of selenium, the MAD dataset (Table 7) was used. The anomalous difference Fourier map showed 15 peaks greater than 3.5σ which correspond to selenium atoms of selenomethionine, indicating the correctness of the structure.

3.3.5 Comparison with the Webster-, Raghunathan- and Matsumoto-*EcoSSB* model

The present *EcoSSB* model was compared to the Webster-, Raghunathan- and Matsumoto-*EcoSSB* model and the resulting r.m.s.d. values based on C_{α} atoms were 3.4 Å, 1.0 Å and 1.9 Å, respectively. The significant r.m.s.d. to the Webster- and Matsumoto-*EcoSSB* model is due to a registration error starting at residue Ser-92 in each monomer. The structures reported by Webster *et al.* (1997) and Raghunathan *et al.* (1997) did not report the identification of the C-termini; however, Matsumoto *et al.* (2000) claimed to resolve the termini to 2.2 Å and illustrated the identification of C-termini (upto residue Gln-140), which have different conformation in all four monomers. The proposed model is inconsistent with our structure factors. Even the detailed analysis of vis-à-vis their and our model or structure factors (extracted from the PDB) did not substantiate this finding. If the structures are compared, the

present model shows 96% in the most favourable region of the Ramachandran plot, whereas 74% of the Raghunathan and Matsumoto *EcoSSB* model and only 43% of Webster *EcoSSB* model are in the most favourable region of Ramachandran plot. Therefore, it can be concluded that quality-wise, the present model is much better than the other three structures. Hence, the present *EcoSSB* model will be discussed in detail and considered as a reference model for comparison with other SSB structures.

3.3.6 Overall structure of *EcoSSB*

Structure of *EcoSSB* monomer and dimer

The N-terminal domain (residues 1 to 112) (Figure 24) of each monomer consists of six β -strands and a small helix which is connected by a third and fourth β -strand, whereas the C-terminal domain (from residue 112) is disordered. The core of the monomer structure is hydrophobic in nature. There are two extended loops which form β -hairpins. Density for Loop II (residues 40-50) and Loop III (residues 85–100) were well defined in three subunits (A-, B- and D-monomers) but these parts were poorly defined in the C-monomer.

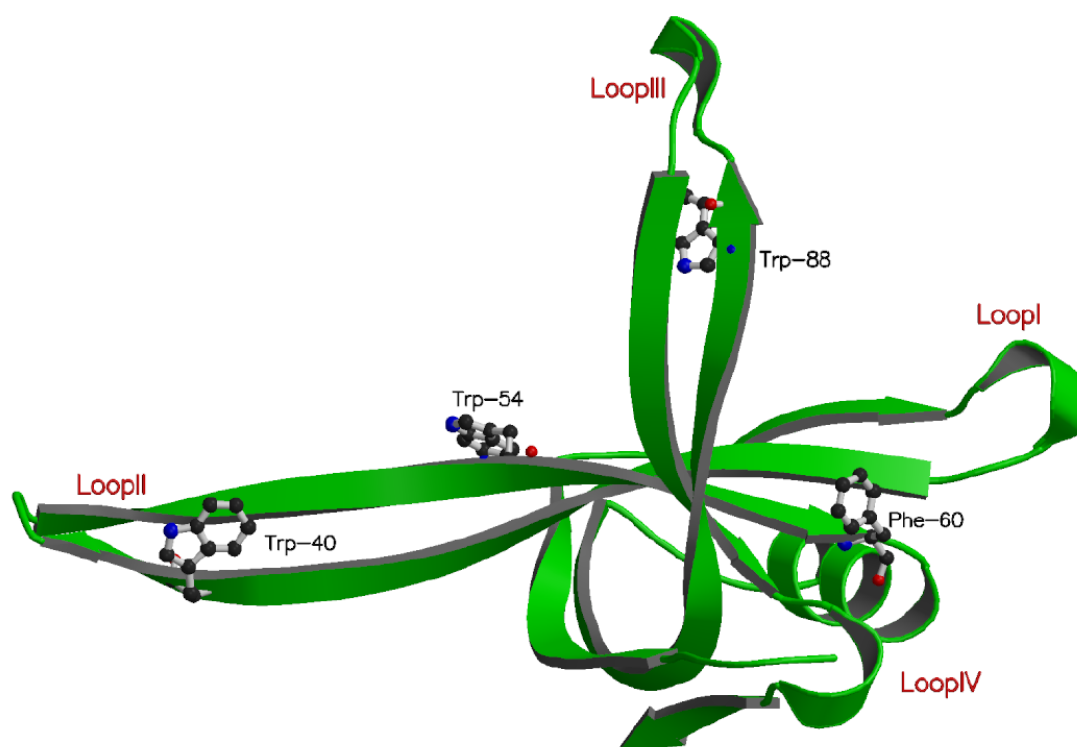


Figure 24 Ribbon diagram of the *EcoSSB* B-monomer showing (ball-and-stick) the aromatic residues involved in DNA-binding.

The *ssb-3* mutant, a Gly-15 to Asp mutation, renders the cell extremely sensitive to UV (Schmellik-Sandage *et al.*, 1990). In the model of the *EcoSSB* structure, the backbone of Gly-15 is only 3.6 Å from the side chain of an essential Trp-54 residue that is one of the key determinants of ssDNA-binding.

The mutation would produce a striking clash with the Trp (Webster *et al.*, 1997). Recently, the *Eco*SSB-DNA complex structure was published and it was shown that the Gly-15 is within 3.5 Å of the phosphate backbone of DNA (Ragunathan *et al.*, 2000), therefore the mutation to Asp may also sterically hinder ssDNA-binding.

***Eco*SSB dimer**

The monomer-monomer interface consists primarily of main chain hydrogen bonds involving two intermolecular antiparallel β -strands (residues 5-11). A large β -pleated sheet is formed with six antiparallel strands. In a front view, the shape of the large β -pleated sheet is flat (Figure 25); the shape appears like an arch in a side view. The total surface area buried upon dimer formation is 2180 Å². The two monomers in the dimer are related by a 2-fold axis. This conformation is probably beneficial to the formation of the homotetramer.

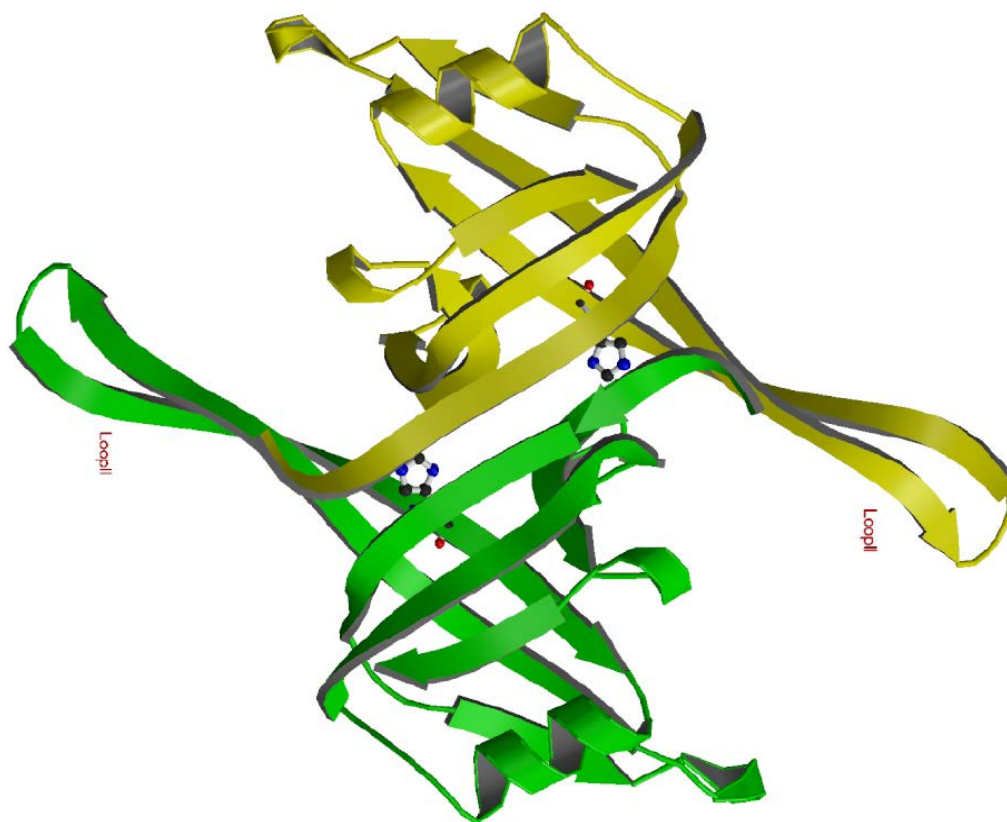


Figure 25 Ribbon diagram of an *Eco*SSB CD-dimer. Residue His-55 involved in tetramerisation, is shown as ball-and-stick model for both monomers.

His-55 has been shown to be involved in tetramerisation. Although substituting His-55 with phenylalanine or isoleucine does not alter the properties of protein, substituting with a tyrosine or lysine causes the tetramer to dissociate into monomers or dimers (Curth *et al.*, 1991). Another residue (which may also be involved in the tetramer destabilisation) is Tyr-78.

When substituted by Arg, *EcoSSB* forms dimers (Ute Curth personal communication). This can be understood in the light of the present structure.

His-55 of a monomer is involved in hydrogen bonding contacts with the main chain of Leu-34, the side chain of Thr-36 within the same monomer and water W31 (Figure 26). The water molecule forms a hydrogen bonding contact with other two water molecules (W45 and W21), which are involved in an extensive hydrogen-bonding network. These water molecules continue the hydrogen bonding network and mediate the contacts between His-55 and Glu-53, Trp-54, Thr-99, Tyr-97 from the same monomer as well as Thr-85, Thr-99, Tyr-97 from the other monomer. The hydrogen bond network of water molecules and residues involved in contacts is shown in Figure 26 and hydrogen bond distances are listed in Table 10.

His-55 is tightly enclosed in a hydrophobic environment. Therefore destabilisation of the monomer-monomer interface by substitution with the bulkier tyrosine side chain could result from increased steric hindrance and/or changes in the nature of contact surfaces.

Tyr-78 is involved in hydrogen bonding contacts with the side chain of Gln-110 and Glu-80 within the same monomer. This residue is situated in the hydrophobic pocket, which consists of Leu-10, Val-11 and Val-101. In this case, destabilisation of the dimer-dimer interface by substitution with Arg could result from a direct clash with the equi-positioned Arg side chain of the adjacent monomer. Therefore, there is a chance of a conformation change occurring (i.e. tetramer to dimer).

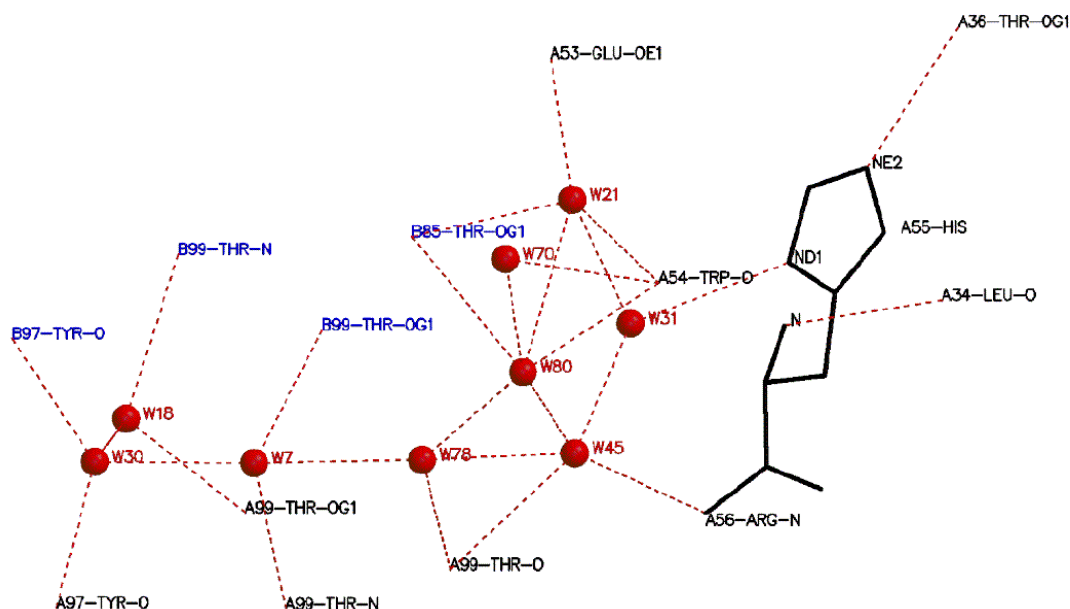


Figure 26 Schematic diagram of the hydrogen-bonding network at the monomer-monomer interface of the *EcoSSB*.

Table 10 Hydrogen-bonding distances between various donor-acceptors atoms as shown in Figure 26

Donor	Acceptor	H-bond distance (Å)
His-A55-N	Leu-A34-O	2.73
Leu-A34-N	His-A55-O	2.87
His-A55-NE2	Thr-A36-OG1	2.88
His-A55-ND1	Wat-W31-O	2.84
Wat-W21-O	Wat-W31-O	2.81
Wat-W21-O	Wat-W80-O	3.50
Wat-W21-O	Glu-A53-OE1	2.47
Wat-W21-O	Trp-A54-O	3.03
Thr-B85-OG1	Wat-W21-O	2.76
Wat-W45-O	Wat-W80-O	2.90
Wat-W70-O	Wat-W80-O	2.58
Wat-W70-O	Trp-A54-O	2.90
Wat-W78-O	Wat-W80-O	2.46
Wat-W80-O	Trp-A54-O	2.76
Thr-B85-OG1	Wat-W80-O	3.48
Arg-A56-N	Wat-W45-O	3.01
Wat-W45-O	Thr-A99-O	2.97
Wat-W78-O	Thr-A99-O	2.91
Wat-W7-O	Wat-W78-O	2.92
Wat-W45-O	Wat-78-O	2.91
Thr-A99-N	Wat-W7-O	2.96
Thr-B99-OG1	Wat-W7-O	2.89
Wat-W7-O	Wat-W30-O	2.68
Wat-W18-O	Wat-W30-O	2.82
Wat-W30-O	Tyr-A97-O	2.69
Wat-W30-O	Tyr-B97-O	2.61
Thr-A99-OG1	Wat-W18-O	2.80
Thr-B99-N	Wat-W18-O	2.91

Architecture of the *EcoSSB* tetramer

EcoSSB is known to form tetramers in solution (William *et al.*, 1984). Two different types of tetramers are possible. Which of the two possible tetramers configurations observed in the crystal reflects the configuration of *SSB* tetramer in solution? Figure 27a shows a tetramer of type I, which consists of two dimers sharing a non-crystallographic interface formed by two Loop II (residues 85-100) contacts, while Figure 27b shows a type II tetramer consisting of two dimers facing each other.

In a type I tetramer, two dimers come together in an offset tail-to-tail arrangement. One Loop III per dimer is involved in forming the parallel dimer-dimer interface and interacting with Loop II from the other dimer. The Loop III mediated dimer-dimer interface extends over a surface area of about 700 Å². This tetramer configuration is seen in the asymmetric unit of the crystal.

In a type II tetramer, the six-stranded β -sheet mediated tetramer, two dimers with a non-crystallographic two-fold axis, form a head-to-head tetramer with D₂ symmetry. This tetramer

configuration has been also reported for HsmtSSB (Yang *et al.*, 1997). The amount of contact surface between dimers (1843 \AA^2) is greater than that observed for the type I tetramer configuration. The structure make-up of the interface is different. Contacts involve mainly residue side chain rather than main chain atoms.

The interface is mixed in its physico-chemical composition with a very limited hydrophobic core surrounded by a network of hydrogen bonded tyrosine, glutamic acid, lysin, glutamine, and valine residues (Lys-7, Tyr-78, Glu-80, Val-5, Gln-110). Lys-7 and Glu-80 form a salt bridge, a feature also observed in the HsmtSSB structure (Yang *et al.*, 1997). It is more likely that this configuration of tetramers represents the *Eco*SSB tetramer in solution.

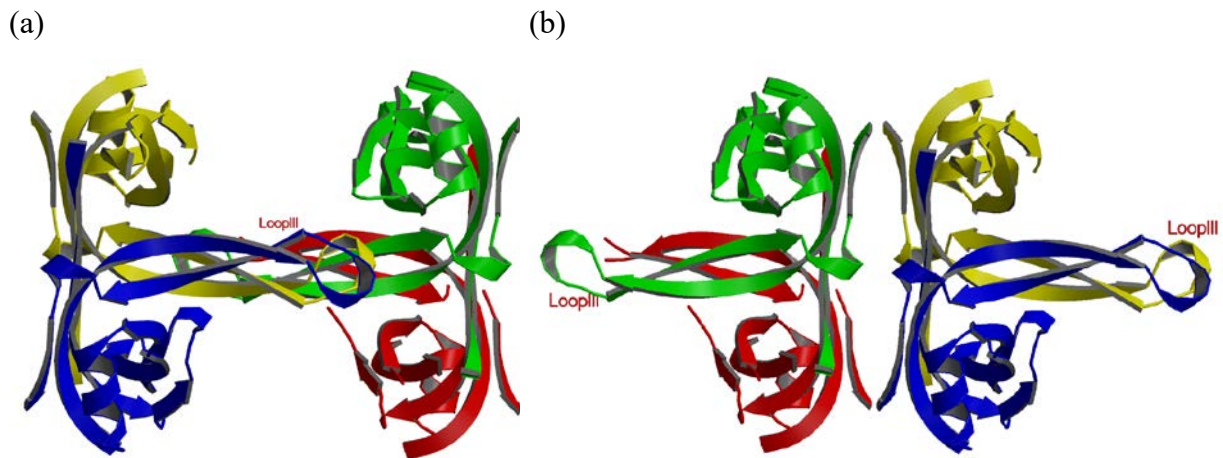


Figure 27 Ribbon diagram of two types of the *Eco*SSB tetramers (a) Type I tetramer (Loop III mediated) (b) Type II tetramer (six stranded β -sheet mediated).

Residues involved in DNA-binding

Chemical modification studies suggested that lysine and tryptophan residues are important for ssDNA-binding, whereas surface arginine and tyrosine residues do not seem to play a major role (Chase *et al.*, 1983; Anderson *et al.*, 1975). The implication of tryptophan residues is consistent with the high degree of SSB tryptophan fluorescence quenching upon binding of ssDNA (Lohman *et al.*, 1985; Overmann *et al.*, 1988; Anderson *et al.*, 1975; Bandyopadhyay *et al.*, 1987). There are four tyrosine and four tryptophan residues per SSB monomer (Sancar *et al.*, 1981), however, as most often is the case, the fluorescence emission spectrum of the protein shows contributions only from tryptophan (Bandyopadhyay *et al.*, 1987). Spectroscopic evidence suggests that both Trp-40 and Trp-54 are involved in stacking interactions with the nucleic acids bases (Khamis *et al.*, 1987). However, the possibility that mutations at these residues may also decrease the stability of SSB tetramers in a manner similar to the *ssb-1* mutation at residue His-55 has not been addressed. UV-crosslinking of *Eco*SSB to oligonucleotides, e.g. (dT)₈, has been observed to occur almost exclusively at

phenylalanine-60 (Merrill *et al.*, 1984). Recently, the structure of the *Eco*SSB-DNA complex has been solved which gives a detailed insight about the residues which are involved in DNA-binding (Raghunathan *et al.*, 2000). It confirms that the residues described above are involved in DNA-binding.

3.3.7 Crystal Packing of the *Eco*SSB structure

The *Eco*SSB crystals contain four monomers in the C2 asymmetric unit (Figure 27). The four monomers (A-, B-, C-, and D-monomer) are related by non-crystallographic symmetry (NCS). A superposition of any two monomers in the asymmetric unit gives r.m.s.d. values from 0.29 to 0.70 Å on the basis of C α pairs (85-97 atoms) (Table A4). The largest difference for main chain atoms (> 1.3 Å) is confined to the regions of Loop I (residues 23-27) and Loop II (residues 40-50) in each monomer. This large discrepancy may be a consequence of the different crystalline environment. Loop II of the D-monomer is in close proximity to Loop I of the C-monomer (Figure 28, 53a). Interestingly, Loop III (residues 85-100) shows high B-factors as does Loop I and II (Figure A2); however, there is little structural difference among the monomers in this area. Loop III from the B-monomer and the C-monomer is involved in the dimer-dimer interface of the Type I-tetramer (Figure 27a, 53b). Loop II of the B-monomer is in crystal contact with Loop III of the D-monomer (Figure 53c). Detail intermolecular hydrogen-bonding contacts are given in Table A14.

In the other parts of the chain the magnitude of deviation is roughly what is expected from the estimated co-ordinate errors and the structural difference between the four independent monomers in the asymmetric unit should be regarded as only negligible.

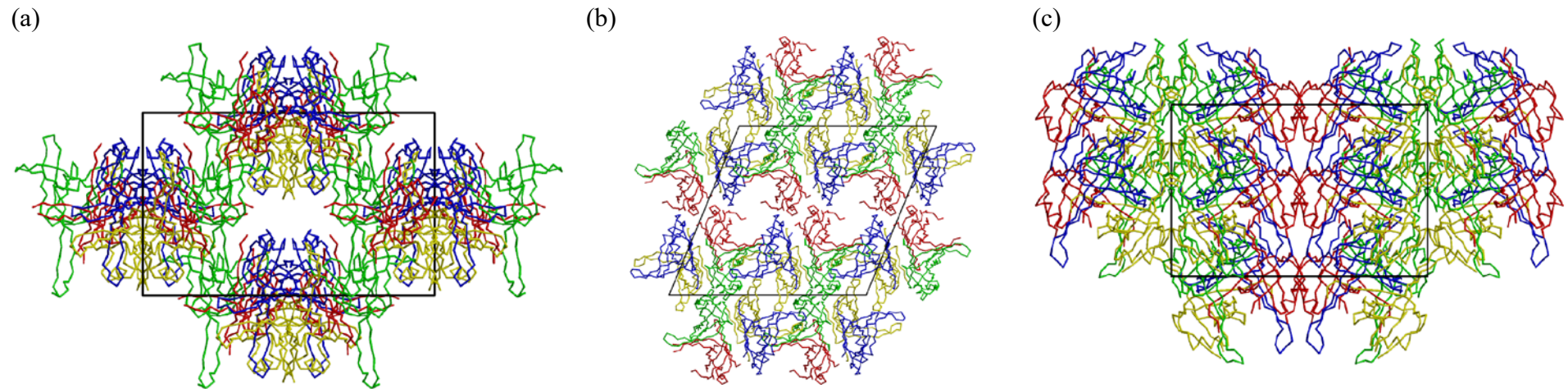


Figure 28 Crystal packing of the *EcoSSB*, each monomer shown in different colour. The unit cell is shown in black. (a) View along the crystallographic z-axis, (b) view along the crystallographic y-axis (c) view along the crystallographic x-axis.

3.4 Crystal structure of *BabSSB*

3.4.1 Crystallisation of *BabSSB*

Crystallisation trials were conducted using the hanging-drop technique at a temperature of 4°C. An initial search for crystallisation conditions was undertaken using sparse matrix sampling (Jancarik & Kim, 1991). Crystals were obtained within two days using Crystal screen I (Hampton Research) solutions containing sodium acetate and sodium cacodylate. At this stage, crystallisation conditions were optimised in order to improve the size of the crystals by decreasing the concentration of sodium acetate and using the additive screen I (Hampton Research). Slightly bigger (0.2x0.1x0.1 mm³) crystals were obtained in the presence of dioxane in hanging drop. However, a high number (~100) of crystals were obtained in the drop. In order to control the nucleation, vapour-diffusion experiments were set-up applying a 0.3 ml oil layer (Oil mixture, Table 4) above the reservoir solution (Chayen *et al.*, 1997). Such optimised crystallisation conditions were defined as 0.1 M sodium cacodylate, 12.5 mM sodium acetate pH 6.5 and 30% (v/v) dioxane in the presence of 17 mg/ml *BabSSB* in Protein storage solution B (Table 4). Under these conditions, three to four crystals were observed in each drop within 1 week. The crystals (Figure 29) were rod shaped and reached 1 mm in length and 0.25x0.25 mm² in cross section. Analysis by polyacrylamide gel electrophoresis of the protein from washed and dissolved crystals indicates that *BabSSB* crystallised as intact protein (Figure 20).

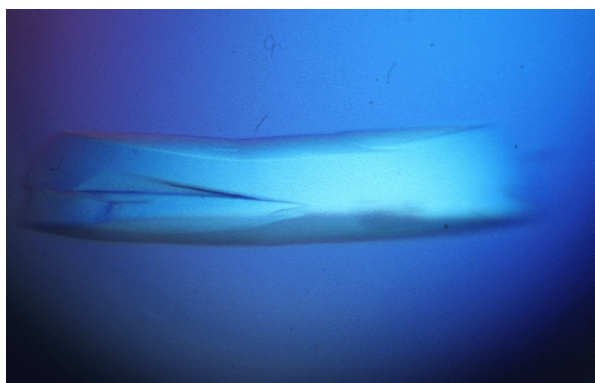


Figure 29 Tetragonal crystal of *BabSSB*.

3.4.2 Native data collection

For the collection of diffraction data at 100K, a *BabSSB* crystal was treated with dried paraffin oil as cryoprotectant. The crystal diffracted to 2.5 Å resolution on our rotating anode. To improve the diffraction quality, the crystal was treated with PanjellyTM before flash-cooling. The crystal diffracted to a maximum resolution 1.78 Å at a wavelength of 1.0 Å on the X-ray diffraction beam line at the ELETTRA synchrotron in Trieste, Italy. A complete dataset was collected using a Mar 18 cm image plate. All diffraction images were processed

using the HKL suite of programs (Otwinowski & Minor 1997) to yield a scaled set of diffraction intensities (Table 8). Analysis of the diffraction images showed that the crystals were tetragonal, belonging to space group $P4_32_12$ with unit cell dimensions $a = b = 113.13 \text{ \AA}$, $c = 52.10 \text{ \AA}$. A Wilson plot (Wilson, 1949) gave an estimate of the overall B-factor of 33 \AA^2 .

3.4.3 Structure determination of *BabSSB*

The structure of *BabSSB* was determined by the method of molecular replacement using the program AMoRe (Navaza, 1994) with a search model based on the 1.95 \AA resolution *EcoSSB* structure. Since from packing consideration (Matthews, 1968) the content of one asymmetric unit of *BabSSB* crystals was assumed to be two intact molecules, one monomer of the *EcoSSB* structure was used as a search model. *BabSSB_dataset1* (Table 8) was used for molecular replacement. Data in the range $10\text{-}4 \text{ \AA}$ were used in the rotation function search. A list of 50 rotation function peaks was obtained using a Patterson cut-off radius of 30 \AA with the top peak having the correlation coefficient of 0.118. The translation function gave two non-symmetric solutions with correlation coefficients of 0.156 and 0.257, respectively. During the data collection there was some ambiguity regarding the space group assignment for *BabSSB* crystals. In order to resolve this problem, the translation function search and rigid-body refinement were performed separately for space groups $P4_12_12$ and $P4_32_12$. The correlation coefficient for the latter space group was significantly higher, and remained so, after rigid-body minimisation using the program AMoRe (0.191 and 0.44, respectively), indicating that $P4_32_12$ was the correct space group. The molecular replacement solution was checked using the program O (Jones *et al.*, 1991) for good packing contacts before proceeding to refinement. The summary of the molecular replacement solution of *BabSSB* structure is given in Table 11.

Table 11 The molecular replacement procedure for *BabSSB*

Resolution range	10.0 - 4.0 \AA
Rotation & Translation function (1 st Monomer)	
Best solution	$\alpha = 36.75^\circ$ $\beta = 140.86^\circ$ $\gamma = 77.46^\circ$ tx = 0.1915 ty = 0.1867 tz = 0.1875
Correlation coefficient	0.156
R-factor	55.0%
Rotation & Translation function (2 nd Monomer)	
Best solution	$\alpha = 2.70^\circ$ $\beta = 56.94^\circ$ $\gamma = 326.68^\circ$ tx = 0.5926 ty = 0.6968 tz = 0.3885

Correlation coefficient	0.257
R-factor	51.8%
Refinement of combined solution:	
Monomer1:	$\alpha = 31.87^\circ$ $\beta = 138.09^\circ$ $\gamma = 73.72^\circ$ tx = 0.1891 ty = 0.1868 tz = 0.1864
Monomer2:	$\alpha = 6.59^\circ$ $\beta = 55.37^\circ$ $\gamma = 324.88^\circ$ tx = 0.5960 ty = 0.6939 tz = 0.3909
Correlation coefficient	0.440
R-factor	45.3%

3.4.4 Refinement and model building of *BabSSB* structure

The molecular replacement solution was subjected to rigid-body refinement using CNS (Brünger *et al.*, 1998) utilising the *BabSSB_dataset1* (Table 8) between 20-3 Å. This resulted in an R-factor 45.6%. For 5% of the reflections against which the model was not refined, the R_{free} was 46.8%. The first round of positional and temperature factor refinement lowered the R-factor to 40% and the R_{free} to 41.8%. At this stage, SigmaA-weighted maps were calculated using the program SIGMAA (Read, 1986) and careful examination of the maps allowed corrections to be made to the model. There was no density for the residues 40-50 and 85-100 in each monomer in the initial maps. Therefore, these residues were removed and the crystallographic refinement was continued against the high resolution data (*BabSSB_dataset2*, Table 8) using the maximum likelihood program REFMAC (Murshudov *et al.*, 1997) coupled to ARP (Lamzin & Wilson 1997). After each round of refinement the 2Fo-Fc and Fo-Fc map were calculated and the model was checked using the program O (Jones *et al.*, 1991) in order to build missing residues. All waters were deleted from the PDB file before the start of a new refinement. After 23 cycles of refinement, the R-factor was 25.7% and the R_{free} 28.3%. Various refinement protocols, including unrestrained REFMAC/ARP refinement, were used in order to localise the region 40-50 in chain B and 85-100 in chain A and chain B and few missing side chains. In the later stages, the model was refined in CNS including a bulk-solvent correction. The refinement protocol included the positional, B-factor and slowcool refinement. The model was inspected in the 3Fo-2Fc and Fo-Fc maps; all missing side chains were built. However, the regions 40-50 in chain B and 85-100 in both chains could not be localised in the density, suggesting that they are disordered. The final R-factor and R_{free} are 25.1% and 26.2%, respectively for all data within the range of 100-1.8 Å resolution. The refinement statistics are summarised in Table 9.

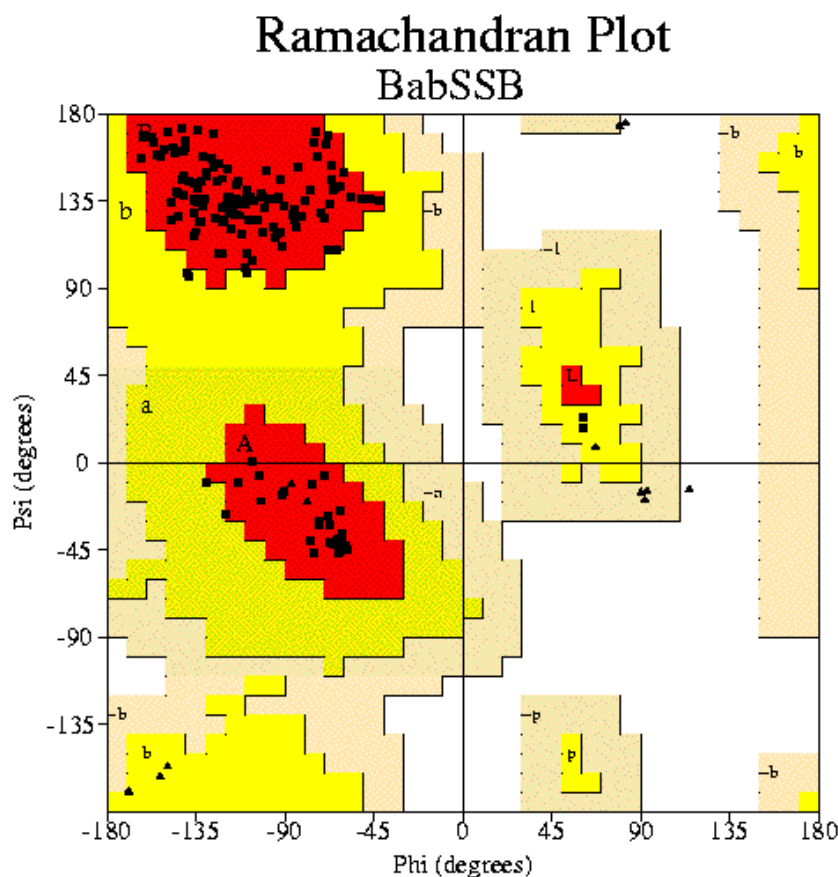
3.4.5 Quality of *BabSSB* model

Figure 30 Ramachandran plot for *BabSSB*, created by PROCHECK (Laskowski *et al.*, 1993). Glycine and proline residues are denoted by triangles and all other residues by squares. The different demarcated regions are labelled as A, B, L, -most favoured regions; a, b, l, p, -allowed; and -a, -b, -l, -p, -generously allowed.

The structure was refined to an R-factor of 25.1% for 29478 reflections. The R_{free} calculated for the set of 1262 reflections was 26.2 %. The model consists of two chains and has 1482 non-hydrogen atoms and 113 water molecules. Residues 1 and 2 as well as all residues beyond 115 are disordered.

The superposition of the 88 C_{α} atoms between monomers gave a r.m.s.d. value of 0.26 Å. The average B-factor of the main chain and side chain atoms is 37.7 and 38.0 Å², respectively. In the final model, 95.1% of the residues are located within the most favourable regions and 4.9% in the allowed region of the Ramachandran plot (Figure 30). A Luzatti plot gave an estimated error of the co-ordinates of 0.28 Å (Luzatti, 1952).

3.4.6 Comparison between the *BabSSB* and *EcoSSB* model

In contrast to the *EcoSSB* crystals, the asymmetric unit of *BabSSB* crystals contains two monomers. However, both proteins form homotetramer (Figure 30). The final model of *BabSSB* includes 189 residues in two polypeptide chains of *BabSSB*. Loop III (residues 85-100, *EcoSSB* numbering; Figure 11) in both monomers and Loop II (residues 40-50) in one monomer are ill defined. Therefore these loops were excluded from all structural comparison. The C-terminal domain (last 60 residues) is also disordered as in the *EcoSSB* structure.

The fold of the polypeptide chain in *BabSSB* is similar to that found in the *EcoSSB*. The *BabSSB* structure was compared with *EcoSSB* using the LSQKAB program of CCP4 suite of programs (Collaborative Computational Project, Number 4, 1994).

When 355 C α atoms of all residues of one tetramer, including all those involved in regular secondary structure are superimposed, an r.m.s.d. of 1.63 Å between *BabSSB* and *EcoSSB* is observed. The superimposed structures of *BabSSB* and *EcoSSB* are shown in Figure 31. A high r.m.s.d. can be observed in the region of Loop I (residues 22-28), Loop II (residues 40-50), a helix (residues 60-70) and a Loop IV (residues 103-107). When *EcoSSB* monomers were compared to each other, it was found that there is also some structural differences in the region of Loop I, II, III and IV (Figure 47). However, there is very little structural difference in the helix region. In the case of *BabSSB*, the helix region shows a high r.m.s.d. value (when compared to *EcoSSB*) due to the insertion of one residue between residue 62 and 63 (*EcoSSB* numbering) in the *BabSSB* structure.

The residue His-55 of *EcoSSB* is also conserved in *BabSSB* and in other SSBs (Figure 11). Corresponding to *EcoSSB*, His-55 also plays an important role in *BabSSB* structure to stabilise the tetramer. In this structure, His-55 is involved in hydrogen bonding contact with the main chain of Ile-34 and side chain of Thr-36 of the same monomer and with water W11. This water is in contact with another two waters (W33 and W72) which are involved in hydrogen bonding contacts with the main chain of Ser-56 and Lys-100 (in case of W33) and the main chain of Trp-54 as well as the side chain of Glu-53 (in case of W72). It can be seen that positions, orientations and contacts of these waters are very similar in both structures. However, in contrast to *EcoSSB*, these waters do not continue the hydrogen-bonding network at the monomer-monomer interface.

As described above (section 3.3.6), Tyr-78 is involved in stabilisation of the dimer-dimer interface of *EcoSSB* tetramer. This residue is also conserved in *BabSSB* (Figure 11). Similar to *EcoSSB*, Tyr-78 also plays an important role at the dimer-dimer interface.

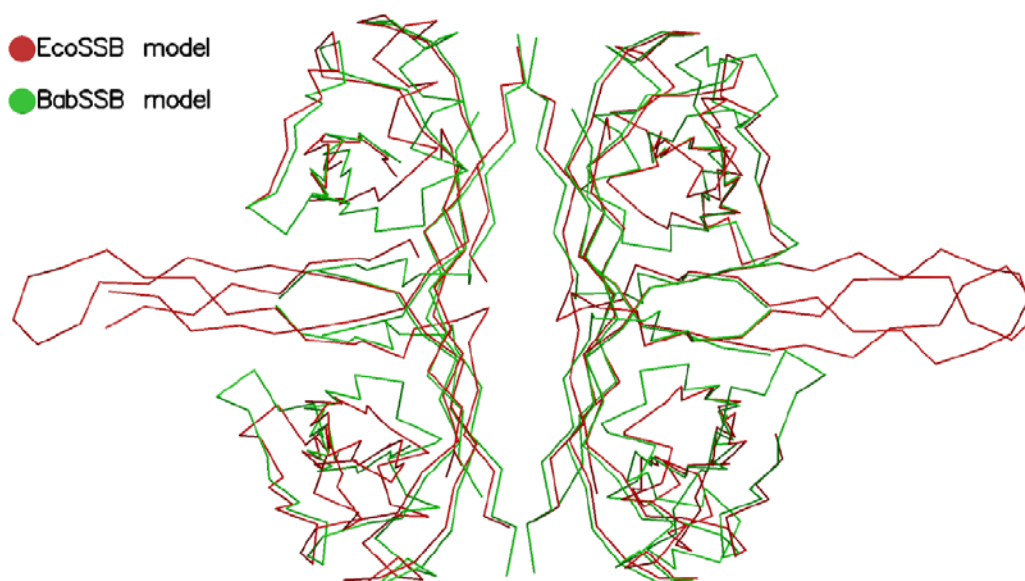


Figure 31 Superposed C α trace of the *BabSSB* and *EcoSSB* tetramers.

Four salt bridges exist at the dimer-dimer interface; two of them are symmetrical (Lys-A7:Glu-B81 and Lys-B7:Glu-A81). Lys-A7 makes three salt bridges, one with Glu-B81 and two with Glu-B110.

