

**Characterization of *Sso*SSB, *Sso*1450,
*Sso*2001 Proteins and Analysis of CRISPR
and *cas* Genes from *Sulfolobus solfataricus* P2**

Dissertation

zur Erlangung des Grades Doktor der Naturwissenschaften

- Dr. rer. nat.-

der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorlegt von

Dong Han

aus Shandong, China

Bayreuth 2007

Die vorliegende Arbeit wurde in der Zeit von Juni 2001 bis März 2007 am Lehrstuhl für Biochemie der Universität Bayreuth unter der Leitung von Herrn Prof Dr. Gerhard Krauss angefertigt.

Vollständiger Abdruck der von der Fakultät Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zu Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.).

Promotionsgesuch eingereicht am : January 16, 2008

Tag des wissenschaftlichen Kolloquiums: May 07, 2008

Prüfungsausschuss:

Prof. Dr. Gerhard Krauss (Erster Gutachter)

Prof. Dr. Wolfgang Schumann (Zweiter Gutachter)

Prof. Dr. Carlo Unverzagt (Vorsitzender)

Prof. Dr. Franz Meußdoerfer

Table of contents

TABLE OF CONTENTS	I
ABBREVIATIONS	V
1. INTRODUCTION	1
1.1 ARCHAEA <i>SULFOLOBUS SOLFATARICUS</i> P2 STRAIN.....	1
1.1.1 <i>Archaea: ubiquitous but unique</i>	1
1.1.2 <i>Sulfolobus solfataricus, a model system in crenarchaea</i>	3
1.2 SSBs.....	4
1.2.1 <i>General introduction of SSBs</i>	4
1.2.2 <i>Bacterial and human mitochondrial SSBs</i>	5
1.2.3 <i>Replication protein A (RPA), the eukaryotic SSBs</i>	6
1.2.4 <i>Archaeal SSB, the ancient SSB?</i>	8
1.2.5 <i>Other SSBs</i>	10
1.3 CRISPR AND CRISPR-ASSOCIATED PROTEINS	11
1.3.1 <i>General introduction of CRISPR</i>	11
1.3.2 <i>CRISPR-associated proteins</i>	12
1.3.3 <i>The biological roles of CRISPRs and Cas proteins</i>	13
1.3.4 <i>CASS in Sulfolobus solfataricus</i>	15
1.3.5 <i>Prospect</i>	17
1.4 AIM OF THE PRESENT WORK.....	18
2. MATERIALS AND METHODS	20
2.1 MATERIALS	20
2.1.1 <i>Chemicals, enzymes and proteins</i>	20
2.1.1.1 <i>Chemicals</i>	20
2.1.1.2 <i>Enzymes and proteins</i>	20
2.1.2 <i>Bacterial and archaeal strains, media and antibiotics</i>	21
2.1.2.1 <i>Bacterial strains</i>	21
2.1.2.2 <i>Media, inducer and antibiotics</i>	21
2.1.3 <i>Plasmids and phage</i>	21
2.1.4 <i>Oligonucleotides and tRNAs</i>	22
2.1.4.1 <i>PCR primers</i>	22
2.1.4.2 <i>Primers for mutation</i>	23
2.1.4.3 <i>Substrates for SsoSSB</i>	23
2.1.4.4 <i>Substrates for nuclease assay</i>	24
2.1.4.5 <i>Substrates for Sso1450C6H</i>	24
2.1.5 <i>Buffers and solutions</i>	25
2.1.6 <i>Commercial kits</i>	27
2.1.7 <i>Instruments and materials</i>	27
2.1.8 <i>Chromatographic materials</i>	28
2.1.9 <i>Softwares</i>	28
2.2 STANDARD METHODS	28
2.2.1 <i>Spectrophotometric determination</i>	28
2.2.1.1 <i>Determination of protein concentration</i>	28
2.2.1.2 <i>Determination of nucleic acid concentration</i>	29
2.2.1.3 <i>Determination of bacterial cell density</i>	29
2.2.2 <i>Gel electrophoresis</i>	29
2.2.2.1 <i>Agarose gel electrophoresis</i>	29
2.2.2.2 <i>Native polyacrylamide gel electrophoresis</i>	30
2.2.2.3 <i>Denaturing polyacrylamide gel electrophoresis</i>	30
2.2.2.4 <i>SDS polyacrylamide gel electrophoresis (SDS PAGE)</i>	30
2.2.3 <i>Detection of radioactively labeled nucleic acids in gel</i>	31

Table of contents

2.2.4	Detection of unlabeled nucleic acids in gel.....	31
2.2.5	Staining of protein gels.....	31
2.3	MOLECULAR BIOLOGY METHODS.....	31
2.3.1	Preparation and transformation of competent cells.....	31
2.3.2	Culture of bacterial strains.....	32
2.3.3	Extraction of nucleic acids with phenol:chloroform.....	33
2.3.4	Nucleic acid precipitation by ethanol.....	33
2.3.5	Dephosphorylation of nucleic acid.....	33
2.3.6	5'-end labeling of oligonucleotides by [γ - ³² P]-ATP.....	33
2.3.7	Hybridization of oligonucleotides.....	34
2.3.8	Small scale preparation of plasmid DNA.....	34
2.3.8.1	Growth of the bacterial culture.....	34
2.3.8.2	Cell harvest and plasmid DNA purification.....	34
2.3.9	Preparation of bacteriophage M13 ssDNA.....	35
2.3.9.1	Infection of bacterial <i>E.coli</i> XL1-bule cells by bacteriophage M13 ssDNA and culture of the cells ..	35
2.3.9.2	Harvest of M13 ssDNA.....	35
2.3.10	Polymerase chain reaction (PCR).....	35
2.3.10.1	Gene amplification from <i>Sulfolobus solfataricus</i> (<i>Sso</i>) P2 genomic DNA.....	36
2.3.10.2	Gene amplification from plasmid vector.....	36
2.3.10.3	Colony PCR.....	36
2.3.10.4	PCR product purification.....	36
2.3.10.5	Point mutation by quick change method.....	37
2.3.11	DNA digestion by restriction endonucleases.....	37
2.3.12	DNA ligation.....	37
2.4	COMPUTATIONAL ANALYSIS OF GENES FROM <i>Sso</i> P2 STAIN.....	38
2.4.1	Homology analysis.....	38
2.4.2	Gene location and operon analysis in <i>Sso</i> P2.....	38
2.5	PREPARATION OF PROTEINS.....	39
2.5.1	Overexpression and purification of <i>SsoSSB</i>	39
2.5.2	Solubility analysis of Proteins from <i>Sso</i> P2 expressed in <i>E.coli</i>	40
2.5.2.1	Solubility analysis of single <i>Sso</i> putative repair proteins during small-scale expression.....	40
2.5.2.2	Solubility analysis of <i>Sso</i> proteins in a small-scale coexpression system.....	41
2.5.3	Protein refolding.....	41
2.5.3.1	Protein purification under denaturing conditions.....	41
2.5.3.2	On-column refolding assay.....	42
2.5.3.3	Rapid dilution refolding assay.....	42
2.5.3.4	Dialysis refolding assay.....	43
2.5.3.5	High-throughput refolding assay.....	43
2.5.3.6	Co-refolding assay.....	44
2.5.4	Western Blotting.....	44
2.5.5	Preparation of <i>Sso2001Est</i> fusion protein.....	45
2.5.5.1	Cloning, expression and purification of <i>Sso2001Est</i> fusion protein.....	45
2.5.5.2	Esterase activity detection assay.....	46
2.5.6	Preparation of <i>Sso1450C6H</i> fusion protein.....	47
2.6	NUCLEIC ACID-PROTEIN INTERACTION.....	48
2.6.1	Electrophoretic mobility shift assay (EMSA).....	48
2.6.1	Nuclease assay.....	48
2.6.2	ATPase assay (thin layer chromatography assay).....	48
2.6.3	Fluorescence anisotropy assay.....	49
2.6.3.1	Fluorescence anisotropy measurements.....	50
2.6.3.2	Competition titration.....	50
2.6.3.3	Data analysis.....	50
2.6.4	Atomic force microscopy (AFM).....	51
2.6.4.1	DNA substrate for AMF.....	52
2.6.4.2	AFM measurements.....	53
3.	RESULTS.....	54
3.1	DNA BINDING PROPERTIES OF SINGLE-STRANDED DNA BINDING PROTEIN FROM <i>Sso</i> P2 (<i>SsoSSB</i>) ..	54