Table 12 Salt bridges in the *BabSSB* structure

Donor	Acceptor	Salt bridge (Å)
Lys-A7-NZ	Glu-B81-OE1	2.91
Lys-A7-NZ	Glu-B110-OE1	3.22
Lys-A7-NZ	Glu-B110-OE2	3.19
Lys-B7-NZ	Glu-A81-OE2	3.18
Lys-A49-NZ	Glu-B62-OE1	3.18
Lys-A49-NZ	Asp-B27-OD1	2.49
Arg-A108-NH2	Glu-A38-OE1	3.33
Arg-B108-NH2	Glu-B38-OE1	3.55
Arg-B22-NH2	Glu-A47-OE1	3.40

This is not the case with Lys-B7, which makes only one salt bridge with Glu-A81, as indicated by its different orientation in the electron density. There are a total of 9 salt bridges in this structure: they are listed in Table 12.

Gly-15 of *EcoSSB* is also conserved in *BabSSB* (Figure 11) and its mutation to Asp accordingly modulates the ssDNA-binding properties.

3.4.7 Crystal packing of the *BabSSB* structure

The *BabSSB* crystals belong to the space group P4₃2₁2. In this space group, the unit cell contains eight asymmetric units; hence there are eight symmetry operators (Table A13). The asymmetric unit of *BabSSB* crystals contains two independent monomers. These two molecules are related by non-crystallographic symmetry: a rotation by 90.3° followed by a translation 26 Å along the rotation axis. A superimposition of the two independent monomers

in the asymmetric unit gives an r.m.s.d. of 0.26 Å using 88 C_α atom pairs. In contrast to the *EcoSSB* monomers, *BabSSB* monomers show very little structural difference between two monomers including all loops.

BabSSB is probably also biologically active as a tetramer. It can be seen in the crystal packing that even though the *BabSSB* is a dimer in the asymmetric unit, it forms a crystallographic tetramer. In Figure 32, the A-monomer is coloured in red and B-monomer in green. Each monomer of *BabSSB* forms a dimer with its symmetry mate. The AA-dimer (A') and BB-dimer (B') are shown in red and green, respectively in Figure 32. These dimers are arranged in two specific ways. The first pattern is A'B'B'A' and the second is B'A'A'B'. A'B' forms a head-to-head tetramer where as B'B' or A'A' forms a tail-to-tail tetramer.

Tail-to-tail tetramers are arranged in such a way that their Loops III (modelled using the *EcoSSB* structure) come together (Figure 32). In the tail-to-tail arrangement, the tips of Loop III from the B'-dimer clash with the adjacent symmetry related B'-dimer and the distance between their centres of mass is 53.5 Å. However, this is not the case with the Loops III of the A'-dimer. The distance between the A'-dimer and its symmetry related A'-dimer from their centre of mass is 70.5 Å and its loop is in very close contact with the Loop III of the adjacent B-monomer. Therefore it can be suggested that the Loop III is flexible in this area, which is supported by the high B-factor of Loop III in *EcoSSB* (Figure A2). Due to the high flexibility of the Loop III, in the *BabSSB* structure, it is completely disordered (Figure 32).

Loop II (residues 40-50) is disordered in the B-monomer. It has a different crystalline environment than Loop II from A-monomer. It clashes with tip of the Loop I (residues 24-27) of symmetry mates of A-monomer (Figure 32). This indicates that Loop II of B-monomer is also mobile and hence it is disordered in the structure.

The ordered Loop II of the A-monomer and its symmetry mates are arranged in such way that they do not clash but interact via salt bridges. One of the salt bridges is between Lys-49 of this loop and Asp-27 of an adjacent symmetry mate of the B-monomer (Figure 53h). The area where Loops II of A-monomer and its symmetry mates are packed shows a relatively large empty space (Figure 32). The tip of the Loop IV (residues 105 –108) of the B-monomer is in close contact with the tip of Loop II of the A-monomer. Loop I interacts with its symmetry mates and also forms a few crystal contacts with Loop II (Table A15). The packing interactions involve mostly salt bridges or hydrogen bonds. Intermolecular hydrogen bonds are made mainly either between the protein atoms or protein atoms and water molecules (Table A15).

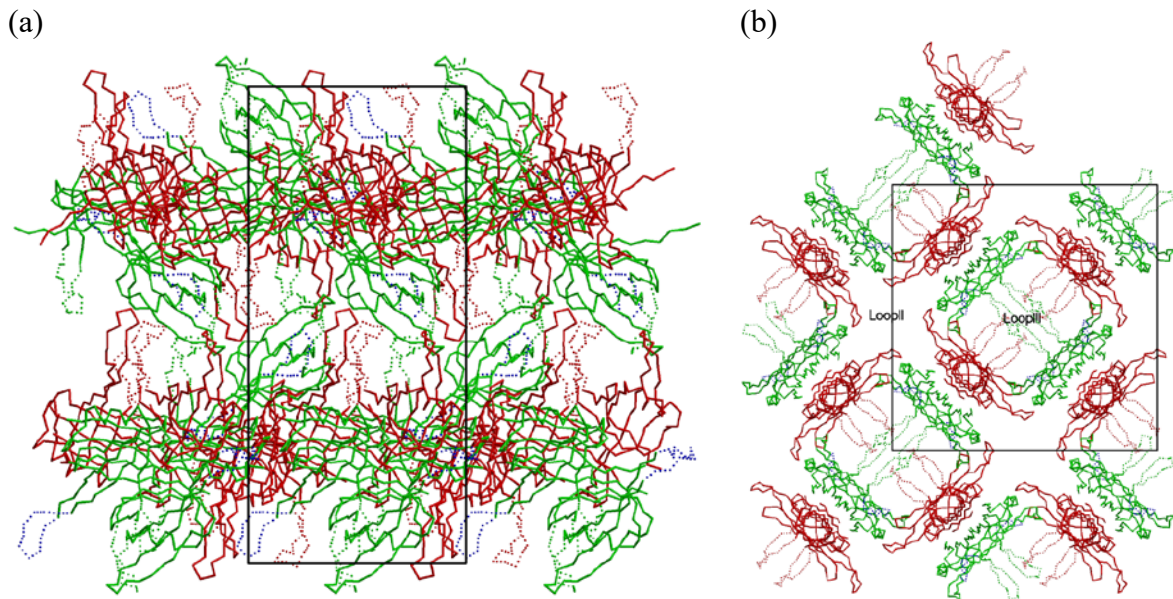


Figure 32 Crystal packing of *BabSSB*. The A- monomer is shown in red and the B-monomer in green, Loop III is disordered in both monomers, and is shown by a dotted line. Loop II of the B-monomer is also disorder and shown as dotted line in blue. (a) View along the crystallographic x-axis, (b) view along the crystallographic z-axis.

3.4.8 Model of *BabSSB*-ssDNA complex

The crystal structure of *BabSSB*-ssDNA complex has not been solved yet, therefore a theoretical model of this complex need to be built and examined. On the basis of the crystal structure of the *EcoSSB*-ssDNA complex (Ragunathan *et al.*, 2000), it can be proposed that *BabSSB* should also bind to ssDNA in a similar manner as *EcoSSB* binds to ssDNA. However some minor changes can be expected in the crystal structure of *BabSSB*-DNA complex. The model of *BabSSB*-ssDNA complex was prepared by superimposing the *EcoSSB*-DNA complex onto the *BabSSB* tetramer structure (Figure 33). When 428 C_{α} atoms of the tetramer are superimposed, an r.m.s.d. of 2.3 Å is obtained (for 295 C_{α} -positions of the tetramer excluding loops and helices, the r.m.s.d. value is 1.2 Å) between the *BabSSB* and the *EcoSSB*-ssDNA complex structure. As already mentioned in the Introduction section, most conserved aromatic residues and charged residues are involved in the ssDNA-binding. Those residues are also conserved in the *BabSSB* (i.e. Trp-40, Trp-54, Tyr-71, Arg-21, Phe-60; *BabSSB* numbering; Figure 51). They are involved in ssDNA-binding as can be corroborated from the theoretical model. Apart from these residues, there are a few more residues (Ser-37, Glu-38, Asp-50, Thr-52, Ser-56, Asn-63, Gln-70, Glu-101, Gln-105, Lys-106, Glu-110 and Asp-115; *BabSSB* numbering) in *BabSSB*, which are proximal to the ssDNA backbone, suggesting an involvement in DNA-binding. It is interesting to note, that Gln-105 and Lys-106 are not conserved in *EcoSSB* (Figure 51) and the residues at this place in *EcoSSB* are not

at all involved in ssDNA-binding. From this observation it can be explained, why there is an insertion of one residue between 104 and 105 (*EcoSSB* numbering, Figure 11) in *BabSSB*.

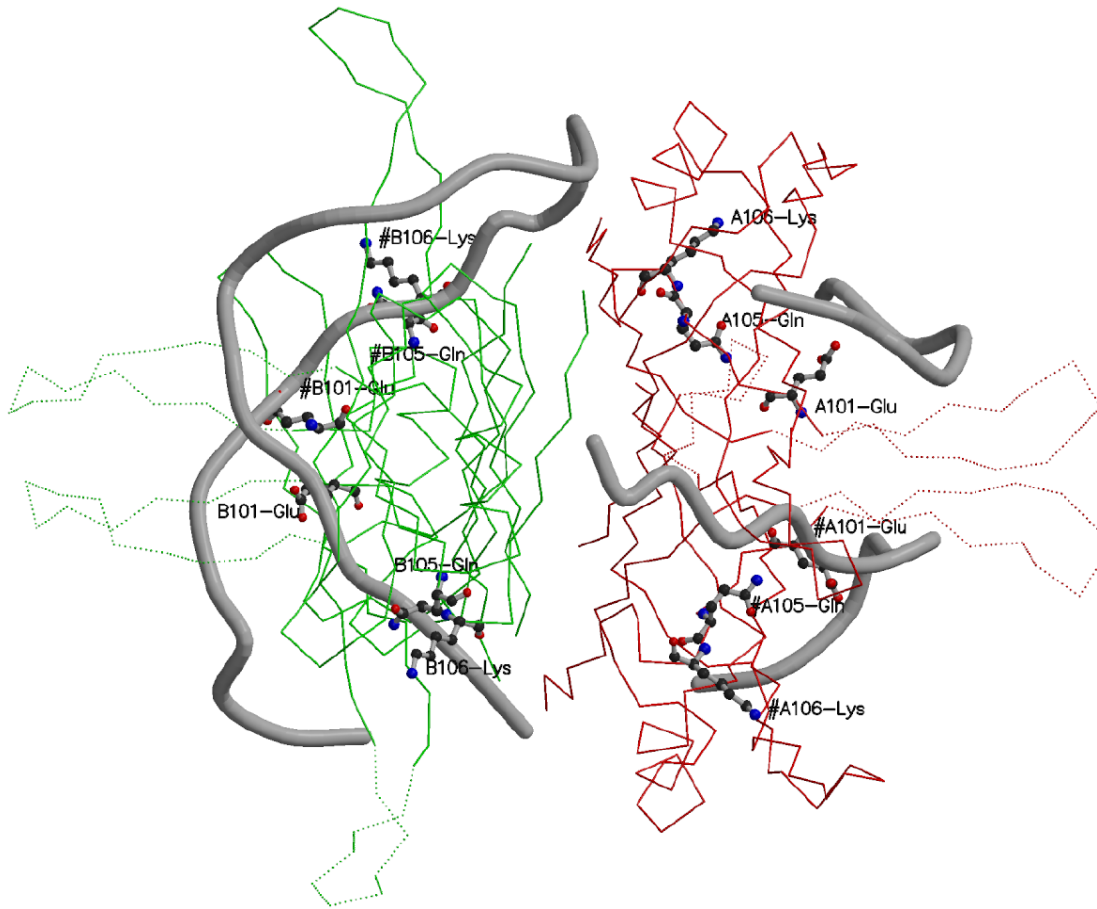


Figure 33 Theoretical model of the *BabSSB*-ssDNA complex. The C_{α} trace indicated by the dotted line shows that these parts are disordered in the structure. The colour convention is similar as for Figure 32. The non-conserved residues (Gln-105 and Lys-106) involved in the ssDNA-binding from each monomer are shown as ball-and-stick model.

These two residues (Gln-105 and Lys-106) belong to the Loop IV, which in *BabSSB* shows significant deviation compared to Loop IV of *EcoSSB* (Figure 47). This is probably due to the involvement of the loop in the *BabSSB* structure in ssDNA-binding.

3.5 Crystal structure of *PmiSSB*

3.5.1 Crystallisation of *PmiSSB*-(dT)₂

Crystallisation trials were conducted using the hanging-drop technique at 25°C. An initial search for crystallisation conditions was undertaken by using sparse matrix sampling (Jancarik & Kim, 1991). Crystals were obtained after five weeks in Crystal screen NATRIX HR2-116 (Hampton Research) solutions containing magnesium sulphate, MES and PEG4000. The crystallisation conditions were optimised in order to improve the size of crystals by decreasing the percentage of PEG4000. The optimised crystallisation conditions were found to be 5 mM magnesium sulphate, 50 mM MES pH 6.5 and 3% (w/v) PEG4000 at 16.5 mg/ml *PmiSSB* and 0.356 mM (dT)₂ (1:1 ratio) in Protein storage solution A (Table 4). Under these conditions, only three or four bigger crystals were observed in each drop after six weeks. The crystals were of rectangular shape and reached a size of 0.5x0.3x0.3 mm³ (Figure 34) Analysis by polyacrylamide gel electrophoresis of the protein from washed and dissolved crystals indicates that crystallised *PmiSSB*-(dT)₂ was partially cleaved (Figure 19).

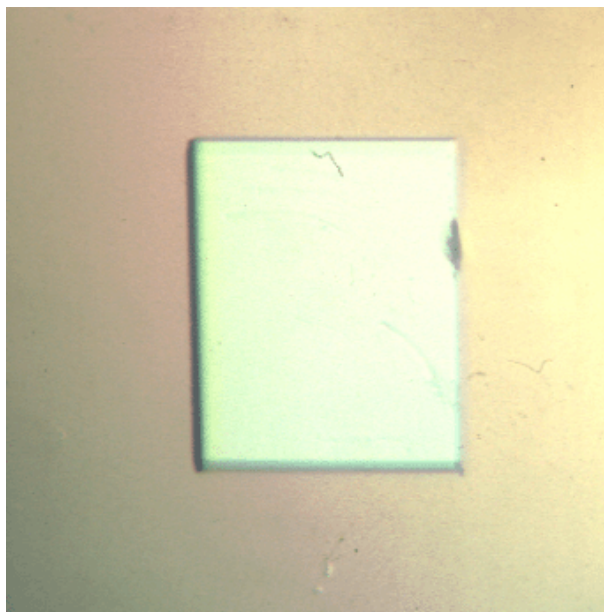


Figure 34 Orthorhombic crystal of *PmiSSB*-(dT)₂.

3.5.2. Native data collection

Crystals of *PmiSSB* were prepared for data collection at 100 K using Panjelly™. The crystal diffracted to 2.5 Å at the X-ray diffraction beamline, ELETTRA, Trieste. Data were collected at a wavelength of 1.0 Å using a Mar 30 cm imaging plate and processed using the HKL program package (Otwinowski & Minor, 1997), which yielded a scaled set of diffraction intensities (Table 8) with 98.9% completeness and an R_{merge} of 7.4%. The analysis of the

diffraction images showed that the crystals were orthorhombic, belonging to space group $P2_12_12$ with unit cell dimensions $a = 146.77 \text{ \AA}$, $b = 153.32 \text{ \AA}$, $c = 60.96 \text{ \AA}$.

3.5.3 Structure determination of *PmiSSB*

The program AMoRe (Navaja, 1994) was used for molecular replacement. Reflections with $F_0 > 2\sigma$ were used throughout the molecular replacement stage. Initial trials to solve the *PmiSSB* structure by molecular replacement were conducted using monomeric *EcoSSB* as a search model. This failed to give a prominent solution, as was the case with a dimeric *EcoSSB* model. This is not surprising since a monomeric/dimeric search model constitutes only a fraction of the scattering matter in the asymmetric unit. The tetrameric model of *EcoSSB* was employed successfully as a search model.

The *PmiSSB_dataset1* (Table 8) was used during the molecular replacement procedure. Data in the region 10-4 \AA were used for the rotation function and in the range 8-4 \AA for the translation function. Using a Patterson cut-off radius of 35 \AA , a list of 10 rotation function peaks was obtained, with the top peak having a correlation coefficient of 0.183. The translation function was calculated for all 10 rotation function peaks. The best solution obtained had a correlation coefficient of 0.196. Furthermore a second translation function search was carried out by fixing the first solution which yielded a solution with a correlation coefficient of 0.393. After performing rigid body refinement using the program AMoRe (Navaja, 1994) for two solutions, the correlation coefficient was 0.46. These two solutions were related by a 65.87° rotation along the z-axis, thereby indicating that the asymmetric unit contains two tetramers (solvent content of 49.4%). A summary of the molecular replacement solution is given in Table 13.

Table 13 The molecular replacement procedure for *PmiSSB*

Resolution range	10.0 - 4.0 \AA
Rotation & Translation function (1 st Tetramer)	
Best solution	$\alpha = 37.49^\circ$ $\beta = 80.84^\circ$ $\gamma = 204.01^\circ$ tx = 0.2369 ty = 0.4175 tz = 0.1173
Correlation coefficient	0.196
R-factor	53.6%
Rotation & Translation function (2 nd Tetramer)	
Best solution	$\alpha = 104.80^\circ$ $\beta = 80.64^\circ$ $\gamma = 204.58^\circ$ tx = 0.5747 ty = 0.3220 tz = 0.1599
Correlation coefficient	0.393

R-factor	48.4%
Refinement of combined solution:	
Tetramer 1:	$\alpha = 37.61^\circ$ $\beta = 80.52^\circ$ $\gamma = 203.67^\circ$ tx = 0.2371 ty = 0.4169 tz = 0.1172
Tetramer 2:	$\alpha = 103.40^\circ$ $\beta = 80.52^\circ$ $\gamma = 204.13^\circ$ tx = 0.5752 ty = 0.3222 tz = 0.1612
Correlation coefficient	0.46
R-factor	46.2%

3.5.4 Refinement and model building of *PmiSSB* structure

The output model from AMoRe was subjected to rigid body refinement using CNS (Brünger, *et al.*, 1998) using the *PmiSSB_dataset1* (Table 8) in the resolution range 100-3.0 Å. A random set containing 4% of the total data was excluded from the refinement, and the agreement between calculated and observed structure factors corresponding to these reflections (R_{free}) was used to monitor the course of the refinement (Brünger, 1992). The crystallographic free R-factor (R_{free}) was 40.4%. Data were then used in the refinement in the resolution range 100-2.5 Å. First, a round of positional and temperature factor refinement was performed which lowered the R-factor to 29.3% and the R_{free} to 34.4%. At this stage, sigmaA-weighted maps were calculated using the program SIGMAA (Read, 1986) and careful examination of the maps allowed corrections to be incorporated into the model. There was only weak density for the Loop I (residues 23-27), Loop II (residues 40-50) and Loop III (residues 85-100) in each monomer in the initial maps. Therefore, these residues from the loops region were removed from the model. Alternating cycles of manual rebuilding, conventional positional refinement and simulated annealing, using the slowcool protocol as implement in CNS (Brünger *et al.*, 1998), allowed some of the missing residues to be placed in the density.

3.5.5 Quality of *PmiSSB* model

The refined model of *PmiSSB* has an R-factor 24.4% for 46031 reflections to 2.5 Å resolution (no low-resolution and sigma cut-off was applied). The R_{free} calculated using 1946 reflections omitted from the refinement, is 29.3%. The model contains eight monomers and has 5908 non-hydrogen atoms plus 137 water molecules. The electron density is well defined for each chain. Loop I (residues 23-27) of the B-, C-, F- and G-monomer, the Loop II (residues 40-50) of all except the H-monomer as well as Loop III (residues 85-100) of the A-, C-, D-, and H-monomer show weak electron density. Only Loop I of the F-monomer, Loop II

of A-monomer and Loop III of the H-monomer could be partially build in the density and were included in the final model. The structure quality assessment together with relevant crystallographic statistics is reported in Table 9. Among 646 non-glycine and non-proline residues in the eight monomers in the asymmetric unit, the number of residues lying in the most favoured and additionally allowed region in the Ramachandran plot is 600 (92.9%) and 46 (7.1%), respectively (Figure 35). A Luzzati plot yields an estimated error on the coordinates of 0.4 Å (Luzzati, 1952).

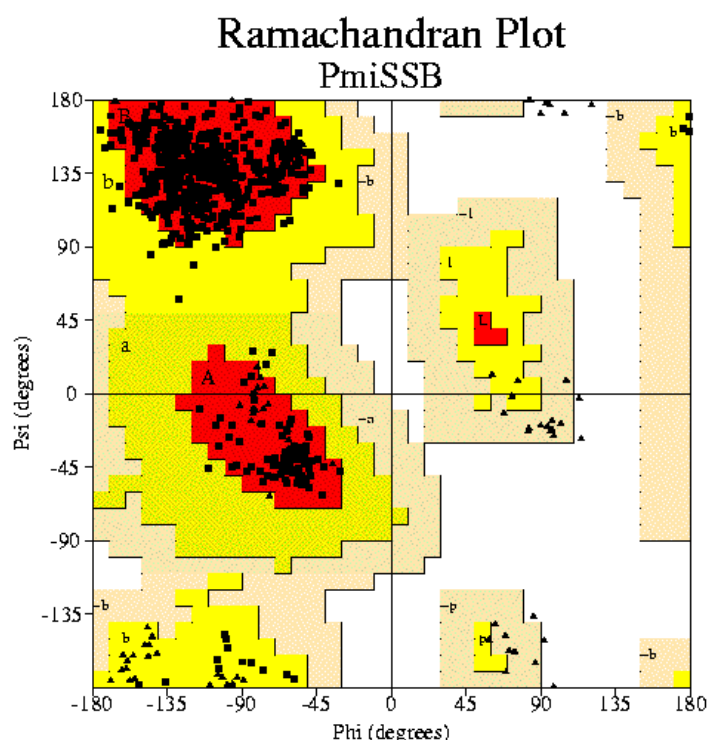


Figure 35 Ramachandran plot for *PmiSSB*, created by PROCHECK (Laskowski *et al.*, 1993). Glycine and proline residues are denoted by triangles and all other residues by squares. The different demarcated regions are labelled as A, B, L, - most favoured regions; a, b, l, p, -allowed; and -a, -b, -l, -p, -generously allowed.

3.5.6 Comparison between the *PmiSSB* and *EcoSSB* model

Despite of the high sequence identity (81%) between *EcoSSB* and *PmiSSB*, both SSBs were crystallised in different space groups. *PmiSSB* belongs to space group $P2_12_12$ and contains two tetramers in the asymmetric unit (Figure 36), whereas *EcoSSB* belongs to space group $C2$ and contains only one tetramer in the asymmetric unit (Figure 27). The fold of the polypeptide chain in *PmiSSB* is similar to that found in the *EcoSSB*. Each tetramer of *PmiSSB* was compared with the *EcoSSB* tetramer. When 369 C_α atoms of each tetramer, including all those, which are involved in regular secondary structure elements, are superimposed with tetrameric *EcoSSB*, a r.m.s.d. value of 1.4 Å for the first tetramer and 1.2 Å for the second

tetramer (Figure 37) of *PmiSSB* is obtained. In the r.m.s.d. calculation Loop II (residues 40-50) was not included since this part is only available in the H-monomer of *PmiSSB*. Loop III (residues 85-100) of all monomers of *PmiSSB* tetramer shows only a small r.m.s.d. to Loop III of *EcoSSB*. Significant differences are found near the first β -strand (residues 12-39) and the helix region (residues 65-74) in each monomer of *PmiSSB*. In this region the A-, B-, C-, D-monomer of *PmiSSB*, show higher r.m.s.d. values than the E-, F-, G-, and H-monomer to the corresponding structure of *EcoSSB*.

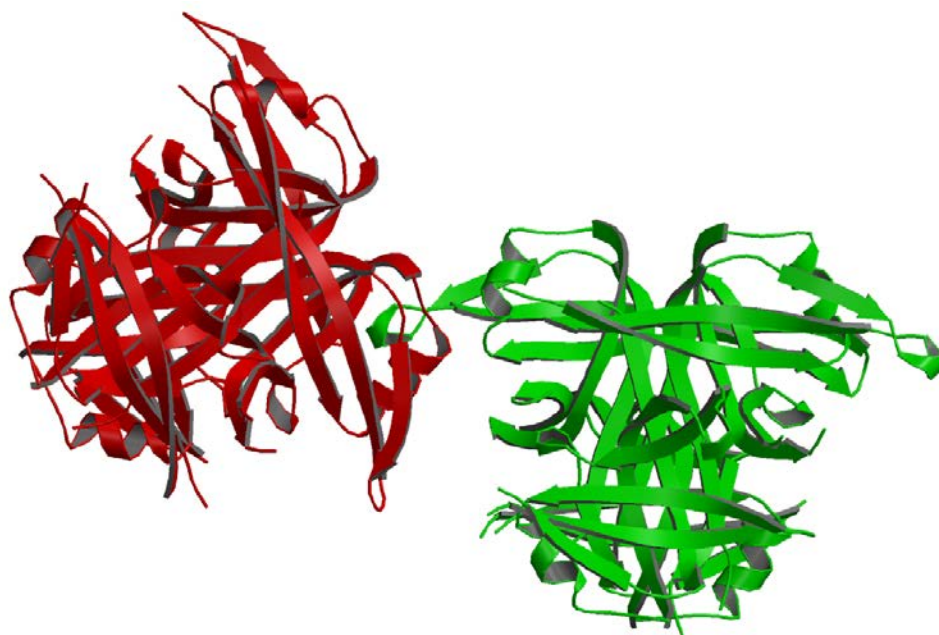


Figure 36 Ribbon representation of the two tetramers of *PmiSSB* drawn using MOLSCRIPT (Kraulis, 1991). The first and last four subunits of *PmiSSB* are shown in red and green, respectively. The view is along the z-axis.

In both structures (*EcoSSB* and *PmiSSB*), the polypeptide chains show no secondary structures beyond residue 112. The conserved residue His-55 plays a similar role in the *PmiSSB* as in *EcoSSB*. In each monomer of this structure, His-55 is also involved in hydrogen bonding with the main chain of Leu-34 and side chain of Thr-36 of the same monomer and a water molecule. The water mediates hydrogen bonding to two other water molecules. One of these two waters is in contact with the main chain of Arg-56 and Thr-99 of the same monomer and the other forms hydrogen bonds with the side chain of Glu-53 (same monomer) and Thr-85 (adjacent monomer).

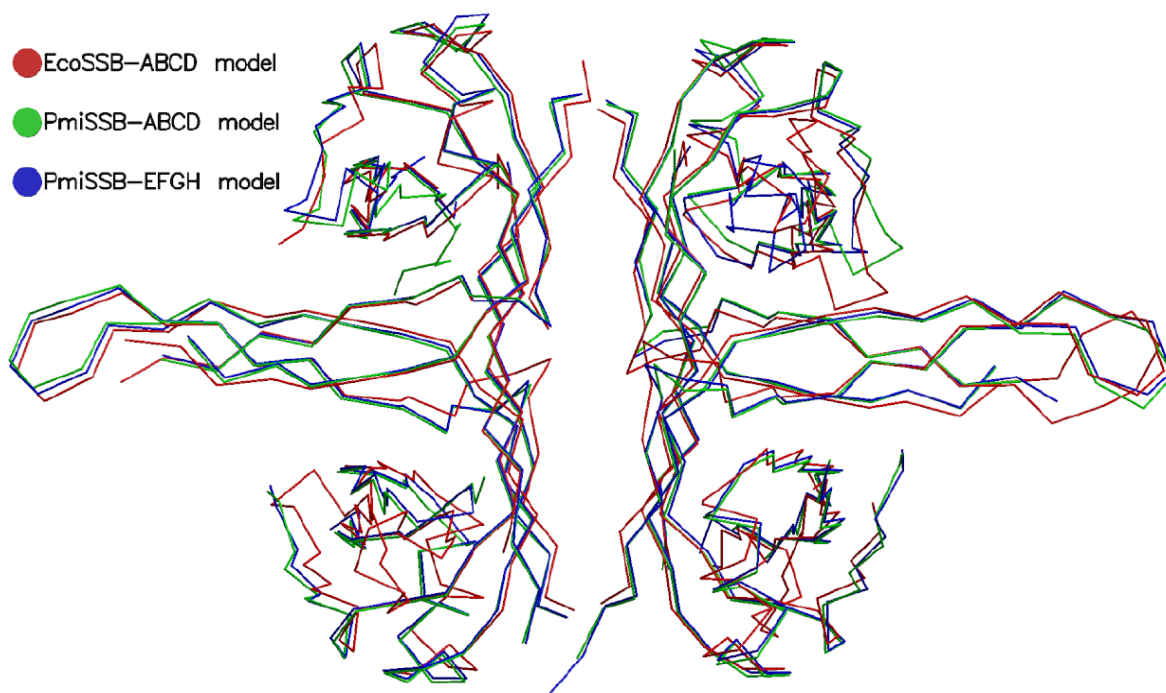


Figure 37 Superposition of the C_{α} atoms of *EcoSSB*-ABCD (red line) *PmiSSB*-ABCD (green line), and *PmiSSB*-EFGH tetramer (blue line) drawn using MOLSCRIPT (Kraulis, 1991).

In contrast to *EcoSSB*, these water molecules do not continue the hydrogen-bonding network. Probably due to the limited resolution of the *PmiSSB* structure, not all water molecules were detected. It is interesting to note that the orientation and position of these three waters in each monomer of *PmiSSB* are same as in *EcoSSB*. Tyr-78 and Gly-15 of *PmiSSB* also plays similar role as in *EcoSSB*.

There are total 21 salt bridges in this structure. Lys-7 and Glu-80 are involved in the formation of salt bridges in each monomer. Lys-7 is involved in salt bridge formation with Glu-80 of both the same and the adjacent monomer at the dimer-dimer interface.

3.5.7 The dithymidene (dT)₂ electron density

When the *PmiSSB* model was completely refined, electron density was inspected near the aromatic residues which are generally involved in ssDNA-binding. Density for the nucleotides appeared near the aromatic residues Trp-88, where Loop III was completely built. This Loop is ordered only in the B-, C-, F- and G-monomers. However, the density of the nucleotides was too weak to refine the dinucleotide molecules. Therefore this part of the model was not included in the structure. The density was confirmed by the theoretical model of *PmiSSB*/ssDNA complex. The theoretical model was prepared for *PmiSSB* in a similar way as the *BabSSB*-ssDNA model. This part of the electron density is arranged in a specific way. In Figure 38, Trp-88 and Lys-87 of the G- and F-monomer are shown. The side chains of the two Trp-88 are tilted towards the electron density of the dinucleotides, showing an

involvement in the nucleotide binding while Lys-87 points to the opposite direction. The residues of *Pmi*SSB which are involved in ssDNA-binding are sequentially and structurally conserved, therefore it can be suggested that *Pmi*SSB might bind ssDNA in a similar manner as *Eco*SSB binds to ssDNA.

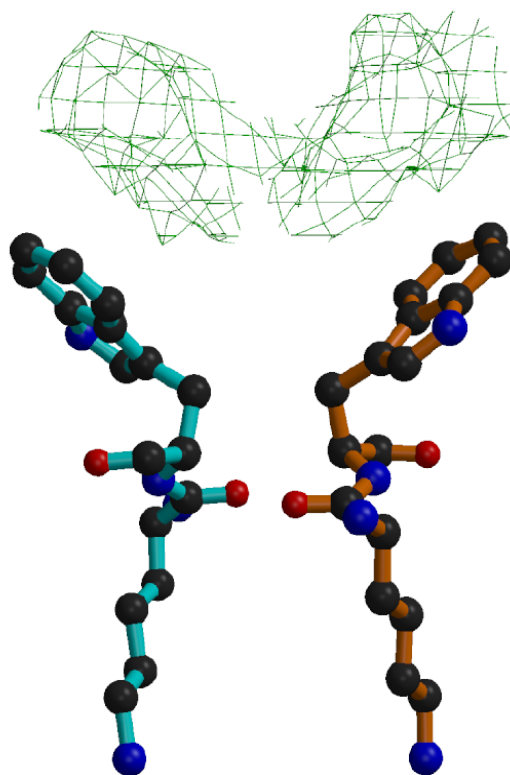


Figure 38 Ball-and-stick model of Trp-88 and Lys-87 from G-monomer and symmetry mate of F-monomer with (F_o-F_c)-map at 1.5 σ for the nucleotides.

3.5.8 Crystal packing of the *Pmi*SSB structure

There are eight monomers of *Pmi*SSB in the asymmetric unit. These eight monomers are essentially of the same structure. The r.m.s.d. between any two monomers varies from 0.21 to 0.47 Å for the 83-98 C α pairs (Table A4). The first four monomers (A-, B-, C-, and D-monomer) form one tetramer (ABCD-tetramer) and the last four monomers (E-, F-, G-, and H-monomer) comprise another tetramer (EFGH-tetramer). These two tetramers are related by a 65.87° rotation along the z-axis. The r.m.s.d. between the two tetramers is 0.55 Å for 339 C α pairs, and 0.33 Å for 256 C α pairs (when loop parts were excluded in the r.m.s.d. calculation). This difference in the r.m.s.d. values indicates that there is some variation in the loop regions. It is probably due to crystal packing effects. The ABCD-tetramer is surrounded by three EFGH-tetramers in the crystal packing (in the xy-plane); similarly a EFGH-tetramer is surrounded by three ABCD-tetramers (Figure 39). Therefore there are three main contact

regions for each tetramer: (i) the first contact region shows the involvement of the helix (Tyr-70) of the D-monomer with the fifth β -strand (Met-111) of the E-monomer and vice versa. (ii) The second, between the helix (Glu-69, Arg-72) of the C-monomer and the fifth β -strand (Met-111, Leu-112) of the G-monomer and (iii) the third, between the helix (Glu-65) of the C-monomer and the end part of the second β -strand (Ser-39) of F-monomer.

The involvement of Loop I, II and III in crystal packing for each tetramer of *PmiSSB* (53d, e, f) is similar as found in the crystal packing of *EcoSSB* tetramer (Figure 28).

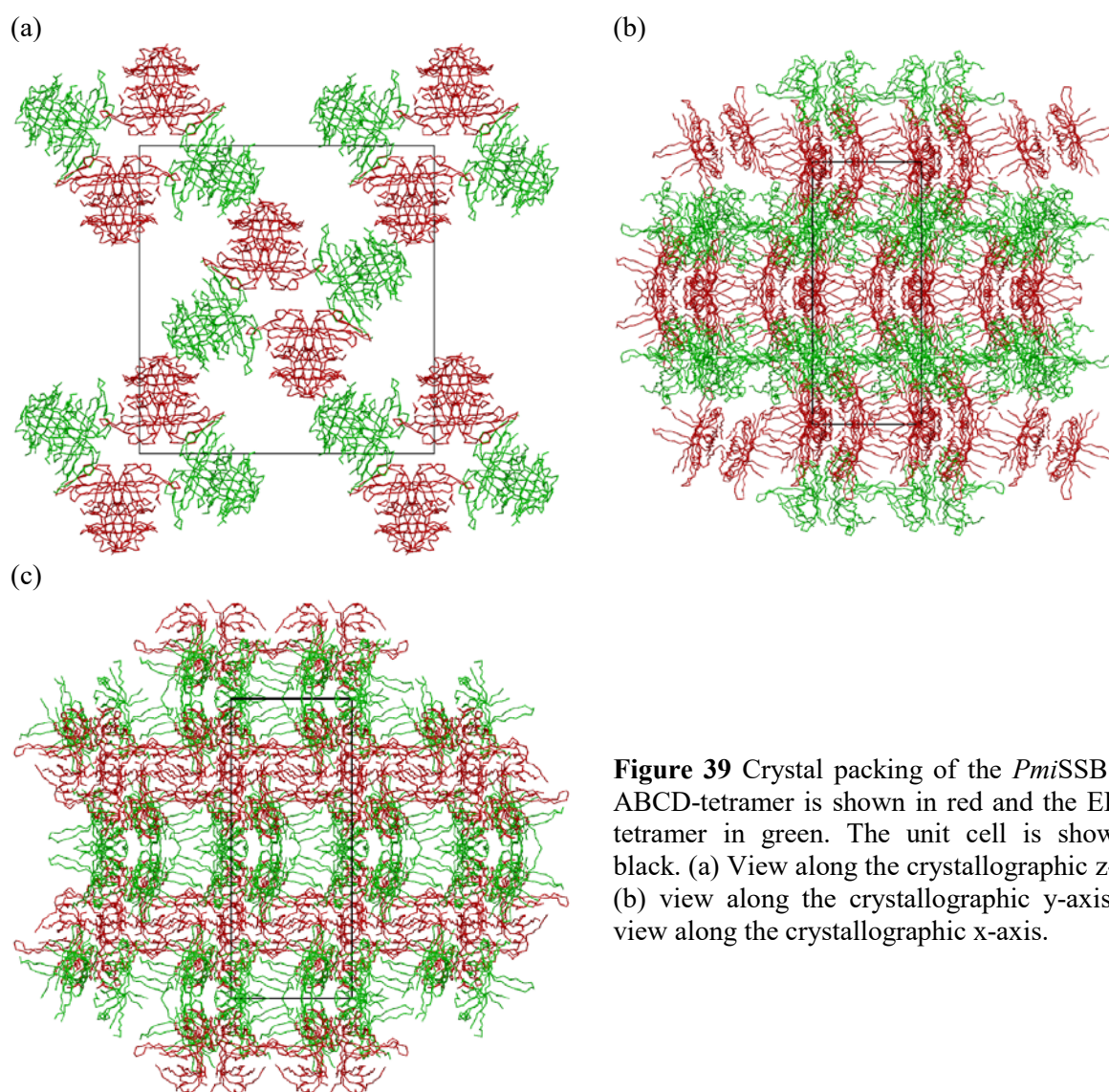


Figure 39 Crystal packing of the *PmiSSB*. The ABCD-tetramer is shown in red and the EFGH-tetramer in green. The unit cell is shown in black. (a) View along the crystallographic z-axis, (b) view along the crystallographic y-axis (c) view along the crystallographic x-axis.

3.6 Crystal structure of *Sma*SSB

3.6.1 Crystallisation of *Sma*SSB

The protein solution was made up to a concentration of 12.2 mg/ml in Protein storage solution B (Table 4). Initial crystallisation trials employing the hanging-drop vapour diffusion method were performed using Hampton crystal NATRIX screen at 4°C. Each well of a Sarstedt tissue-culture plate contained 500 µl precipitant buffer and 4 µl of a 1:1 mixture of protein and precipitant buffer hanging on a siliconised coverslip. Within three days, micro crystals appeared along with few big crystals in the NATRIX screen solution number 45 (0.025 M Magnesium sulphate, 0.05 M Tris HCl pH 8.5 and 1.8 M Ammonium sulphate). These bigger crystals exhibited two or three strong diffraction spots, indicating that they were salt crystals. The tray was brought to room temperature for mounting the crystal and later it was placed again at 4°C. One week later, when the drops were checked again, no more micro crystals was present and the salt crystals were dissolved. Protein crystals appeared in the drop in bipyramidal form and continued to grow. The optimal size (0.2x0.2x0.2 mm³) was reached in 20 days.

The crystals were reproduced (Figure 40) by simply keeping the crystallisation tray, first at 4°C and let the micro crystals and salt crystals grow for one day and then transferring it to room temperature. When the micro crystals were dissolved, the tray was placed again at 4°C.

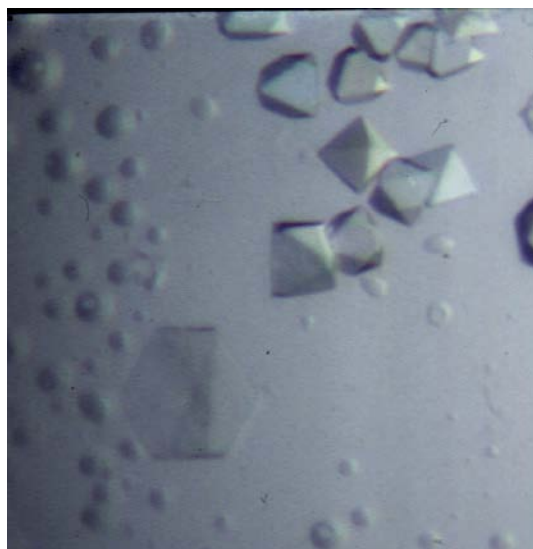


Figure 40 Tetragonal crystals of *Sma*SSB.

3.6.2 Native data collection

The bipyramidal crystals of *Sma*SSB were sensitive to temperature. For cryoprotection, the crystal was transferred into Panjelly™ at 4°C which was placed on a microscope cover slip. The slip was brought to the mounting station from 4°C. The crystal was fished with a cryo-

loop and transferred to the cryo-stream. The first diffraction pattern was achieved using a 10 minutes exposure at a rotating anode. The spots extended only to 6.5 Å resolution. The cooled crystal was brought into dried paraffin oil for washing and was then again transferred into the cryo-stream. At the same exposure, the crystal diffracted to 4.5 Å and with 1 hour exposure to 3.6 Å. A complete dataset was then collected from the crystal in two days on our rotating anode. After finishing the data collection, the crystal was stored in liquid nitrogen for the use at a synchrotron source.

The data were integrated, reduced and scaled using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997). The analysis of diffraction pattern showed that the crystals belong to a tetragonal space group with unit cell dimensions of $a=b=111.25$ Å, and $c=141.76$ Å.

A complete dataset was collected at BM14 beamline ESRF (Grenoble) from the same cryo-cooled crystal. There, the crystal diffracted to 2.8 Å. The statistics of the data collection are given in Table 8.

3.6.3 Structure determination of *Sma*SSB

The structure of *Sma*SSB was determined by molecular replacement using the program AMoRe (Navaza, 1994) as implemented in the CCP4 program package with a search model based on the 1.95 Å resolution *Eco*SSB structure. The initial estimation of the solvent content indicated that six molecules were present in the asymmetric unit with a fractional solvent content of 39% ($V_m = 2.0$ Å³/Da). Therefore, a single dimer of the tetrameric *Eco*SSB structure was used as a search model. The *Sma*SSB_dataset1 (Table 8) was used during the molecular replacement. Data in the range 10-4 Å were exploited for the rotation function. Using a Patterson cut-off radius of 41 Å, a list of 31 rotation function peaks was obtained, with the top peak having a correlation coefficient of 0.159. Translation searches were conducted separately with 8-4 Å data for the first dimer in the space groups $P4_22_12$, $P4_12_12$, and $P4_32_12$. The correlation coefficient for the first space group was significantly higher than for the other two (0.256, 0.189 and 0.18, respectively), which established $P4_22_12$ as the correct space group. Fixing the first solution, and searching for a second translation vector resulted in a correlation coefficient of 0.399 and an R-factor of 47.8%. Afterwards the first and the second solutions were fixed to search for the presumed third dimer. This yielded a correlation coefficient of 0.48 and an R-factor of 43.8%. The Eulerian angles for the first two solutions were the same, while the third solution was related by 35.5° along the z-axis. However, rigid-body refinement did not improve the correlation coefficient and R-factor (0.382 and 47.4%, respectively). The packing of the models from these solutions were checked using the program O (Jones *et al.*,

1991). No significant steric clashes between the first two dimers was observed, also they were found to assemble into a tetramer in exactly the same way as the *EcoSSB* tetramer. The third solution also formed a tetramer. However, in this tetramer, each monomer was overlaid by another monomer. This position was carefully checked and it was found that this tetramer is placed on the crystallographic two-fold axis; therefore only a single monomer was required in the third solution to form a tetramer. This means that asymmetric unit of *SmaSSB* crystals contains five molecules only (solvent content 49%, $V_m=2.4 \text{ \AA}^3/\text{Da}$). This was the reason, why rigid-body refinement couldn't improve the correlation coefficient and the R-factor. A summary of the molecular replacement solution of the *SmaSSB* structure in the correct space group is given in Table 14.

Table 14 The molecular replacement procedure for *SmaSSB*

Resolution range	8.0 - 4.0 Å
Rotation & Translation function (1 st Dimer)	
Best solution	$\alpha = 72.62^\circ$ $\beta = 90.00^\circ$ $\gamma = 312.77^\circ$ tx = 0.2123 ty = 0.3147 tz = 0.1631
Correlation coefficient	0.256
R-factor	51.2%
Rotation & Translation function (2 nd Dimer)	
Best solution	$\alpha = 72.62^\circ$ $\beta = 90.00^\circ$ $\gamma = 312.77^\circ$ tx = 0.7193 ty = 0.4549 tz = 0.1617
Correlation coefficient	0.399
R-factor	47.8%
Rotation & Translation function (3 rd Dimer)	
Best solution	$\alpha = 42.73^\circ$ $\beta = 88.25^\circ$ $\gamma = 312.43^\circ$ tx = 0.9093 ty = 0.9207 tz = 0.9931
Correlation coefficient	0.48
R-factor	43.8%
Refinement of combined solution:	
Dimer 1:	$\alpha = 75.16^\circ$ $\beta = 89.17^\circ$ $\gamma = 313.45^\circ$ tx = 0.2133 ty = 0.3143 tz = 0.1634
Dimer 2:	$\alpha = 75.79^\circ$ $\beta = 89.27^\circ$ $\gamma = 312.37^\circ$ tx = 0.7220 ty = 0.4547 tz = 0.1614
Dimer 3:	$\alpha = 40.25^\circ$ $\beta = 87.21^\circ$ $\gamma = 311.82^\circ$ tx = 0.9096 ty = 0.9201 tz = 0.9918

Correlation coefficient	0.382
R-factor	47.4%

3.6.4 Refinement and model building of *Sma*SSB structure

Initial refinement was carried out against data in the range 100-3.6 Å, with 728 (6.7%) of the reflections used for cross-validation (Brünger, 1992). The properly oriented model was subjected to rigid body refinement using CNS (Brünger *et al.*, 1998) which led to a reasonable improvement in R-factor (36.6%) and R_{free} (43.2%). After inspection of the $3F_o-2F_c$ and F_o-F_c maps at this point Loop I (residues 22-28), Loop II (residues 40-50) and Loop III (residues 85-100) were deleted from the model, because they were out of electron density. The model which included five molecules, then underwent several rounds of a standard refinement protocol with restrained non-crystallographic symmetry (NCS). Each round consisted of Powell minimisation, simulated annealing from 3000K, and individual temperature-factor refinement. Each round of refinement was followed by model building using O (Jones *et al.*, 1991). At the point, at which the R-factor and R_{free} fell below 28.2% and 32.5%, respectively. Loop I was rebuilt in D- and E-monomers and non-conserved residues were mutated in the density.

When synchrotron data became available, the structure was refined to 3.0 Å in a similar manner as with the low resolution dataset. At this resolution, Loop III was built in the A-, B-, D-, and E-monomers and Loop I was built in the A-monomer. In the final round of refinement, the R-factor and R_{free} dropped to 26.7% and 29.8%, respectively.

3.6.5 Quality of *Sma*SSB model

The refined model contains 3637 non-hydrogen protein atoms and eight ordered water molecules, accounting for 496 residues in the five independent molecules of *Sma*SSB in the asymmetric unit (Table 9). It gives an R-factor of 26.7% for 17347 unique reflections in the range 100-3.0 Å. Loop II in each monomer and Loop I of B-, and C-monomer could not be seen in the electron density and were not included in the model. There were few surface residues whose side chains could not be traced because of their poor electron density, therefore, these side chain were also not included in the model.

The r.m.s.d. for ideal stereochemistry are 0.009 Å for bond lengths, and 1.52° for bond angles. The average positional error is estimated to be 0.5 Å from a Luzzati plot (Luzzati 1952). Among 402 non-glycine and non-proline residues in the five molecules in the asymmetric unit, the numbers of residues lying in the most favoured, additionally allowed, and generously allowed in the Ramachandran plot are 353 (87.8%), 46 (11.4%), and three (0.7%), respectively (Figure 41). For all molecules, the average B-factors are 54.5 Å² for 496 main-

chain atoms and 55.6 \AA^2 for 437 side-chain atoms. The B-factors for the eight solvent molecules range from 15 to 52.7 \AA^2 , with the average being 33.6 \AA^2 .

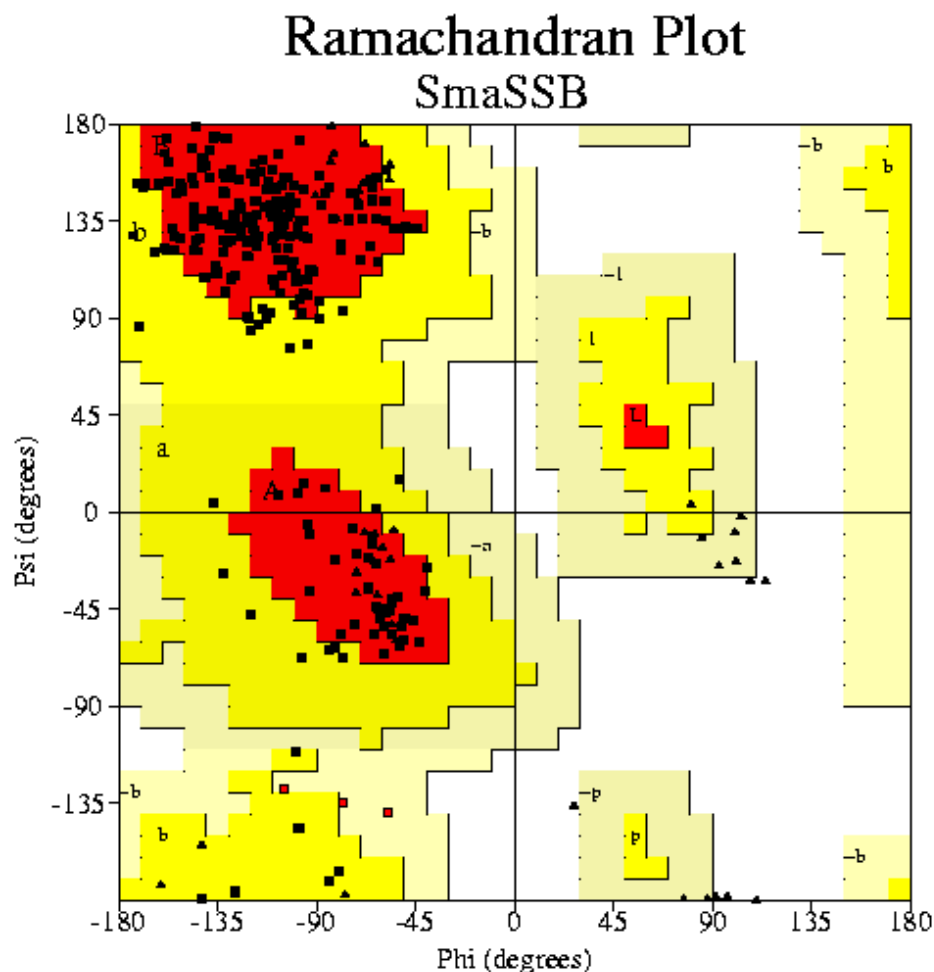


Figure 41 Ramachandran plot for *SmaSSB*, created by PROCHECK (Laskowski *et al.*, 1993). Glycine and proline residues are denoted by triangles and all other residues by squares. The different demarcated regions are labelled as A, B, L, -most favoured regions; a, b, l, p, -allowed; and -a, -b, -l, -p, -generously allowed.

3.6.6 Comparison between the *SmaSSB* and *EcoSSB* model

The *SmaSSB* shows high sequence identity (89%) with *EcoSSB*. In the N-terminal domain of *SmaSSB*, there are only seven residues at positions 48, 51, 70, 82, 84, 89 and 101 (*EcoSSB* numbering) that are different in *EcoSSB*, whereas eleven residues are different in the C-terminal domain. Despite the high conservation, *SmaSSB* has been crystallised in a different space group ($P4_22_12$). Another striking difference is that the asymmetric unit of *SmaSSB* crystals contains five independent molecules (1.25 tetramers); whereas the asymmetric unit of *EcoSSB* contains four molecules (1 tetramer).

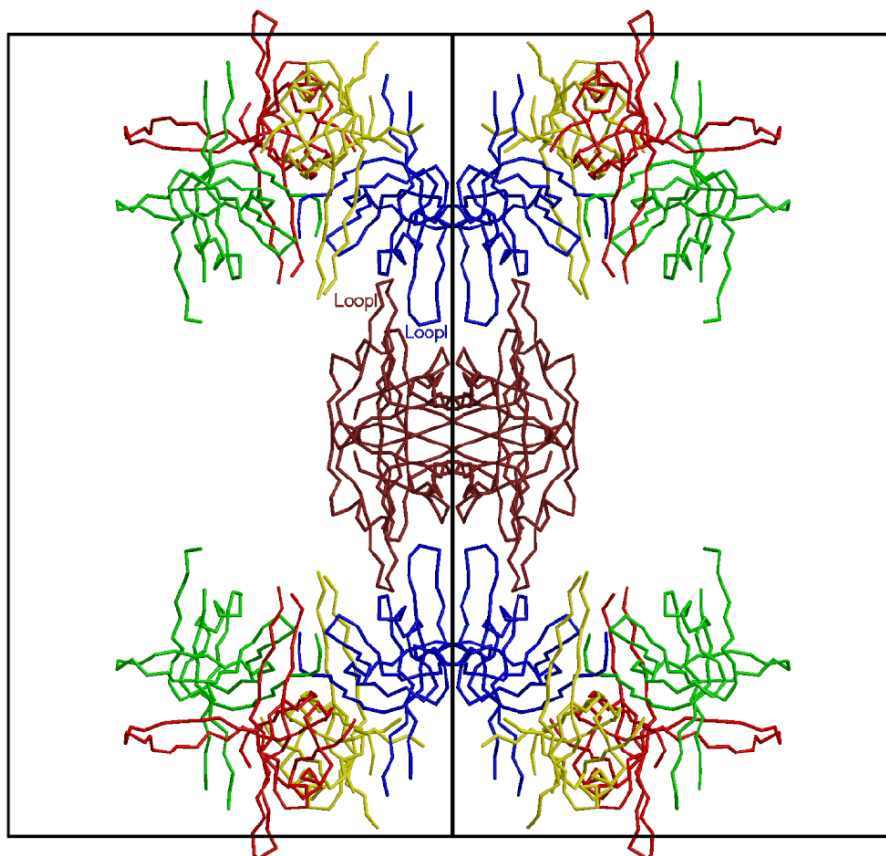


Figure 42 C_{α} trace of non-crystallographic and crystallographic tetramers of *SmaSSB*.

The overall fold of the polypeptide chain in *SmaSSB* is similar to that found in *EcoSSB*. The C-terminal domain (last 60 residues) of *SmaSSB* is also disordered as in the *EcoSSB* structure. In the *SmaSSB* model, four monomers out of five form an NCS-tetramer similar to *EcoSSB*. The fifth monomer forms a second tetramer with its crystallographic symmetry mates (Figure 42). The packing arrangement of this tetramer is specific in a way, which will be discussed in the next section. This one crystallographic tetramer interacts with four other NCS-tetramer via Loop I (Figure 42); consequently there are two crystallographic tetramers and eight NCS-tetramers in the unit cell. Here, our interest is to compare the NCS-tetramer and the crystallographic tetramer of *SmaSSB* with the *EcoSSB*-tetramer in order to see the differences, as this is likely due to a packing effect. When 371 C_{α} atoms of the NCS-tetramer and 394 C_{α} atoms of crystallographic tetramer of *SmaSSB* are superimposed, an r.m.s.d of 1.4 Å between the NCS tetramer of *SmaSSB* and the *EcoSSB* tetramer is observed. The r.m.s.d. between the crystallographic tetramer of *SmaSSB* and *EcoSSB* tetramer is 2.3 Å. After neglecting all loops, 275 C_{α} atoms of each tetramer of *SmaSSB* were compared separately with the *EcoSSB* tetramer, an average displacement of 1.0 Å for the NCS-tetramer and 1.5 Å for the crystallographic tetramer of *SmaSSB* is observed. Therefore, it is obvious that these two tetramers are different. They have a different crystallographic environment and their

loops have different conformations. To investigate this further, each of the five monomers of *Sma*SSB was compared with the individual monomers of *Eco*SSB (A-, B-, C-, and D-monomers) using the program ALIGN (Cohen, 1997). The average of the four r.m.s.d. values obtained on comparison of a single *Sma*SSB monomer with each of the four *Eco*SSB tetramers was taken to avoid r.m.s.d. bias arising from single comparison. To compare, the structural change of each monomer of *Sma*SSB, the r.m.s.d. value is plotted against the residue number (residues 2-112) for every monomer (Figure 43).

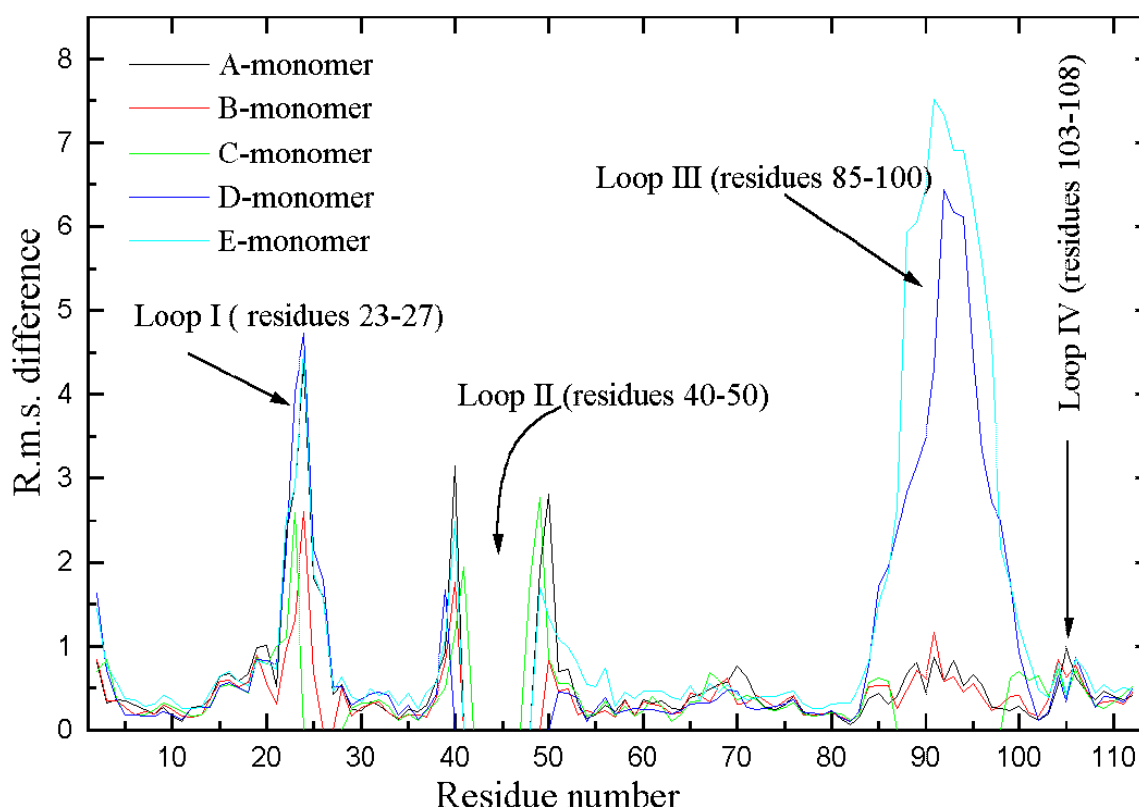


Figure 43 A plot of r.m.s.d. (Å) for the C_{α} atoms in the *Sma*SSB monomers.

In the graph, the r.m.s.d. plot (Figure 43) of the A-, B-, C-, D-, and E-monomers is shown in black, red, green, blue and cyan, respectively. Arrows denote the loop areas. Loop II (residues 40-50) is disordered in *Sma*SSB; therefore, it was not included in the r.m.s.d. calculation. The A-, B-, and C-monomers show very little variation near Loop III and Loop IV including the core of the monomer. However, these monomers show significant difference near Loop I, and just before and after Loop II. Relatively large structural differences are found near Loop I and Loop III of the D-, and E-monomers (Figure 43 and Figure 44). These structural changes in the loop regions might be due to the arrangement of the fifth molecule in the asymmetric unit. They share Loop III in the tail-to-tail interface. This will be discussed in more detail in the next section.

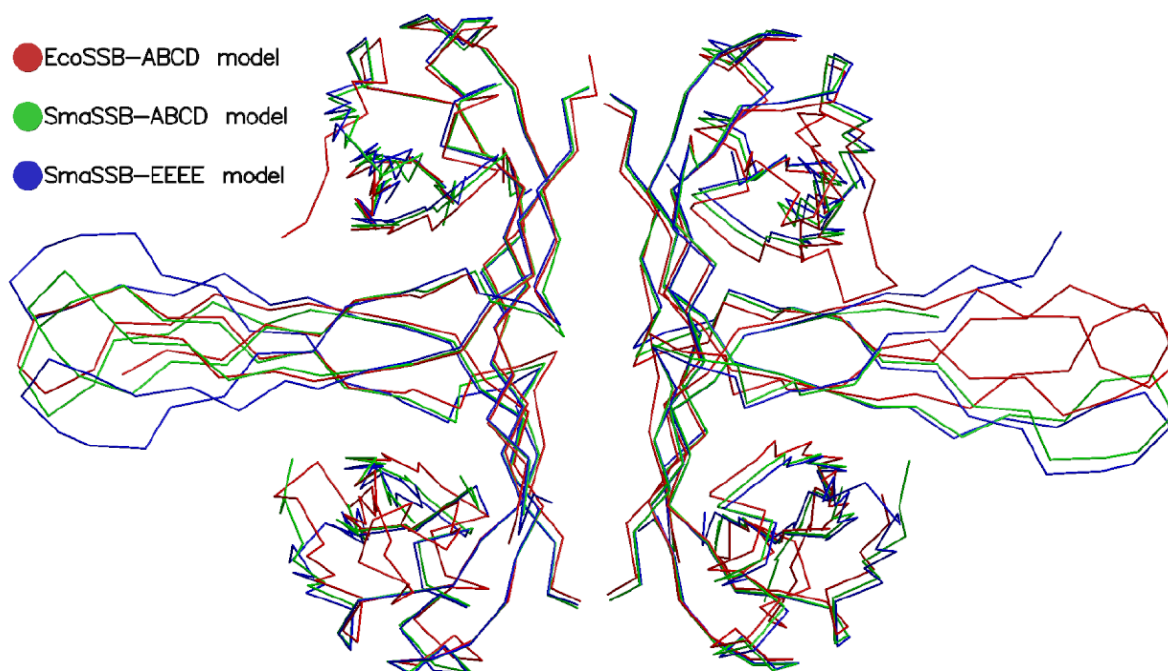


Figure 44 Superposition of the C_{α} traces of the NCS *EcoSSB*-tetramer (red), the NCS *SmaSSB*-tetramer (green) and the crystallographic *SmaSSB* tetramer (blue), drawn using MOLSCRIPT (Kraulis, 1991).

3.6.8 Crystal packing of the *SmaSSB* structure

There are five copies of the *SmaSSB* molecules in the $P4_22_12$ asymmetric unit, which is unusual for bacterial SSBs, since most of the bacterial SSB proteins have been crystallised in a different space group and contain an even number of molecules in the asymmetric unit. (i.e. *BabSSB*, *EcoSSB* and *PmiSSB* contain two, four and eight molecules, respectively in their asymmetric units). Among the five monomers, three monomers (A-, B- and C-monomers) are essentially identical and the other two monomers (D- and E-monomers) are different only in their Loop III regions. The r.m.s.d. between *SmaSSB* monomers based on C_{α} pairs (75-100 atoms) range from 0.22 to 0.40 Å (Table A4). However, for Loop III (residues 85 – 100) of the D- and E-monomers compared pairwise to the A-, B-, and C-monomers, the r.m.s.d. values range from 0.80 to 6.08 Å, whereas inter-comparison among the A-, B- and C-monomer in this region gave r.m.s.d. values varying only from 0.20 to 0.98 Å.

The different conformation of the D- and E-monomers are due to crystal packing. Within the asymmetric unit, the D-monomer belongs to a non-crystallographic tetramer whereas the E-monomer belongs to a crystallographic tetramer (Figure 42). The E-monomer forms a crystallographic tetramer with its symmetry mates at the centre and each corner of the unit cell involving the 2-fold and 4_2 -axis, and the NCS tetramers are arranged in the unit cell about the 2_1 axis (Figure 45, 46). The interface between the A- and E-monomers involves

contacts between Loops III (Figure 53f). Similar contacts can be found between B- and D-monomers. Loop I of the A-monomer interacts with Loop I of its symmetry mates (Figure 53g). The E-monomer is involved in intermolecular contacts in its crystallographic tetramer. The C-monomer is not at all involved in crystal contacts. The intermolecular hydrogen bonding contacts of *Sma*SSB are listed in the Table A17.

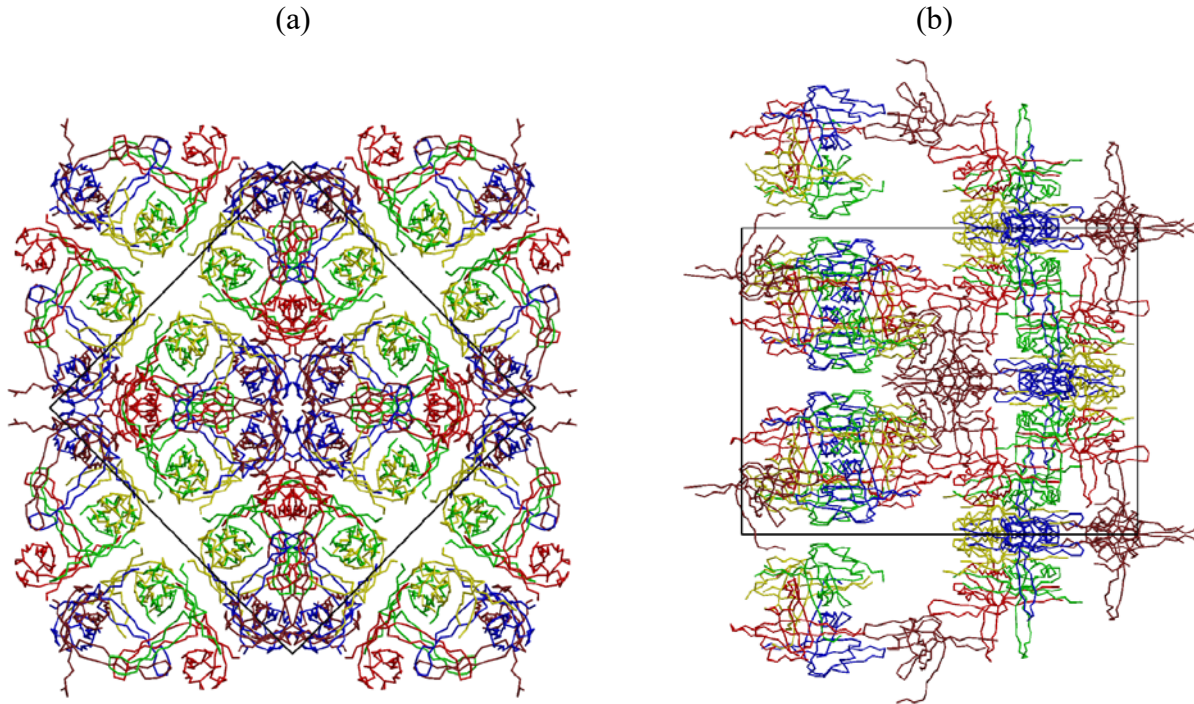


Figure 45 Crystal packing of *Sma*SSB. A-, B-, C-, D-, and E-monomers are shown in red, green, yellow, blue and brown, respectively. (a) View along the crystallographic z-axis, (b) view along the crystallographic y-axis.

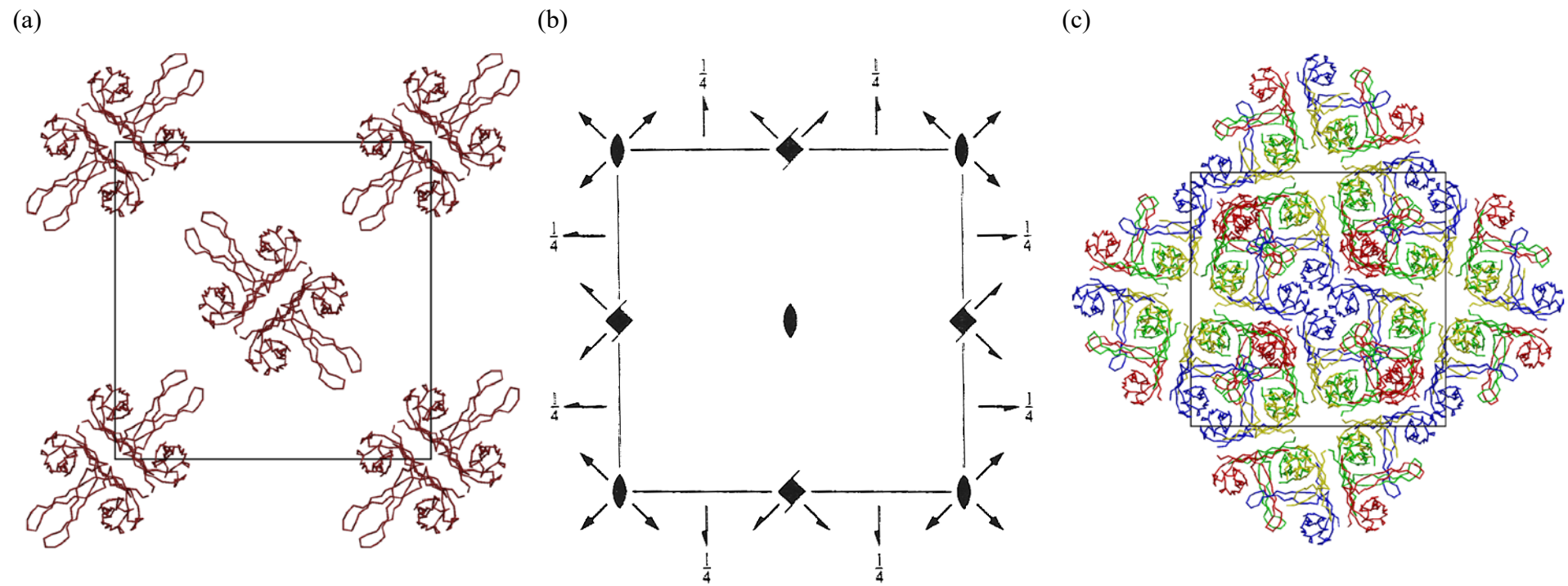


Figure 46 Crystal packing of the crystallographic and the non-crystallographic *SmaSSB* tetramer along the z-axis.

3.7 Crystallisation of *EcoSSB-d(T)*₃₅

d(T)₃₅ was prepared at 0.377 mM concentration in the Protein storage solution A (Table 4). *EcoSSB* protein was dialysed against Protein storage solution A (Table 4). The dialysed protein was then concentrated by centrifugation until an approximate protein concentration of 7.5 mg/ml (0.377 mM) was achieved. A stock solution was prepared for use by mixing equal volumes of *EcoSSB* and the d(T)₃₅ solution. The stock solution was used for the crystallisation trials.

An initial screen for crystallisation conditions were carried out using the Ammonium sulphate screen (Hampton Research) using the hanging-drop method. A single thin plate type crystal appeared after three months in the condition 0.8 M Ammonium sulphate and 0.1 M HEPES pH 7.0. To optimise the crystallisation condition, the crystallisation tray was setup at different ammonium sulphate concentrations (0.375 - 0.7 M) keeping the same buffer. One crystallisation drop was made using 2 µl of *EcoSSB-d(T)*₃₅ complex, 2 µl of a 100 mM spermidine solution. After five days, the reservoir was replaced by 60% ammonium sulphate, 20 mM MES pH 6.5 and 3 mM NaN₃. Crystals grown using these condition had a square like morphology and grew to approximately 0.25x0.25x0.15 mm³ in six months.

To verify that the crystals contained both DNA and protein, a number of them were harvested and washed thoroughly in Protein storage solution A before they were dissolved in water and subjected to a 6% native polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide to test for the presence of DNA and then with Coomassie Blue to probe for protein (data not shown). A coincident band for protein and DNA and the comparison of this band with the native protein and DNA controls showed the species to be of a higher molecular weight than either native *EcoSSB* or the DNA.

Crystal preparation and data collection from *EcoSSB-d(T)*₃₅ crystal

The crystal was transferred into PanjellyTM and then transferred to the cryo-stream in the cryo-loop. After a few minutes, the crystal was brought into paraffin oil to wash the PanjellyTM layer from the crystal and again cryo-cooled in the cryo-stream.

The suitability for diffraction of the crystal was checked using a Cu-K_α rotating anode. The first diffraction pattern of the crystal showed long streaks. When the crystal was rotated by 90°, the diffraction pattern showed separated spots, distributed in a hexagonal pattern. The first diffraction image, in which the long streaks were recorded, raised the suspicion that one of the axes of the unit cell must be extremely long. To resolve the long streak, it was mandatory to collect the diffraction data at a synchrotron. To resolve the long streaks, the

CCD detector was moved back to 750 mm at the X13 beamline (newly built beam line of IMB Jena at DESY, Hamburg). 360 images were recorded, using 0.5° rotation per image to a resolution 10 Å. Another 90 images were collected to a resolution 6 Å, from the same crystal from that area where diffraction pattern showed separated spots (1° rotation per image).

Unit cell determination and data processing of *EcoSSB-d(T)*₃₅ data

To determine the space group and the unit cell parameters, the program DENZO was first used, since this program offers two indexing methods: automatic and interactive. All possible ways were tried to index the *EcoSSB-d(T)*₃₅ diffraction images, but all attempts to determine the unit cell parameters and space group failed. Finally, the XDS program (Kabsch, 1988) was tried to index diffraction images.

To index the *EcoSSB-d(T)*₃₅ diffraction images by XDS, the first twenty and last twenty frames were used. The analysis of the diffraction images showed that the crystals were of unit cell dimensions $a = b = 74.55$ Å, $c = 356.41$ Å and $\gamma = 120^\circ$ and space group possibly hexagonal ($P6_x$ or $P6_x22$) or trigonal ($P3_x$, $P3_x21$, or $P3_x12$). Due to poor reproducibility of the crystals and the practical difficulties involved in further analysis of the diffraction data, subsequent attempts for structure solution were abandoned.

3.8 Overall comparison between SSBs

3.8.1 Common axis

Dimensions of one SSB-tetramer of $a=104.4$ Å, $b=60.5$ Å, and $c=96.5$ Å resulted in the observation that all crystals exhibits at least one axis of a unit cell of 55 ± 5 Å or a multiple of this value (Table 15).

Table 15 SSB crystal forms

SSB	Space group	Number of molecules†	Unit cell parameters				V_m (Å ³ /Da)
			a (Å)	b (Å)	c (Å)	β (°)	
<i>Eco</i> SSB	C2	4	104.4	60.5	96.5	112.5	2.3
<i>Bab</i> SSB	P4 ₃ 2 ₁ 2	2	113.3	113.3	52.2	90	2.6
<i>Pmi</i> SSB	P2 ₁ 2 ₁ 2	8	146.7	153.2	61.0	90	2.4
<i>Sma</i> SSB	P4 ₂ 2 ₁ 2	5	111.3	111.3	141.8	90	2.4
<i>Hsmt</i> SSB	P4 ₁ 2 ₁ 2	2	51.8	51.8	184.2	90	2.3

† Per asymmetric unit. V_m is the crystal packing density (Matthews, 1968).

The unit cell dimension of 55 ± 5 Å or a multiple of this value is shown in bold.

3.8.2 Comparison of SSB monomers

The structures were superimposed by aligning their β -sheet segments using the program ALIGN (Cohen, 1997). The resulting diagram is presented in Figure 47. R.m.s.d. values between monomers of different SSBs are listed in Table A4. Comparison of individual loops of SSBs will be discussed in section 3.8.6.

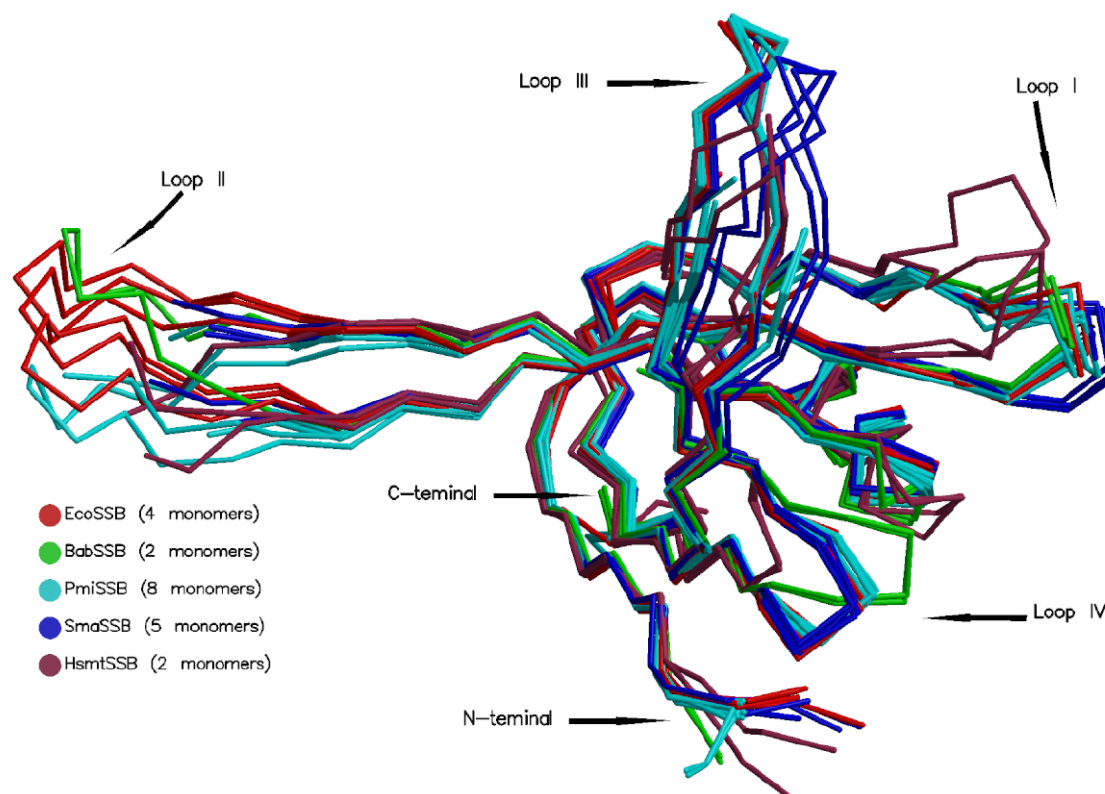


Figure 47 Superposition of 21 monomers of SSBs.