3.1.1 Overexpression of <i>ssoSSB</i> gene	54
3.1.2 Purification of <i>SsoSSB</i> protein.....	54
3.1.3 Characterization of <i>SsoSSB</i>	55
3.1.3.1 Gel retardation of DNA-protein complex.....	56
3.1.3.2 Binding complex detection of <i>SsoSSB</i> by AFM	58
3.1.3.3 Stoichiometry of <i>SsoSSB</i> -ssDNA complexes as followed by fluorescence anisotropy.....	60
3.1.3.4 Dissociation constant of <i>SsoSSB</i> -ssDNA complexes.....	61
3.2 COMPUTATIONAL ANALYSIS OF THE PUTATIVE, DNA REPAIR RELATED, SPECIFICALLY CONSERVED GENE CLUSTERS IN <i>SULFOLOBUS SOLFATARICUS</i> P2	63
3.2.1 Homology analysis.....	63
3.2.2 Operon analysis	65
3.3 EXPRESSION SCREENING AND FUNCTIONAL SCANNING OF PUTATIVE <i>Sso</i> PROTEINS IN <i>E. COLI</i>	67
3.3.1 Expression screening and solubility detection	67
3.3.2 Further soluble expression, refolding efficiency and functional screening	68
3.4 EXPRESSION, PURIFICATION AND CHARACTERIZATION OF <i>Sso2001</i> PROTEIN FROM <i>Sso</i> P2 WITH ESTERASE AS A FUSION PARTNER	73
3.4.1 Gene amplification from <i>Sulfolobus solfataricus</i> P2.....	73
3.4.2 <i>Sso2001</i> fusion protein gene expression	73
3.4.3 Purification of <i>Sso2001Est</i> fusion protein	74
3.4.4 Nuclease activity detection.....	76
3.4.5 Reaction condition investigation.....	77
3.4.5.1 Temperature	77
3.4.5.2 Optimal pH.....	77
3.4.5.3 Effect of bivalent metal ions.....	78
3.4.6 Protein active site determination by mutants	79
3.4.7 Substrate specificity of nuclease activity.....	81
3.4.8 Steady state kinetics of nuclease activity.....	84
3.5 EXPRESSION, PURIFICATION AND CHARACTERIZATION OF <i>Sso1450C6H</i> FUSION PROTEIN FROM <i>Sso</i> P2	86
3.5.1 Protein gene expression	86
3.5.2 Protein purification.....	87
3.5.3 Binding of <i>Sso1450C6H</i> on nucleic acid substrates	88
3.5.4 Binding mode of <i>Sso1450C6H</i> on nucleic acid substrates	90
3.5.5 Renaturation activity of <i>Sso1450C6H</i> for denatured dsDNA.....	91
4. DISCUSSION.....	93
4.1 CHARACTERIZATION OF <i>SsoSSB</i> PROTEIN	93
4.1.1 <i>SsoSSB</i> protein binds single-stranded DNA with high affinity.....	93
4.1.2 <i>SsoSSB</i> protein is a monomer in solution.....	94
4.1.3 Binding mode of <i>SsoSSB</i> to ssDNA	94
4.1.4 The binding site size and DNA-protein interaction.....	96
4.1.5 Does <i>SsoSSB</i> represent the ancestral SSB of three domains?.....	97
4.2 COMPUTATIONAL ANALYSIS AND EXPRESSION SCREENING OF THE PUTATIVE PROTEIN GENES FROM <i>Sso</i> P2	98
4.2.1 Scanning and collecting putative protein genes in <i>Sso</i> P2.....	98
4.2.2 Do these genes represent a novel repair system?	99
4.2.3 Operon analysis of putative genes in <i>Sso</i> P2 genome.....	100
4.2.4 <i>cas</i> gene expression, <i>Cas</i> protein refolding and enzymatic activity detection	101
4.3 EXPRESSION AND CHARACTERIZATION OF <i>Sso2001EST</i> PROTEIN	103
4.3.1 Soluble expression of <i>Sso2001</i> with esterase	103
4.3.2 <i>Sso2001</i> is as an endonuclease	104
4.3.3 Characterization of <i>Sso2001</i> nuclease.....	105
4.3.4 Is <i>Sso2001</i> a HD-superfamily enzyme?.....	106
4.4 CHARACTERIZATION OF <i>Sso1450C6H</i>	108
4.4.1 Nonspecific binding of <i>Sso1450C6H</i> protein to nucleic acid.....	108
4.4.2 <i>Sso1450C6H</i> promotes annealing of complementary DNA strands.....	108
4.4.3 What might be the role of <i>Sso1450</i> in vivo?.....	109

Table of contents

5. SUMMARY	111
6. ZUSAMMENFASSUNG.....	113
7. REFERENCES	115
8. ACKNOWLEDGEMENT	130
ERKLÄRUNG	131

Abbreviations

A	in DNA, Adenosine; in protein, Alanine
AcOH	acetic acid
AFM	Atomic Force Microscopy
Amp	Ampicillin
Approx.	approximately
APS	Ammoniumperoxodisulfate
Arg	Arginine
ATP	Adenosine-5'-triphosphate
ATPase	Adenosine-5'-triphosphatase
bp	base pair (s)
BRE	Transcription Factor B recognition element
	Base excision repair
BSA	Bovine Serum Albumin
C	Cytidine
cas	CRISPR-associated
Cass	CRISPR-associated system
CHES	2-(N-Cyclohexylamino)ethane Sulfonic Acid
Ci	Curie (1 Ci=3.7x10 ¹⁰ Bq)
COG	Clusters of orthologous groups
CRISPR	Clustered regularly interspaced short palindromic repeats
D	Aspartic acid
dATP	2'-deoxyadenosine-5'triphosphate
dCTP	2'-deoxycytidine-5'triphosphate
ddH ₂ O	Double distilled water
DNA	Deoxyribonucleic acid
dGTP	2'-deoxyguanosine-5'triphosphate
DMF	N,N-dimethyl foramide
dNTPs	mixture of dATP, dCTP, dGTP and dTTP
dsDNA	double-stranded Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	2'-deoxythymidine-5'triphosphate
E	Glutamic acid
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetate
EMSA	Electrophoretic mobility shift assay
FA	Fluorescence Anisotropy
FPLC	Fast performance liquid chromatography
g	gram, gravity acceleration in centrifugation
G	Guanosine
GSH/GSSG	reduced/oxidized forms of Glutathione
gp32	T4 gene32 protein
H	Histidine
h	hour (s)

Abbreviations

HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HTG	Horizontally transferred genetic
IPTG	Isopropyl- β -D-thiogalactoside
Kd	Dissociation constant
kDa	Kilodalton
l	litre (dm ³)
LB	Luria-Bertani medium
LUCA	last universal common ancestor
M	molar concentration (mol/l), molecular weight marker
mA	milli Ampere
MCM	mini-chromosome maintenance proteins
MES	2-(N-morpholino)ethanesulfonic acid
mg	milligram
MSH	mercaptoethanol
min	minute (s)
N	any nucleotide
NDSB	Non-Detergent Sulphobetaines
Ni-NTA	Ni ²⁺ -nitriloacetic acid
nm	nanometer
nt	nucleotide (s)
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylenglycol
PMSF	Phenylmethylsulfonylfluoride
psiRNA	prokaryotic siRNA
r	Anisotropy
R	purine in DNA
RAMP	Repeat-Associated Mysterious Proteins
RNA	Ribonucleic acid
RNAi	RNA interference
RPA	Replication protein A
rpm	revolution per minute
S	Serine
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecylsulfate
siRNA	Small interfering RNA
SSB	Single strand binding protein
ssDNA	Single-stranded Deoxyribonucleic acid
<i>Sso</i>	<i>Sulfolobus solfataricus</i>
T	Thymine
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFIIB	Transcription Factor II B

TFK	Trifluoromethyl ketone
TLC	Thin layer chromatography
Tris	Tris-(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
UNG2	uracil-DNA glycosylase
UV	Ultraviolet light
V	voltage
v/v	volume per volume
W	A or T in DNA
w/v	weight per volume
XPA	Xeroderma pigmentosum group A
XPC	Xeroderma pigmentosum group C
X-gal	5-bromo-4-chloro-indolyl- β -D-galactoside
Y	pyrimidine in DNA

1. Introduction

1.1 Archaea *Sulfolobus solfataricus* P2 strain

1.1.1 Archaea: ubiquitous but unique

The domains Archaea, Eukarya and Bacteria represent three distinct phylogenetic lineages that encompass all known life on earth (Woese et al., 1990). Archaea are the last one which was recognized as a distinct phylogenetic group about 30 years ago (Woese and Fox, 1977). When these microorganisms were first discovered in 1977, they were considered to be bacteria. However, when their 16S rRNAs were analyzed, it became clear that there was no close relationship to the bacteria. In fact, they were more closely related to the eukarya (Gutell et al., 1985), although they share some features with bacteria, as both are prokaryotes. Archaea are single cell forms with bacteria-like cell walls and flagella. They have circular chromosomes and lack a nuclear membrane. Based on 16S rRNA analysis, the archaeal domain is split into four subdomains: the euryarchaea, the crenarchaea, the korarchaea and the nanoarchaea (Figure 1.1).