3.8.2.1 N-terminus (first 10 residues)

The first two or three residues are disordered in *EcoSSB*, *BabSSB*, *PmiSSB* and *SmaSSB*; whereas the first 10 residues are disordered in HsmtSSB. In the bacterial SSBs this segment exhibits a β -strand structure. In *EcoSSB*, the C_{α} trace is almost identical to *SmaSSB*, but different from *BabSSB* and *PmiSSB*. These differences could be due to the packing effects.

3.8.2.2 C-terminus

The C-terminal domain is disordered in the bacterial SSB structures; the polypeptide chain could not be traced beyond residues 112 (bacterial SSBs, *EcoSSB* numbering). However, the ultimate 10 residues of the last 60 residues in bacterial SSBs of the intact chain can have a regular secondary structure (perhaps α -helical) (Meyer & Laine, 1990), since there are a high proportion of negatively charged side chains in this segment. The other 50 residues of the bacterial SSBs contain a high percentage of glycine and/or proline residues (Figure 11), and would probably exhibit no secondary structure. This region is thought to modulate the strength of DNA binding, possibly by distancing the highly negatively charged C-terminal end from the DNA binding domain. The acidic tail has been suggested to be important in the interaction of SSBs with various replication proteins (i.e. DNA polymerase, exonuclease I) and is not essential for DNA binding (Kelman *et al.*, 1998; Genschel *et al.*, 2000). Interestingly, it has been shown that removal of this negatively charged C-terminal region leads to an enhanced affinity of the resulting truncated *EcoSSB* protein for ssDNA (Williams *et al.*, 1983).

3.8.2.3 Helix (Residues 61 to70)

The helix of SSB is one of the components of OB-fold (Figure 1). It is capped between the third and fourth β -strand. In the alignment of the 3D structures, the helix of *EcoSSB* shows little difference with that of *PmiSSB* and *SmaSSB*, but a larger deviation from that of *BabSSB* and HsmtSSB due to a single residue insertion (residues between 62 and 63, *EcoSSB* numbering) at the beginning of the helix. The helix may have important bearing on the ssDNA binding activity of the SSBs, since Phe-60 (equivalent to Phe-74 of HsmtSSB), the Ncap residue has been implicated by crosslinking experiment (Merrill *et al.*, 1984) and mutational experiments (Casas-Finate *et al.*, 1987; Bayer *et al.*, 1989).

3.8.2.4 Protein core

The core of the SSBs of each monomer consists of residues 4 to 20 (β -strand I), 30-37 (β -strand II), 51-60 (β -strand III), 61-70 (α -helix), 74-84 (β -strand V) and 109-112 (β -strand V') (*EcoSSB* numbering). There are number of hydrophobic residues (mainly Val, Ile, Leu and Ala) which may be crucial for the formation of the fold (Table A12). The overall fold (OB-

fold) of the SSBs is conserved (Figure 48). Only two residues (Trp-54, Phe-60) of the core are involved in ssDNA-binding, otherwise most of the residues involved in the ssDNA-binding are found on the surface of the protein. It is interesting to note that all residues in each structure, which are involved in the formation of the monomer-monomer and the dimer-dimer interface, belong to the main body of the protein.

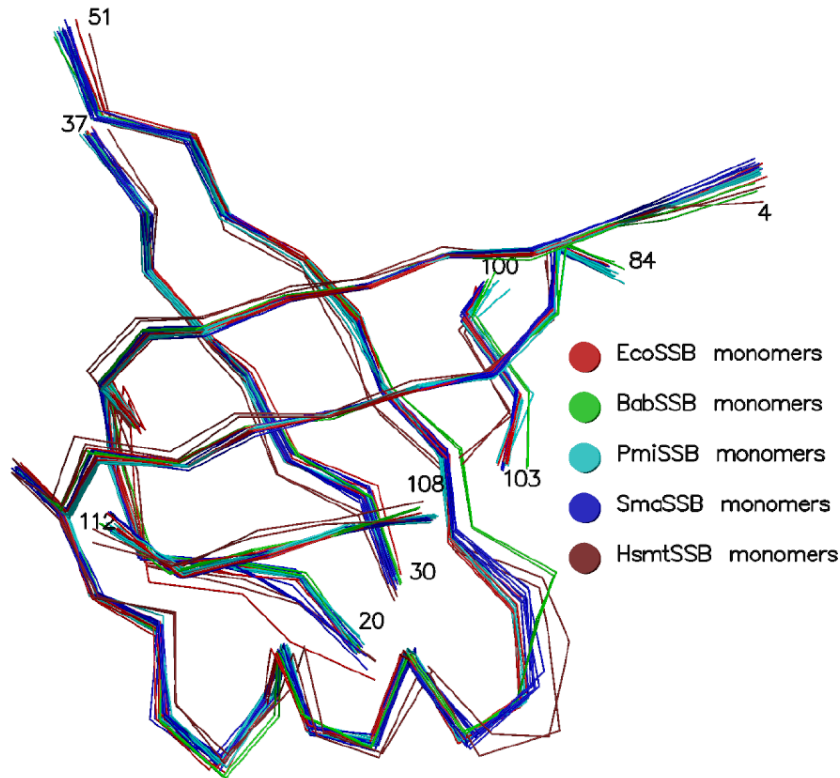


Figure 48 Superposition of the C_{α} atom traces from the SSBs core.

3.8.3 Interface

3.8.3.1 Monomer-monomer interface

Monomer-monomer interface in SSBs is formed, mediated by the β -sheet (Figure 25). This dimer consists of two antiparallel β -sheets from two SSB monomers. At least a few residues from each β -strands are buried upon dimer formation (Table A11). The monomer-monomer interface primarily consists of main chain hydrogen bonds involving residues 5-11 (*EcoSSB* numbering) of strand I in both monomers. The average surface area buried upon dimer formation for bacterial SSBs is 2013 \AA^2 , whereas for HsmtSSB, it is 2777 \AA^2 (Table A10). The buried surface area of HsmtSSB is significant; this is because it contains higher number of bulkier residues than the bacterial SSBs which are buried upon dimer formation (Table A11). His-55 (*EcoSSB*) is an important residue at this interface, which is conserved in other bacterial SSBs and it is equivalent to His-69 of HsmtSSB. The role of this residue in tetramerisation is described in detail in section 3.3.6.

3.8.3.2 Dimer-dimer interface

The make up of the dimer-dimer interface of SSBs is different than that of the monomer-monomer interface. However, there are some common residues in SSBs which are involved in both interfaces. These residues are mainly found on β -strand 1 (i.e Gly-4, Val-5, Lys-7, Ile-9 and Val-11; *Eco*SSB numbering). They exhibit a larger buried surface area upon dimer formation than tetramer formation (Table A7 and Table A8). Other than these residues (from β -strand 1), there are few residues which are involved in the formation of dimer-dimer interface from the β -strand IV and β -strand V. Participating residues from β -strand IV in the interface are shown in Figure 49 for each SSB structure.

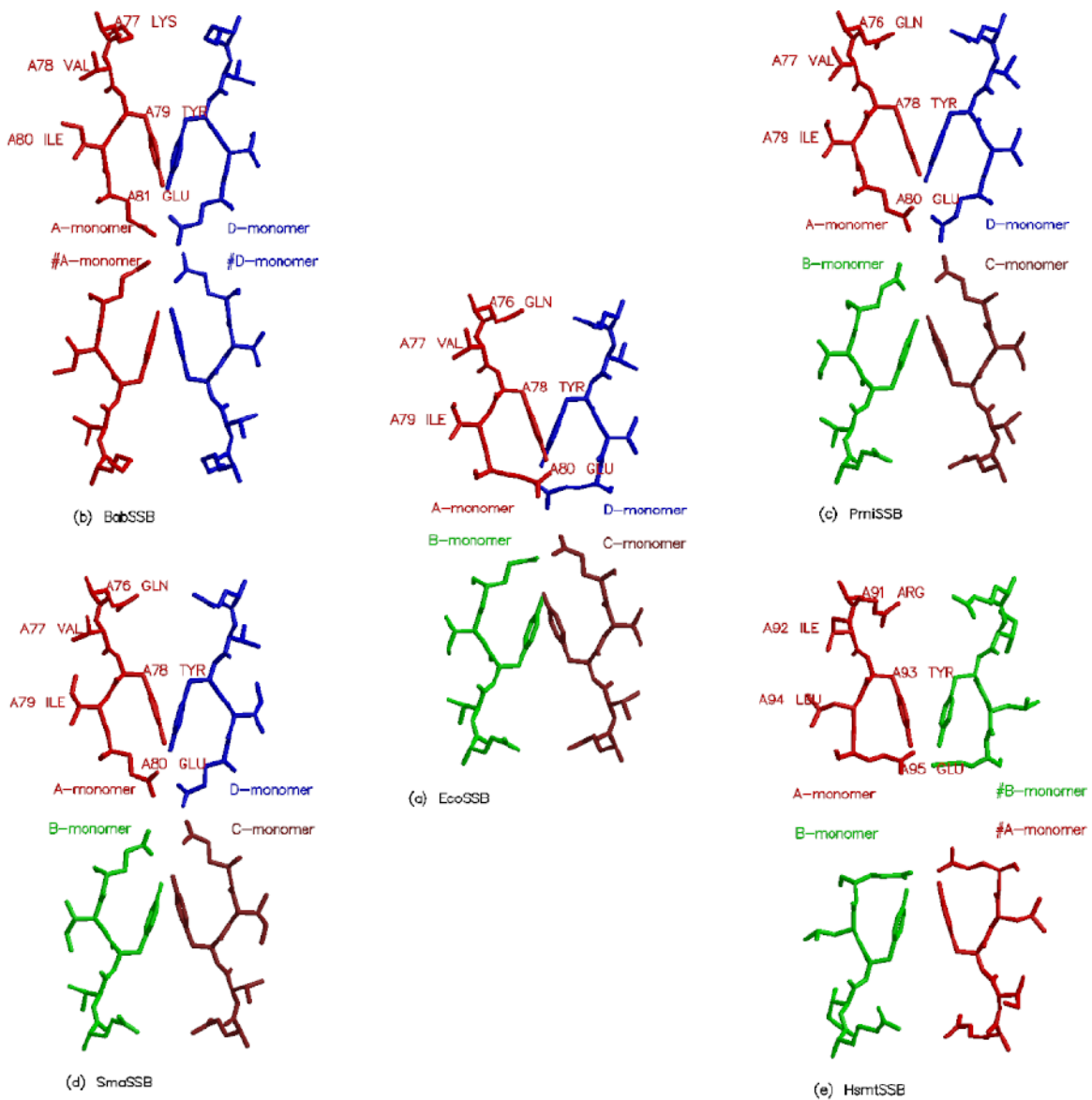
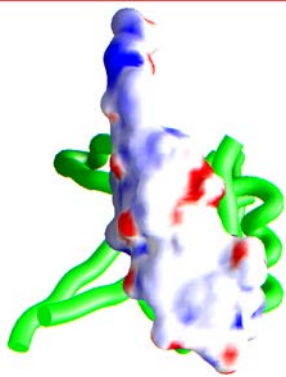
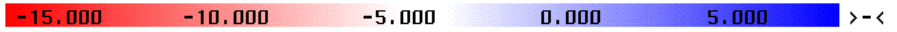
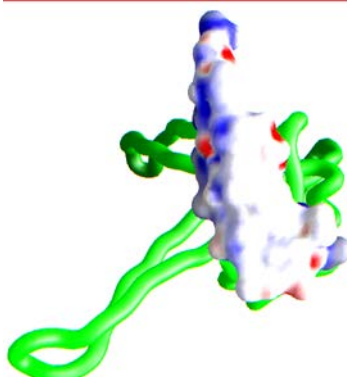


Figure 49 Partial view of the tetramer interface of SSBs embedding residues (e.g. Tyr-78) from β -strand IV.

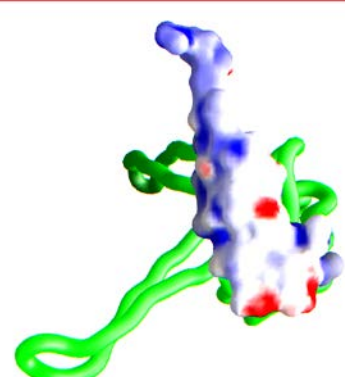
Surface Potential



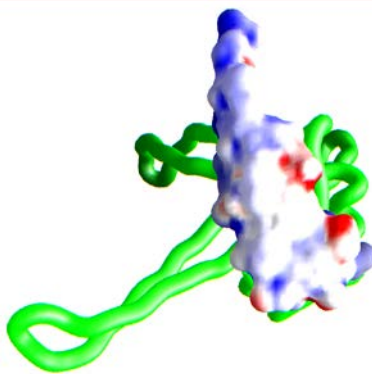
(a) A-monomer of *EcoSSB*



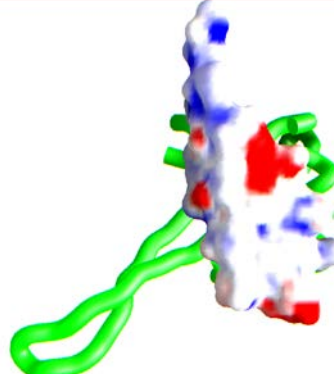
(b) B-monomer of *EcoSSB*



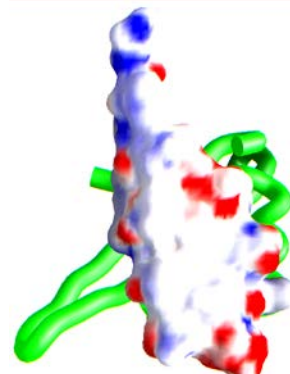
(c) C-monomer of *EcoSSB*



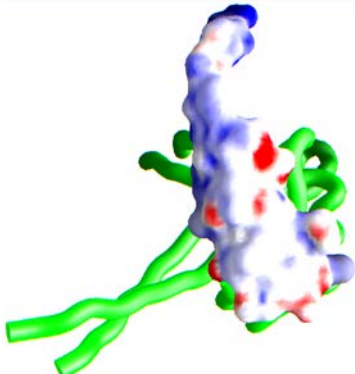
(d) D-monomer of *EcoSSB*



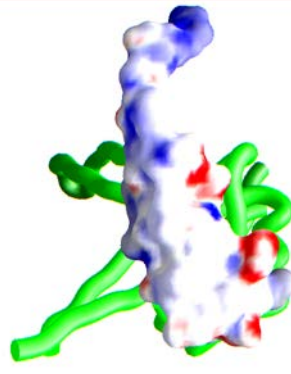
(e) A-monomer of *BabSSB*



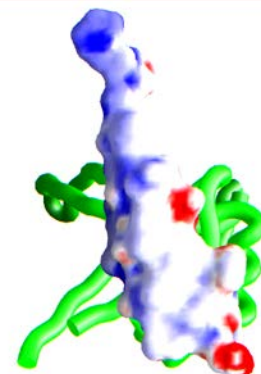
(f) B-monomer of *BabSSB*



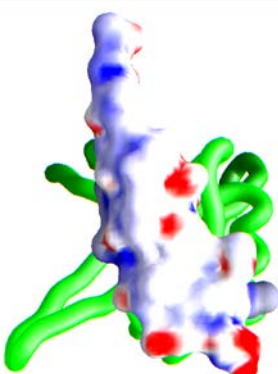
(g) A-monomer of *PmiSSB*



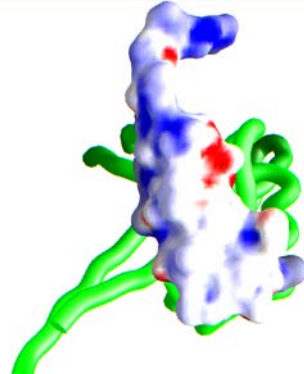
(h) B-monomer of *PmiSSB*



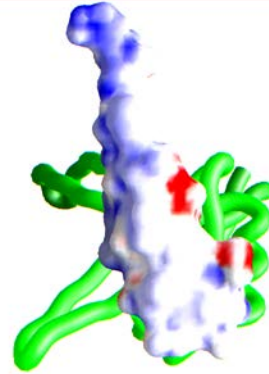
(i) C-monomer of *PmiSSB*



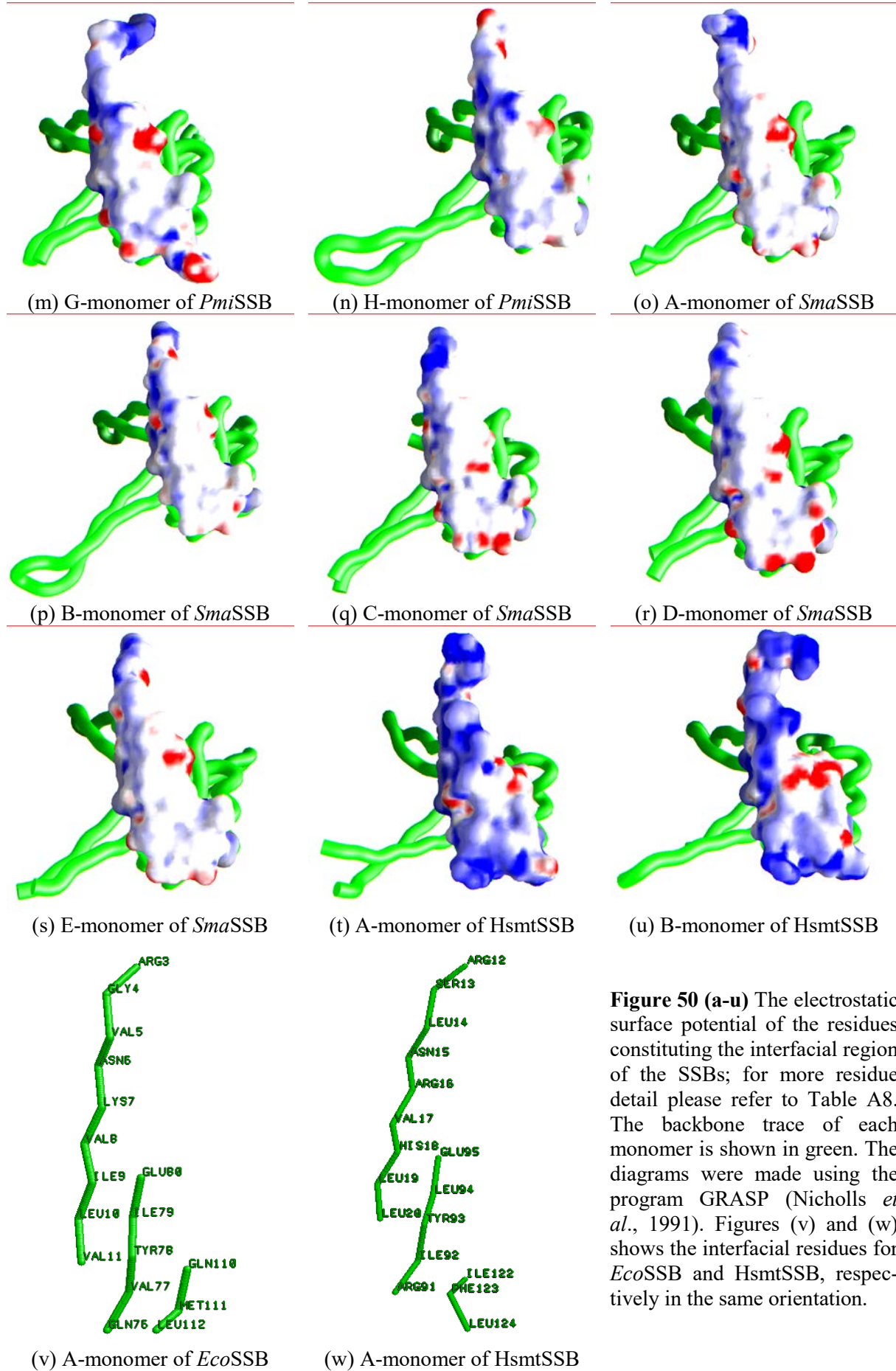
(j) D-monomer of *PmiSSB*



(k) E-monomer of *PmiSSB*



(l) F-monomer of *PmiSSB*



Among these residues, Tyr-78 is important for the dimer-dimer interface stabilisation. This residue is also conserved in other bacterial SSB as well as in HsmtSSB. The importance of this residue is described in detail in section 3.3.6. The distance between the residues at this dimer-dimer interface (which is shown in Figure 49) is tabulated in the Table A9. Buried surface area upon the tetramer formation in HsmtSSB (2300 Å²) is larger than in the bacterial SSBs (1890 Å²). The interface is reinforced by a network of hydrogen bonded tyrosine, glutamic acid, lysine, and glutamine residues [Lys-7, Gln-76, Tyr-78, Glu-80, Gln-110 (*EcoSSB* numbering)]. Lys-7 and Glu-80 form a salt bridge, which is conserved in all SSBs. In light of the similarities between various SSBs one would have expected similar electrostatic potential properties on the interfacial surfaces across different species. However, a look at Figure 50 suggests that there are substantial variations. For example, in HsmtSSB, the interfacial surface is mostly positively charged, however, it is strikingly opposite for *BabSSB*. Even within the bacterial SSBs there are noticeable distinctions; these are also perceptible across non-crystallographic monomers from the same species. For instance, the distribution of the highly-negatively charged regions of the B-monomer from *BabSSB*, is rather disperse when compared to its A-monomer. This is true for almost all the non-crystallographic monomers from the SSBs. Considering the fact that the residues are mostly conserved in the bacterial SSBs (Figure 11), it appears that the placement of the side chains (Figure 51) are a major cause of altering the surface electrostatic properties.

	1	10	20																										
1: <i>EcoSSB</i>	A	S	R	G	V	N	K	V	I	L	V	G	N	L	G	Q	D	P	E	V	R	Y	M	P	N	G	G	A	V
2: <i>StnaSSB</i>	A	S	R	G	V	N	K	V	I	L	V	G	N	L	G	Q	D	P	E	V	R	Y	M	P	N	G	G	A	V
3: <i>PmiSSB</i>	A	S	R	G	V	N	K	V	I	L	V	G	N	L	G	Q	D	P	E	I	R	Y	M	P	S	G	G	A	V
4: <i>BabSSB</i>	M	A	G	S	V	N	K	V	I	L	V	G	N	L	G	A	D	P	E	I	R	R	L	N	S	G	D	M	V
	30	40	50																										
1: <i>EcoSSB</i>	A	N	I	T	L	A	T	S	E	S	W	R	D	K	A	T	G	E	M	K	E	Q	T	E	W	H	R	V	V
2: <i>StnaSSB</i>	A	N	I	T	L	A	T	S	E	S	W	R	D	K	A	T	G	E	Q	K	E	K	T	E	W	H	R	V	V
3: <i>PmiSSB</i>	A	N	L	T	L	A	T	S	E	S	W	R	D	K	Q	T	G	E	M	K	E	K	T	E	W	H	R	V	V
4: <i>BabSSB</i>	A	N	L	R	I	A	T	S	E	S	W	R	D	R	Q	T	G	E	R	K	D	R	T	E	W	H	S	V	V
	60	70	80																										
1: <i>EcoSSB</i>	L	F	G	K	L	A	E	V	A	S	E	Y	L	R	K	G	S	Q	V	Y	I	E	G	Q	L	R	T	R	
2: <i>StnaSSB</i>	L	F	G	K	L	A	E	V	A	G	E	Y	L	R	K	G	S	Q	V	Y	I	E	G	S	L	Q	T	R	
3: <i>PmiSSB</i>	I	F	G	K	L	A	E	I	A	G	E	Y	L	R	K	G	S	Q	V	Y	I	E	G	Q	L	Q	T	R	
4: <i>BabSSB</i>	I	F	N	E	N	L	A	K	V	A	E	Q	Y	L	K	K	G	A	K	V	Y	I	E	G	A	L	Q	T	R
	90	100	110																										
1: <i>EcoSSB</i>	K	W	T	D	Q	S	G	Q	D	R	Y	T	T	E	V	V	V	N	V	G	G	T	M	Q	M	L	G		
2: <i>StnaSSB</i>	K	W	Q	D	Q	S	G	Q	D	R	Y	T	T	E	I	V	V	N	V	G	G	T	M	Q	M	L	G		
3: <i>PmiSSB</i>	K	W	Q	D	Q	S	G	Q	D	R	Y	S	T	E	V	V	V	N	I	G	G	T	M	Q	M	L	G		
4: <i>BabSSB</i>	K	W	Q	D	Q	N	G	N	D	R	Y	S	K	E	I	V	L	Q	K	F	R	G	E	L	Q	M	L	D	

Figure 51 A part of the Figure 11, was reproduced to indicate the residues that have different side chain conformations for the conserved regions of polypeptide chains in at least one of the SSBs (indicated by a box). The numbering scheme is according to *BabSSB* numbering.

3.8.4 Comparison of tetramers

The r.m.s.d. values on the basis of C_{α} pairs of crystallographic and non-crystallographic tetramers of different SSBs were calculated in three ways: (i) either including all residues which are involved in secondary structure formation, (ii) excluding the loops regions and (iii) excluding all loop regions except Loop III (Table A5).

The r.m.s.d. value varies from 0.54 Å (between two non-crystallographic tetramers of *Pmi*SSB) to 3.34 Å (between crystallographic tetramer of *Sma*SSB and *Hsmt*SSB) when the whole length of the polypeptide chains were used for the calculations. In general the *Hsmt*SSB exhibits more variation in fold when compared to bacterial SSBs. Among the bacterial SSBs, the crystallographic tetramer of *Sma*SSB is dissimilar; this is due to substantial variation of the polypeptide segments in the loop regions, mostly in Loop III (Figure 52)

On excluding all loops from the calculations above, the r.m.s.d. values among various pairs were reduced considerably [ranging from 0.33 Å (between two non-crystallographic tetramers of *Pmi*SSB) to 2.19 Å (between the crystallographic tetramer of *Bab*SSB and *Hsmt*SSB)]. This is quite natural since the loop regions of proteins are known to be more variable than the core.

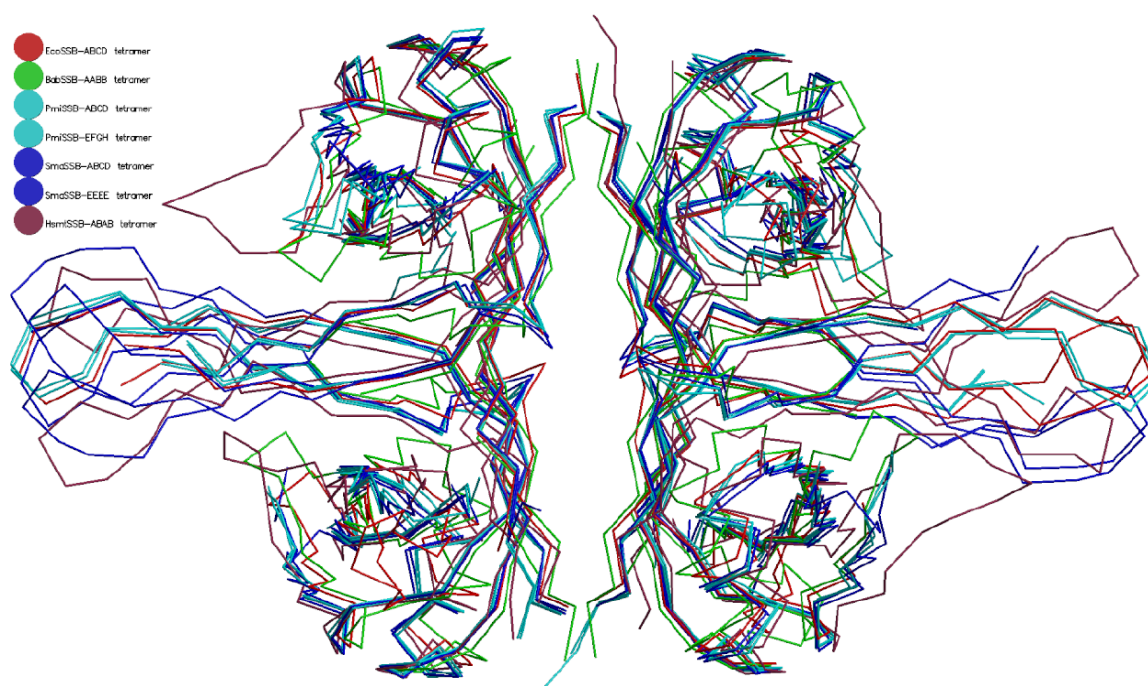


Figure 52 Superposition of SSB-tetramers.

As Loop III is known to be directly involved in ssDNA-binding, it is of interest to discern how conserved the conformation of this loop is? A third pair of r.m.s.d. values were calculated for this purpose excluding all loops except Loop III. The results (Table A5) suggest

that the conformation of the loop is not quite preserved. Possibly the distortion in conformation is induced by intermolecular contacts in crystal lattice. It can also be suggested that the flexibility of this loop is an attribute the protein uses to bind the ssDNA *in vivo* optimally. This is highlighted by the fact that this loop region is completely disordered in *BabSSB*.

3.8.5 Conservation of water molecules in SSB structures

In order to speculate on the role of the conserve water molecules in SSB structures, the non-crystallographic *EcoSSB* tetramer, the crystallographic *BabSSB* and two non-cystallographic *PmiSSB* tetramers were compared; they contain 104, 226 and 90 water molecules, respectively.

Two other SSB structures (*SmaSSB* and *HsmtSSB*) were not considered in this regard, because of the limited resolution of these structures.

The conservation of water molecules was calculated from the following criteria: First, the tetramers (*BabSSB*, *PmiSSB*) were superimposed onto the *EcoSSB* tetramer after a pair-wise alignment of the C α atoms as described above. The structurally equiositioned water molecules were identified as those being closer than a cut-off radius of 1.0 Å. The results from this alignment of water molecules in the different SSB structures using cut-off radius of 1.0 Å are listed in Table A6.

It is interesting to note that the conserved water molecules in these structures belong to the core of the protein. Water molecules near Asn-6, Thr-36, Glu-53, His-55, Arg-56 and Tyr-97 (*EcoSSB* numbering) are conserved almost in all monomers of each SSB tetramer (Table A6). These water-bound residues are found at the monomer-monomer interface in each SSB tetramer. His-55 is well known for its role in the tetramerisation of *EcoSSB*, and this residue is important in the other bacterial SSBs as well for the same purpose. Therefore it is likely that water molecules which are at the hydrogen bond distance from these residues, play a crucial role in the stabilisation of momomer-monomer interface in SSBs structures. Three inter hydrogen-bonded water molecules (W31, W21 and W45, example from *EcoSSB* (Figure 26)) proximal to the ND1 atom of the conserved His-55 could be located in the well-resolved crystal structures.

The analysis in general showed that a large fraction of the water molecules although not conserved across all SSBs introduced during the refinement of the model could be regarded as structural, in the sense that they are important for the folding and the function of the protein.

3.8.6 Comparison of crystal packing of SSBs

Although the overall structure of bacterial SSBs is much the same to that of *HsmtSSB*, there are some significant differences between these structures, which can be ascribed to the

different intermolecular contact resulting from the different crystal packing of these structures: *EcoSSB* in C2 monoclinic space group, *BabSSB*, *SmaSSB* and *HsmtSSB* in tetragonal space group and *PmiSSB* in orthorhombic space group (Table 15). These SSBs exhibits different number of molecules in the asymmetric unit, however the packing density is similar in each case (Table 15).

The crystal packing details of each SSB show an identical head-to-head tetramer formation across different species, suggesting their conservation in prokaryotes; a comparison revealed a similar arrangement in human mitochondrial SSB. A tail-to-tail tetramer is not common in these SSBs (i.e. *HsmtSSB*, *BabSSB*), indicating that the head-to-head tetramer may be the biological entity (Figure 28, 32, 39, and 46).

Despite the SSB structures being rich in β -strands, mostly the loops are involved in crystal packing. In each SSB structure high temperature factors for the loop regions (Figure A2, A3, and A4) can be observed.

Table 16 Nomenclature of loops

	Bacterial SSB (<i>EcoSSB</i> numbering)	HsmtSSB
Loop I	Residues 23 to 27	Residues 31 to 37
Loop II	Residues 40 to 50	Residues 50 to 60
Loop III	Residues 85 to 100	Residues 95 to 110
Loop IV	Residues 103-108	Not present

Loop I of *HsmtSSB* shows high r.m.s.d. between its own monomers. However, the loop in bacterial SSBs shows only small structural difference among their own monomers. The fold of this loop in *EcoSSB* is similar to that of *BabSSB* and *SmaSSB* but different from the *HsmtSSB* and *PmiSSB* (Figure 47).

Residues in Loop II exhibit the greatest difference in the backbone conformations (Figure 47). This loop is not visible in *HsmtSSB* and is disordered in the bacterial SSBs (A-monomer of *EcoSSB*, the B-monomer of *BabSSB*, and all monomers except the H-monomer of *PmiSSB* and all monomers of *SmaSSB*).

Loop III is the most important and the longest loop in the tetrameric SSBs. It is involved in ssDNA-binding and in the formation of tail-to-tail tetramer in bacterial SSBs. This loop of any *EcoSSB* monomer is similar to its other monomers. All monomers of *PmiSSB*, and three

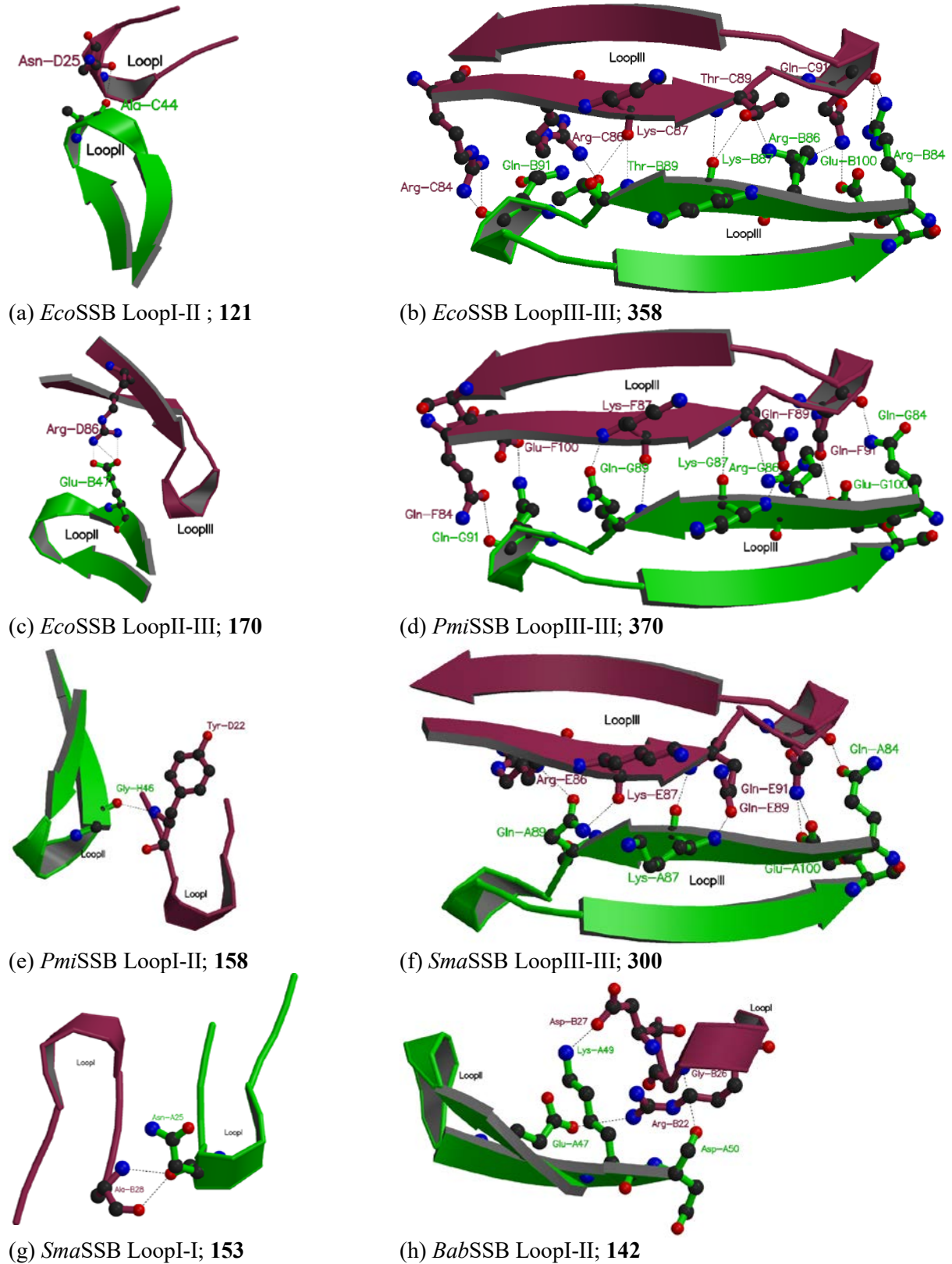


Figure 53 Diagram showing hydrogen-bonding interactions between various loop regions in bacterial SSBs. The buried surface area (\AA^2) associated with loop-interaction is shown in bold in Figure sub-legend. Further details involving hydrogen-bonding contacts are given in Table A14-A17.

monomers (A-, B-, and C-monomers) of the *Sma*SSB, this loop is similar to any *Eco*SSB monomer. The loop in HsmtSSB and two monomers (D- and E-monomers) of the *Sma*SSB are different.

Between *Eco*SSB, *Pmi*SSB and *Sma*SSB, Loop IV shows very little structural differences, however, it deviates significantly in the case of *Bab*SSB, because of the insertion of a single residue between 104 and 105 (*Eco*SSB numbering) (Figure 47).

The structural deviations in loops in each monomer of the different structures are probably due to the difference in crystal contacts. In the *Eco*SSB, *Bab*SSB and *Pmi*SSB structure, Loop I of one monomer packs against Loop II of the other monomer, however the type of contacts are different in these three structures (Figure 53a, h, e); whereas in the HsmtSSB, Loop I of one monomer is in association with the β -strand III of the other monomer and this loop of one monomer interacts with its symmetry mate in *Sma*SSB (Figure 53g).

Loop III of the *Eco*SSB, *Pmi*SSB and *Sma*SSB is profoundly involved in the crystal contact. It interacts with Loop III and II of another monomer (Figure 53 (b-f)). Gln-91 is located at the tip of Loop III forms a hydrogen bond with Asn-31 located at the β -strand III in all three SSBs structures. As mentioned above, this loop is completely disordered in the *Bab*SSB. In contrast, it shows only few crystal contacts via salt bridges in the HsmtSSB structure (Arg-B16:Glu-B95; Arg-A22:Asp-B105).

Loop IV is not involved in any crystal contact in the bacterial SSBs (except in *Bab*SSB), and this loop is absent in the HsmtSSB. The loop of *Bab*SSB interacts with the base of Loop II via salt bridges (Glu-A38:Arg-A108; Glu-B38:Arg-B108).

Since *Bab*SSB and HsmtSSB contain two and *Sma*SSB contains five monomers in their asymmetric units, crystal contacts between the β -strands of both monomer of the *Bab*SSB, and HsmtSSB and the fifth monomer of *Sma*SSB upon tetramer formation are observed (Table A15, A18 and A17, respectively).

It appears that different non-covalent forces that steer crystal growth and kinetics are different in SSBs due to which they crystallise in different space group. A large part of the SSB sequences are similar in amino acid and fold (especially the core regions). The loop regions in the SSBs are quite variable and most of them form crystal contacts which are involved in different packing arrangements. Loop III especially, forms major crystal contacts (Table A14-A17) and this region in SSBs is highly conserved (sequentially, Figure 11).

An easy way to estimate variations in non-covalent forces that arise due to differences in the placements of the polar and non-polar residues in the three dimensional protein molecule may be obtained by measuring the hydrophobic moment of the structure. The concept first proposed by Eisenberg *et al.* (1984) can be explained as follows:

Hydrophobic moment μ is the vector sum of the individual hydrophobicities (Kyte & Doolite, 1982). Hydrophobicity is treated as a vector or a quantity with both magnitude and direction. If all side chains are hydrophobic, then the vectors cancel, and the hydrophobic moment is low. If one side is hydrophilic, as in an amphipathic helix, its vectors are negative in magnitude, and reinforce the positive hydrophobic vectors on the opposite side. A high hydrophobic moment therefore indicates considerable heterogeneity of the protein milieu and various thermodynamic processes that the protein molecule takes part in should be substantially influenced by the differential interplay of the resultant non-covalent forces.

The calculation of the hydrophobic moment of various SSB tetramers was done using all the co-ordinates, and repeated, removing all the loop regions. The results are listed in Table 17.

As we have seen in earlier sections, the core region of these SSBs is conserved with minor differences; however, they exhibit significant hydrophobic moments. The differences in the moment magnitudes between various pairs of tetramers are also considerable indicating that even though the amino acid positions in the sequence are conserved, the side-chain orientation is sometimes different (Figure 51). Since a majority of the side-chain atoms make up the protein surface which form packing contacts, it is obvious their variations cause differences in surface properties resulting in differences in protein-protein interaction and the resultant crystal packing.

Table 17 Hydrophobic moment of SSBs

	¹ μ	² μ
<i>Eco</i> SSB-ABCD	6.2	4.9
<i>Bab</i> SSB-AABB	2.4	0.4
<i>Pmi</i> SSB-ABCD	9.9	4.9
<i>Pmi</i> SSB-EFGH	8.2	5.3
<i>Sma</i> SSB-ABCD	6.1	4.9
<i>Sma</i> SSB-EEEE	0.5	0.5
<i>Hsmt</i> SSB-ABAB	8.7	3.3

μ Hydrophobic moment

1 core of the protein and including loop regions

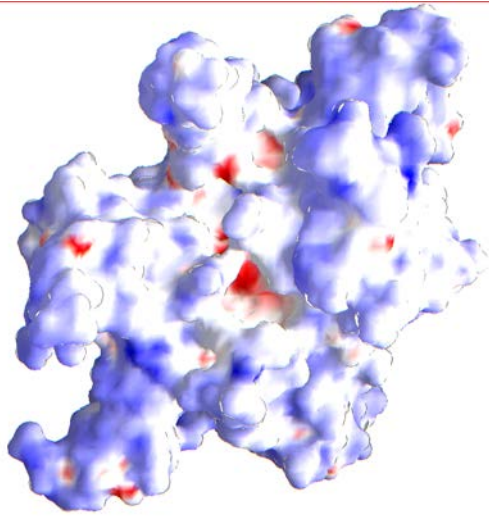
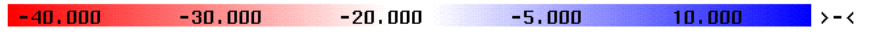
2 core of the protein

The tetrameric SSB is indicated along with the chain ids (separated by a hyphen). Repetition of the same chain id indicates that the tetramer was formed using symmetry related molecules.

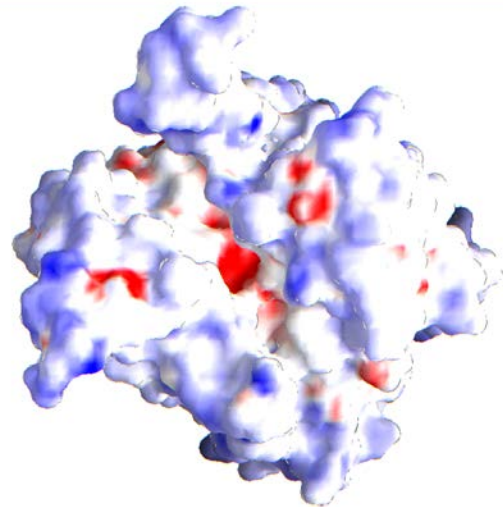
This was again confirmed using calculations of the electrostatic surface potential of each SSB tetramer. As evident from Figure 54, there are indeed wide variations in the

Results and Discussion

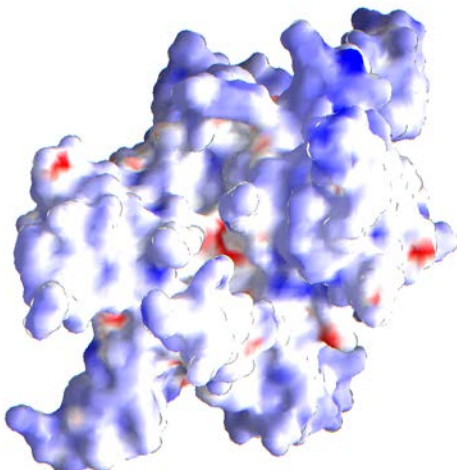
Surface Potential



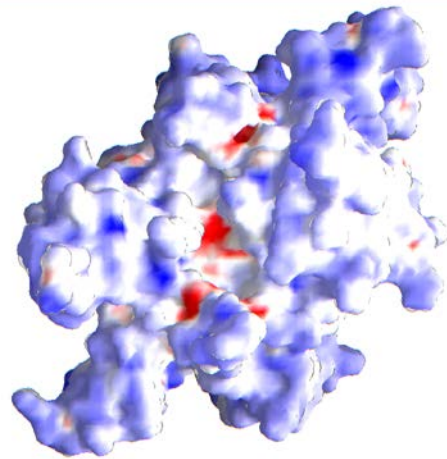
(a) *EcoSSB-ABCD*



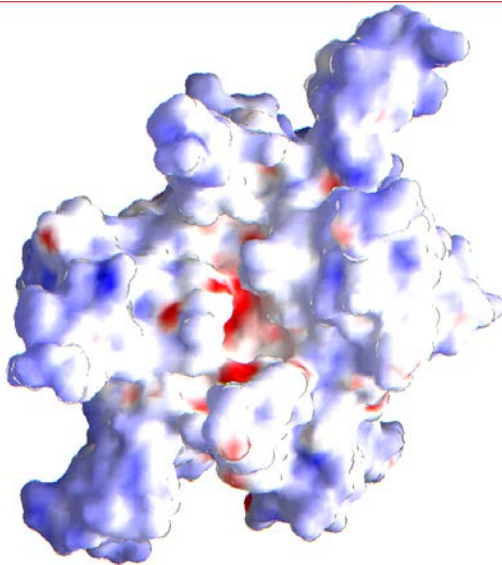
(b) *BabSSB-AABB*



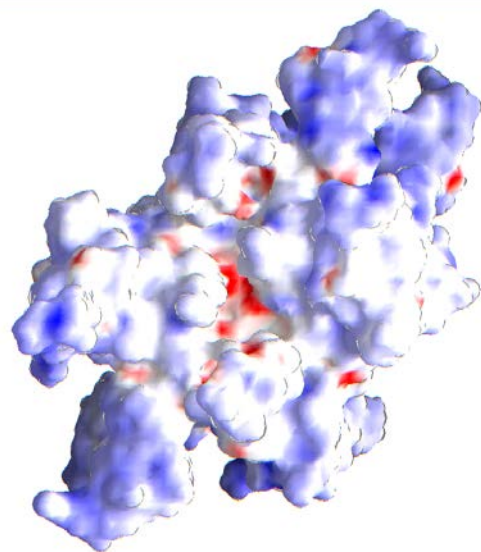
(c) *PmiSSB-ABCD*



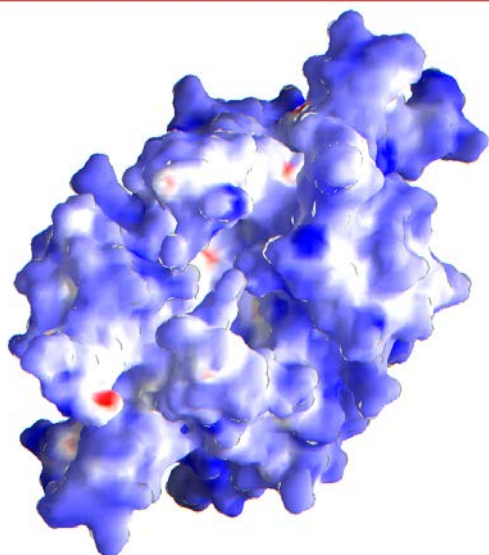
(d) *PmiSSB-EFGH*



(e) *SmaSSB-ABCD*



(f) *SmaSSB-EEEE*



(g) HsmtSSB-ABAB

Figure 54 The electrostatic surface potential of the residues of SSB tetramers are shown; the diagrams were made using the program GRASP (Nicholls *et al.*, 1991). The tetrameric SSB is indicated along with the chain ids (separated by a hyphen). Repetition of the same chain id indicates that the tetramer was formed using symmetry related molecules.

distribution of the charges on the surfaces in a very heterogeneous manner. Although a major part of the tetramer appears blue, indicating a positive potential, the white to red patches are widely distributed indicating negative potential. It is logical that the resultant forces arising out of the interplay between the force vectors originating from these varied surfaces would differ both in magnitude and direction steering different crystallogenesis processes which would crystallise proteins in distinct space groups. It is also instructive to note that electrostatic surface potential of the human mitochondrial SSB is significantly different from the bacterial SSBs (Figure 54g); the surface is almost totally covered by a positive potential. The highly negatively charged groove regions are also inconspicuous.

3.9 Global comparison of OB-fold containing proteins

As described in Introduction, SSBs contain a common fold, which is known as OB-fold. There are other proteins which also contain this fold. These are DNA-binding proteins, tRNA synthetases, Pertussis toxin, heat labile enterotoxins, Verotoxin-1, staphylococcal nuclease and Pyrophosphate phosphohydrolase. These proteins bind to either DNA or sugar.

In Figure 55, it is shown how these proteins are related to each other on the basis of r.m.s.d. of their C_{α} trace. The r.m.s.d. value for each pair for these proteins (using monomeric models) was calculated (Table A19) using the program SARF2 (Alexandrov *et al.*, 1995), and the dendrogram (Figure 55) was then drawn based on the r.m.s.d. values obtained by the program CLUSTER.

It is interesting to note that the homotetrameric SSBs form a cluster (*A*), with the r.m.s.d. values varying between 0.8 and 1.7 Å (Table A19). The monomeric class of SSBs forms a distinct cluster (*B*) from the homotetrameric SSBs. The maximum r.m.s.d. value between

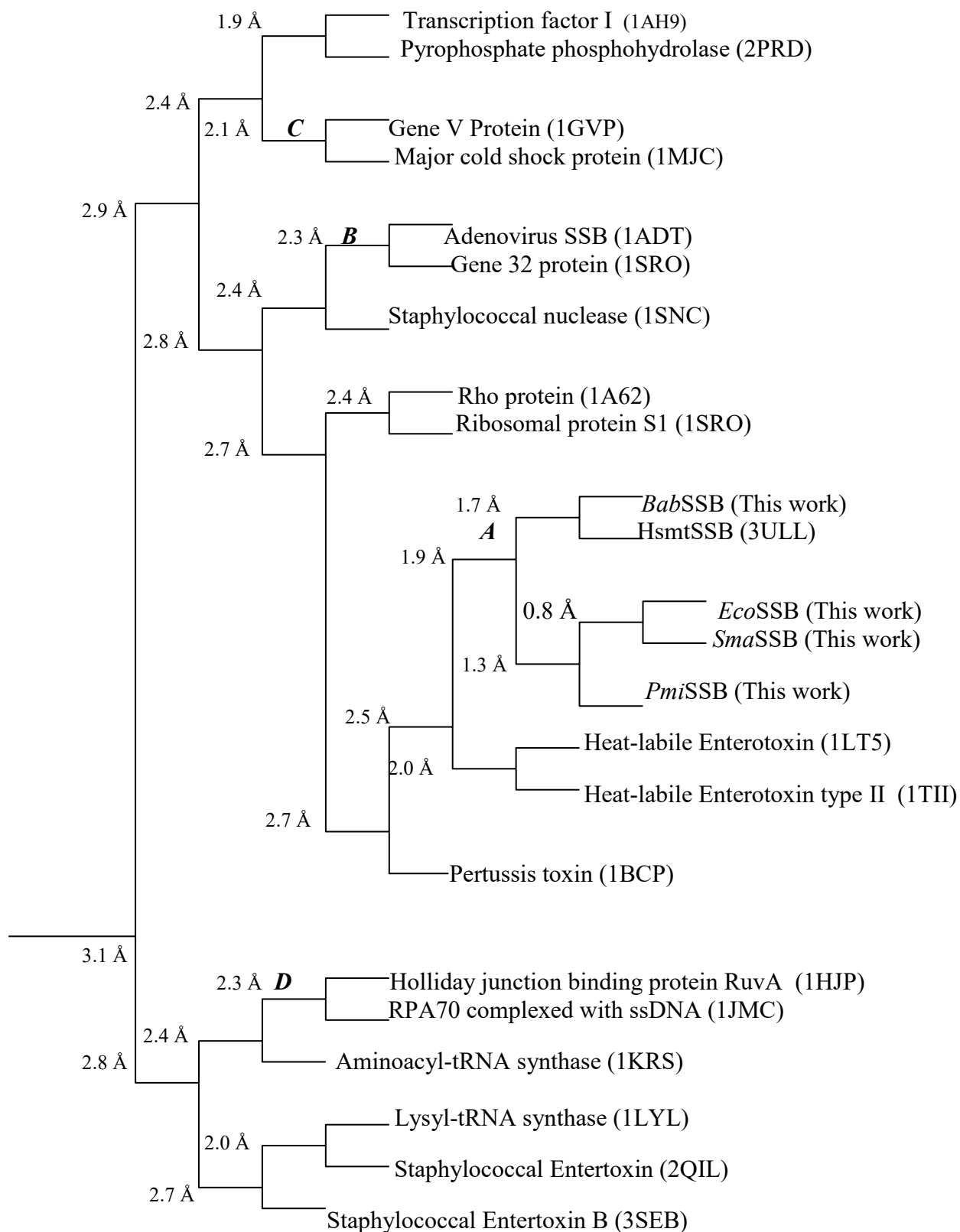


Figure 55 Dendrogram of the OB fold containing proteins. The cluster analysis was performed using hierarchical complete linkage algorithm (Johnson & Wichern, 1996; as implemented in program CLUSTER, written by Dr. Debnath Pal, IMB Jena). PDB codes of the respective proteins are given in parenthesis and few representative clusters of SSBs are shown in italics bold letters.

these two clusters is 2.8 Å. Dimeric SSBs (GVP) form a cluster (C) comprising the Major cold shock protein which shows a deviation from the monomeric and the homotetrameric SSB. Heterotrimeric SSB (RPA70) is more closely related to the Holliday junction binding protein RuvA, tRNA synthase and Staphylococcal Entertoxin than to other SSB classes.

The fold relationship in various proteins can arise in two different ways: (i) the proteins have a common ancestor (sometimes referred to as homologous fold or as members of the same super family), or (ii) the proteins has converged into a common fold and have no common evolutionary history ('analogous' fold relationship or just a fold relationship) (Russell *et al.*, 1998). It appears that SSBs may have originated from a common ancestor; this is because the fold seems to be highly conserved.

It can be suggested that the precise folding of these SSBs may be a primary requirement for function. From Figure 55, one can see that all the homotetrameric SSBs have clustered in a single region (cluster A), although not all of them have very good pair wise sequence identity (30% to 89%).

When the overall sequence similarity among polypeptide sequences is not large enough, proper juxtaposition of residues at important functional sites is essential for its activity. The SSBs seem to satisfy all these requirements (Prasad & Chiu, 1987). It has earlier been suggested that heterotrimeric eukaryotic RPA (PDB code: 1JMC) and homotetrameric SSBs are derived from a common heterotetrameric ancestor protein by gene duplication, merging and reshuffling (Suck *et al.*, 1997). However, as can be seen from Figure 55, they have quite substantial variation in the fold and are more related to Holliday junction binding protein RuvA (PDB code: 1HJP). The inference from the figure supports the contention that the proposition is rather speculative (Suck, 1997). The homotetrameric SSBs seem to share some structural similarity with heat-labile enterotoxins notwithstanding their distinct features. The enterotoxins functions by binding sugars on the cell surface, one can use the result obtained here as a clue to investigate whether SSBs too have other functions; however, it appears unlikely that they share a common ancestor.

4 Summary

The architecture of bacterial single-stranded DNA binding proteins has been investigated by elucidating the structures of SSBs from *Brucella abortus*, *Proteus mirabilis*, *Serratia marsecens* by X-ray crystallographic methods. The structure of the SSB from *Escherichia coli* has been refined to a high resolution.

Two cryo-techniques: (i) repeated-annealing using dried paraffin oil and (ii) cryo-cooling using Panjelly™ were developed to overcome the problematic bottlenecks arising from poorly diffracting crystals and premature crystal decay arising from radiation damage.

The essential structure of SSBs across all species consists of five β -strands, one α -helix interspersed by four loops. It is similar to the already described Oligonucleotide/Oligosaccharide Binding (OB) fold. Out of the four loops, two loops involving residues 40 to 50 and 85 to 100 (*E. coli* numbering) are always involved in DNA-binding. The core regions comprising the entire regular secondary structure is almost identical between the different SSBs; the loop regions exhibit variable conformation. These loops are often involved in crystal packing giving rise to different combinations of loop-loop interactions. Different SSBs show identical head-to-head tetramer formation similar to the situation in human mitochondrial SSB. The principal role of the conserved His-55 residue in tetramerisation, as earlier demonstrated by biochemical experiments could be rationalised. The conserved Tyr-78 residue at the dimer-dimer interface may be also essential for tetramer formation. The conserved salt bridge between Lys-7 and Glu-80 at the same interface indicates its dominant role in stabilising the quaternary structure. Three inter hydrogen-bonded water molecules proximal to the ND1 atom of the conserved His-55 could be located in the well-resolved crystal structures; these water molecules may be functionally important. The DNA-binding aromatic residues (Trp-40, 54, 88, Phe-60) are sequentially, structurally and conformationally conserved corroborating their specific role in binding ssDNAs. Three-dimensional structural comparison of the homotetrameric SSBs along with other proteins containing the common OB-fold revealed that the fold of the monomeric SSB (Gene 32 protein) is more similar to the homotetrameric form than to the dimeric (Gene V protein) or heterotrimeric form (human replication protein 70).

5 References

- Alberts, B. M., Amodio, F.J., Jenkins, M., Gutmann, E. D., and Ferris, F. L. (1968). Studies with DNA-cellulose chromatography. I. DNA-binding proteins from *Escherichia coli*. Cold Spring Harb. Symp. Quant. Biol. **33**, 289-305.
- Alberts, B. M., and Frey, L. (1970). T4 bacteriophage Gene 32: a structural protein in the replication and recombination of DNA. Nature **227**, 1313-1318.
- Alberts, B. M., and Herrik, G. (1971). DNA-cellulose chromatography. Meth. Enzymol. **21**, 198-217.
- Alberts, B. M., and Sternglanz, R. (1977). Recent excitement in the DNA replication problem. Nature **269**, 655-661.
- Alexandrov, N. N., Nussinov R., and Zimmer R. M. (1995). Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. Pacific Symposium on Biocomputing 1996 (Lawrence Hunter & Teri E. Klein, Eds) World Scientific Publishing Co., Singapore, 53-72.
- Allen G. C., and Kornberg A. (1993). Assembly of the primosome of DNA replication in *Escherichia coli*. J. Biol. Chem. **268**, 19204-19209.
- Baker, T. A., Sekimizu, K., Funnell, B. E., and Kornberg, A. (1986). Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. Cell **45**, 53-64.
- Baker, T. A., Funnell, B. E., and Kornberg, A. (1987). Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin in vitro. J. Biol. Chem. **262**, 6877-6885.
- Baluch, J., Chase, J. W., and Sussman, R. (1980). Synthesis of recA protein and induction of bacteriophage lambda in single-strand deoxyribonucleic acid-binding protein mutants of *Escherichia coli*. J. Bacteriol. **144**, 489-498.
- Balwind, J. M (1980). The structure of human carbonmonoxy haemoglobin at 2.7 Å resolution. J. Mol. Biol. **106**, 103-128.
- Barton, G. J. (1993). ALSCRIPT a tool to format multiple sequence alignments. Protein Eng. **6**, 37-40.
- Berchtold, H., and Hilgenfeld, R. (1999). Binding of phenol to R6 insulin hexamers. Biopolymers. **51**, 165-172.
- Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., and Bourne P. E. (2000). The Protein Data Bank. Nucl. Ac. Res. **28**, 235-242.
- Benz, E. W., Reinberg, D., Vicuna, R., and Hurwitz, J. (1980). Initiation of DNA replication by the dnaG protein. J. Biol. Chem. **255**, 1096-106.
- Blundell, T. L., and Johnson, L. N. (1976). Protein crystallography. Academic press.
- Bobst, E. V., Bobst, A. M., Perrino, F. W., Meyer, R. R., and Rein, D. C. (1985). Variability in the nucleic acid binding site size and the amount of single-stranded DNA-binding protein in *Escherichia coli*. FEBS Lett. **181**, 133-137.
- Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature **385**, 176-181.
- Bochkarev, A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1999). The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. EMBO J. **18**, 4498-4504.
- Bochkareva, E., Frappier, L., Edwards, A. M., and Bochkarev, A. (1998). The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain. J. Biol. Chem. **273**, 3932-3947.

References

- Bouche, J. P., Zechel, K., and Kornberg, A. (1975). DnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. *J. Biol. Chem.* **250**, 5995-6001.
- Brayer, G. D., and McPherson, A. (1983). Refined structure of the gene 5 DNA binding protein from bacteriophage fd. *J. Mol. Biol.* **169**, 565-596.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization and dynamics calculation. *J. Comp. Chem.* **4**, 187-217.
- Brünger, A. T., Krukowski, A., and Erickson, J. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* **A46**, 585-593.
- Brünger, A. T. (1992b). The Free-R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921.
- Bujalowski, W., and Lohman, T. M. (1986). *Escherichia coli* single strand binding protein forms multiple, distinct complexes with single stranded DNA. *Biochemistry* **25**, 7799-7802.
- Bujalowski, W., and Lohman, T. M. (1987). Limited cooperativity in protein-nucleic acid interactions. A thermodynamic model for the interactions of *Escherichia coli* single strand binding protein with single stranded nucleic acids in the "beaded", (SSB)₆₅ Mode. *J. Mol. Biol.* **195**, 897-907.
- Bujalowski, W., Overman, L. B., and Lohman, T. M. (1988). Binding mode transitions of *Escherichia coli* single strand binding protein-single stranded DNA complexes cation, anion, pH and binding density effects. *J. Biol. Chem.* **263**, 4629-4640.
- Bujalowski, W., and Lohman, T. M. (1989). Negative cooperativity in *Escherichia coli* single strand binding protein - oligonucleotide interactions, evidence and a quantitative model. *J. Mol. Biol.* **207**, 249-268.
- Bujalowski, W., and Lohman, T. M. (1989). Negative cooperativity in *Escherichia coli* single strand binding protein-oligonucleotide interactions. II. salt, temperature and oligonucleotide length effects. *J. Mol. Biol.* **207**, 269-288.
- Bujalowski, W., Lohman, T. M., and Anderson, C. F. (1989). On the cooperative binding of large ligands to a one-dimensional homogeneous lattice: The generalized three-state lattice model. *Biopolymers* **28**, 1637-1643.
- Carlini, L. E., Curth, U., Kindler, B., Urbanke, C., and Porter, R. D. (1998). Identification of amino acids stabilizing the tetramerization of the single stranded DNA binding protein from *Escherichia coli* FEBS Lett. **430**, 197-200.
- Casas-Finet, J. R., Khamis, M. I., Maki, A. H., and Chase, J. W. (1987). Tryptophan 54 and phenylalanine 60 are involved synergistically in the binding of *Escherichia coli* SSB protein to single-stranded polynucleotides. FEBS Lett. **220**, 347-352.
- Cha, T. A., and Alberts, B. M. (1988). In vitro studies of the T4 bacteriophage DNA replication system. *Cancer cells* **6**, 1-10.
- Cha, T. A., and Alberts, B. M. (1989). The bacteriophage T4 DNA replication fork. Only DNA helicase is required for leading strand DNA synthesis by the DNA polymerase holoenzyme. *J. Biol. Chem.* **264**, 12220-12225.
- Chayen, N. E. (1997). A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Cryst.* **30**, 198-202.

References

- Chase, J. W., Merrill B. M., and Williams, K. R. (1983). F sex factor encodes a single-stranded DNA binding protein (SSB) with extensive sequence homology to *Escherichia coli* SSB. Proc. Natl. Acad. Sci. USA. **80**, 5480-5484.
- Chase, J. W., L'Italien, J. J., Murphy, J. B., Spicer, E. K., and Williams, K. R., (1984). Characterization of the *Escherichia coli* SSB-113 mutant single-stranded DNA-binding protein. Cloning of the gene, DNA and protein sequence analysis, high pressure liquid chromatography peptide mapping, and DNA-binding studies. J. Biol. Chem. **259**, 805-814.
- Chase J. W., and Williams K. R. (1986). Single-stranded DNA binding proteins required for DNA replication. Ann. Rev. Biochem. **55**, 103-136.
- Chen, J., Smith, D. L, and Griep, M. A. (1998). The role of the 6 lysines and the terminal amine of *Escherichia coli* single-strand binding protein in its binding of single-stranded DNA. Protein Sci. **8**, 1781-1788.
- Chrysogelos, S., and Griffith, J. (1982). *Escherichia coli* single-stranded binding protein organizes single-stranded DNA in nucleosome-like units. Proc. Natl. Acad. Sci. USA **79**, 5803-5807.
- Cohen, G. E., (1997). ALIGN: a program to superimpose protein co-ordinates, accounting for insertion and deletions. J. Appl. Cryst. **30**, 1160-1161.
- Craig, N. L., and Roberts J. W. (1980). *E. coli* recA protein-directed cleavage of phage λ repressor requires polynucleotide. Nature **283**, 26-30.
- Curth, U., Bayer, I., Greipel, J., Mayer, F., Urbanke, C. and Maass, G. (1991). Amino Acid 55 plays a central role in tetramerization and function of *Escherichia coli* single stranded DNA binding protein. Eur. J. Biochem. **196**, 87-93.
- Curth, U., Greipel, J., Urbanke, C. and Maass, G. (1993). Multiple binding modes of the single-stranded DNA binding protein from *Escherichia coli* as detected by tryptophan fluorescence and site-directed mutagenesis. Biochemistry **32**, 2585-2591.
- Curth, U., Urbanke, C., Greipel, J., Gerberding, H., Tiranti, V. and Zeviani, M. (1994). Single-stranded-DNA-binding proteins from human mitochondria and *Escherichia coli* have analogous physicochemical properties. Eur. J. Biochem. **221**, 435-443.
- Curth, U., Genschel, J., Urbanke, C. and Greipel, J. (1996). In vitro and in vivo function of the carboxyterminus of *Escherichia coli* single-stranded DNA binding protein. Nucleic Acids Res. **24**, 2706-2711.
- Deluis. H., Mantell, N. J., and Alberts, B. M. (1972). Characterization by electron microscopy of the complex formed between T4 bacteriophage Gene 32-protein and DNA. J. Mol. Biol. **67**, 341-350.
- de Vries, J., and Wackernagel, W. (1993). Cloning and sequencing of the *Serratia marcescens* gene encoding a single-stranded DNA-binding protein (SSB) and its promoter region. Gene **127**, 39-45.
- de Vries, J., Genschel, J., Urbanke, C., Thole, H., and Wackernagel, W. (1994). The single-stranded-DNA-binding proteins (SSB) of *Proteus mirabilis* and *Serratia marcescens*. Eur. J. Biochem. **224**, 613-622.
- de Vries, J., and Wackernagel, W. (1994). Cloning and sequencing of the *Proteus mirabilis* Gene for a single-stranded DNA-binding protein (SSB) and complementation of *Escherichia coli* ssb point and deletion mutations. Microbiology **140**, 889-895.
- Diederichs K., and Karplus P. A. (1997). Improved R-factor for diffraction data analysis in macromolecular crystallography. Nature Struc. Biol. **5**, 269-275.
- Dornreiter, I., Erdile, L. F., Gillbert, I. U., von Winkler, D., Kelly, T. J., and Fanning, E. (1992). Interaction of DNA polymerase α -primase with cellular replication protein A and SV40 T antigen. EMBO J. **11**, 769-776.
- Eisenberg, D., Weiss, R. M., and Terwilliger T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. USA **81**, 140-144.

References

- Engh, R. A., and Huber, R. (1991). Accurate bond angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.
- Ennis, D. G., Amundsen, S. K., and Smith, G. R. (1987). Genetic functions promoting homologous recombination in *Escherichia coli*: a study of inversions in phage lambda. *Genetics* **115**, 11-24.
- Esnouf, R. M. (1999). Further additions to Molscript version 1.4, including reading and contouring electron density maps. *Acta Cryst.* **D55**, 938-940.
- Ferrari, M. E., Bujalowski, W., and Lohman, T. M. (1994). Cooperative binding of *Escherichia coli* SSB tetramers to single stranded DNA in the (SSB)₃₅ binding mode. *J. Mol. Biol.* **236**, 106-123.
- Ferrari, M. E., Fang, J., and Lohman, T. M. (1997). A Mutation in *Escherichia coli* SSB protein (W54S) alters intra-tetramer negative cooperativity and inter-tetramer positive cooperativity for single stranded DNA binding. *Biophys. chem.* **64**, 235-251.
- Folkers, P. J., Nilges, M., Folmer, R. H., Konings R. N., and Hilbers C. W. (1994). The solution structure of the Tyr41-->His mutant of the single-stranded DNA binding protein encoded by gene V of the filamentous bacteriophage M13. *J. Mol. Biol.* **236**, 229-246.
- Folmer, R. H., Nilges, M., Konings, R. N., and Hilbers C.W. (1995). Solution structure of the single-stranded DNA binding protein of the filamentous Pseudomonas phage Pf3: similarity to other proteins binding to single-stranded nucleic acids. *EMBO J.* **14**, 4132-4142.
- French, G. S., and Wilson K. S. (1978). On the treatment of negative intensity observations. *Acta. Cryst.* **A34**, 517-525.
- Friedber, E. C. (1985). DNA repair. W. H. Freeman & Co., San Francisco.
- Fulford, W., and Model P. (1988). Bacteriophage fl DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. *J. Mol. Biol.* **203**, 39-48.
- Fuller, R. S., Kaguni, J. M., and Kornberg, A. (1981). Enzymatic replication of the origin of the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. USA* **78**, 7370-7374.
- Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* **38**, 889-900.
- Funnell, B. E., Baker, T. A., and Kornberg, A. (1987). In vitro assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **262**, 10327-10334.
- Garman, E. F., and Schneider, T. R. (1997). Macromolecular cryocrystallography. *J. Appl. Cryst.* **30**, 211-237.
- Geider. K., and Kornberg A. (1974). Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins. *J. Biol. Chem.* **249**, 3999-4005.
- Geider, K. (1978). Interaction of DNA with DNA-binding proteins: protein exchange and complex stability. *Eur. J. Biochem.* **87**, 617-622.
- Genschel, J. Curth, U., and Urbanke C. (2000). Interaction of *E. coli* single-stranded DNA binding protein (SSB) with exonuclease I. The carboxy-terminus of SSB is the recognition site for the nuclease. *Biol. Chem.* **381**, 183-192.
- Giedroc, D. P., Keating, K. M., Williams. K. R., and Coleman J. E. (1987). The function of zinc in gene 32 protein from T4. *Biochemistry.* **26**, 5251-5259.
- Glassberg, J., Meyer, R. R., Kornberg, A. (1979). Mutant single-strand binding protein of *Escherichia coli*: genetic and physiological characterization. *J. Bacteriol.* **140**, 14-19.
- Glover, B. P., and McHenry, C. S. (1998). The χ ψ subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J. Biol. Chem.* **273**, 23476-23484.

References

- Golub, E.I., and Low, K. B. (1983). Indirect stimulation of genetic recombination. *Proc. Natl. Acad. Sci. USA* **80**, 1401-1405.
- Gomes, X. V., and Wold, M. S. (1995). Structural analysis of human replication protein A. Mapping functional domains of the 70 kDa subunit. *J. Biol. Chem.* **270**, 4534-4543.
- Gomes, X. V., and Wold, M. S. (1996). Functional domains of the 70-kilodalton subunit of human replication protein A. *Biochemistry* **35**, 10558-10568.
- Harp, J. M., Timm, D. E., and Bunick, G. J. (1998). Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Cryst.* **D54**, 622-628.
- Harp, J. M., Hanson, B. L., Timm, D. E., and Bunick, G. J. (1999). Macromolecular crystal annealing: evaluation of techniques and variables. *Acta Cryst.* **D55**, 1329-1334.
- Hahn, T. ed. (1995). *International tables for crystallography, Volume A*. Kluwer Academic Publishers.
- Hendrickson, W. A., Smith, J. L., Phizackerley, R. P., and Merrit, E. A. (1988). Crystallographic structure analysis of lamprey haemoglobin from anomalous dispersion of synchrotron radiation. *Proteins* **4**, 77-88.
- Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665-1672.
- Hilgenfeld, R., Saenger, W., Schomburg, U., and Krauss, G. (1984). Novel crystal forms of a proteolytic core of the single-stranded DNA binding protein (SSB) from *E. coli*. *FEBS Lett.* **170**, 143-146.
- Hope, H. (1990). Crystallography of biological macromolecules at ultra-low temperature. *Ann. Rev. Biophys. Chem.* **19**, 107-126.
- Hubbard, S. J., Campbell, S. F., and Thronton, J. M. (1991). Molecular recognition. conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220**, 507-530.
- Jack, A., and Levitt, M. (1978). Refinement of large structures by simultaneous minimization of energy and R Factor. *Acta Cryst.* **A34**, 931-935.
- Jancarik, J., and Kim, S. H. (1991). Sparse-matrix sampling - a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409-411.
- Johnson, B. F. (1984). Two-dimensional electrophoretic analysis of the regulation of SOS proteins in three ssb mutants. *Arch. Microbiol.* **138**, 106-112.
- Johnson R. A., and Wichern D. W. (1996). *Applied multivariable statistical analysis*. Prentice Hall of India.
- Jones, T. A., Zou, J. Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of error in these models. *Acta Cryst.* **A47**, 110-119.
- Jensen, L. (1985). Overview of refinement in macromolecular structure analysis. *Meth. Enzymol.* **115**, 227-234.
- Kabsch, W. (1988). Evaluation of single crystal x-ray data from a position sensitive detector. *J. Appl. Cryst.* **21**, 67-71.
- Kelman, Z., Yuzhakov, A., Andjelkovic, J., and O'Donnell, M. (1998). Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.* **17**, 2436-2449.
- Khamis, M. I., Casas-Finet, J. R., Maki, A. H., Murphy, J. B., and Chase, J. W. (1987). Role of tryptophan 54 in the binding of *Escherichia coli* single-stranded DNA-binding protein to single-stranded polynucleotides. *FEBS Lett.* **211**, 155-159.
- Kim, Y. T., Tabor, S., Churchich, J. E., and Richardson, C. C. (1992). Interaction of gene 2.5 protein and DNA polymerase of bacteriophage T7. *J. Biol. Chem.* **267**, 15032-15040.