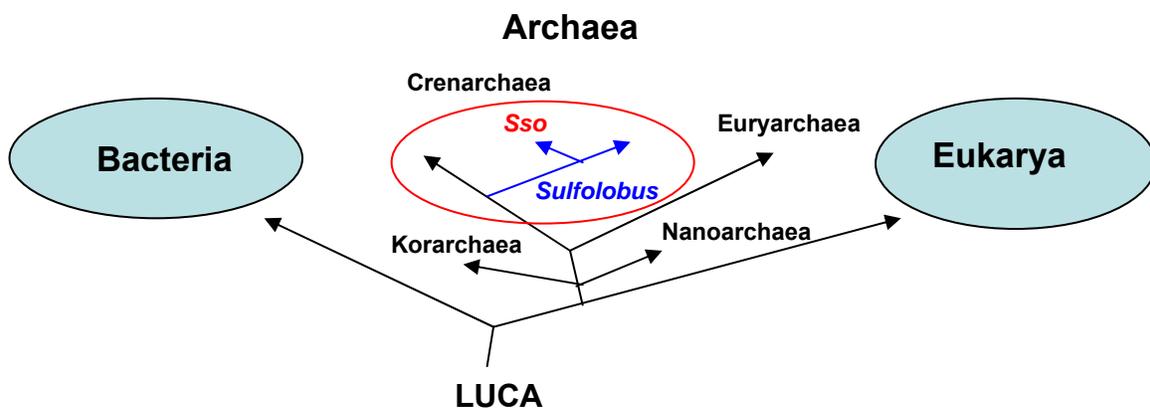


Figure 1.1: Phylogenetic tree of three domains of life. All cellular life on earth can be classified as belonging to one of three domains: Eukarya, Bacteria and Archaea. The archaea are split into four subdomains. *Sulfolobus* (blue lines) is a branch of the Crenarchaea subdomain (red circle) in Archaea, where *solfataricus* (*Sso*) is a twig on it. LUCA: last universal common ancestor.

Archaea are known to live in extreme conditions, like high temperature, often above 100°C. They are found in hot springs, black smokers or oil wells. Some are found in very cold places and others in highly salty, acidic or alkaline water. Whereas, it is now apparent that archaea are ubiquitous, they exist not only in harsh environments, but are also present in normal habitat conditions (Rother and Metcalf, 2005).

Besides the environments they inhabit, Archaea do share unique features not typically found in both other domains. Although their morphology resembles bacteria and unicellular eukaryotes, archaea possess glycerol-based phospholipids with own characters. First, in the archaeal lipids, the stereochemistry of the glycerol moiety is the reverse of that found in bacteria and eukaryotes. This is strong evidence for a different biosynthetic pathway. Second, most bacteria and eukaryotes have membranes composed mainly of glycerol-ester lipids, whereas archaea have membranes composed of glycerol-ether lipid, including mesophilic archaea. Even when bacteria have ether-like lipids, the stereochemistry of the glycerol is of the bacterial type. Third, archaeal lipids are based on isoprenoid building blocks. This is common in rubber and as a component of some bacterial and eukaryotic vitamins. However, only the archaea incorporate these compounds into their cellular lipids (White, 1995; Delong and Pace, 2001; van, V et al., 1998; White, 1995). The archaeal cell wall and flagella are also unusual (Howland, 2000). Another example is the histone protein. Archaea and eukaryotes have histones, whereas these are not found in bacteria (Caetano-Anolles and Caetano-Anolles, 2003). Archaeal histones possess a minimal histone fold structure and show a higher flexibility (Sandman and Reeve, 2005; Decanniere et al., 2000; Li et al., 2003). In the presence of DNA, archaeal histone dimers aggregate further to form tetramers or hexamers (Marc et al., 2002). By contrast, the eukaryotic nucleosome core histones have additional sequences that extend N- and/or C-terminal from their histone fold (Sullivan et al., 2002; Luger et al., 1997b). Furthermore, eukaryotic histones have four conserved subunits that assemble into the histone octamer (Luger et al., 1997a).

Individual character comparison of archaeal, bacterial and eukaryotic cells reveals that, despite of their unique features, archaea are similar to other prokaryotes in most aspects of cell structure and metabolism. However, their genetic processes do not show many typical bacterial features, and are in many aspects similar to those of eukaryotes

(Bell and Jackson, 1998; Bell, 2005; Geiduschek and Ouhammouch, 2005; Myllykallio et al., 2000). Some of the eukaryotic and bacterial traits of archaea are collected in Table 1.1.

Table 1.1: Summary of some bacterial and eukaryotic traits of archaea.

Eukaryotic traits	Bacterial traits
DNA replication machinery Histones Nucleosome-like structures Transcription machinery RNA polymerase TFIIB TATA-binding protein (TBP) Translation machinery Initiation factors Ribosomal proteins Elongation factors Poisoned by diphtheria toxin	Single, circular chromosome Operons Bacterial-type membrane transport channels Many metabolic processes Energy production Nitrogen-fixation Polysaccharide synthesis

There are many interesting and exciting features that make the archaea an attractive domain to study. The simplicity of its eukarya-like metabolism machinery provides an opportunity for investigating the eukaryotic machinery in a simple way. Its mixed features of eukarya and bacteria narrow the gap between these two domains and promote phylogenetic analysis. Its tolerance of harsh environments (for instance, temperature, pressure, salinity and pH) is interesting for industrial purposes.

1.1.2 *Sulfolobus solfataricus*, a model system in crenarchaea

Among the archaea, the genus *Sulfolobus* has been well studied. It comprises different strains isolated from acidic, solfataric fields all over the world. They grow at temperatures between 60°C and 95°C and at pH of 1 to 5. They are Gram-negative, irregularly shaped. Most *Sulfolobus* strains are able to gain their energy by oxidizing sulfide to sulfate, and many strains are able to oxidize ferrous iron (Brock, 1978). These organisms, belonging to the phylum Crenarchaea (Figure 1.1) have been chosen as a model system for biochemical and genetic studies for several reasons: they can be easily grown on appropriate liquid and solid media both as single colonies and as lawn; they are aerobic and thermophilic; gene transfer can occur by conjugation, transduction and

transformation. Moreover, various genomes have been completely sequenced (She et al., 2001; Chen et al., 2005; Kawarabayasi et al., 2001).

In the present work, the investigation focused on *Sulfolobus solfataricus* (*Sso*) P2 strain that had been completely sequenced previously (She et al., 2001; Kawarabayasi et al., 2001). *Sso* P2 is an aerobic crenarchaeon that grows optimally at 80°C and pH 2- 4. Its genomic DNA contains about 3×10^6 bp, encoding about 2900 proteins, 33% of which are found only in *Sulfolobus*, whereas 40% have homologs in archaea, 12% have homologs in bacteria but not in eukarya, and 2.3% in eukarya but not in bacteria. 25% are shared with both bacteria and eukarya. It is the most widely studied organism of the crenarchaeal branch of the archaea and a model for research on the mechanisms of DNA replication, repair, the cell cycle, chromosomal integration, transcription, RNA processing, and translation. These further studies reveal common features in archaea, even between three domains (Ciaramella et al., 2002; Kawarabayasi et al., 2001). These are expected to contribute new discoveries in the near future.