References

- Kim, Y. T. and Richardson, C. C. (1994). Acidic carboxyl-terminal domain of gene 2.5 protein of bacteriophage T7 is essential for protein-protein interactions. *J. Biol. Chem.* **269**, 5270-5278.
- Klessig, D. F., and Grodzicker, T. (1979). Mutations that allow human Ad2 and Ad5 to express late genes in monkey cells map in the viral gene encoding the 72K DNA binding protein. *Cell* **17**, 957-966.
- Kleywegt, G. J., and Jones, T. A. (1997). Good model-building and refinement practice, *Meth. Enzymol.* **277**, 208-230.
- Kornberg, R. D., and Lorch Y. (1992). Chromatin structure and transcription. *Ann. Rev. Cell Biol.* **8**, 563-87.
- Kowalczykowski, S. C. (1987). Mechanistic aspects of the DNA strand exchange activity of *E. coli* RecA protein. *Trends Biochem. Sci.* **12**, 141-145.
- Kowalczykowski, S. C., Clow, J. C., Somani, R., and Varghese, A. (1987). Effects of the *Escherichia coli* SSB protein on the binding of *Escherichia coli* RecA protein to single-stranded DNA: Demonstration of competitive binding and the lack of a specific protein-protein interaction. *J. Mol. Biol.* **193**, 81-95.
- Kowalczykowski, S. C., and Krupp, R. A. (1987). Effects of the *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA protein: Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J. Mol. Biol.* **193**, 97-113.
- Kowalczykowski, S. C. (1991). Biochemistry of genetic recombination: Energetics and mechanism of DNA strand exchange. *Ann. Rev. Biophys. Biophys. Chem.* **20**, 539-575.
- Kowalczykowski, S. C. (2000). Some assembly required. *Nat. Struc. Biol.* **7**, 1087-1089.
- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
- Kruijjer, W., Van Schaik, F. M., and Sussenbach, J. S. (1982). Nucleotide sequence of the gene encoding adenovirus type 2 DNA binding protein. *Nucl. Ac. Res.* **10**, 4493-5000.
- Kyte, J. and Doolite, R. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
- Kwong, P. D., and Liu, Y. (1999). Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. *J. Appl. Cryst.* **32**, 102-105.
- Lämmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Lahue, R. S., Au K. G., and Modrich, P. (1989). DNA mismatch correction in a defined system. *Science.* **245**, 160-164.
- Lamzin, V. S., and Wilson, K. S. (1997). Automated refinement for protein crystallography. *Meth. Enzymol.* **277**, 269-305.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures *J. Appl. Cryst.* **26**, 283-291.
- Lavery, P. E., and Kowalczykowski, S. C. (1992). Enhancement of RecA protein-promoted DNA strand exchange activity by volume-occupying agents. *J. Biol. Chem.* **267**, 9307-9314.
- Lavery, P. E., and Kowalczykowski, S. C. (1992). A postsynaptic role for single-stranded DNA-binding protein in RecA protein-promoted DNA strand exchange. *J. Biol. Chem.* **267**, 9315-9320.
- Lavery, P. E., and Kowalczykowski, S. C. (1990). Properties of RecA441 protein-catalyzed DNA strand exchange can be attributed to an enhanced ability to compete with SSB protein. *J. Biol. Chem.* **265**, 4004-4010.
- LeBowitz, J. H., and McMacken, R. (1986). The *Escherichia coli* dnaB replication protein is a DNA helicase. *J. Biol. Chem.* **261**, 4738-4748.

References

- Lerš, N., Salaj-Šmic, E. and Trgovcevic Z. (1989). Overproduction of SSB protein enhances the capacity for photorepair in *Escherichia coli* recA cells. *Photochem. Photobiol.* **49**, 225-227.
- Little, J. W., and Mount, D. W. (1982). The SOS regulatory systems of *Escherichia coli* Cell **29**, 11-22.
- Lin, L. L., and Little J. W. (1989). Autodigestion and RecA-dependent cleavage of Ind⁻mutant LexA proteins. *J. Mol. Biol.* **210**, 439-452.
- Lohman, T. M., and Overman, L. B. (1985). Two binding modes in *Escherichia coli* single strand binding protein (SSB) – single-stranded DNA complexes: modulation by NaCl concentration. *J. Biol. Chem.* **260**, 3594-3603.
- Lohman, T. M., Overman, L. B., and Datta, S. (1986). Salt dependent changes in the DNA binding cooperativity of *Escherichia coli* single strand binding protein. *J. Mol. Biol.* **187**, 603-615.
- Lohman, T. M., and Bujalowski, W. (1988). Negative co-operativity within individual tetramers of *Escherichia coli* SSB protein is responsible for the transition between the (SSB)₃₅ and (SSB)₅₆ DNA binding modes. *Biochemistry* **27**, 2260-2265.
- Lohman, T. M., Bujalowski, W., and Overman, L. B. (1988). *Escherichia coli* single strand binding protein: A new look at helix destabilizing proteins. *Trends Biochem. Sci.* **13**, 250-255.
- Lohman, T. M., and Bujalowski, W. (1994). Effects of base composition on the negative cooperativity and binding mode transitions of *Escherichia coli* SSB-ssDNA Complexes. *Biochemistry* **33**, 6167-6176.
- Lohman, T. M., Overman, L. B., Ferrari M. E. and Kozlov, A. G. (1996). A highly salt-dependent enthalpy change for *Escherichia coli* SSB protein-nucleic acid binding due to ion-protein interactions. *Biochemistry* **35**, 5272-5279.
- Low, R. L., Shlomai, J., and Kornberg, A. (1982). Protein n, a primosomal DNA replication protein of *Escherichia coli*. Purification and characterization. *J. Biol. Chem.* **257**, 6242-6250.
- Lu, C., Scheuermann, R. H., and Echols, H. (1986). Capacity of RecA protein to bind preferentially to UV lesions and inhibit the editing subunit (ψ) of DNA polymerase III: a possible mechanism for SOS-induced targeted mutagenesis. *Proc. Natl. Acad. Sci. USA* **83**, 619-623.
- Luzzati, V. (1952). Traitement statistique des errors dans la determination des structures cristallines. *Acta Cryst.* **5**, 802-810.
- Mackay, V., and Linn, S. (1976). Selective inhibition of the dnase activity of the recBC enzyme by the DNA binding protein from *Escherichia coli*. *J. Biol. Chem.* **251**, 3716-3719.
- Marians, K. J. (1992). Prokaryotic DNA replication. *Ann. Rev. Biochem.* **61**, 673-719.
- Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T. Shimamoto, N., Tsukihara, T., and Yasuoka, N. (2000). Roles of functional loops and the C-terminal segment of a single-stranded DNA binding protein elucidated by X-ray structure analysis. *J. Biochem.* **127**, 329-335.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- Mazin, A.V., and Kowalczykowski, S. C. (1996). The specificity of the secondary DNA binding site of RecA protein defines its role in DNA strand exchange. *Proc. Natl. Acad. Sci. USA* **93**, 10673-10678.
- Mazin, A.V., and Kowalczykowski, S. C. (1997). The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *EMBO J.* **17**, 1161-1168.
- McHenry, C. S. (1988). DNA polymerase III holoenzyme of *Escherichia coli*. *Ann. Rev. Biochem.* **57**, 519-550.
- McPherson, A. (1982). *The preparation and analysis of protein crystals*. Wiley New York.

References

- Merrill, B. M., Williams, K. R., Chase, J. W., and Konigsberg, W. H. (1984). Photochemical cross-linking of the *Escherichia coli* single-stranded DNA-binding protein to oligodeoxynucleotides. Identification of phenylalanine 60 as the site of cross-linking. *J. Biol. Chem.* **259**, 10850-10856.
- Merritt, E. A., and Murphy, M. E. P. (1994). Raster3D Version 2.0 - A Program for Photorealistic Molecular Graphics. *Acta Cryst.* **D50**, 869-873.
- Merritt, E. A., and Bacon, D. J. (1997). Raster3D Photorealistic Molecular Graphics. *Meth. Enzymol.* **277**, 505-524.
- Meyer, R. R., Glassberg, J., and Kornberg, A. (1979). An *Escherichia coli* mutant defective in single-strand binding protein is defective in DNA replication. *Proc. Natl. Acad. Sci. USA* **76**, 1702-1705.
- Meyer, R. R., Glassberg, J., Scott J. V., and Kornberg, A. (1980). A temperature-sensitive single-stranded DNA-binding protein from *Escherichia coli*. *J. Biol. Chem.* **255**, 2897-2901.
- Meyer, R. R., and Laine, P. S. (1990). The single-stranded DNA-binding protein of *E. coli*. *Microbiol. Rev.* **54**, 342-380.
- Model, P., and Russel, M. (1988). *The Bacteriophages*, Calender, R. ed., Plenum, New York, 2099-2107.
- Molineux, I. J., Friedman, S., and Gefter, M. L. (1974). Purification and properties of the *Escherichia coli* deoxyribonucleic acid-unwinding protein. Effects on deoxyribonucleic acid synthesis in vitro. *J. Biol. Chem.* **249**, 6090-6098.
- Molineux, I. J., and Gefter, M. L. (1975). Properties of the *Escherichia coli* DNA-binding (unwinding) protein interaction with nucleolytic enzymes and DNA. *J. Mol. Biol.* **98**, 811-825.
- Monzinger, A. F., and Christiansen, C. (1983). Crystallization of single-strand DNA-binding protein. *J. Mol. Biol.* **170**, 801- 801.
- Morin N., Delsert C., and Klessig D. F. (1989). Nuclear localization of the adenovirus DNA-binding protein: requirement for two signals and complementation during viral infection. *Mol. Cell Biol.* **9**, 4372-4380.
- Muniyappa, K., Shaner, S. L., Tsang, S. S., and Radding, C. M. (1984). Mechanism of the concerted action of RecA protein and helix destabilizing proteins in homologous recombination. *Proc. Natl. Acad. Sci. USA* **81**, 2757-2761.
- Murshudov, G.N., Vagin A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.
- Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson K. S., and Dodson E. J. (1999). Efficient anisotropic refinement of Macromolecular structures using FFT. *Acta Cryst.* **D55**, 247-255.
- Murthy, H. M., Hendrickson, W. A., Orme-Johnson, W. H., Merrit, E. A., and Phizackerley, R. P. (1988). Crystal structure of clostridium acidu-urici ferredoxin at 5 Å resolution based on measurement of anomalous x-ray scattering at multiple wavelengths. *J. Biol. Chem.* **263**, 18430-18436.
- Murzin, G. A. (1993). OB (oligonucleotide/oligosaccharide binding)- fold: common structural and functional solution for non-homologous sequence. *EMBO J.* **12**, 861-867.
- Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.
- Ng, J. D., and McPherson A. (1989). Preliminary crystallographic analysis of a proteolytically modified form of *Escherichia coli* single stranded DNA binding protein. *J. Biomol. Struct. Dyn.* **6**, 1071-1076.
- Nicholls, A., Sharo, K., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties. *Proteins* **11**, 281-296.

References

- Ollis, D., Brick, P., Abdel-Meguid, S. S., Murthy, K., Chase, J. W., Steitz, T. A. (1983). Crystals of *Escherichia coli* single-strand DNA-binding protein show that the tetramer has D₂ symmetry. *J. Mol. Biol.* **170**, 797- 801.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307-326.
- Overman, L. B., Bujalowski, W., and Lohman, T. M. (1988). Equilibrium binding of *Escherichia coli* single strand binding protein to single stranded nucleic acids in the (SSB)₆₅ binding mode. cation and anion effects and polynucleotide specificity. *Biochemistry* **27**, 456-471.
- Overman, L. B., and Lohman, T. M. (1994). Linkage of pH, anion and cation effects in protein-nucleic acid equilibria. *Escherichia coli* SSB protein-single stranded nucleic acid interactions. *J. Mol. Biol.* **236**, 165-178.
- Perrino, F. W., Meyer, R. R., Bobst, A. M., and Rein, D. C. (1988). Interaction of a folded chromosome-associated protein with single-stranded DNA-binding protein of *Escherichia coli*, identified by affinity chromatography. *J. Biol. Chem.* **263**, 11833-11839.
- Philipova, D., Mullen, J. R., Maniar, H. S., Lu J., Gu C., and Brill S. J. (1996). A hierarchy of SSB protomers in replication protein A. *Genes Dev.* **10**, 2222-33.
- Porter, R. D., Black, S., Pannuri, S., and Carison, A., (1990). Use of the *E. coli* SSB gene to prevent bioreactor takeover by plasmidless cells. *Biotechnology (NY)* **8**, 47-51.
- Prasad, B. V., and Chiu, W. (1987). Sequence comparison of single-stranded DNA binding proteins and its structural implications. *J. Mol. Biol.* **193**, 579-584.
- Prigodich, R. V., Shamoo Y., Williams, K. R., Chase, J. W., Konigsberg, W. H., and Coleman, J. E. (1986). 1H NMR (500 MHz) identification of aromatic residues of gene 32 protein involved in DNA binding by use of protein containing perdeuterated aromatic residues and by site-directed mutagenesis. *Biochemistry.* **25**, 3666-72.
- Raghunathan, S., Ricard, C. S., Lohman, T. M., and Waksman G. (1997). Crystal Structure of the homo-tetrameric single-stranded DNA binding domain of *Escherichia coli* single-stranded DNA binding protein determined by multiwavelength X-ray diffraction on the seleno-methionyl protein at 2.9 Å resolution. *Proc. Natl. Acad. Sci. USA* **94**, 6652-6657.
- Raghunathan, S., Kozlov A.G., Lohman, T. M., and Waksman, G. (2000). Structure of the DNA binding domain of *Escherichia coli* SSB bound to ssDNA. *Nat. Struc. Biol.* **7**, 648-652.
- Ramachandran, G. N., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-437.
- Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M., and Kroon, J. (1997). STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. *J. Appl. Cryst.* **30**, 551-554.
- Read, R. J. (1986). Improved fourier coefficient for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.
- Resnick, J., and Sussman, R. (1982). *Escherichia coli* single-strand DNA binding protein from wild type and lexC113 mutant affects in vitro proteolytic cleavage of phage lambda repressor. *Proc. Natl. Acad. Sci. USA* **79**, 2832-2835.
- Riboldi-Tunnicliffe, A., and Hilgenfeld, R. (1999). Cryocrystallography with oil - an old idea revived. *J. Appl. Cryst.* **32**, 1003-1005.
- Riddles, P. W., and Lehman, I. R. (1985). The formation of plectonemic joints by the RecA protein of *Escherichia coli*. *J. Biol. Chem.* **260**, 170-173 .
- Roberts, J. W., Phizicky, E. M., Burbee, D. G., Roberts, C. W., and Moreau, P. L. (1982). A brief consideration of the SOS inducing signal. *Biochimie* **64**, 805-7.
- Rodgers, D. W. (1994). Cryocrystallography. *Structure* **2**, 1135-1140.

References

- Rodgers, D. W. (1997). Practical cryocrystallography. *Meth. Enzymol.* **276**, 183-203.
- Rossmann, M. G., and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.
- Russell, R. B., Saqi, M. A., Bates, P. A., Sayle, R. A., and Sternberg, M. J. (1998). Recognition of analogous and homologous fold protein folds- assessment of prediction success and associated alignment accuracy using empirical substitutions matrices. *Protein Eng.* **11**, 1-9.
- Ruyechan, W. T., and Wetmur, J. G. (1975). Studies on the cooperative binding of the *Escherichia coli* DNA unwinding protein to single-stranded DNA. *Biochemistry.* **14**, 5529-5534.
- Ruyechan, W. T., and Wetmur, J. G. (1976). Studies on the noncooperative binding of the *Escherichia coli* DNA unwinding protein to single-stranded nucleic acids. *Biochemistry* **15**, 5057-5064.
- Sancar, A. Kacinski, B. M., Mott, D. L., and Rupp, W. D. (1981). Identification of the uvrC gene product. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 5450-5454.
- Samygina, V. R., Antonyuk, S. V., Lamzin, V. S., and Popov, A. N. (2000). Improving the X-ray resolution by reversible flash-cooling combined with concentration screening, as exemplified with PPase. *Acta Cryst.* **D56**, 595-603.
- Sauer, U. H., and Ceska, T. A. (1997). A simple method for making reproducible fibre loops for protein cryocrystallography. *J. Appl. Cryst.* **30**, 71-72.
- Schekman, R., Weiner, J. H., Weiner, A., and Kornberg, A. (1975). Ten proteins required for conversion of phiX174 single-stranded DNA to duplex form in vitro. Resolution and reconstitution. *J. Biol. Chem.* **250**, 5859-5865.
- Schmellik-Sandage, C. S., and Tessman, E. S. (1990). Signal strains that can detect certain DNA replication and membrane mutants of *Escherichia coli*: isolation of a new ssb allele, ssb-3. *J Bacteriol.* **172**, 4378-4385.
- Shamoo, Y., Williams, K. R., and Konigsberg W. H. (1988). Photochemical crosslinking of bacteriophage T4 single-stranded DNA-binding protein (gp32) to oligo-p(dt)8: identification of phenylalanine-183 as the site of crosslinking. *Proteins* **4**, 1-6.
- Shamoo, Y., Ghosaini, L. R., Keating, K. M., Williams, K. R., Sturtevant, J. M., and Konigsberg, W. H. (1989). Site-specific mutagenesis of T4 gene 32: the role of tyrosine residues in protein-nucleic acid interactions. *Biochemistry* **28**, 7409-7417.
- Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg W. H., and Steiz T. A. (1995). Crystal structure of replication fork single-stranded DNA binding (T4gp32) complexed to DNA. *Nature* **376**, 362-366.
- Shwartz, H., Shavitt, O., and Livneh, Z. (1988). The role of exonucleolytic processing and polymerase-DNA association in bypass of lesions during replication in vitro. Significance for SOS-targeted mutagenesis. *J. Biol. Chem.* **263**, 18277-18285.
- Sigal, N., Delius, H., Kornberg, T., Gefter, M. L., and Alberts, B. M. (1972). A DNA-unwinding protein isolated from *Escherichia coli*: its interaction with DNA and with DNA polymerases. *Proc. Natl. Acad. Sci. USA* **69**, 3537-3541.
- Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet R. M., Gray, C. W., Konings, R. N., Wang, A. H., and Terwilliger, T. C. (1994). Structure of the gene V protein of bacteriophage phi determined by multiwavelength x-ray diffraction on the selenomethionyl protein. *Proc. Natl. Acad. Sci. USA* **91**, 2071-2075.
- Suck, D. (1997). Common fold, common function, common origin? *Nature Struct. Biol.* **4**, 161-165.
- Teng, T. Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Cryst.* **23**, 387-391.
- The CCP4 Suite. Programs for Protein crystallography (1994). *Acta Cryst.* **D50**, 760-763.

References

- Thorn, J., M., Carr, P. D., Chase, J. W., Dixon, N. E. and Ollis, D. L. (1994). Crystallization and low temperature diffraction studies of the DNA binding domain of the single-stranded DNA J. Mol. Biol. **240**, 396- 399.
- Tong, L., and Rossmann, M. G. (1990). The locked rotation function. Acta Cryst. **A46**, 783-792.
- Tsernoglou, D., Tsugita, A., Tucker, A. D., and van der Vliet P. C. (1985). Characterization of the chymotryptic core of the adenovirus DNA-binding protein. FEBS Lett. **188**, 248-252.
- Tucker, P. A., Tsernoglou, D., Tucker, A. D., Coenjaerts, F. E. J., Leenders, H., and van de Vliet P. C. (1994). Crystal structure of the adenovirus DNA binding protein reveals a hook-on model for cooperative DNA binding. EMBO J. **13**, 2994-3002.
- Urbanke, C., and Schaper A. (1990) Kinetics of binding of single-stranded DNA binding protein from *Escherichia coli* to single-stranded nuclei acids. Biochemistry. **29**, 1744-1749.
- Vales, L. D., Chase J. W., and Murphy J. B. (1980). Effect of ssb-1 and lexC113 mutations on lambda prophage induction, bacteriophage growth, and cell survival. J. Bacteriol. **143**, 887-896.
- van der Ende, A., Baker, T. A., Ogawa, T., and Kornberg, A. (1985). Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: primase as the sole priming enzyme. Proc. Natl. Acad. Sci. USA **82**, 3954-3958.
- van Duynhoven, J. (1992). Ph.D. thesis (University of Nijmegen) Nijmegen, The Netherlands.
- Wahle, E., Lasken, R. S., and Kornberg, A. (1989). The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. J. Biol. Chem. **264**, 2463-2468.
- Wahle, E., Lasken, R. S., and Kornberg, A. (1989). The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. J. Biol. Chem. **264**, 2469-2475.
- Walker, G. C. (1985). Inducible DNA repair systems. Ann. Rev. Biochem. **54**, 425-457.
- Wang, T. C., and Smith, K. C. (1982). Effects of the ssb-1 and ssb-113 mutations on survival and DNA repair in UV-irradiated delta uvrB strains of *Escherichia coli* K-12. J. Bacteriol. **151**, 186-192.
- Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C., and Hilgenfeld, R. (1997). A common core for binding single-stranded DNA: Structural comparison of the single-stranded DNA-binding proteins from *Escherichia coli* and human mitochondria. FEBS Lett. **411**, 313-316.
- Weiner, J. H., Bertsch, L. L., and Kornberg, A. (1975). The deoxyribonucleic acid unwinding protein of *Escherichia coli*. Properties and functions in replication. J. Biol. Chem. **250**, 1972-1980.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. Jr., and Weiner, P. (1984). A new forcefield for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. **106**, 765-784.
- Weinstock, G. M., and McEntee, K. (1981). RecA protein-dependent proteolysis of bacteriophage lambda repressor, characterization of the reaction and stimulation by DNA-binding proteins. J. Biol. Chem. **256**, 10883-10888.
- Weiss, M. S., and Hilgenfeld, R. (1997). On the use of the merging R factor as a quality indicator for X-ray data. J. Appl. Cryst. **30**, 203-205.
- Weiss, M. S. (2001). Global indicators of X-ray data quality. J. Appl. Cryst. **34**, 130-135.
- West, S. C., Cassuto, E., and Howard-Flanders, P. (1982). Role of SSB protein in RecA promoted branch migration reactions. Mol. Gen. Genet. **186**, 333-338.
- West, S. C. (1988). Protein-DNA interactions in genetic recombination. Trends Genet. **4**, 8-13.

References

- Whittier, R. F., and Chase, J. W. (1981). DNA repair in *E. coli* strains deficient in single-strand DNA binding protein. *Mol. Gen. Genet.* **183**, 341-347.
- Whittier, R. F., Chase, J. W., and Masker, W. E. (1983). Repair resynthesis in *Escherichia coli* mutants deficient in single-stranded DNA-binding protein. *Mutat. Res.* **112**, 275-286.
- Williams, K. R., Spicer, E. K., Lopresti, M. B., Guggenheimer, R. A., and Chase, J. W. (1983). Limited proteolysis studies on the *Escherichia coli* single-stranded DNA binding protein. Evidence for a functionally homologous domain in both the *Escherichia coli* and T4 DNA binding proteins. *J. Biol. Chem.* **258**, 3346-3355.
- Williams, K. R., Murphy, J. B., and Chase, J. W. (1984). Characterization of the structural and functional defect in the *Escherichia coli* single-stranded DNA binding protein encoded by the *ssb-1* mutant gene expression of the *ssb-1* gene under lambda pL regulation. *J. Biol. Chem.* **259**, 11804-11811.
- Wickner, S., Wickne, R. B., Wright, M., Berkower, I., and Hurwitz, J. (1974). Conversion of $\psi\chi$ 174 and fd DNA to their replicative forms by two enzyme systems in *Escherichia coli*. *Basic Life Sci.* **3**, 395-409.
- Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318-321.
- Wobbe, C. R., Weissbach, L., Borowiec, J. A., Dean, F. B., Murakami, Y., Bullock, P., and Hurwitz, J. (1987). Replication of simian virus 40 origin-containing DNA in vitro with purified proteins. *Proc. Natl. Acad. Sci. USA* **84**, 1834-1838.
- Wold, M. S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Ann. Rev. Biochem.* **66**, 61-92.
- Yang, C., Curth, U., Urbanke, C., and Kang, C. H. (1997). Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 Å resolution *Nature Struct. Biol.* **4**, 153-157.
- Yeh, J. I. and Hol, W. G. J. (1998). A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Cryst.* **D54**, 479-480.
- Zhang, S., and Grosse, F. (1992). A complex between replication factor A (SSB) and DNA helicase stimulates DNA synthesis of DNA polymerase α on double-stranded DNA. *FEBS Lett.* **312**, 143-146.
- Zhu, Y., Oliveira, S. C., Splitter, G. A. (1993). Isolation of *Brucella abortus* *ssb* and *uvrA* genes from a genomic library by use of lymphocytes as probes. *Infect. Immun.* **61**, 5339-5344.

6 Appendix

6.1 Cryo-cooling of protein crystals

Table A1 Protein crystals used to study the effects of repeated-annealing in dried paraffin oil

Protein	Crystallisation condition	Solvent content (%)	Space group	Unit cell	MW/ AU	V_m (Å ³ /Da)
<i>EcoSSB</i>	4% (v/v) PEG 400, 10 mM β -mercapto ethenol, 40 mM Na cacodylate, pH 6.5	46	C2	a=104.40 Å b=62.46 Å c=96.70 Å β =112.67°	64	2.27
<i>BabSSB</i>	12.5 mM Na acetate, 0.1 M Na cacodylate, pH 6.0	53	P4 ₃ 2 ₁ 2	a=113.33 Å c=52.26 Å	32	2.62
CPS-1	6% (v/v) PEG 15 K, 0.1 M Na cacodylate, 0.1 M KCl, 0.1 M NH ₄ Cl 0.01 M MnCl ₂ , pH 6.1	47	P3 ₂	a=176.90 Å c=331.60 Å γ =120°	1280	2.34
Alliinase	Data not provided	35	C222	a=55.7 Å b=97.7 Å c=153.3 Å	55	1.90
IF2-tRNA	2M (NH ₄) ₂ SO ₄ , 100mM HEPES/KOH 3 mM MgCl ₂	50	I222/I2 ₁ 2 ₁ 2 ₁	a=74.7 Å b=116.2 Å c=227.9 Å	100	2.50
Des-PheB1-Insulin (phenol complex)	0.75 M (NH ₄) ₂ SO ₄ , 20 mM Zn acetate, 120 mM NaCl, 3 mM NaN ₃ , 0.1% (v/v) Phenol, pH 8.0	37	R3	a=79.0 Å c=72.5 Å γ =120°	23	1.96
Green fluorescent protein (GFP)	2.1 M (NH ₄) ₂ SO ₄ , 100 mM Tris, pH 8.5	53	P6 ₁ 22	a=87.30 Å c=129.54 Å γ =120°	27	2.64

MW/AU is the total molecular weight of protein per asymmetric unit. V_m is the crystal packing density (Matthews, 1968).

CSP-1, Alliinase, Insulin, GFP and IF2-tRNA crystals were kindly provided by Astrid Rau, Bartholomeus Küttner, Dr. Ursula Schell, Dr. Gottfried Palm and Prof. Udo Heinemann, respectively.

Appendix

Table A2 Comparison of diffraction properties of crystals of seven different proteins at 293 K, after flash-cooling and after repeated-annealing in dried paraffin oil

Protein name	Data at 293K			Data after flash-cooling using oil			Data after repeated-annealing		
	Number of crystals tested	Resolution (Å)	Mosaicity (°)	Number of crystals tested	Resolution (Å)	Mosaicity (°)	Number of crystals tested	Resolution (Å)	Mosaicity (°)
<i>EcoSSB</i>	10	2.8	0.4 – 0.5	4	2.5	0.6-0.8	3	2.4	0.5-0.55
<i>BabSSB</i>	5	>8.0	n.d.	5	2.5	0.3-0.4	6	2.2	0.3-0.4
CPS-1	8	8.0	n.d.	5	>8.0	n.d.	2	4.5	0.7
Alliinase	5	>8.0	n.d.	5	>8.0	n.d.	2	3.2	1.2
IF2-tRNA	4	4.5	n.d	2	>4.5	n.d	2	2.5	-
DesPheB1 insulin	8	2.7	n.d.*)	2	3.0	1.4-1.7	3	2.5	1.0-0.9
GFP	4	2.1	0.2 –0.3	3	2.4	0.6-0.8	2	2.2	0.4-0.5

*) Collection of more than a small fraction of the data was impossible owing to severe radiation damage. n.d.: not determined

Table A3 Comparison of diffraction properties of crystals of nine different proteins after flash-cooling using dried paraffin oil and in Panjelly™

Protein	Dried paraffin oil		Panjelly™	
	Resolution at 100K (Å)	Mosaicity (°)	Resolution at 100K(Å)	Mosaicity (°)
GFP	2.40	0.40	1.60	0.60
Alliinase	3.20	1.20	2.70	0.90
DesPheB1 insulin	2.50	1.00	2.50	0.80
Mip-mutant	>8.00	n.d.	2.50	0.70
<i>EcoSSB</i>	2.20	0.60	1.80	0.60
<i>BabSSB</i>	2.20	0.30	1.78	0.30
<i>PmiSSB</i>	3.50	1.30	2.50	0.40
<i>SmaSSB</i>	>8.00	n.d.	2.70	0.70
HcV-protease	>8.00	n.d.	3.00	0.40

n.d.: not determine

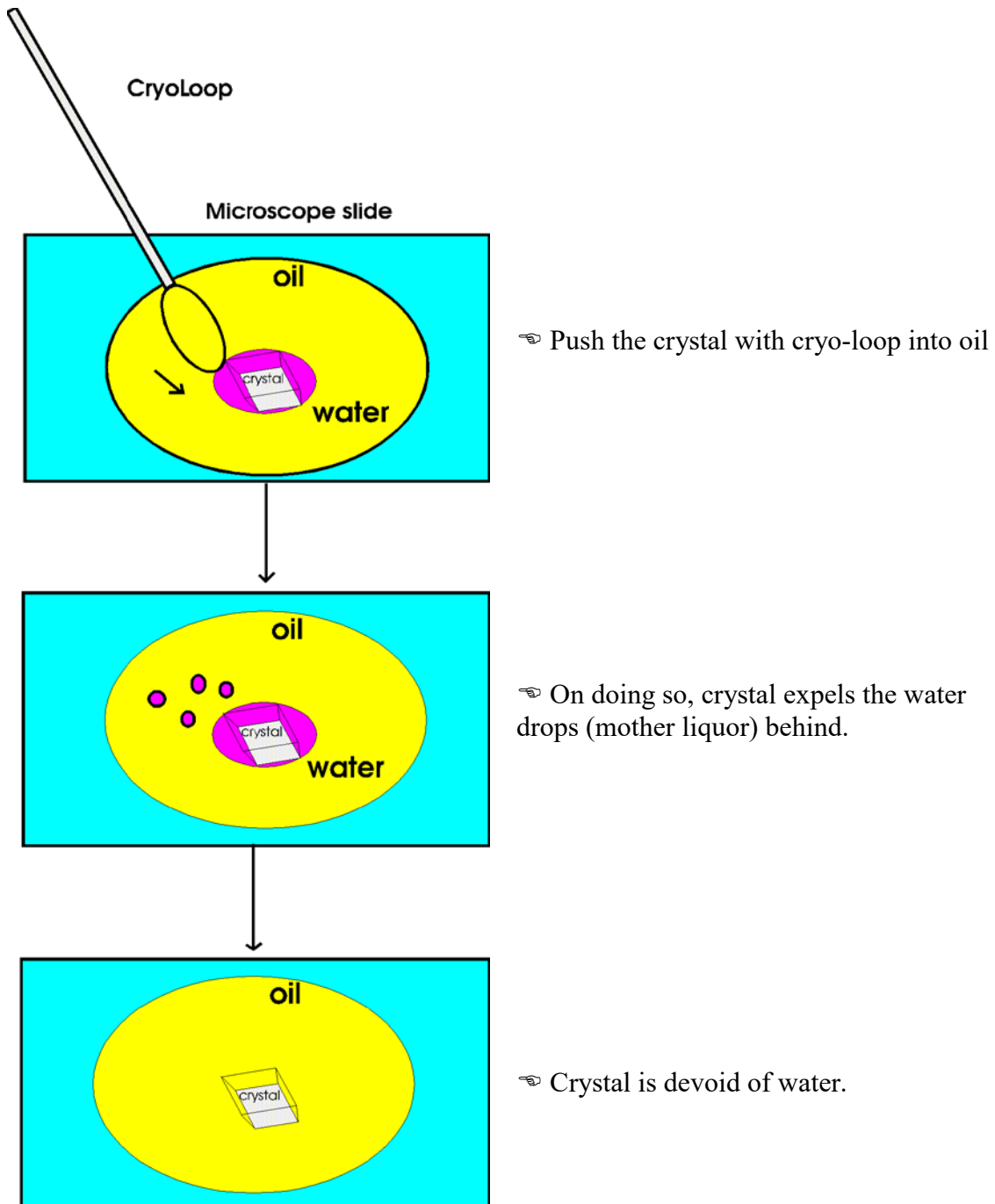


Figure A1 Washing procedure of protein crystal in dried paraffin oil.

6.2 Comparison of SSB's monomers/tetramer

Table A4 R.m.s. deviations between monomers of SSBs. (see footnotes)

	A_eco	B_eco	C_eco	D_eco	A_bab	B_bab	A_pmi	B_pmi	C_pmi	D_pmi	E_pmi	F_pmi	G_pmi	H_pmi	A_sma	B_sma	C_sma	D_sma	E_sma	A_hsm	B_hsm
A_eco		0.70	0.64	0.40	0.76	0.79	0.71	0.72	0.62	0.70	0.64	0.68	0.68	0.68	0.58	0.64	0.64	0.70	0.69	1.25	1.46
B_eco	87		0.29	0.39	0.98	0.64	0.61	0.38	0.56	0.64	0.58	0.40	0.52	0.84	0.45	0.43	0.39	0.28	1.25	1.38	1.25
C_eco	88	97		0.30	0.89	0.63	0.52	0.43	0.52	0.47	0.42	0.42	0.45	1.33	0.42	0.42	0.42	0.37	0.40	1.31	1.21
D_eco	85	94	91		0.70	0.78	0.65	0.58	0.52	0.48	0.48	0.59	0.60	0.95	0.47	0.37	0.42	0.34	0.50	1.32	1.48
A_bab	79	92	86	83		0.26	0.79	0.54	0.76	0.74	0.66	0.77	0.67	0.47	0.59	0.56	0.54	0.65	0.58	1.40	1.34
B_bab	79	77	76	83	88		0.83	0.59	0.80	0.83	0.69	0.81	0.75	0.64	0.61	0.60	0.53	0.53	0.54	1.30	1.17
A_pmi	87	86	84	94	86	84		0.28	0.52	0.29	0.32	0.28	0.43	0.40	0.48	0.60	0.46	0.42	0.43	1.49	1.57
B_pmi	87	89	94	97	78	77	82		0.43	0.31	0.30	0.28	0.35	0.28	0.64	0.55	0.46	0.98	1.45	1.14	1.24
C_pmi	88	92	93	94	81	81	85	93		0.53	0.46	0.35	0.23	0.47	0.57	0.49	0.41	0.45	1.53	0.99	1.43
D_pmi	87	92	86	91	86	85	89	83	88		0.21	0.33	0.48	0.34	0.39	0.44	0.47	0.42	0.41	1.33	1.50
E_pmi	87	89	85	92	84	82	92	84	85	86		0.30	0.47	0.29	0.42	0.39	0.41	0.46	0.44	1.44	1.48
F_pmi	87	92	94	99	85	84	83	94	91	87	83		0.37	0.27	0.45	0.38	0.41	0.36	1.43	1.19	1.19
G_pmi	89	94	93	97	80	81	83	94	91	89	87	98		0.41	0.52	0.49	0.52	1.02	1.51	1.18	1.27
H_pmi	86	95	102	104	73	77	93	79	83	86	83	81	83		0.72	0.72	0.50	0.54	0.52	1.09	1.16
A_sma	84	95	91	98	76	75	85	94	91	85	85	90	92	95		0.30	0.32	0.24	0.31	1.47	1.32
B_sma	89	95	95	96	77	76	88	94	90	86	84	90	92	95	94		0.22	0.25	0.35	1.18	1.26
C_sma	81	81	85	86	77	75	81	84	75	83	83	82	81	81	78	77		0.22	0.24	1.07	1.09
D_sma	79	76	78	77	77	73	78	91	74	80	81	77	91	79	82	77	75		0.40	1.45	1.51
E_sma	78	97	77	83	75	74	78	97	94	79	79	97	97	78	85	80	75	100		1.53	1.98
A_hsm	84	87	87	88	86	83	93	85	81	86	89	85	85	86	90	85	79	88	91		0.60
B_hsm	87	87	87	95	85	80	92	89	90	88	88	87	88	85	89	89	78	88	93	89	

The first capital letter denotes the chain id. of the monomer and last three letters denote the name of the SSB, respectively (eco: *EcoSSB*, bab: *BabSSB*, pmi: *PmiSSB*, sma: *SmaSSB*, hsm: *HsmtSSB*). The upper triangle and the lower triangle in the above matrix show the r.m.s. deviation and the number of C_α pair used for alignment respectively. The lowest and highest r.m.s. difference are shown in the bold letters.

Appendix

Table A5 R.m.s. deviations between SSB-tetramers (see footnotes)

	<i>Eco</i> SSB-ABCD	<i>Bab</i> SSB-AABB	<i>Pmi</i> SSB-ABCD	<i>Pmi</i> SSB-EFGH	<i>Sma</i> SSB-ABCD	<i>Sma</i> SSB-EEEE	HsmtSSB-ABAB
<i>Eco</i> SSB-ABCD		1.63 0.99 -	1.42 1.35 1.30	1.36 1.17 1.15	1.45 1.06 1.27	2.27 1.46 2.17	2.52 1.63 1.91
<i>Bab</i> SSB-AABB	355 256 -		2.00 1.94 -	1.87 1.76 -	2.00 1.66 -	2.48 2.07 -	2.43 2.19 -
<i>Pmi</i> SSB-ABCD	369 256 306	339 256 -		0.54 0.33 0.36	1.11 0.48 0.68	1.62 0.46 1.57	2.28 1.37 1.74
<i>Pmi</i> SSB-EFGH	379 256 311	340 256 -	383 256 306		1.24 0.42 0.96	1.85 0.57 1.79	2.57 1.29 1.67
<i>Sma</i> SSB-ABCD	370 256 303	336 256 -	363 256 294	372 256 310		1.49 0.54 1.45	2.56 1.36 1.78
<i>Sma</i> SSB-EEEE	392 256 315	340 256 -	361 256 306	386 256 298	384 256 308		3.34 1.42 1.93
HsmtSSB-ABAB	369 256 308	298 256 -	365 256 304	375 256 304	358 256 305	384 256 312	

The tetrameric SSB is indicated along with the chain ids (separated by a hyphen). Repetition of the same chain id indicates that the tetramer was formed using symmetry related molecules. In the lower triangular matrix, each cell contains three values. The first (top) value indicates the maximum number of C α pairs used that could be built in the electron density maps for the polypeptide chains. The second (middle value) indicates the same number when the co-ordinates from all loop regions are excluded. The third (bottom value) indicates also the same number but when the co-ordinates from all loop regions except Loop III are excluded. The bottom value is not calculated for the *Bab*SSB tetramer because Loop III is disorder in this structure hence indicated by hyphen. The corresponding r.m.s. difference values for the two different cases are indicated in the upper triangular matrix at equivalent positions.

6.3 Analysis of water molecules of SSBs

Table A6 Conserved water molecules, their b-factors and their r.m.s. deviations

<i>Eco</i> SSB-ABCD				<i>Bab</i> SSB-AABB			<i>Pmi</i> SSB-ABCD			<i>Pmi</i> SSB-EFGH		
Water molecules	B-factor (Å ²)	H-bonding		Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)
		Atom involved	Distance(Å)									
W11	26.6	Thr-C36-OG1 Asn-D6-OD1 Wat-W81-O Wat-W86-O	2.56 3.17 2.74 3.01	W4	26.0	0.27	-	-	-	W58	35.7	0.54
W16	28.7	His-C55-ND1 Wat-W2-O Wat-W10-O	2.88 2.75 2.87	W11	40.3	0.62	W23	32.8	0.48	W40	41.7	0.66
W5	28.8	Thr-A36-OG1 Asn-B6-OD Wat-W57-O	2.75 3.11 2.80	#W39	34.7	0.62	W86	52.6	0.71	W74	48.5	0.91
W3	30.6	Thr-D36-OG1 Asn-C6-OD Wat-W17-O Wat-W61-O	2.68 3.14 2.82 2.71	#W4	26.0	0.36	W9	35.8	0.94	W4	37.5	0.66
W6	30.8	His-B55-ND1 Wat-W20-O Wat-W74-O	3.02 2.80 2.85	W15	39.1	0.74	W70	54.5	0.76	W73	49.3	0.33
W31	30.8	His-A55-ND1 Wat-W21-O Wat-W45-O	2.84 2.81 2.58	#W15	39.1	0.40	W67	56.5	0.89	W2	29.7	0.32
W42	31.1	His-D55-ND1 Wat-W52-O Wat-W63-O	2.83 2.75 2.69	#W11	40.3	0.09	W7	32.6	0.30	W1	28.7	0.31
W36	33.2	Thr-B36-OG1 Asn-A6-ND2	2.61 3.25	-	-	-	W24	33.8	0.48	W55	31.6	0.83
W54	33.7	Arg-C56-NH2 Thr-C99-N Thr-D99-OG1 Wat-W19-O Wat-W25-O	3.02 2.88 2.67 3.38 2.70	-	-	-	W3	38.8	0.99	-	-	-
W4	34.8	Tyr-D97-N	3.15	-	-	-	W75	40.4	0.78	-	-	-

Appendix

<i>EcoSSB-ABCD</i>				<i>BabSSB-AABB</i>			<i>PmiSSB-ABCD</i>			<i>PmiSSB-EFGH</i>		
Water molecules	B-factor (Å ²)	H-bonding		Equivalent water molecules	B-factor (Å ²)	d [†] (Å)	Equivalent water molecules	B-factor (Å ²)	d [†] (Å)	Equivalent water molecules	B-factor (Å ²)	d [†] (Å)
		Atom involved	Distance(Å)									
		Wat-W26-O	2.95									
W29	35.2	Glu-D53-OE1	2.59	#W40	38.4	0.56	-	-	-	-	-	-
		Trp-D54-N	3.03									
W10	35.3	Arg-C56-N	2.82	W33	46.0	0.49	W31	40.4	0.78	W13	32.0	0.87
		Thr-C99-O	2.96									
		Wat-W16-O	2.87									
W37	35.8	Asn-B104-N	2.99	W48	41.7	1.01	-	-	-	-	-	-
		Gln-B82-NE2	3.15									
		Wat-W43-O	3.05									
W25	36.3	Tyr-C97-O	2.54	-	-	-	-	-	-	W76	46.3	0.57
		Tyr-D97-O	2.76									
		Wat-W19-O	2.67									
		Wat-W54-O	2.70									
W18	36.3	Thr-A99-OG1	2.80	-	-	-	-	-	-	W88	48.2	0.31
		Thr-B99-N	2.91									
		Wat-W30-O	2.82									
W14	36.5	Glu-C53-OE1	2.93	W40	38.4	0.80	-	-	-	-	-	-
		Trp-C54-N	3.02									
		Wat-W101-O	2.70									
W75	36.8	Arg-B56-N	2.80	-	-	-	W29	51.2	0.60	-	-	-
		Lys-C87-O	2.94									
W17	37.0	Gln-D51-O	2.77	#W74	49.5	0.33	-	-	-	-	-	-
		Wat-W3-O	2.82									
		Wat-W88-O	2.76									
W45	37.2	Thr-A99-O	2.97	#W52	50.3	0.81	-	-	-	W27	41.1	0.62
		Arg-A56-N	3.01									
		Wat-W31-O	2.58									
		Wat-W77-O	2.91									
		Wat-W79-O	2.90									
W40	37.3	B6-Asn-N	2.99	W51-	45.3	0.04	-	-	-	-	-	-
W9	37.6	Arg-A96-NH2	2.65	-	-	-	-	-	-	W8	42.0	0.42
		Tyr-B97-N	3.14									
		Wat-W38-O	2.91									
W32	37.7	Asn-C6-N	2.97	W16	35.4	0.85	-	-	-	W32	38.8	0.83

Appendix

<i>EcoSSB-ABCD</i>				<i>BabSSB-AABB</i>			<i>PmiSSB-ABCD</i>			<i>PmiSSB-EFGH</i>		
Water molecules	B-factor (Å ²)	H-bonding		Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)
		Atom involved	Distance(Å)									
		Wat-W103-O	2.33									
		Wat-W61-O	3.09									
W2	38.9	Glu-C53-OE1	2.56	W72	49.9	0.75	W30	45.2	0.13	W48	40.3	0.24
		Trp-C54-O	3.00									
		Thr-D85-OG1	2.63									
		Wat-W16-O	2.75									
W52	38.9	Arg-D56-N	2.81	#W33	46.0	0.35	W33	57.8	0.78	W25	44.7	0.37
		Thr-D91-O	3.04									
		Wat-W42-O	2.75									
W35	39.6	Glu-C80-OE1	2.59	-	-	-	W11	37.7	0.67	W38	36.9	0.82
		Tyr-C78-OH	2.76									
W76	40.0	Lys-87-O	2.94	W55	48.8	0.98	-	-	-	-	-	-
		Gln-B82-OE1	3.07									
		Leu-B83-O	2.62									
		Wat-W62-O	2.68									
W20	41.0	Glu-B53-OE1	2.56	-	-	-	W65	44.5	0.78	-	-	-
		Thr-A85-OG1	2.79									
		Trp-B54-O	2.77									
		Wat-W6-O	2.80									
		Wat-W15-O	3.48									
W21	41.7	Glu-A53-OE1	2.47	-	-	-	W71	41.4	0.90	W18	49.1	0.75
		Trp-A54-O	3.03									
		Thr-B85-OG1	2.76									
		Wat-W31-O	2.81									
		W79-Wat-O	3.50									
W57	42.0	Gln-A51-O	2.99	#W87	51.3	0.85	-	-	-	-	-	-
		Wat-W5-O	2.80									
		Wat-W62-O	2.97									
W64	42.1	Thr-33C-N	2.82	W20	35.1	0.93	-	-	-	-	-	-
		Asn-C16-O	2.91									
W63	43.9	Glu-D53-OE1	2.54	#W72	49.9	0.20	-	-	-	W35	38.1	0.57
		Thr-C85-OG1	2.56									
		Trp-D54-O	3.07									
		Wat-W42-O	2.69									
W60	45.5	Asn-A6-N	2.90	#W51	45.3	0.68	-	-	-	-	-	-

Appendix

<i>EcoSSB-ABCD</i>				<i>BabSSB-AABB</i>			<i>PmiSSB-ABCD</i>			<i>PmiSSB-EFGH</i>		
Water molecules	B-factor (Å ²)	H-bonding		Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)	Equivalent water molecules	B-factor (Å ²)	d† (Å)
		Atom involved	Distance(Å)									
W55	45.5	Val-A57-O Asn-C31-OD1	2.77 3.15	W82	47.5	0.90	-	-	-	-	-	-
W86	46.0	Lys-D87-N Gln-C51-O Wat-W11-O Wat-W97-O	2.89 2.63 3.01 2.73	W74	49.5	0.50	-	-	-	W84	60.8	0.83
W53	46.2	Asn-D31-OD1 Ile-D32-O Val-D57-O	2.95 3.21 3.07	#W82	47.5	0.55	-	-	-	-	-	-
W74	47.5	Thr-B85-O Arg-B56-N Thr-B99-O Wat-W6-O	2.86 2.80 2.97 2.85	W52	50.3	0.93	W34	45.3	0.544	W62	45.8	0.37
W91	50.5	Thr-D52-N Asn-D13-OD1 Gln-D91-NE2	2.82 2.78 2.68	-	-	-	-	-	-	W85	47.6	0.86
W71	53.6	Gln-D76-OE1	2.56	#W23	39.6	0.80	-	-	-	-	-	-
W102	54.4	Lys-B87-O Asn-B13-OD1	2.95 2.87	-	-	-	W61	53.1	0.56	-	-	-
W38	55.4	Tyr-A97-N Wat-W9-O Wat-W59-O	3.58 2.91 2.48	-	-	-	-	-	-	W39	59.5	1.00
W104	56.5	Trp-C54-O Asn-D6-N W81-Wat-O	2.89 2.71 3.43	#W16	35.4	0.89	-	-	-	-	-	-

† Distance between two structurally equivalent waters..

symmetry mates.

The tetrameric SSB is indicated along with the chain ids (separated by a hyphen). Repetition of the same chain id indicates that the tetramer was formed using symmetry related molecules.

6.4 Interface of SSBs

Table A7 Dimer-dimer interfaces of SSB-tetramers and their extent of burial

Dimer-dimer interface	Buried surface area (Å ²)
Dimer (AB)-dimer(CD) interface of <i>Eco</i> SSB	1843
Dimer (A#A)-dimer(B#B) interface of <i>Bab</i> SSB	1972
Dimer (AB)-dimer(CD) interface of <i>Pmi</i> SSB	1885
Dimer (EF)-dimer(GH) interface of <i>Pmi</i> SSB	1936
Dimer (AB)-dimer(CD) interface of <i>Sma</i> SSB	1873
Dimer (E#E)-dimer(#E#E) interface of <i>Sma</i> SSB	1830
Dimer (AB)-dimer(#A#B) interface of <i>Hsmt</i> SSB	2299

Symmetry mate.

The capital letter in parenthesis indicates the chain id. of the monomer used to build the dimer.

Table A8 Buried surface area of individual residues in each monomer upon tetramer formation

<i>Eco</i> SSB				
Residues	Buried surface area (Å ²)			
	A-monomer	B-monomer	C-monomer	D-monomer
Ser-2	21	17		17
Arg-3			10	
Gly-4	19	21	23	19
Val-5	53	56	48	51
Lys-7	38	42	44	36
Ile-9	36	35	37	39
Val-11	20	24	25	24
Gln-76	40	35	39	36
Tyr-78	47	59	60	48
Glu-80	23	26	26	28
Gly-106	5	25	6	21
Gln-110	57	79	79	70
Leu-112	60	61	64	62
<i>Bab</i> SSB				
	A-monomer	#A-monomer	B-monomer	#B-monomer
Gly-3	45	45	51	51
Ser-4	21	21	17	17
Val-5	47	47	47	47
Lys-7	55	55	50	50
Ile-9	35	35	33	33
Val-11	24	24	24	24
Lys-77	35	35	37	37
Tyr-79	42	42	44	44
Glu-81	17	17	13	13
Arg-108			14.	14
Glu-110	28	28	16	16
Gln-112	47	47	41	41
Met-113	15	15	16	16
Leu-114	79	79	77	77
<i>Pmi</i> SSB-ABCD				
	A-monomer	B-monomer	C-monomer	D-monomer
Ala-1	30			
Arg-3	9	12	16	9.
Gly-4	10	17	17	18

Appendix

	A-monomer	B-monomer	C-monomer	D-monomer
Val-5	38	36	37	37
Lys-7	43	46	47	43
Ile-9	34	35	42	33
Ile-11	41	41	43	41
Gln-76	41	29	31	32
Tyr-78	51	53	50	52
Glu-80	21	17	19	18
Gln-110	73	63	77	66
Leu-112	76	71	44	42
<i>PmiSSB-EFGH</i>				
	E-monomer	F-monomer	G-monomer	H-monomer
Ala-1	49			
Ser-2	41			
Arg-3	11	19	11	15
Gly-4	10	20	20	18
Val-5	40	37	34	36
Lys-7	45	44	45	47
Ile-9	33	36	34	34
Ile-11	43	41	37	37
Gln-76	37	35	29	36
Tyr-78	51	49	51	51
Glu-80	20	22	19	20
Gln-110	63	64	83	68
Leu-112	64	70	42	69
<i>SmaSSB-ABCD</i>				
	A-monomer	B-monomer	C-monomerr	D-monomer
Ser-2	15	15	8	18
Arg-3	45			
Gly-4	10	19	19	20
Val-5	41	40	38	38
Lys-7	50	50	51	51
Ile-9	33	34	34	33
Val-11	25	25	26	28
Gln-76	34	22	34	36
Tyr-78	63	62	62	62
Glu-80	21	20	19	21
Gln-110	74	74	89	77
Leu-112	63	62	66	64
<i>SmaSSB-EEEE</i>				
	E-monomer	#E-monomer	#E-monomer	#E-monomer
Ser-2	16	16	16	16
Arg-3	2	2	2	2
Gly-4	16	16	16	16
Val-5	39	39	39	39
Lys-7	50	50	50	50
Ile-9	35	35	35	35
Val-11	27	27	27	27
Gln-76	34	34	34	34
Tyr-78	61	61	61	61
Glu-80	19	19	19	19
Gln-110	81	81	81	81
Leu-112	68	68	68	68
<i>HsmtSSB</i>				
	A-monomer	B-monomer	#A-monomer	#B-monomer

Appendix

	A-monomer	B-monomer	#A-monomer	#B-monomer
Arg-12	68	119	68	119
Leu-14	40	29	40	29
Arg-16	45	59	45	59
His-18	27	16	27	16
Leu-20	39	31	39	31
Arg-91	45	42	45	42
Tyr-93	33	42	33	42
Glu-95	17	31	17	31
Asp-119	1	21	1	21
Asn-120	2	18	2	18
Ile-121	83	92	83	92
Ile-122	23	16	23	16
Phe-123	63	60	63	60
Leu-124	33	1	33	1

Appendix

Table A9 Distance between the side chain oxygen atoms from the residues pointing towards each other from two monomers at the dimer-dimer interface (shown in Figure 49)

<i>EcoSSB</i>			<i>BabSSB</i>			<i>PmiSSB</i>			<i>SmaSSB</i>			<i>HsmtSSB</i>		
Atom1	d† (Å)	Atom2	Atom1	d† (Å)	Atom2	Atom1	d† (Å)	Atom2	Atom1	d† (Å)	Atom2	Atom1	d† (Å)	Atom2
Gln-A76-OE1	4.4	Gln-D76- OE1	Lys-A77-OE1	10.9	Lys-D77-OE1	Gln-A76-OE1	5.9	Gln-D76- OE1	Gln-A76-OE1	5.4	Gln-D76- OE1	Arg-A91-NH1	11.6	Arg-#B91-NH1
Tyr-A78- OH	7.6	Tyr-D78- OH	Tyr-A79-OH	9.0	Tyr-D79-OH	Tyr-A78- OH	7.2	Tyr-D78- OH	Tyr-A78- OH	7.1	Tyr-D78- OH	Tyr-A93-OH	4.8	Tyr-#B93-OH
Glu-A80-OE2	12.1	Glu-D80- OE2	Glu-A81-OE2	11.6	Glu-D81-OE2	Glu-A80-OE2	14.3	Glu-D80- OE2	Glu-A80-OE2	14.1	Glu-D80- OE2	Glu-A95-OE2	9.5	Glu-#B95-OE2
Glu-A80-OE1	4.8	Glu-C80-OE1	Glu-A81-OE1	5.2	Glu-#A81-OE1	Glu-A80-OE1	3.5	Glu-C80-OE1	Glu-A80-OE1	6.0	Glu-C80-OE1	Glu-A95-OE1	7.3	Glu-#A95-OE2
Glu-B80-OE1	4.4	Glu-D80-OE1	Glu-D81-OE1	5.2	Glu-#D81-OE1	Glu-B80-OE1	3.8	Glu-D80-OE1	Glu-B80-OE1	5.9	Glu-D80-OE1	Glu-#B95-OE1	9.4	Glu-B95-OE1
Gln-B76-OE1	4.5	Gln-C76-OE1	Lys-#A77-OE1	10.9	Lys-#D77-OE1	Gln-B76-OE1	5.6	Gln-C76-OE1	Gln-B76-OE1	5.2	Gln-C76-OE1	Arg-B91-OE1	11.6	Arg-#A91-OE1
Tyr-B78-OH	6.6	Tyr-C78-OH	Tyr-#A79-OH	9.0	Tyr-#D79-OH	Tyr-B78-OH	7.2	Tyr-C78-OH	Tyr-B78-OH	7.5	Tyr-C78-OH	Tyr-B93-OH	4.8	Tyr-#A93-OH
Glu-B80-OE2	10.3	Glu-C80-OE2	Glu-#A81-OE2	11.6	Glu-#D81-OE2	Glu-B80-OE2	14.2	Glu-C80-OE2	Glu-B80-OE2	11.9	Glu-C80-OE2	Glu-B95-OE2	9.5	Glu-#A95-OE2

† Atom1-atom2 distance

Table A10 Monomer-monomer interface of SSB-dimers and their extent burial

Monomer-monomer interface	Buried surface area (Å ²)
Monomer (A)-monomer (B) interface of EcoSSB	2207
Monomer (C)-monomer (D) interface of EcoSSB	2180
Monomer (A)-monomer (#A) interface of BabSSB	1766
Monomer (B)-monomer (#B) interface of BabSSB	1846
Monomer (A)-monomer (B) interface of PmiSSB	2114
Monomer (C)-monomer (D) interface of PmiSSB	1894
Monomer (E)-monomer (F) interface of PmiSSB	2136
Monomer (G)-monomer (H) interface of PmiSSB	2097
Monomer (A)-monomer (B) interface of SmaSSB	2178
Monomer (C)-monomer (D) interface of SmaSSB	1844
Monomer (E)-monomer (#E) interface of SmaSSB	1880
Monomer (A)-monomer (B) interface of HsmtSSB	2777

symmetry mate.

The capital letter in parenthesis indicates the chain id. of the monomer.

Table A11 Buried surface area of individual residues in each monomer upon dimer formation

<i>EcoSSB</i>				
Residues	Buried surface area (Å ²)			
	AB-dimer		CD-dimer	
	A-monomer	B-monomer	C-monomer	D-monomer
Ala-1	62			
Ser-2	25	44		55
Ala-3	49	49	60	50
Gly-4	35	33	40	38
Val-5	78	75	76	76
Asn-6	41	46	44	43
Lys-7	45	50	52	47
Val-8	19	21	21	18
Ile-9	66	60	59	67
Leu-10	16	17	17	20
Val-11	48	47	47	47
Thr-36	34	34	33	33
Ser-37	45	43	45	44
Glu-38	69	63	46	36
Ser-39		40	17	
Gln-51	15	20		16
Glu-53	67	81	70	70
His-55	30	30	30	30
Arg-56	16		32	39
Gln-82		33	28	28
Leu-83	82	82	82	86
Arg-84	79	29	27	104
Thr-85	26	25	47	43
ASP-95	17	13	15	13
Arg-96	66	30	63	67
Tyr-97	09	41	18	11
Thr-99	44	44	42	42

Appendix

<i>BabSSB</i>				
Residues	Buried surface area (Å ²)			
	A#A-dimer		B#B-dimer	
	A-monomer	#A-monomer	B-monomer	#B-monomer
Gly-3	33	33	36	36
Ser-4	75	75	71	71
Val-5	73	73	74	74
Asn-6	55	55	44	44
Lys-7	51	51	52	52
Val-8	24	24	23	23
Ile-9	64	64	66	66
Leu-10	15	15	16	16
Val-11	51	51	52	52
Thr-36	30	30	29	29
Ser-37	19	19	19	19
Glu-38	46	46	38	38
Arg-51	02	02	55	55
Glu-53	80	80	47	47
His-55	27	27	28	28
Lys-77	21	21	20	20
Leu-84	83	83	83	83
Gln-85	12	12	52	52
Thr-86	31	31		
Ala-100	35	35	44	44
Gln-105			18	18
Arg-108	30	30	33	33
<i>PmiSSB</i>				
	AB-dimer		CD-dimer	
	A-monomer	B-monomer	C-monomer	D-monomer
Ala-1	19			
Ser-2	32			
Arg-3	130	119	99	55
Gly-4	31	38	36	42
Val-5	81	79	80	79
Asn-6	44	46	44	42
Lys-7	54	52	51	54
Val-8	22	23	22	21
Ile-9	66	64	64	65
Leu-10	16	19	17	18
Ile-11	50	53	52	48
Thr-36	31	31	26	31
Ser-37	36	51	40	38
Glu-38	44	40		39
Ser-39	24	28		
Glu-50	21	12		
Glu-53	68	65	63	69
His-55	32	32	34	29
Arg-56				26
Gln-82	36	35	34	
Leu-83	88	81	87	88
Gln-84	13	17	12	23
Thr-85	26	24	25	27

Appendix

Thr-99	45	49	47	45
<i>PmiSSB</i>				
Residues	Buried surface area (Å ²)			
	EF-dimer		GH-dimer	
	E-monomer	F-monomer	G-monomer	H-monomer
Ala-1	16			
Ser-2	35			
Arg-3	121	10	113	60
Gly-4	30	36	35	35
Val-5	81	83	82	84
Asn-6	43	43	42	45
Lys-7	54	54	54	52
Val-8	20	22	18	21
Ile-9	65	64	63	66
Leu-10	17	19	19	17
Ile-11	48	50	53	52
Thr-36	32	30	28	27
Ser-37	44	37	38	46
Glu-38	37	51	37	42
Ser-39		31	05	33
Glu-53	71	60	69	65
His-55	30	34	29	30
Gln-76	26	37	06	07
Gln-82	30	33	33	34
Leu-83	84	86	87	86
Gln-84	43	23	33	23
Thr-85	24	25	26	24
Tyr-97	08	08	23	13
Thr-99	45	47	48	45
<i>SmaSSB</i>				
	AB-dimer		CD-dimer	
	A-monomer	B-monomer	C-monomer	D-monomer
Ala-1		63		
Ser-2	61	33	82	64
Arg-3	101	50	44	51
Gly-4	29	34	35	33
Val-5	74	74	76	75
Asn-6	49	51	49	48
Lys-7	51	51	51	51
Val-8	20	21	20	22
Ile-9	62	62	63	62
Leu-10	15	14	16	16
Val-11	48	48	47	47
Thr-36	33	32	33	34
Ser-37	42	44	43	40
Glu-38	50	27	32	32
Ser-39	46	17	24	27
Glu-53	69	68	76	64
His-55	30	32	32	33
Leu-83	80	83	80	83
Gln-84	16	20	11	26
Thr-85	22	25	28	32

Appendix

Asp-95	12	15		
Arg-96	60	65		
Tyr-97	30	25		
Thr-99	40	39	44	40
<i>SmaSSB</i>				
Residues	Buried surface area (Å ²)			
	E#E-dimer			
	E-monomer	#E-monomer		
Ala-1	65	65		
Ser-2	25	25		
Ala-3	52	52		
Gly-4	35	35		
Val-5	74	74		
Asn-6	48	48		
Lys-7	50	50		
Val-8	21	21		
Ile-9	62	62		
Leu-10	16	16		
Val-11	47	47		
Thr-36	32	32		
Ser-37	44	44		
Glu-38	28	28		
Ser-39	46	46		
Glu-53	74	74		
His-55	29	29		
Leu-83	82	82		
Gln-83	14	14		
Thr-85	37	37		
Arg-96	11	11		
<i>HsmtSSB</i>				
Residues	Buried surface area (Å ²)			
	AB-dimer			
	A-monomer	B-monomer		
Leu-10	03	58		
Glu-11	64	84		
Arg-12	46	27		
Ser-13	90	57		
Leu-14	89	95		
Asn-15	57	47		
Arg-16	43	38		
Val-17	25	27		
His-18	77	78		
Leu-19	17	22		
Leu-20	46	49		
Arg-22	26	30		
Thr-45	30	21		
Asn-46	79	17		
Glu-47	28	52		
Gln-64	14			
Lys-65	31			
Thr-67	39	46		
Trp-68	19	19		

Appendix

His-69	38	35		
Arg-91	47	63		
Lys-97	20	51		
Ile-98	71	55		
Asp-99	02	27		
Tyr-100	82	94		
Arg-110	76	34		
Arg-111	95	111		
Ala-113	47	51		

6.5 Hydrophobic residues of SSBs

Table A12 Hydrophobic residues belonging to the SSBs core (see footnotes)

<i>Eco</i> SSB	<i>Bab</i> SSB	<i>Pmi</i> SSB	<i>Sma</i> SSB	HsmtSSB
Val-5	Val-5	Val-5	Val-5	Leu-14
Val-8	Val-8	Val-8	Val-8	Val-17
Ile-9	Ile-9	Ile-9	Ile-9	
Leu-10	Leu-10	Leu-10	Leu-10	Leu-19
Val-11	Val-11	Ile-11	Val-11	Leu-20
Leu-14	Leu-14	Leu-14	Leu-14	Val-23
	Ala-16			
Val-20	Ile-20	Ile-20	Val-20	Leu-29
Ala-30	Ala-30	Ala-30	Ala-30	
Ile-32	Leu-32	Leu-32	Ile-32	Phe-41
Leu-34	Ile-34	Leu-34	Leu-34	Leu-43
Ala-35	Ala-35	Ala-35	Ala-35	Ala-44
Trp-54	Trp-54	Trp-54	Trp-54	Trp-68
Val-57	Val-57	Val-57	Val-57	Ile-71
Val-58	Val-58	Val-58	Val-58	
Leu-59	Ile-59	Ile-59	Leu-59	Val-73
Phe-60	Phe-60	Phe-60	Phe-60	Phe-74
Leu-63	Leu-64	Leu-63	Leu-63	Leu-78
Ala-64	Ala-65	Ala-64	Ala-64	
Val-66	Val-67	Ile-66	Val-66	Val-81
Ala-67	Ala-68	Ala-67	Ala-67	Ala-82
Tyr-70	Tyr-71	Tyr-70	Tyr-70	Tyr-85
Leu-71	Leu-72	Leu-71	Leu-71	Val-86
Val-77	Val-78	Val-77	Val-77	Ile-92
Tyr-78	Tyr-79	Tyr-78	Tyr-78	Tyr-93
Ile-79	Ile-80	Ile-79	Ile-79	Leu-94
	Ala-83			
Leu-83	Leu-84	Leu-83	Leu-83	Ile-98
Val-101	Ile-102	Val-101	Ile-101	
Val-102	Val-103	Val-102	Val-102	Ile-116
Val-103	Leu-104	Val-103	Val-103	Ile-117
				Ala-118
Met-109	Leu-111	Met-109	Met-109	
				Ile-121
Met-111	Met-113	Met-111	Met-111	Ile-122
Leu-112	Leu-114	Leu-112	Leu-112	Phe-123
				Leu-124

Residues in the same row are in structurally equivalent position in the protein. A blank indicates that there are no structurally equivalent residues.

6.6 Symmetry operators and crystals contacts of SSBs

Table A13 Symmetry operators (see footnotes)

<i>EcoSSB</i> (C2)				
Alphabetical numbering		Symmetry operator		
a	1:	-X,	Y,	-Z
b	2:	X+1/2,	Y+1/2,	Z
c	3:	-X+1/2,	Y+1/2,	-Z
d	4:	X,	Y,	Z
<i>BabSSB</i> (P4 ₃ 2 ₁ 2)				
e	1:	-Y+1/2,	X+1/2,	Z+3/4
f	2:	-X,	-Y,	Z+1/2
g	3:	Y+1/2,	-X+1/2,	Z+1/4
h	4:	Y,	X,	-Z
i	5:	X+1/2,	-Y+1/2,	-Z+1/4
j	6:	-Y,	-X,	-Z+1/2
k	7:	-X+1/2,	Y+1/2,	-Z+3/4
d	8:	X,	Y,	Z
<i>PmiSSB</i> (P2 ₁ 2 ₁ 2)				
l	1:	X+1/2,	-Y+1/2,	-Z
m	2:	-X+1/2,	Y+1/2,	-Z
n	3:	-X,	-Y,	Z
d	4:	X,	Y,	Z
<i>SmaSSB</i> (P4 ₂ 2 ₁ 2)				
o	1:	-X,	-Y,	Z
p	2:	-Y+1/2,	X+1/2,	Z+1/2
q	3:	Y+1/2,	-X+1/2,	Z+1/2
r	4:	-X+1/2,	Y+1/2,	-Z+1/2
s	5:	X+1/2,	-Y+1/2,	-Z+1/2
h	6:	Y,	X,	-Z
t	7:	-Y,	-X,	-Z
d	8:	X,	Y,	Z
<i>HsmtSSB</i> (P4 ₁ 2 ₁ 2)				
u	1:	-X,	-Y,	Z+1/2
v	2:	-Y+1/2,	X+1/2,	Z+1/4
w	3:	Y+1/2,	-X+1/2,	Z+3/4
x	4:	-X+1/2,	Y+1/2,	-Z+1/4
y	5:	X+1/2,	-Y+1/2,	-Z+3/4
h	6:	Y,	X,	-Z
z	7:	-Y,	-X,	-Z+1/2
d	8:	X,	Y,	Z

The alphabetical numbering of symmetry operators is used in Table A14, A15, A16, A17, A18.

Table A14 Intermolecular hydrogen-bonding contacts of *EcoSSB*

Acceptor	Donor	H-bond distance(Å)	Symmetry mates*
Glu-A69-O	Wat-W79-O	2.71	a
Arg-A72-NE	LeuA112-O	2.92	a
Arg-A72-NH1	Met-A111-O	2.84	a
Asn-B31-OD1	Gln-C91-OE1	2.77	b
Asn-B31-OD1	GlnC91-NE2	2.40	b
Asn-B31-ND2	Gln-C91-OE1	2.57	b
Gly-B46-O	GlnD91-OE1	3.19	b
Glu-B47-OE1	Arg-D86-NH1	2.82	b
Glu-B47-OE1	Arg-D86-NH2	3.21	b
Glu-B47-OE2	Arg-D86-NH2	2.85	b
Lys-B49-NZ	Glu-D19-OE2	3.43	b
Ser-B68-OG	Arg-C72-NH2	3.42	c
Ser-B68-OG	Arg-C72-NE	2.74	c
Ser-B68-O	Arg-C72-NH2	3.18	c
Arg-B72-NH2	Glu-C69-O	3.33	c
Arg-B84-NH1	Gln-C91-O	2.57	b
Arg-B84-NH2	Gln-C91-O	3.01	b
Arg-B86-NH1	Thr-C89-O	2.93	b
Arg-B86-NH2	Gln-C91-NE2	3.43	b
Lys-B87-O	Thr-C89-N	3.08	b
Lys-B87-O	Wat-W102-O	2.95	b
Lys-B87-O	Thr-C89-OG1	3.43	b
Thr-B89-N	Lys-C87-O	3.05	b
Thr-B89-OG1	Lys-C87-O	3.48	b
Thr-B89-OG1	Wat-W76-O	3.36	b
Thr-B89-O	Arg-C86-NH1	3.02	b
Gln-B91-OE1	Asn-C31-OD1	2.52	b
Gln-B91-OE1	Asn-C31-ND2	2.60	b
Gln-B91-NE2	Asn-C31-ND2	3.42	b
Gln-B91-O	Arg-C84-NH1	2.84	b
Gln-B91-O	Arg-C84-NH2	2.61	b
Gly-B93-N	Arg-C84-NH2	3.51	b
Glu-B100-OE2	Gln-C91-NE2	2.97	b
Arg-C3-NH2	Wat-W1-O	3.27	a
Ala-C44-O	Asn-D25-OD1	2.95	a
AlaC44-O	Asn-D25-N	3.13	a
Glu-C50-OE1	Lys-D87-NZ	3.38	a
Lys-C87-N	Wat-W12-O	2.66	b
Thr-C89-OG1	Wat-W34-O	2.68	b
Wat-W27-O	Wat-W65-O	3.19	b

* See Table A13.

Table A15 Intermolecular hydrogen-bonding contacts of *BabSSB*

Acceptor	Donor	H-bond distance (Å)	Symmetry mates*
Val-A5-N	Val-A11-O	2.90	h
Val-A5-O	Val-A11-N	3.00	h
Val-A5-O	Val-A11-O	3.51	h
Lys-A7-N	Ile-A9-O	3.02	h
Lys-A7-O	Ile-A9-N	2.84	h
Lys-A7-O	Ile-A9-O	3.38	h
Val-B5-N	Val-B11-O	2.84	h
Val-B5-O	Val-B11-O	3.40	h
Val-B5-O	Val-B11-N	2.97	h
Lys-B7-N	Ile-B9-O	2.93	h
Lys-B7-O	Ile-B9-N	2.81	h
Lys-B7-O	Ile-B9-O	3.34	h
Asp-A50-O	Gly-B26-N	3.53	e,g
Thr-A52-N	Asn-B24-O	2.75	e
Gly-A3-O	Lys-A77-NZ	3.21	h
Gly-B3-O	Lys-B77-NZ	3.01	h
Ser-A4-OG	Ser-A37-N	2.96	h
Ser-A4-OG	Ser-A37-O	3.17	h
Ser-B4-OG	Ser-B37-N	2.93	h
Ser-B4-OG	Ser-B37-O	2.91	h
Ser-A4-O	Gln-B112-OE1	3.39	h
Ser-A4-N	Gln-B112-OE1	3.20	h
Ser-B4-N	Gln-A112-OE1	3.27	h
Ser-B4-O	Gln-A112-OE1	3.35	h
Gly-B3-N	Gln-A112-OE1	2.60	h
Trp-A40-NE1	Ser-B25-O	3.04	e
Glu-A53-OE1	Thr-A86-N	2.93	h
Glu-A53-OE2	Thr-A86-N	3.37	h
Ala-A68-O	Lys-A73-NZ	3.13	h
Glu-A69-O	Lys-A73-NZ	2.85	h
Leu-A72-O	Lys-A73-NZ	2.85	h
Lys-A7-NZ	Glu-B81-OE2	2.91	h
Lys-A7-NZ	Glu-B110-OE1	3.22	h
Lys-A7-NZ	Glu-B110-OE2	3.19	h
Lys-B7-NZ	Glu-A81-OE1	2.50	h
Lys-B7-NZ	Glu-A81-OE2	3.18	h
Lys-B7-NZ	Glu-A110-OE1	3.31	h
Glu-A38-OE1	Arg-A108-NH1	2.57	h
Glu-A38-OE1	Arg-A108-NH2	3.33	h
Glu-B38-OE2	Arg-B108-NH1	2.62	h
Glu-B38-OE2	Arg-B108-NH2	3.55	h
Glu-A47-OE1	Arg-B22-NH2	3.40	e,g
Lys-A49-NZ	Asp-B27-OD1	2.49	e
Lys-A49-NZ	Glu-B62-OE1	3.18	e
Lys-A49-NZ	Glu-B62-OE2	2.66	e
Glu-A53-OE1	Thr-A86-OG1	2.53	h
Glu-A53-OE2	Thr-A86-OG1	3.33	h
Trp-A54-NE1	Asn-B24-OD1	2.93	e
Arg-B51-NH1	Gln-B85-OE1	3.15	h
Wat-W9-O	Tyr-A71-O	2.77	h
Wat-W12-O	Gln-A70-O	3.44	h
Wat-W17-O	Leu-B23-O	2.89	e
Wat-W17-O	Asn-B24-O	3.47	e,g
Wat-W29-O	Ser-B37-O	3.53	h
Wat-W44-O	Pro-A18-O	2.63	h
Wat-W46-O	Leu-B23-O	3.55	e

Appendix

Acceptor	Donor	H-bond distance (Å)	Symmetry mates*
Wat-W46-O	Asn-B24-N	3.54	e,g
Wat-W46-O	Arg-B22-O	3.15	e,g
Wat-W54-O	Gly-B3-N	3.35	h
Wat-W80-O	Ser-A4-O	3.45	h
Wat-W94-O	Gln-A70-O	2.59	h
Wat-W100-O	Ser-A4-N	3.39	h
Wat-W107-O	Ser-A4-O	3.34	h
Wat-W110-O	Ser-B37-O	2.68	h
Wat-W113-O	Glu-B38-N	3.56	h
Wat-W113-O	Ser-B39-N	3.51	h
Wat-W80-O	Ser-A4-OG	3.43	h
Wat-W100-O	Ser-A4-OG	2.83	h
Wat-W37-O	Ser-A4-OG	3.27	h
Wat-W4-O	Asn-A6-ND2	3.29	h
Wat-W39-O	Asn-B6-ND2	3.39	h
Wat-W96-O	Glu-A47-OE1	2.97	e
Wat-W96-O	Glu-A47-OE2	2.73	e
Wat-W91-O	Arg-A51-NE	2.77	e
Wat-W91-O	Arg-A51-NH2	3.14	e
Wat-W55-O	Arg-B51-NH2	3.19	h
Wat-W97-O	Gln-A85-NE2	3.14	h
Wat-W40-O	Thr-A86-OG1	3.11	h
Wat-W65-O	Glu-A110-OE1	3.20	h
Wat-W65-O	Glu-A110-OE2	3.14	h
Wat-W110-O	Gln-A112-OE1	3.07	h
Wat-W113-O	Ser-B37-OG	2.86	h
Wat-W113-O	Ser-B39-OG	2.83	h
Wat-W46-O	Asn-B24-OD1	2.80	e
Wat-W69-O	Glu-B62-OE2	2.83	e
Wat-W5-O	Wat-W31-O	2.97	h
Wat-W9-O	Wat-W10-O	2.67	h
Wat-W16-O	Wat-W107-O	2.80	h
Wat-W17-O	Wat-W58-O	2.79	e
Wat-W29-O	Wat-W39-O	3.06	h
Wat-W35-O	Wat-W82-O	2.65	h
Wat-W53-O	Wat-W82-O	3.46	h

* See Table A13.