1.2 SSBs

1.2.1 General introduction of SSBs

Single-stranded DNA-binding proteins (SSBs) are indispensable elements in all living organism cells. They have little in common at the protein sequence level and subunit composition, but more at the functional and structural levels. The common structural feature is a conserved domain called an oligonucleotide/oligosaccharide-binding (OB) fold that binds single-stranded DNA (ssDNA) (Murzin, 1993). OB folds bind ssDNA in a cleft formed primarily by β -strands, by using aromatic residues that stack against nucleotide bases, and positively charged residues that form ionic interactions with the DNA backbone (Bochkarev et al., 1997; Raghunathan et al., 2000; Shamoo et al., 1995; Matsumoto et al., 2000). SSBs are usually present in stoichiometric quantities with the corresponding ssDNA substrates, and protect the transiently formed ssDNA regions against nuclease attack, and they prevent the formation of secondary structures (Perales et al., 2003). In this way, SSBs participate in many aspects of nucleic acid metabolism, including DNA replication, recombination, repair, chromosome maintenance and

transcription, although the function details are not fully understood (Aravind et al., 1999; De et al., 2004; Aravind and Koonin, 1999; Komori and Ishino, 2001; Carpentieri et al., 2002; Dionne et al., 2003).

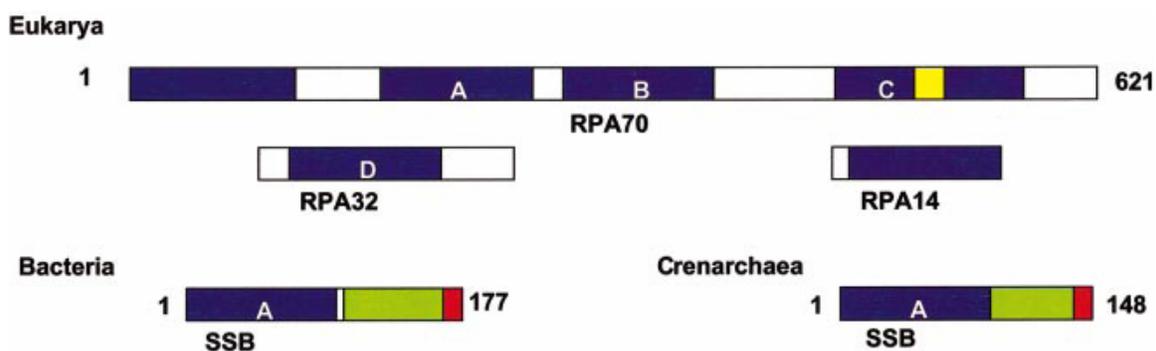


Figure 1.2: Domain organization of SSB proteins in eukarya, bacteria and archaea. Blue rectangles represent the OB folds; A, B, C and D represent the OB folds participating ssDNA binding events; the red rectangles represent C-terminal domains; the yellow rectangle represents zinc-finger.

1.2.2 Bacterial and human mitochondrial SSBs

Most bacterial SSBs, as well as mitochondrial SSBs (mtSSBs) have similarities in sequence and structure. They are monomers. Each monomer comprises single N-terminal ssDNA binding domain, OB fold, and a less structured C-terminal tail. They assemble into active homotetramers in the absence of DNA (Figure 1.2B). Among these proteins, *E. coli* SSB (EcoSSB) has been most extensively studied. Its crystal structure in the presence and absence of ssDNA has been solved and the key aromatic residues in the OB fold have been defined (Raghunathan et al., 2000; Raghunathan et al., 1997) (Figure 1.3A). Interestingly, *EcoSSB* binds ssDNA cooperatively in two modes: (SSB)₃₅ in which, the ssDNA interacts with two protomers in the tetramer, and (SSB)₆₅, in which, the ssDNA interacts with all four protomers in the tetramer (numbers represent the binding site size on ssDNA substrates), depending upon the monovalent salt concentration (with distinct effects of both cation and anion types), as well as divalent cations, polyamines, temperature and pH (Lohman and Ferrari, 1994) (Figure 1.3B). The (SSB)₃₅ complex is

formed below 10 mM NaCl whereas (SSB)₆₅ complex forms above 200 mM up to 5 M NaCl. In the range of 10-200 mM, the site size expands continuously with the concentration of NaCl (Lohman et al., 1986). *In vivo* changes in the ionic strength may play an important role in regulating the alternation of these modes during its various functions (Lohman and Overman, 1985).

It is worth to note that the highly conserved, acidic residue-rich C-terminal region of *Eco*SSB is not present in mtSSB. This region is neither essential for DNA binding nor for homotetramer formation, whereas it is required for *in vivo* functions suggesting a role in interactions with other proteins. The region between the N-terminal OB fold and the acidic C-terminus probably functions only as a spacer, keeping the negative charges away from the DNA bound to SSB (Webster et al., 1997).

1.2.3 Replication protein A (RPA), the eukaryotic SSBs

In eukaryotes, RPA acts as a SSB protein. It is heterotrimer possessing three different subunits, RPA70, RPA32 and RPA14. The largest subunit RPA70 contains four OB folds including an N-terminal domain and three DNA binding domains (DBDs, DBD-A, DBD-B and DBD-C). The subunit RPA32 contains DBD-D flanked by N-terminal phosphorylation sites. The small subunit RPA14 folds into an OB structure and plays a role in trimerization (Figure 1.3C). In the heterotrimer of human RPA, four of the six OB folds participate in the DNA binding process, namely DBD-A, -B, -C of RPA70 and DBD-D of RPA32, (Iftode et al., 1999).

Similar to bacterial SSBs, RPA binds ssDNA in two alternative modes along with significantly conformational change, probably depending upon the salt concentration (Pfuetzner et al., 1997). During the binding process, RPA binds first to 8-10 nucleotides (nt) in an unstable manner via DBD-A and B, the major DNA binding domains that harbour most of the binding activity of the full trimer (Blackwell and Borowiec, 1994; Walther et al., 1999). The second binding step is associated with the ssDNA binding of two minor DNA binding domains, DBD-C and D. In this manner, all four DBDs directly contact the ssDNA substrate, occluding a total of 30 nt (Brill and Bastin-Shanower, 1998; Bastin-Shanower and Brill, 2001; Bochkareva et al., 1998; Kim et al., 1992b). Various

factors contribute to the latter step, for instance, the zinc ribbon in DBD-C stabilizes the trimer through a cooperative manner (Bochkareva et al., 2000). The switching from 8-10 nt to the 30 nt mode is mediated by DNA binding of the trimerization core (Bochkareva et al., 2002) (Figure 1.3D). The binding of adjacent trimers occurs with low cooperativity (Kim and Wold, 1995).