Table A16 Intermolecular hydrogen-bonding contacts of *PmiSSB*

Acceptor	Donor	H-bond distance (Å)	Symmetry mates*
Arg-A3-NE	Tyr-G22-OH	3.28	l
Arg-A3-NH1	Tyr-G22-OH	3.13	l
Arg-A3-NH2	Tyr-G22-OH	2.76	l
Arg-A3-NH2	Glu-G65-OE2	2.43	l
Met-A48-N	Ile-E20-O	3.09	n
Arg-B3-NH2	Gly-E68-O	2.66	n
Asn-B31-OD1	Gln-C91-NE2	3.09	d
Ser-B39-OG	Glu-G65-OE2	3.28	l
Gln-B84-NE2	Gln-C91-O	3.19	d
Lys-B87-N	Gln-C89-OE1	2.74	d
Lys-B87-O	Gln-C89-N	3.17	d
Gln-B89-N	Lys-C87-O	2.99	d
Gln-B89-OE1	Lys-C87-N	2.93	d
Gln-B89-O	Arg-C86-NH1	3.37	d

Appendix

Acceptor	Donor	H-bond distance (Å)	Symmetry mates*
Gln-B91-N	Arg-C86-NH1	3.24	d
Gln-B91-OE1	Asn-C31-ND2	3.29	d
Gln-B91-NE2	Asn-C31-OD1	2.90	d
Gln-B91-NE2	Glu-C100-OE2	3.29	d
Gln-B91-O	Gln-C84-OE1	3.54	d
Gln-B91-O	Gln-C84-NE2	3.34	d
Glu-B100-OE1	Gln-C91-OE1	2.86	d
Glu-B100-OE2	Gln-C91-OE1	3.25	d
Glu-C65-OE1	Ser-F39-OG	3.41	l
Glu-C65-OE2	Ser-F39-OG	2.58	l
Glu-C69-N	Ala-E1-O	3.54	l
Glu-C69-OE1	Met-G111-N	2.86	l
Arg-C72-NH2	Leu-G112-O	3.53	l
Arg-C72-NH2	Gly-G113-O	3.24	l
Met-C111-N	Glu-G69-OE1	3.05	l
Leu-C112-O	Arg-G72-NH1	3.59	l
Gly-C113-O	Arg-G72-NE	3.55	l
Gly-C113-O	Arg-G72-NH1	3.33	l
Gly-C113-O	Arg-G72-NH2	3.23	l
Ile-D20-O	Met-H48-N	3.01	n
Tyr-D22-N	Gly-H46-O	2.80	n
Gly-D68-O	Arg-G3-NH2	3.23	n
Tyr-D70-OH	Met-E111-N	3.02	n
Met-D109-O	Wat-W51-O	3.18	n
Met-D111-N	Tyr-E70-OH	3.21	n
Ile-E20-O	Met-A48-N	3.09	n
Arg-F21-NH2	Gln-G91-OE1	3.41	d
Asn-F31-OD1	Gln-G91-NE2	2.96	d
Gln-F84-OE1	Gln-G91-O	3.22	d
Arg-F86-NE	Gln-G91-N	3.44	d
Arg-F86-NE	Gln-G89-O	3.24	d
Lys-F87-N	Gln-G89-OE1	2.71	d
Lys-F87-O	Gln-G89-N	3.11	d
Gln-F89-N	Lys-G87-O	3.09	d
Gln-F89-OE1	Lys-G87-N	2.71	d
Gln-F89-O	Arg-G86-NH1	3.27	d
Gln-F91-OE1	Glu-G100-OE2	3.46	d
Gln-F91-OE1	Asn-G31-OD1	2.67	d
Gln-F91-OE1	Asn-G31-ND2	3.51	d
Gln-F91-O	Gln-G84-NE2	2.84	d
Glu-F100-OE2	Gln-G91-NE2	3.09	d
Tyr-G70-O	Wat-W41-O	2.72	d
Gln-G91-O	Wat-W79-O	3.41	d
Ala-H69-O	Arg-H72-NH2	3.28	n
Arg-H72-NE	Wat-W26-O	3.34	n
Arg-H72-NH1	Wat-W26-O	2.65	n
Arg-H72-NH1	Wat-W42-O	2.68	n
Arg-H72-NH2	Wat-W42-O	2.88	n
Wat-W26-O	Wat-W46-O	2.77	n

* See Table A13.

Appendix

Table A17 Intermolecular hydrogen-bonding contacts of *SmaSSB*

Acceptor	Donor	H-bond distance (Å)	Symmetry mates*
Asn-A25-O	Ala-A28-O	2.69	h
Asn-A25-O	Ala-A28-N	2.71	h
Asn-A25-O	Gly-A61-N	3.09	h
Asn-A31-OD1	Gln-E91-OE1	3.49	r
Asn-A31-OD1	Gln-E91-NE2	3.08	r
Asn-A31-ND2	Gln-E91-OE1	3.43	r
Gln-A84-OE1	Gln-E91-O	2.82	r
Lys-A87-N	Gln-E89-OE1	3.43	r
Lys-A87-O	Gln-E89-N	2.66	r
Gln-A89-OE1	Arg-E86-NH2	3.23	r
Gln-A89-NE2	Lys-E87-O	3.20	r
Gln-A91-O	Arg-E21-NH1	3.45	r
Glu-A100-OE1	Gln-E91-NE2	3.57	r
Glu-A100-OE2	Gln-E91-NE2	2.68	r
Asn-B31-OD1	Gln-D91-NE2	2.55	r
Asn-B31-ND2	Gln-D91-NE2	3.13	r
Gln-B84-NE2	Gln-D91-O	2.63	r
Gln-B84-NE2	Gly-D93-N	3.35	r
Arg-B86-NH1	Gln-D91-OE1	3.16	r
Lys-B87-N	Gln-D89-OE1	2.76	r
Lys-B87-O	Gln-D89-N	3.16	r
Gln-B89-NE2	Ala-D87-O	3.03	r
Glu-B100-OE1	Gln-D91-OE1	2.80	r
Glu-B100-OE2	Gln-D91-OE1	2.90	r
Ala-E1-O	Ser-E39-N	3.04	t
Ala-E1-N	Ser-E39-O	2.45	t
Ser-E2-O	Gln-E110-OE1	3.34	o
Arg-E3-N	Ser-E37-O	2.85	t
Arg-E3-O	Ser-E37-N	2.94	t
Gly-E4-N	Gln-E110-NE2	2.70	o
Gly-E4-O	Gln-E110-OE1	3.49	o
Val-E5-N	Val-E11-O	2.89	t
Val-E5-O	Ile-E9-O	3.37	t
Val-E5-O	Val-E11-N	2.96	t
Asn-E6-ND2	His-E55-NE2	3.43	t
Lys-E7-N	Ile-E9-O	2.94	t
Lys-E7-NZ	Glu-E80-OE1	2.79	o
Lys-E7-NZ	Glu-E80-OE2	2.86	o
Lys-E7-NZ	Tyr-E78-OH	3.21	o
Lys-E7-O	Ile-E9-N	2.75	t
Lys-E7-O	Ile-E9-O	3.30	t
Glu-E53-OE1	Thr-E85-OG1	3.26	t
Glu-E53-OE1	Thr-E85-N	2.30	t
Glu-E53-OE1	Thr-E85-O	3.41	t
Thr-E99-OG1	Wat-W4-O	2.71	t

* See Table A13.

Table A18 Intermolecular hydrogen-bonding contacts of HsmtSSB

Acceptor	Donor	H-bond distance(Å)	Symmetry mates*
Arg-A12-NH1	Ile-A122-N	3.47	z
Arg-A12-NH1	Wat-A131-O	3.08	z
Arg-A12-NH1	Ile-A122-O	3.42	z
Arg-A12-NH2	Ile-A122-O	3.33	z
His-A18-ND1	Wat-B130-O	2.84	z
Arg-A22-NH1	Asp-B105-O	3.29	x
Arg-A22-NH2	Asp-B105-O	2.97	x
Gln-A25-NE2	Glu-B33-OE2	3.20	x
Gln-A25-NE2	Glu-B33-OE1	3.17	x
Arg-A30-NH2	Ser-A63-OG	3.53	x
Glu-A33-O	Thr-B66-OG1	3.44	x
Gly-A34-N	Trp-B68-NE1	3.58	x
Gly-A34-O	Trp-B68-NE1	3.60	x
Ser-A42-OG	Glu-B33-OE1	3.00	x
Trp-A68-NE1	Glu-B33-O	3.27	x
Trp-A68-NE1	Gly-B34-N	3.24	x
Asp-A80-OD2	Tyr-B83-OH	3.24	d
Gln-A84-OE1	Leu-B29-O	3.08	d
Tyr-A85-OH	Arg-B79-NE	2.93	d
Tyr-A85-OH	Asp-B80-OD1	2.66	d
Lys-A88-O	Met-B104-N	2.79	x
Lys-A88-NZ	Gly-B101-O	3.29	x
Lys-A88-NZ	Gln-B112-OE1	2.79	x
Gly-A89-O	Met-B104-O	2.98	x
Arg-A91-NH1	Ile-B122-O	3.07	z
Tyr-A93-OH	Wat-B130-O	2.75	z
Ile-A122-O	Wat-B135-O	3.36	z
Ile-A122-O	Arg-B91-NH2	3.00	z
Leu-A124-O	Wat-B135-O	2.97	z
Leu-A124-O	Arg-B91-NH2	3.54	z
Wat-A128-O	Wat-B140-O	3.49	x
Wat-A141-O	Asp-B119-OD1	3.09	z
Wat-A141-O	Asp-B119-OD2	3.46	z
Arg-B12-NH1	Asn-B120-O	2.87	z
Arg-B12-NH2	Asn-B120-O	3.37	z
Arg-B12-NH2	Asp-B119-OD1	2.68	z
Arg-B12-NH2	Asn-B120-N	2.65	z
Arg-B16-NH1	Glu-B95-OE2	3.09	z
Arg-B16-NH1	His-B18-NE2	3.43	z
Arg-B16-NH2	Glu-B95-OE1	2.75	z
Arg-B16-NH2	Glu-B95-OE2	2.80	z
Lys-B35-NZ	Glu-B102-OE1	3.16	x

* See Table A13

6.7 B-factor plots of SSBs

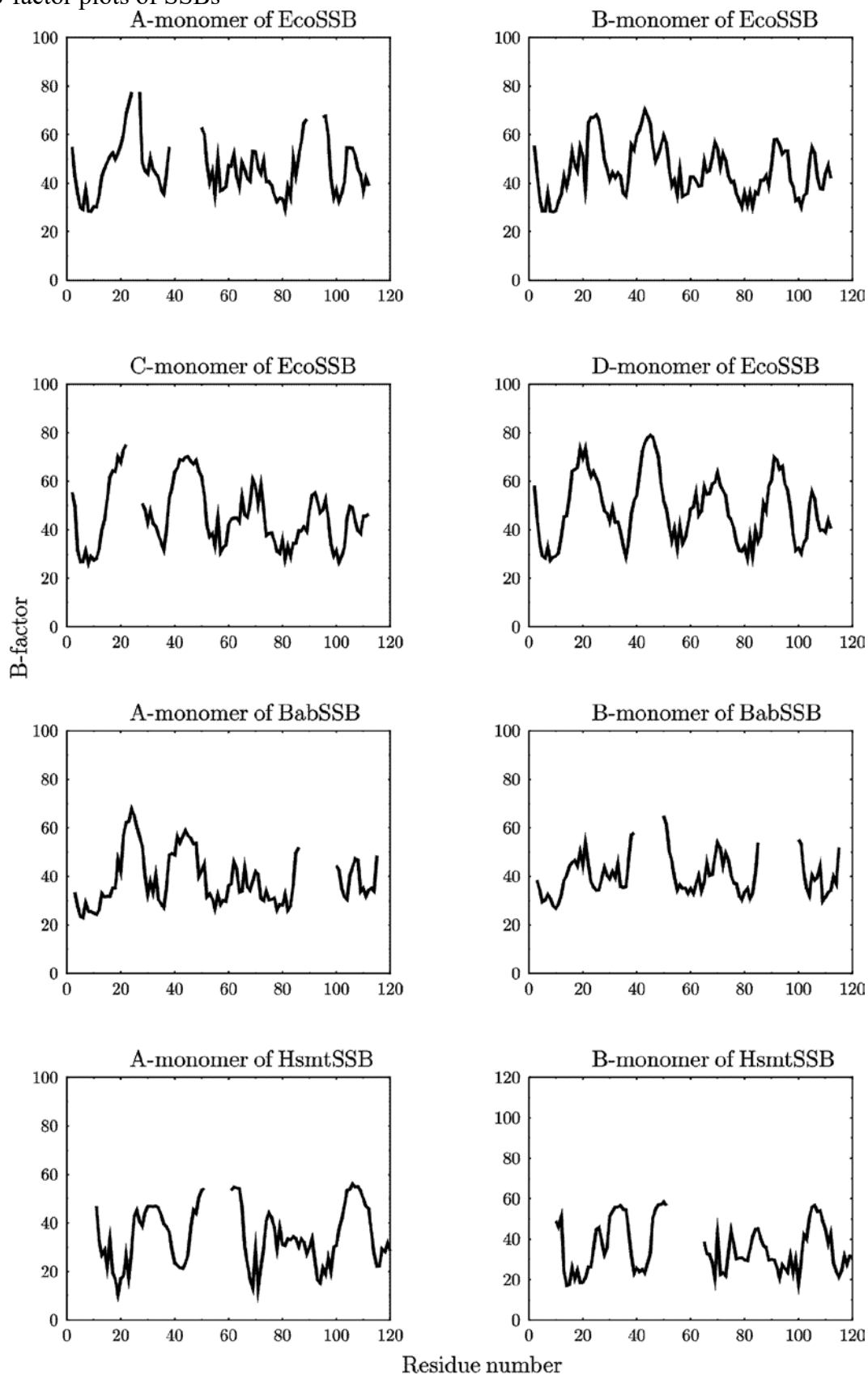


Figure A2 B-factor plot of *EcoSSB*, *BabSSB* and *HsmtSSB* monomers. The average B-factors for the main-chain and side-chain atoms of each residue are plotted against the residue number. Gaps in the plot indicate residues missing in the model. The last four residues of each monomer of *HsmtSSB* are not shown in the plot.

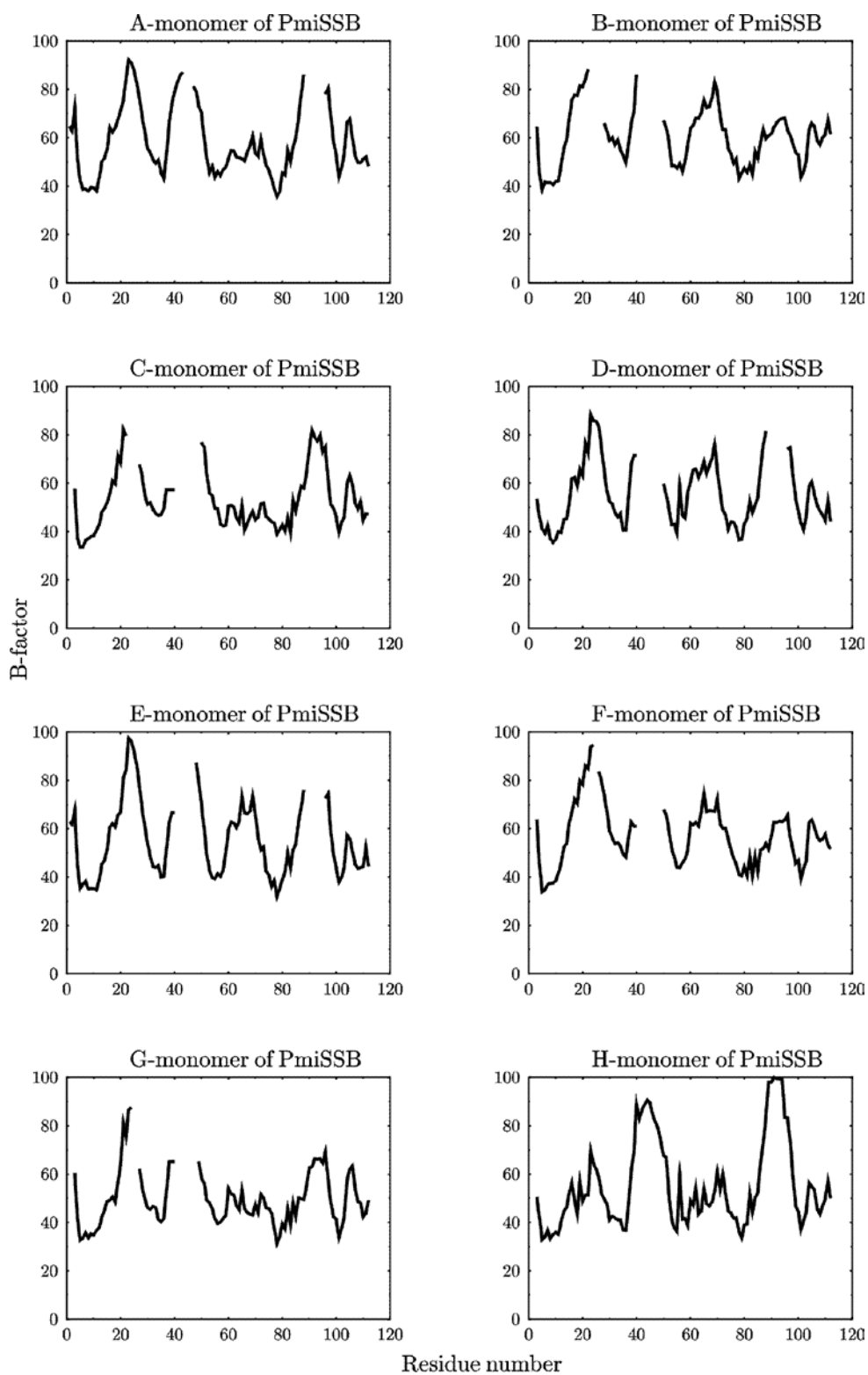


Figure A3 B-factor plot of *PmiSSB* monomers. The average B-factors for the main-chain and side-chain atoms of each residue are plotted against the residue number. Gaps in the plot indicate residues missing in the model.

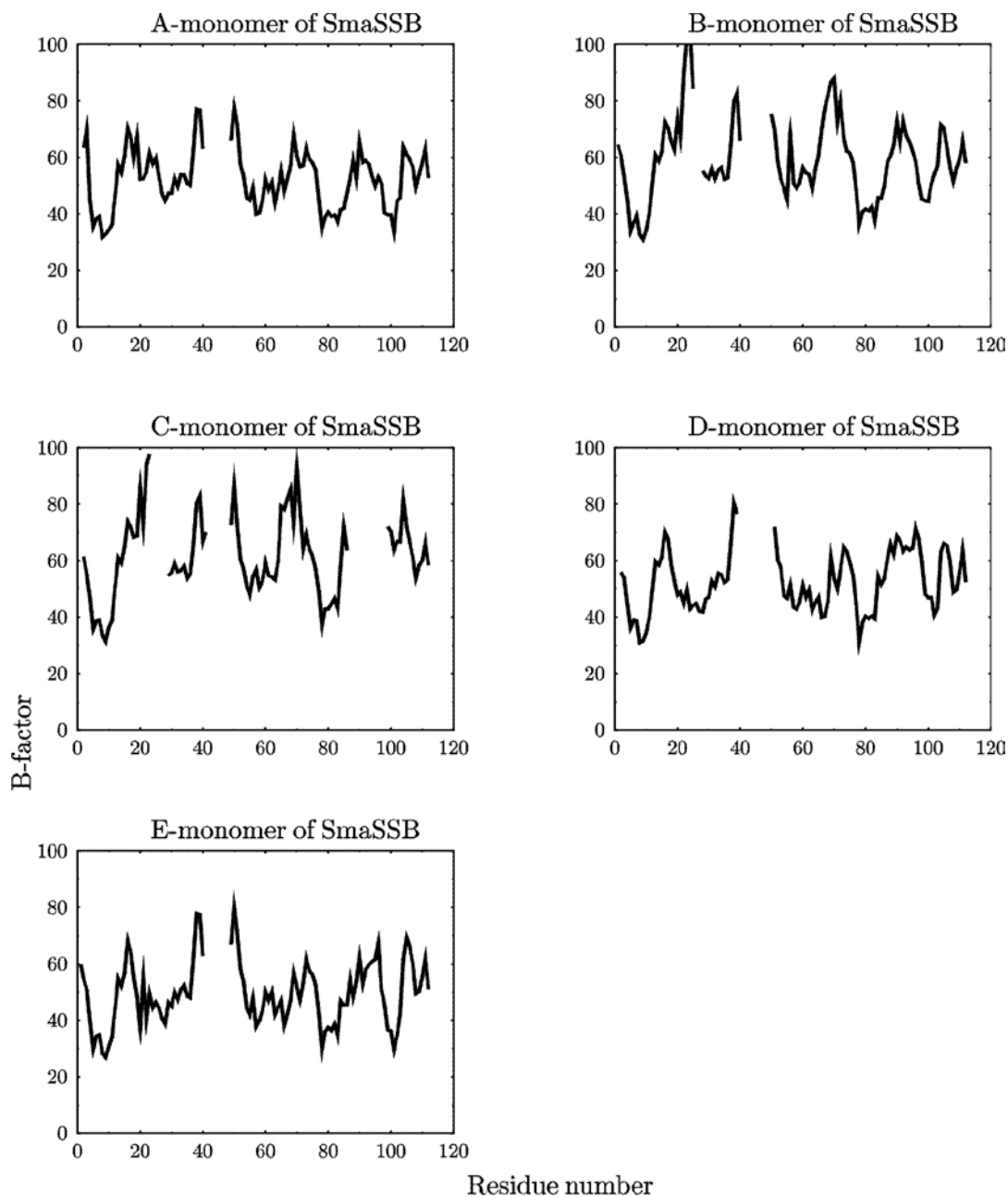


Figure A4 B-factor plot of *SmaSSB* monomers. The average B-factors for the main-chain and side-chain atoms of each residue are plotted against the residue number. Gaps in the plot indicate residues missing in the model.

Table A19 Comparison of OB-fold containing proteins (see footnotes)

PDB	Å (C _α pair)	PDB	PDB	Å (C _α pair)	PDB	PDB	Å (C _α pair)	PDB	PDB	Å (C _α pair)	PDB
1A62	2.59(34)	1ADT	1BAB-A	2.58(72)	1BCP	1GPC	2.43(35)	1SNC	1LT5-D	2.49(66)	1PMI-H
1A62	2.57(46)	1AH9	1BAB-A	1.36(83)	1ECO-A	1GPC	2.52(32)	1SRO	1LT5-D	2.38(65)	1SMA-A
1A62	2.77(56)	1BAB-A	1BAB-A	2.68(57)	1GPC	1GPC	2.79(51)	1TII-D	1LT5-D	2.71(52)	1SNC
1A62	2.74(54)	1BCP	1BAB-A	2.54(52)	1GVP	1GPC	2.78(33)	2PRD	1LT5-D	2.72(47)	1SRO
1A62	2.39(58)	1ECO-A	1BAB-A	2.25(57)	1HJP	1GPC	2.71(39)	2QIL-A	1LT5-D	1.99(87)	1TII-D
1A62	2.68(42)	1GPC	1BAB-A	2.43(80)	1JMC-A	1GPC	2.66(45)	3SEB	1LT5-D	2.81(40)	2PRD
1A62	2.82(40)	1GVP	1BAB-A	2.70(70)	1KRS	1GPC	2.55(60)	3ULL-A	1LT5-D	2.85(61)	2QIL-A
1A62	2.42(53)	1HJP	1BAB-A	2.33(69)	1LT5-D	1GVP	2.58(38)	1HJP	1LT5-D	2.32(57)	3SEB
1A62	2.47(60)	1JMC-A	1BAB-A	2.34(67)	1LYL-A	1GVP	2.71(58)	1JMC-A	1LT5-D	2.45(72)	3ULL-A
1A62	2.65(50)	1KRS	1BAB-A	2.47(55)	1MJC	1GVP	3.03(34)	1KRS	1LYL-A	2.93(63)	1MJC
1A62	2.72(52)	1LT5-D	1BAB-A	1.72(92)	1PMI-H	1GVP	2.48(46)	1LT5-D	1LYL-A	2.31(77)	1PMI-H
1A62	2.82(55)	1LYL-A	1BAB-A	1.90(87)	1SMA-A	1GVP	2.50(45)	1LYL-A	1LYL-A	2.04(76)	1SMA-A
1A62	2.37(57)	1MJC	1BAB-A	2.47(52)	1SNC	1GVP	2.14(45)	1MJC	1LYL-A	2.91(55)	1SNC
1A62	2.44(55)	1PMI-H	1BAB-A	2.57(54)	1SRO	1GVP	2.63(55)	1PMI-H	1LYL-A	2.63(60)	1SRO
1A62	2.61(56)	1SMA-A	1BAB-A	2.42(63)	1TII-D	1GVP	2.64(50)	1SMA-A	1LYL-A	2.70(65)	1TII-D
1A62	2.86(39)	1SNC	1BAB-A	2.44(51)	2PRD	1GVP	2.66(41)	1SNC	1LYL-A	2.50(53)	2PRD
1A62	2.42(56)	1SRO	1BAB-A	2.18(67)	2QIL-A	1GVP	2.89(40)	1SRO	1LYL-A	2.08(62)	2QIL-A
1A62	2.53(47)	1TII-D	1BAB-A	2.65(59)	3SEB	1GVP	2.60(45)	1TII-D	1LYL-A	2.72(65)	3SEB
1A62	2.56(49)	2PRD	1BAB-A	1.80(85)	3ULL-A	1GVP	2.41(37)	2PRD	1LYL-A	2.14(75)	3ULL-A
1A62	2.66(53)	2QIL-A	1BCP	2.53(72)	1ECO-A	1GVP	2.14(50)	2QIL-A	1MJC	2.68(57)	1PMI-H
1A62	2.72(51)	3SEB	1BCP	2.98(54)	1GPC	1GVP	2.62(44)	3SEB	1MJC	2.66(56)	1SMA-A
1A62	2.56(55)	3ULL-A	1BCP	2.35(51)	1GVP	1GVP	2.63(54)	3ULL-A	1MJC	2.22(49)	1SNC
1ADT	2.76(28)	1AH9	1BCP	2.40(59)	1HJP	1HJP	2.35(57)	1JMC-A	1MJC	2.84(63)	1SRO
1ADT	2.55(31)	1BAB-A	1BCP	2.54(70)	1JMC-A	1HJP	2.44(56)	1KRS	1MJC	2.84(48)	1TII-D
1ADT	2.60(47)	1BCP	1BCP	2.38(72)	1KRS	1HJP	2.66(53)	1LT5-D	1MJC	2.35(59)	2PRD
1ADT	2.28(30)	1ECO-A	1BCP	2.37(79)	1LT5-D	1HJP	2.72(53)	1LYL-A	1MJC	2.68(56)	2QIL-A
1ADT	2.34(36)	1GPC	1BCP	2.24(72)	1LYL-A	1HJP	2.59(49)	1MJC	1MJC	2.15(43)	3SEB
1ADT	2.81(27)	1GVP	1BCP	2.62(51)	1MJC	1HJP	2.15(60)	1PMI-H	1MJC	2.39(56)	3ULL-A
1ADT	2.87(43)	1HJP	1BCP	2.34(67)	1PMI-H	1HJP	1.86(60)	1SMA-A	1PMI-H	1.34(100)	1SMA-A
1ADT	2.75(35)	1JMC-A	1BCP	2.24(69)	1SMA-A	1HJP	2.85(57)	1SNC	1PMI-H	2.79(57)	1SNC
1ADT	2.53(31)	1KRS	1BCP	2.44(53)	1SNC	1HJP	2.70(56)	1SRO	1PMI-H	2.75(55)	1SRO
1ADT	2.57(32)	1LT5-D	1BCP	2.66(55)	1SRO	1HJP	2.31(53)	1TII-D	1PMI-H	2.52(62)	1TII-D
1ADT	2.12(41)	1LYL-A	1BCP	2.45(77)	1TII-D	1HJP	2.65(55)	2PRD	1PMI-H	2.56(51)	2PRD
1ADT	2.27(25)	1MJC	1BCP	2.54(42)	2PRD	1HJP	2.26(67)	2QIL-A	1PMI-H	2.13(61)	2QIL-A
1ADT	2.72(29)	1PMI-H	1BCP	2.68(61)	2QIL-A	1HJP	2.75(54)	3SEB	1PMI-H	2.79(56)	3SEB
1ADT	2.75(32)	1SMA-A	1BCP	2.43(67)	3SEB	1HJP	2.25(62)	3ULL-A	1PMI-H	1.41(91)	3ULL-A
1ADT	2.37(34)	1SNC	1BCP	2.31(78)	3ULL-A	1JMC-A	2.37(63)	1KRS	1SMA-A	2.73(57)	1SNC
1ADT	2.83(30)	1SRO	1ECO-A	2.57(54)	1GPC	1JMC-A	2.18(66)	1LT5-D	1SMA-A	2.62(59)	1SRO
1ADT	2.78(38)	1TII-D	1ECO-A	2.78(53)	1GVP	1JMC-A	2.33(91)	1LYL-A	1SMA-A	2.52(64)	1TII-D
1ADT	2.85(36)	2PRD	1ECO-A	2.14(58)	1HJP	1JMC-A	2.49(58)	1MJC	1SMA-A	2.47(53)	2PRD
1ADT	2.40(35)	2QIL-A	1ECO-A	2.26(78)	1JMC-A	1JMC-A	2.66(79)	1PMI-H	1SMA-A	1.99(61)	2QIL-A
1ADT	2.86(37)	3SEB	1ECO-A	2.52(66)	1KRS	1JMC-A	2.34(76)	1SMA-A	1SMA-A	2.66(61)	3SEB
1ADT	2.36(34)	3ULL-A	1ECO-A	2.35(69)	1LT5-D	1JMC-A	2.81(49)	1SNC	1SMA-A	1.86(92)	3ULL-A
1AH9	2.45(50)	1BAB-A	1ECO-A	1.55(73)	1LYL-A	1JMC-A	2.64(54)	1SRO	1SNC	2.61(52)	1SRO
1AH9	2.81(50)	1BCP	1ECO-A	2.65(56)	1MJC	1JMC-A	2.52(63)	1TII-D	1SNC	2.37(52)	1TII-D
1AH9	2.59(45)	1ECO-A	1ECO-A	1.24(90)	1PMI-H	1JMC-A	2.58(55)	2PRD	1SNC	2.90(41)	2PRD
1AH9	2.62(36)	1GPC	1ECO-A	1.50(89)	1SMA-A	1JMC-A	2.42(68)	2QIL-A	1SNC	2.75(47)	2QIL-A
1AH9	2.44(33)	1GVP	1ECO-A	2.76(57)	1SNC	1JMC-A	2.82(56)	3SEB	1SNC	2.57(53)	3SEB
1AH9	2.62(50)	1HJP	1ECO-A	2.66(62)	1SRO	1JMC-A	2.36(75)	3ULL-A	1SNC	2.53(58)	3ULL-A
1AH9	2.60(50)	1JMC-A	1ECO-A	2.50(61)	1TII-D	1KRS	2.69(69)	1LT5-D	1SRO	2.74(47)	1TII-D

Appendix

1AH9	2.72(52)	1KRS	1ECO-A	2.55(54)	2PRD	1KRS	2.77(66)	1LYL-A	1SRO	2.51(54)	2PRD
1AH9	2.37(42)	1LT5-D	1ECO-A	2.17(63)	2QIL-A	1KRS	2.71(50)	1MJC	1SRO	2.65(54)	2QIL-A
1AH9	2.77(51)	1LYL-A	1ECO-A	2.72(62)	3SEB	1KRS	2.39(69)	1PMI-H	1SRO	2.59(57)	3SEB
1AH9	2.47(53)	1MJC	1ECO-A	2.03(88)	3ULL-A	1KRS	2.50(72)	1SMA-A	1SRO	2.80(57)	3ULL-A
1AH9	2.51(44)	1PMI-H	1GPC	2.42(56)	1GVP	1KRS	2.88(53)	1SNC	1TII-D	2.80(41)	2PRD
1AH9	2.48(50)	1SMA-A	1GPC	2.77(33)	1HJP	1KRS	2.83(49)	1SRO	1TII-D	2.34(55)	2QIL-A
1AH9	2.40(42)	1SNC	1GPC	2.44(66)	1JMC-A	1KRS	2.04(66)	1TII-D	1TII-D	3.00(59)	3SEB
1AH9	2.46(56)	1SRO	1GPC	2.63(44)	1KRS	1KRS	2.68(45)	2PRD	1TII-D	2.51(71)	3ULL-A
1AH9	2.62(46)	1TII-D	1GPC	2.66(52)	1LT5-D	1KRS	2.86(63)	2QIL-A	2PRD	2.45(58)	2QIL-A
1AH9	1.91(46)	2PRD	1GPC	2.34(34)	1LYL-A	1KRS	2.62(66)	3SEB	2PRD	2.87(46)	3SEB
1AH9	2.30(50)	2QIL-A	1GPC	2.55(35)	1MJC	1KRS	2.32(68)	3ULL-A	2PRD	2.65(47)	3ULL-A
1AH9	2.67(41)	3SEB	1GPC	2.78(57)	1PMI-H	1LT5-D	2.40(63)	1LYL-A	2QIL-A	2.19(151)	3SEB
1AH9	2.57(58)	3ULL-A	1GPC	2.50(46)	1SMA-A	1LT5-D	2.77(47)	1MJC	2QIL-A	2.58(68)	3ULL-A

Each row of each highlighted column shows the comparison between two proteins. The C_{α} r.m.s.d and the number atoms pairs used for calculation are given in between two PDB code. 1ECO, 1BAB, 1PMI, and 1SMA are signed as hypothetical PDB codes for the *Eco*SSB, *Bab*SSB, *Pmi*SSB and *Sma*SSB structure, respectively. PDB code and chain id separated by a hyphen, indicates that the chain id containing monomer is used for the r.m.s.d. calculations and with no chain id indicates that the protein is a monomer.

6.8 Structure factor and co-ordinate files of SSBs

The following files can be found on the attached CD:

Co-ordinate	Structure factor
E_coli_SSB.pdb	E_coli_SSB.hkl
B_abortus_SSB.pdb	B_abortus_SSB.hkl
P_mirabilis_SSB.pdb	P_mirabilis_SSB.hkl
S_marcescens_SSB.pdb	S_marcescens_SSB.hkl

The PDB files contain the refined co-ordinates of the various SSBs in standard Brookhaven PDB format, and the HKL files contain the Miller indices, the structure factor amplitudes and standard deviations in the format (3I4,2F8.2).

Acknowledgments

This is an attempt to acknowledge all those who have been instrumental in making this project a success.

I would like to thank to Prof. Dr. Rolf Hilgenfeld for giving me an opportunity to work in his group under his guidance. I owe a lot to him for spending great deal of his precious time on this project despite his busy schedule.

My sincere gratitude to Dr. U. Curth, PD Dr. C. Urbanke (Biophysikalische Chemi, Medizinische Hochschule Hannover), Dr. Johann de Vries and Prof. Dr. W. Wackernagel (Oldenburg University, Oldenburg) who have constantly supplied enough protein (SSBs) for crystallisation without which this work would not have been possible.

I am thankful to Dr. M. S. Weiss and Dr. G. Palm for valuable discussions and suggestions throughout the execution of this work. I am especially grateful to Dr. Weiss and Dr. D. Pal for providing valuable inputs which have immensely facilitated my thesis writing effort.

I thank Dr. J. Mesters for his generous support.

A special mention for my colleagues Dr. A. Riboldi-Tunnicliffe, T. Hogg, A. Rau and other members of our laboratory for their encouragement and generous help. They were instrumental in keeping high spirits and a vibrant environment in the laboratory at all times.

Without the support of my wife (Bandana) and her forbearance, this work would not have been such enjoyable. She has intangible, but a deep contribution in facilitating this work to its conclusion.

Last but not the least, I would like to accredit this achievement to my parents who despite their physical absence had been of constant mental strength and inspiration during the entire course of this endeavour.

Presentations (Oral and poster) and publications

A. Oral Presentation: "Crystallographic studies of single-stranded DNA-binding proteins" at

1. 3rd Heart of Europe Meeting on Bio-Crystallography, Poznan, (Poland) 28th - 30th September 2000.

B. Oral Presentation: "Novel cryo-cooling techniques" at

1. 9th DGK Annual Meeting, Bayreuth (Germany), 12-15th March, 2001 (Invited talk).
2. Institut für Kristallographie, Freie Universität Berlin (Germany), 2nd October, 2000 (Invited talk and techniques demonstration).
3. 19th European Crystallographic Meeting (ECM), Nancy (France), 25th-31st August, 2000 (Invited talk).
4. 8th International Conference on the Crystallisation of Biological Macromolecules (ICCBM), Sandestin, Florida (USA), 14-19th May 2000.
5. Advance Methods in Protein Crystallisation, IMB, Jena (Germany), 10-11th December 1999
6. Max-Delbrück-Center for Molecular Medicine, Department of Crystallography, Berlin (Germany), 25-26th November 1999. (Invited talk and techniques demonstration).
7. 2nd Heart of Europe Meeting on Bio-Crystallography, Luebben, Berlin (Germany), Sep 1999.

C. Poster presentation: "Crystallographic studies of single-stranded DNA-binding proteins" at

1. 9th DGK Annual Meeting, Bayreuth (Germany), 12-15th March, 2001.
2. 18th European Crystallographic meeting, Prague, 15-20th August, 1998.

D: Publications

1. S. Panjekar and R. Hilgenfeld, (2001). Turning non-diffracting protein crystals into diffracting ones by repeated-annealing in paraffin oil. In preparation.
2. S. Panjekar and R. Hilgenfeld, (2001). PanjellyTM: a new cryoprotectant for macromolecular crystallography. In preparation.
3. S. Panjekar, U. Curth, C. Urbanke and R. Hilgenfeld, (2001). Crystal structure of the single-stranded DNA-binding protein from *Brucella abortus*. In preparation.
4. S. Panjekar, U. Curth, J. de Vries, C. Urbanke, W. Wackernagel and Rolf Hilgenfeld, (2001). Crystal structures of the single-stranded DNA-binding proteins from three bacteria. In preparation.

E: product

PanjellyTM cryo-kit: Marketed by JenaBioScience GmbH Biotechnologie and Gentechnik, Löbstedter Straße 78, 07749 Jena Germany (www.jenabioscience.com).

Lebenslauf

Santosh Panjekar

- 24.03.1973 geboren in Banka, Bihar (Indien).
Eltern: Sri Makeswar Panjekar und Srimati Ranjana Devi.
- 1977 - Juni 1987 Abschluß an der „R.M.K. high school“, Banka, Bihar (Indien).
- 1987 - Juni 1989 Abitur am A. N. Kollege, Patna, Bihar (Indien).
- 1989/90 - Juni 1993 B.Sc. Studium der Mathematik, Physik und Geologie an der Banaras Hindu Universität, UtterPradesh (Indien).
- Juli 1993 - Juli 1995 M.Sc. Studium der Biotechnologie an der Devi Ahilya Universität, Indore, MadhyaPradesh (Indien).
- Januar 1995- Juni 1995 Anfertigung einer „Master Thesis“ mit dem Titel “Spacial neighbour analysis of proteins” unter der Leitung von Prof. Saraswathi Visweshwara, Molekulare Biophysik und Dr. Margaret Biswas, Bioinformatik Zentrum am IISc, Bangalore (Indien).
- Oktober 1995 - Oktober 1997 Wissenschaftliche Hilfskraft bei Prof. Dr. Rudolf Reichelt am Institut für Medizinische Physik und Biophysik, Münster (Deutschland).
- November 1997- April 2001 Promotionsstudium und Anfertigung der vorliegenden Dissertation in der Arbeitsgruppe Strukturbiologie unter der Leitung von Prof. Dr. Rolf Hilgenfeld am Institut für Molekulare Biotechnologie, Jena (Deutschland).
- Seit April 2001 Postdoktorand an der EMBL Hamburg Aussenstation in der Gruppe von Dr. Paul A. Tucker.

Februar 2002

Santosh Panjekar

Selbständigkeitserklärung

Ich erkläre, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt habe.

Februar 2002

Santosh Panjekar

Erklärung zur Bewerbung

Ich erkläre, daß ich mich mit der vorliegenden Arbeit an keiner anderen Hochschule um den akademischen Grad doctor rerum naturalium beworben habe und, daß ich weder früher noch gegenwärtig die Eröffnung eines Verfahrens zum Erwerb des o. g. akademischen Grades an einer anderen Hochschule beantragt habe.

Februar 2002

Santosh Panjekar