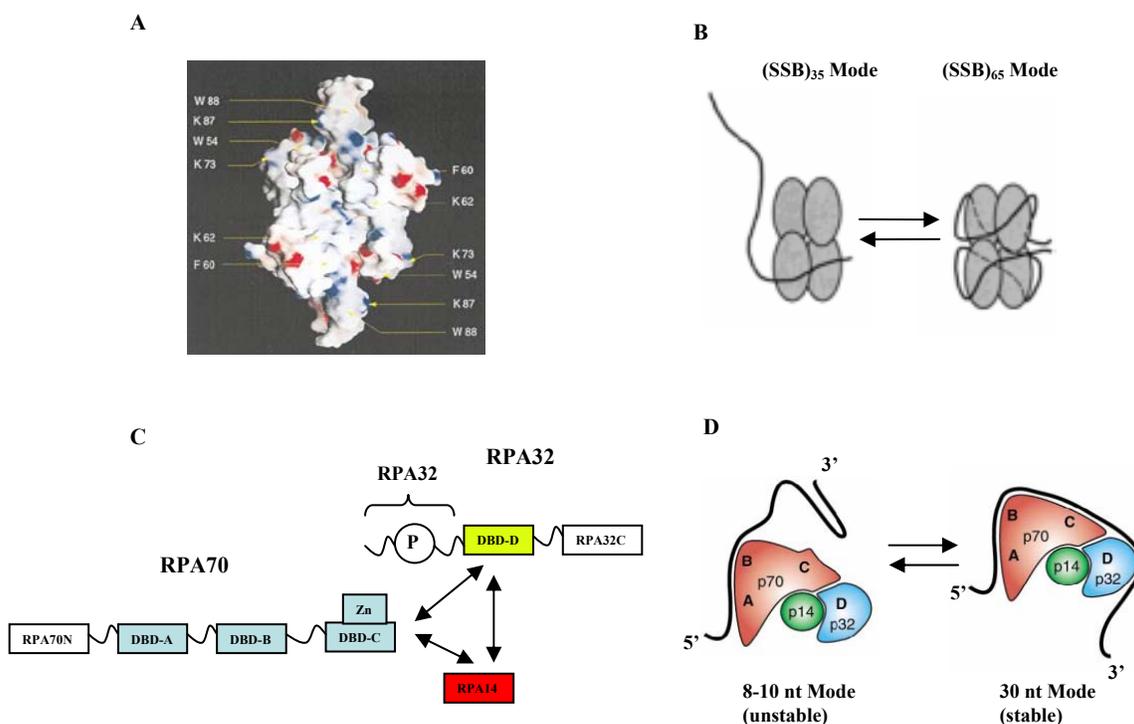


Figure 1.3: Structures and binding modes of bacterial and eukaryotic SSBs. **A**, Protein surfaces involved in binding of *EcoSSB* tetramer to the ssDNA. The surface is colored deep blue in the most positive regions and deep red in the most negative regions. Residues known to be involved in binding are shown by arrows. **B**, Schematic map of the two DNA binding modes of *EcoSSB* tetramer. At low concentration of NaCl (<10 mM), two protomers of *EcoSSB* tetramer interact with the ssDNA to form the (SSB)₃₅ complex, on the contrary, at high concentration of NaCl (>200 mM), the (SSB)₆₅ complex is formed by complete interaction between the ssDNA and all four protomers in the tetramer. Switch between (SSB)₃₅ mode and (SSB)₆₅ mode mostly depends on the salt concentration. **C**, Schematic map of the RPA domain structure. Domains are presented as boxes, their borders are indicated. Zn, the zinc ribbon; P, the unstructured, phosphorylated N-terminus of subunit RPA32. The regions of subunit interaction are indicated by arrows. **D**, Two DNA binding modes of the trimeric RPA molecule (RPA70, RPA32, and RPA14 subunits) with the four DNA binding domains designated as A, B, C, and D. The binding of the ssDNA (thick line) occurs via a multi-step pathway. The initial, unstable 8-nt binding is mediated via domains A and B. A conformational switch then reorients domain C, allowing it (and likely domain D) to make contact with the ssDNA protruding from domain B to attain the stable 30-nt binding mode. The 5 to 3 polarity of DNA engagement by RPA was first reported by de Laat et al..

During DNA-processing, RPA interacts with many nuclear proteins, for example, XPA (Xeroderma pigmentosum group A), XPC (Xeroderma pigmentosum group C), Rad51, p53, DNA polymerase (Iftode et al., 1999). The N-terminus of RPA70 interacts with the tumor suppressor p53 (Lin et al., 1996), and the C-terminal domain of RPA32 with uracil-DNA glycosylase (UNG2) in base excision repair (BER) (Otterlei et al., 1999). It is suggested that DNA binding and protein interaction of RPA are regulated by each other in dynamic way by competing for the same binding sites, but the interaction details remain unclear (Bochkareva et al., 2001).

1.2.4 Archaeal SSB, the ancient SSB?

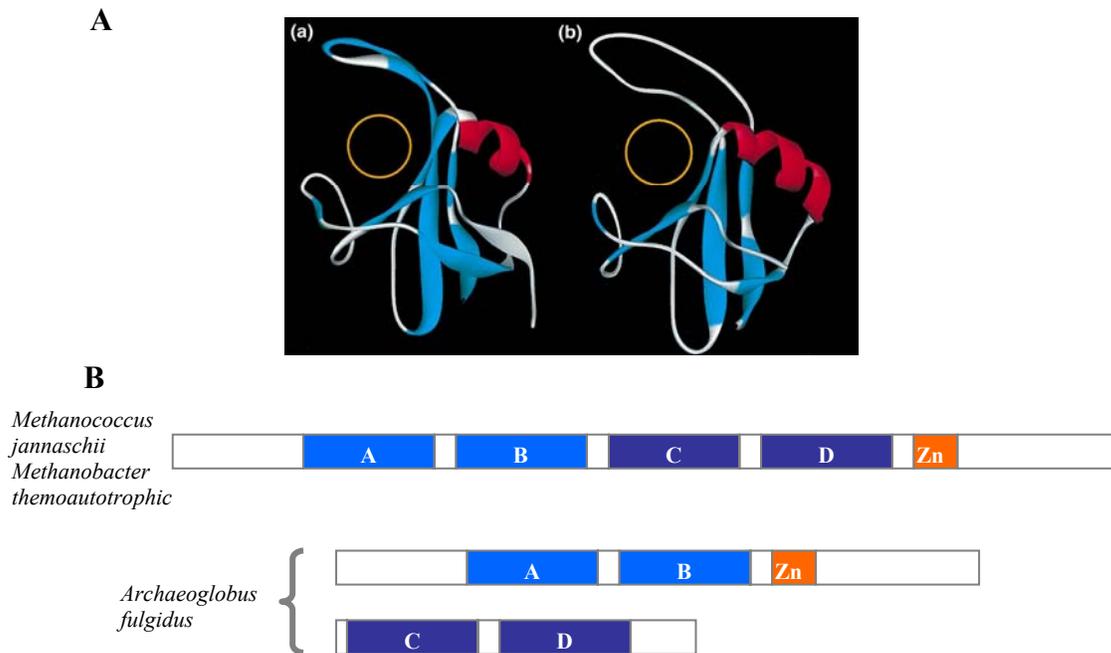


Figure 1.4: A, Structures of the ssDNA-binding domain A between human RPA and Euryarchaeal SSBs. (a) Human RPA70 DBD-A. (b) Euryarchaeal *Methanococcus jannaschii* MJ1159 ssDNA-binding domain A (residues 80–162, OB-A). β -strands are shown in blue; α -helices are shown in red. The structures are shown so that the axis of the channel in which DNA binds is perpendicular to the figure (DNA is shown as an orange circle). **B, Schematic map of OB folds arrangement in euryarchaea.** The upper map represents the OB folds from *Methanococcus jannaschii* and *Methanobacter thermoautotrophicum*, and the lower one, from *Archaeoglobus fulgidus* (within two subunits). Capital letters, A, B, C and D, represent the OB folds; Zn, the zinc finger.

SSBs also exist in the third domain of life, the archaea. Two types of archaeal SSBs have been identified, the euryarchaeal SSBs and crenarchaeal SSBs. SSBs in euryarchaea have a eukaryote-like composition. They contain four OB folds in one gene product or in two. Although there are some variations on the sequence level, euryarchaeal SSBs retain structural similarity to eukaryotic SSBs (Figure 1.4A). The four OB folds all participate in the ssDNA binding event. The zinc finger, which has a counterpart in eukaryotic RPA, is highly conserved and correctly spaced (Chedin et al., 1998) (Figure 1.4B). Therefore, the euryarchaeal SSBs can be treated as OB fold hetrotetramers, which function through the cooperation of the four OB folds.

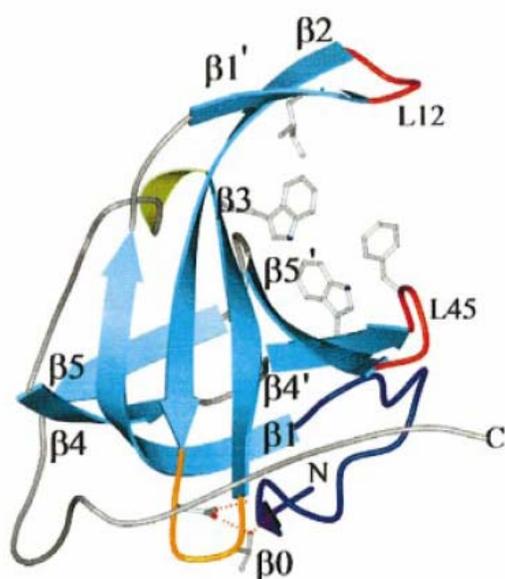


Figure 1.5: The *Sso*SSB monomer. The four aromatic residues suggested in the ssDNA binding in *Sso*SSB are (clockwise from top) Ile30, Phe79, Trp75 and Trp56, coloured in gray. The protein secondary structure is labeled as ribbon-stick mode. The L12 and L45 loops for ssDNA-binding event are coloured red, the L23 loop orange, the capping ‘helical’ region green and the N-terminus royal blue.

All types of OB folds from bacteria, eukarya and archaea show the conservation in structure and key residues. These findings strongly suggest that all SSBs from three domains of life originated from a common ancestral ssDNA-binding protein, and afterwards diverged through evolution. An important question is which of the SSBs comes closest to the ancient SSB ancestor? The answer may lay in the SSB from crenarchaea. *Sulfolobus solfataricus* SSB (*Sso*SSB), the first SSB from crenarchaea, contains a single eukaryote-like OB fold in a monomer, both on sequence and structure level, and a bacteria-like C-terminal tail, without zinc finger in the whole sequence (Wadsworth and White, 2001; Haseltine and Kowalczykowski, 2002). Its crystal structure

indicates the highly conserved key residues in functional loops (Kerr et al., 2003) (Figure 1.5). The oligomer state in the presence/absence of ssDNA in solution is still a matter of debate and its binding mode remains unclear. Many studies have investigated its relationship to mini-chromosome maintenance (MCM) proteins (Carpentieri et al., 2002; Marsh et al., 2006), RNA polymerase (Richard et al., 2004), gyrase (Napoli et al., 2005) and DNA damage detection (Cubeddu and White, 2005) implying the important role of C-terminus of *Sso*SSB in these protein-protein interaction processes, and suggesting the crucial function of *Sso*SSB in DNA replication, transcription, recombination and repair. Taken together, the mixed features of *Sso*SSB from its bacterial and eukaryotic homologues show that *Sso*SSB can be a potential candidate for representing the evolutionary convergence of SSB protein family between three domains of life.

1.2.5 Other SSBs

The properties of some of the typical SSBs mentioned suggest that the evolutionary pathway of SSB protein family is varied. Bacterial phage T4 gene32 protein (gp32) is the first SSB protein to be studied biochemically and biophysically in details. It still can be served as a paradigm for SSBs. Gp32 binds ssDNA mainly as a dimer with high cooperativity of the “unlimited” type that allows the formation of continuous protein clusters that can readily saturate the ssDNA (Williams and Konigsberg, 1978). Some bacterial SSBs from thermophilic species are homodimers, with each monomer encoding two OB folds. The C-terminal domain has nearly all of the key residues binding the ssDNA as that in *Eco*SSB-ssDNA model. But the N-terminal OB fold does not retain numbers of potentially important ssDNA-binding residues. The differences between these two ssDNA-binding domains impose an asymmetry that is likely to affect the DNA binding properties and other functions of each domain (Bernstein et al., 2004). The finding of two OB folds linked by a conserved spacer sequence (such as *Thermus thermophilus* and *Deinococcus radiodurans* SSB proteins and their counterparts from *Deinococcus-Thermus* genera of bacteria) probably is an adaptation of hosts to extreme conditions (Dabrowski et al., 2002; Eggington et al., 2004; Filipkowski et al., 2007; Filipkowski et al., 2006). A common feature appears to be characteristic for the SSBs,

namely that most of SSBs function as a combination of four OB folds. They are found either as a tetramer with a single OB fold per molecule, as a molecular dimer with two OB folds or as a heterotrimer with OB folds unequally distributed in the subunits.

1.3 CRISPR and CRISPR-associated proteins

1.3.1 General introduction of CRISPR

Along with the development of genome sequencing, more and more characteristic features of genomes are discovered. Clustered regularly interspaced short palindromic repeats (CRISPRs) are newly described aspects of genome organization. They were first observed by Ishino and colleagues (Ishino et al., 1987) upstream of *iap* gene in *E.coli*. These sequences share unique features and are now considered as a new family of prokaryotic repeats that is easily distinguishable from any other recurrent motifs. The repeats are typically short partially palindromic sequences of 21- 48 bp, containing inner and terminal inverted repeats that are generally spaced by similarly sized non-repetitive sequences, called “spacers”. The sequence of repeated units is conserved in members of the same phylogenetic group, and there is a high percentage of similarity even among domains (Mojica et al., 2000). Later the rapid progress in whole genome sequencing revealed that CRISPRs are present in about half of the bacterial and most archaeal genomes (Godde and Bickerton, 2006; Jansen et al., 2002a; Lillestol et al., 2006). They represent the most widely distributed family of repeats among prokaryotic genomes, suggesting a biological significance.

In the early studies, this family of the repeats was named differently by the researchers, leading to some confusion. The repeats have been named as TREP (tandem repeats), (Mojica et al., 1995), DVR (direct variant repeats), (van Embden et al., 2000), SRSP (short regularly spaced repeats), (Mojica et al., 2000), LCTR (long cluster of tandem repeat sequences), (She et al., 2001) and SPIDR (spacers interspaced direct repeats), (Jansen et al., 2002b). Based on a systematic characterization in different bacterial and archaeal genomes, Jansen and colleagues (Jansen et al., 2002a) proposed, in agreement with Mojica’s research group, a new name for this family of DNA repeats, which are now generally named as CRISPR.

Since CRISPRs are ubiquitous and peculiar in prokaryotes, it is beginning to attract growing interest and the identification of CRISPR has become an important task for bioinformatic analysis. Very recently, computational tools have been introduced to recognize CRISPR automatically (Durand et al., 2006; Edgar, 2007; Grissa et al., 2007a) and a specific, public CRISPR database is available that is regularly updated (Grissa et al., 2007b) and provides rapid and exact detection, comparison and identification of CRISPR for further research.

1.3.2 CRISPR-associated proteins

Shortly after CRISPR was defined, some putative protein genes flanking CRISPR sequences were identified. Jansen and his colleagues compared the genes flanking the CRISPR loci in the genomes of different prokaryotic species and found a clear homology among four genes. They are always located near to the repeats, and are not present in species without CRISPR loci, or in other words, no homologues of these genes were found in CRISPR-negative genomes. These genes are the so called CRISPR-associated (*cas*) genes and the encoded proteins are therefore called Cas proteins (Jansen et al., 2002a). The *cas* genes usually orient head-to-tail suggesting a coordinated transcription. The most common arrangement is *cas3-cas4-cas1-cas2*. The *cas* gene cluster generally is found within a few hundred of base pairs of the CRISPR locus. Each locus of CRISPR has its own set of *cas* genes indicating that CRISPRs and the accompanying *cas* genes are functionally related. The amino acid sequences of the Cas proteins show some highly conserved amino acid residues or functional domains. Based on the sequence comparison, the four Cas protein groups match perfectly to COG (cluster of orthologous groups) numbers in NCBI database (Makarova et al., 2006). The COG identification number of the Cas1 to Cas4 proteins are COG1468, COG1343, COG1203 and COG1518, respectively (Jansen et al., 2002a). Recently, two new *cas* genes, *cas5* and *cas6* have been defined based on the fact that the particular combination of the core genes (*cas1-4*) and these genes are always found in genomes. Most of the Cas5 proteins match to COG1688, and Cas6 matches to COG1583 (Haft et al., 2005). *cas1-6* are very common in prokaryotes that possess CRISPR loci, and they form the so-called 'core *cas* genes'.

There are some *cas* genes that are not very common, some match the COG protein numbers, some do not, and exist along with CRISPRs with or without function prediction, but until now there has been no experimental characterization on Cas proteins. The functions of the Cas proteins are almost unknown. A functional prediction is possible for only few Cas proteins. For instance, Cas1 proteins are generally highly essential. They are the only Cas proteins found consistently in all species that possess CRISPR loci, and are treated as the marker for Cas protein detection. The Cas1 proteins are predicted as novel nucleases. Cas3 proteins appear to be a helicase, sometime fused with COG2254 proteins which are predicted as HD-family nucleases. Cas4 proteins resemble the RecB family of exonucleases and contain a cysteine-rich motif, suggesting a function in DNA binding. The others remain to be characterized (Makarova et al., 2006).

1.3.3 The biological roles of CRISPRs and Cas proteins

A few years before CRISPRs were found in prokaryotic genomes, the *cas* genes and some of their associated genes had been identified in thermophilic archaea and bacteria and were defined as DNA repair protein genes (Makarova et al., 2002). By that time, the study of DNA damage and repair in archaea was just started. For example, the identification of the archaeal DNA binding protein, SSB in *Sso* (Kerr et al., 2001; Wadsworth and White, 2001) and its interaction with other proteins (Cubeddu and White, 2005) provided new knowledge about DNA damage and repair in archaea, and genomic analysis indicated some putative DNA repair protein genes (Aravind et al., 1999; Grogan, 2000). The prediction of these genes as being involved in DNA repair provided a starting point for the present work because these might constitute a novel DNA repair system. The computational analysis of Koonin's group predicted a series of features of these proteins that were related to DNA damage recognition and repair, such as DNA binding, DNA strand cleavage, DNA degradation, ATP-dependent duplex unwinding and nucleotide polymerization. Accordingly, DNA helicase, ATPase, nuclease and polymerase, were matched in this analysis (Makarova et al., 2002).

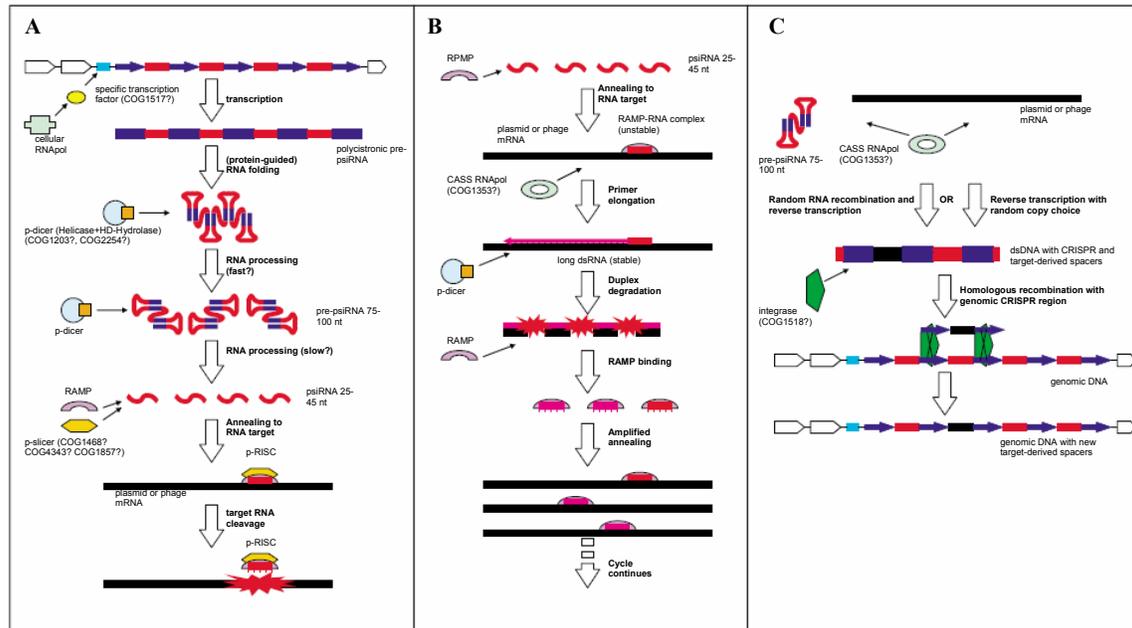


Figure 1.5: The current hypothetical model for CASS functioning and CRISPR formation. **A**, The basic model of CASS functioning. In this model, the transcription of the CRISPR regions is stimulated by the stress of phage or plasmid invasion and is regulated by Cas proteins. The primary transcript encompasses the entire CRISPR repeat region. This transcript is cleaved into 70-100 nt pre-psiRNA (pre-prokaryotic siRNA) by the putative p-dicer, the COG1203 protein. The p-dicer catalyzes the further slower processing step that releases mature psiRNA. The psiRNA molecules then bind RAMPs (Repeat-Associated Mysterious Proteins) in a size-specific manner and anneal to the target mRNA. This complex recruits p-slicer, forming the minimal form of the prokaryotic RNA-induced silencing complex (pRISC) that cleaves the mRNA. pRISC can be recycled to attack the next target molecule, thus silencing the respective gene. **B**, The variant of CASS functioning involving the CASS polymerase. In this pathway, psiRNA is produced in the same way as in the basic one. But in the next step, psiRNA serves as the primer for elongation by the CASS polymerase, yielding an extended dsRNA form of the target. The p-dicer cleaves the dsRNA at the endpoint of the pathway. Or else, the RAMP binds the dsRNA degradation product, forming a complex for annealing to the phage or plasmid mRNA, resulting in amplification of the silencing effect. **C**, Formation of new CRISPR with unique inserts. The path to the creation of new psiRNAs begins just like the response pathway to produce the 70-100 nt psiRNA precursors. At the next step, there are two possible pathways to replace the unique insert within the pre-psiRNA with a new fragment of the foreign RNA. One is the reverse transcription with copy choice whereby a reverse transcriptase, most likely, the CASS polymerase (COG1353) switches from using the pre-psiRNA as a template to using a phage or plasmid mRNA, and then back. The other one is direct, non-homologous RNA recombination between a pre-psiRNA and a foreign mRNA, followed by reverse transcription of the resulting recombinant RNA. The next insertion of the CRISPR fragment with the nascent spacers derived from foreign DNA into host genomic DNA is mediated by an integrase/recombinase, most likely, COG1518.

However, the new reports on the CRISPRs pointed to another biological function of these proteins, although the predicted biochemical functions of the proteins remained mostly unchanged. The new proposed function is that, CRISPR and Cas proteins form a prokaryotic defense system (CASS, CRISPR-associated system) which mimics the

eukaryotic interference RNA system, the RNAi function. There are some bioinformatics data to support this hypothesis (Makarova et al., 2006). First, the *cas* genes are tightly associated with CRISPRs and are conserved both in gene composition and gene order. Furthermore, in the genomes that possess multiple CRISPR loci, each CASS has its own set of unique spacer sequences and *cas* genes, although some of these *cas* genes may not be as common as the core *cas* genes indicating the functional relationship of CRISPR and Cas proteins (Makarova et al., 2006; Jansen et al., 2002a). Second, accumulating homology searches show a similarity of spacer sequences to sequences from viruses, phages, plasmids and transposable elements (Mojica et al., 2005; Godde and Bickerton, 2006; Tyson and Banfield, 2007; Lillestol et al., 2006) suggesting that CASS is involved in resistance against foreign genetic elements. Third, further functional prediction of Cas proteins implied that the putative helicase, nuclease, recombinase, integrase, RNA polymerase probably cooperate at the RNA level (Pourcel et al., 2005; Makarova et al., 2006) or at the gene level where spacer transcripts are annealed directly to a gene, thereby facilitating degradation of the foreign DNA (Lillestol et al., 2006). A hypothetical model for CASS was established by Koonin's group according to *in silico* analysis (Figure 1.5). It intriguingly mimics the eukaryotic RNAi (Makarova et al., 2006), although experimental support for the model is still lacking (Lillestol et al., 2006). The idea is that the spacers of CRISPRs transcribe short RNA sequences that can bind to complementary sequences in messenger RNAs derived from invading DNAs. This would block their translation into proteins and mark them for degradation by Cas proteins.

Very recently, direct evidence has been provided from working with *Streptococcus thermophilus* (Marx, 2007). In that research, the infection of the bacteria with phage leads to incorporation of phage-related spacer sequences with CRISPR region. These bacteria become resistant for further infection by the phage strains from which these sequences were derived. The knocking out of the spacers leads to a loss of resistance. And one *cas* protein gene (*cas7*) is involved in the process, although the mechanism of this process remains unknown.

1.3.4 CASS in *Sulfolobus solfataricus*

Table 1.3: Summary of repeat sequences in *Sso* P2

Repeat sequences	No. of clusters
1. CTTTCAATTCC CTTTTGG GATTAATC	3
2. CTTTCAATTCTATA AGAG ATTATC	2
3. CTTTCAATTCTATA GTAG ATTAGC	2

Conserved bases are marked black.

Table 1.4: *cas* genes in *Sso* P2

No. of repeat sequence	COG No. <i>cas</i> No.	1518 <i>cas1</i>	1343 <i>cas2</i>	1203 <i>cas3</i>	1468 <i>cas4</i>	1688 <i>cas5</i>	1583 <i>cas6</i>	2254	4343 <i>cas1</i>	1857 <i>cas2</i>	2462 <i>cas3</i>	1353 pol.
1	numbers	1	1	1	1	1	1	1	1	1		
2	of Cas	1	1	1	1	1	1	1	1	1	2	1
3	proteins			1		1	1	1		1		1

1.3.5 Prospect

Taken the vast computational analyses and few experimental evidences together, the finding of CASS would deepen our knowledge in several aspects. Study on conservation of palindromic repeats in CRISPRs would reveal the functions of these elements in host cell defense (Mojica et al., 2005), regulation (Haft et al., 2005), chromosomal segregation (Mojica et al., 1995) and rearrangement (DeBoy et al., 2006). The rich diversity of CRISPR spacer sequences in different thermophiles suggests the horizontally transferred genetic elements (HTG elements) existing in the extreme environment (Mongodin et al., 2005). Further studies will help to find out of whether or not there is a correlation between CRISPR spacer sequence and geographic location. The observation of co-transcription and transcription activation of *cas* genes with CRISPR (Shinkai et al., 2007; Viswanathan et al., 2007) would open a door about the specific pathway of trigger and regulation of this co-transcription. Experimental test of the Cas proteins for *in silico* functional prediction (Makarova et al., 2006) will illustrate the mechanism of this novel, peculiar immune system. The direct evidence of CASS in prokaryote could provide a way of gene silencing in prokaryotes like that of RNAi in higher organisms and this could be applied into genetic engineering, such as blocking specific gene activity.

1.4 Aim of the present work

SSB proteins are essential in many biochemical processes. The SSB proteins in bacteria and eukaryotes have been studied in detail and their structures and functions are quite well known. By contrast, we have only incomplete knowledge of the basic properties of the archaeal SSBs. In the present thesis, the SSB from *Sso* P2 strain has been characterized mainly with respect to the following points:

- The natural form of *Sso*SSB in solution in the presence/absence of DNA
- The binding affinity of *Sso*SSB to ssDNA substrates and its cooperativity
- Determination of dissociation constant of *Sso*SSB and binding model with ssDNA
- Direct detection of *Sso*SSB –ssDNA complex by atomic force microscopy (AFM)

It is expected that the clarification of these features would be helpful for investigating the role of the archaeal SSB *in vivo*. And through comparison of *Sso*SSB to bacterial and eukaryotic SSBs, one simple, efficient SSB binding model might be found, indicating the common ancestor SSB for three domains of life.

CRISPR and Cas proteins are a considerably new concept. Interestingly, at the very beginning, because of the similarity of Cas proteins to DNA repair proteins, the lack of observation of some repair systems in archaea and without a functional assignment of CRISPRs, the Cas proteins were misunderstood as novel repair proteins. After finding of ubiquitous CRISPRs and Cas proteins in prokaryotes, in past few years, the study of them concentrated on bioinformatic analysis. Only in the last two years, some experimental evidences were reported. It becomes clear that CASS is a unique prokaryotic immune system, mimicking the RNAi system in eukaryotes. However, the lack of experimental studies leaves most features of this system hypothetical. For instance, how invading DNA triggers the CASS; how CRISPR regulates the function of Cas proteins; what are the roles of Cas proteins; the relationship between Cas proteins; on which level the CASS functions.

Trying to answer some of these questions, the present work primarily investigated the distribution of CRISPR and *cas* genes in crenarchaeal model system, *Sso* P2. Based on

the computational analysis, the *cas* gene expression was scanned in different conditions in *E.coli* expression system and their enzymatic activities were mainly investigated. The characterization of some potential Cas proteins was achieved later on. Some aspects of previous hypothesis were proved, and some new features were observed.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals, enzymes and proteins

2.1.1.1 Chemicals

acrylamide/N,N'-methylenebisacrylamide (19:1)	
acrylamide/N,N'-methylenebisacrylamide (29:1)	
agar, agarose (NEEO), Boric acid, EDTA, ethidiumbromide, HEPES, IPTG, peptone, TEMED, Tricin, Tris, Urea	Roth, Karlsruhe
APS, Arginine, antibiotics, Commasie Brilliant Blue G250, GSH /GSSG (Reduced/Oxidized Glutathione), PMSF, polyethylene glycol (PEG) 8000, RbCl, acetic acid, methanol, dNTPs	Merk, Darmstadt Roche, Mannheim
5-bromo-4 chloro-3-indolyl phosphate (BCIP), Nitro blue tetrazolium chloride (NBT), β -naphthol acetate, Fast Blue BB Salt, Triton X-100	Sigma, Schnelldorf
Yeast extract	Gerbu, Gaiberg
$[\gamma\text{-}^{32}\text{P}]\text{-ATP}$ (5000 Ci/mol)	Hartmann Analytik, Braunschweig

Other chemicals, if not mentioned, were purchased from Merck (Darmstadt). The purity of all chemicals was *pro analysis*.

For fermentations, deionized H₂O was used. For preparation of buffers and other solutions double-distilled (dd) H₂O was used.

2.1.1.2 Enzymes and proteins

Antibodies	
Anti-His antibody	Qiagen, Hilden
Rabbit anti mouse IgG/AP (alkaline phosphatase) conjugate	Pierce, USA
BSA, T4 polynucleotidkinase (T4 PNK), T4 DNA ligase	New England Biolabs, Frankfurt am Main
non-fat dry milk	Roche, Mannheim
Low molecular weight (LMW) protein standard	Pharmacia, Freiburg
Pfu DNA polymerase	Promega, Mannheim

shrimp alkaline phosphatase (SAP) Boehringer, Mannheim
 Taq polymerase Paqlab, Erlangen
 All restriction endonucleases were purchased from New England Biolab, Frankfurt am Main.

2.1.2 Bacterial and archaeal strains, media and antibiotics

2.1.2.1 Bacterial strains

BL21-CodonPlus(DE3)-RIL *E. coli* B F⁻ *ompT hadS(r_B⁻ m_B⁻) dcm⁺ Tet^r gal λ* (DE3) *endA Hte [argU ileY leuW Cam^r]^a* (Stratagene, Heidelberg)
 BL21(DE3) F⁻ *ompT hsdSB(rB – mB–) gal dcm* (DE3)
 Rosetta(DE3)pLysS F⁻ *ompT hadS(r_B⁻ m_B⁻) gal dcm* (DE3) pLysSRARE (Cam^r)
 Rosetta-gami(DE3) Δ (*ara-leu*)7697 Δ *lacX74* Δ *phoA PvuII phoR araD139* *ahpC galE galK rpsL* F'*[lac⁺ lacI^q pro] gor522::Tn10 trxB* pRARE (Cam^R, Kan^R, Str^R, Tet^R) (Novagen, Darmstadt)
 XL-1 blue *supE44 hsdR17 endA1 gyrA96 relA1 thi-1 recA1 lac⁻* [F' *proAB⁺, lacI^q ZΔM15, Tn10 (Tet^r)*] (Stratagene, Heidelberg)

2.1.2.2 Media, inducer and antibiotics

LB–Medium, Luria Bertani Broth	10 g Peptone 5 g yeast extract 5 g NaCl, add 1 l H ₂ O, pH 7.2
Ampicillin	100 mg/ml in ddH ₂ O
Chloramphenicol	34 mg/ml in ethanol
Kanamycin	25 mg/ml in ddH ₂ O
Streptomycin/spectinomycin	10 mg/ml in ddH ₂ O
IPTG	1 M in ddH ₂ O
X-gal	20 mg/ml in DMF

Bacterial media were autoclaved at 121°C (1 Bar) for 20 minutes. For solid media preparation, 1.5% (w/v) agar was added to liquid media. Antibiotics solutions were filter-sterilized.

2.1.3 Plasmids and phage

M13 GTGx	phage for ssDNA production	Kindly gift from Prof. Richard D. Wood Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, UK
pET19b-SsoSSB	Expression vector for SsoSSB	Kindly gift from Prof. M. White, University of St. Andrews, UK
Co-expression vectors		

Materials and methods

pRSFDuet-1 (RSF1030, kan ^r , T7)		
pETDuet-1 (ColE1, Amp ^r , T7)		
pCDFDuet-1 (CloDF13, Strp ^r /Spec ^r , T7)		Novagen, Darmstadt
pET-28C(+)	gene expression vector	Novagen, Darmstadt
pGEM-T	plasmid for gene sequencing	Promega, Erlangen
pIVEX2.3d-Est2.1	vector containing esterase gene	Lehrstuhl Biochemie I, Universität Bayreuth
pUC19	plasmid for dsDNA preparation	

In the present work, the recombinant plasmids for the cloning and expression of the putative repair genes from *Sso* P2 are all derived from the plasmids listed above.

2.1.4 Oligonucleotides and tRNAs

All the oligonucleotide sequences are from 5' to 3', left to right.

2.1.4.1 PCR primers

M13AFM-for	CCGGGATCCGAAGGGATATCAGCTGTTGCCCGTC
M13AFM-rev	AGCTTTCGGCACCGCTTCTGGTGCCGGA
<i>sso</i> 1392-for	CCCCCGCTAGCCCATGGGTGAGAGTGATTGCATGTT
<i>sso</i> 1392-rev	CCCCGAATTCAATCAACATAAGCTAAATTTTC
<i>sso</i> 1402-for	CCGCTAGCCCATGGATGGGATGTGAAGAAAAT
<i>sso</i> 1402-rev	CCGGATCCTCATTTCCTCAACCTCAACCC
<i>sso</i> 1439-for	CCGCTAGCCCATGGTTGATCAAGCCTTGTGCTT
<i>sso</i> 1439-rev	CCGGATCCTCATAGAGTGGAACCTCCATT
<i>sso</i> 1440-for	CGTAGCCCATGGTTGAAGTTGAGGC- TAAGGAAGCCTATGAGGTCGGCGGG
<i>sso</i> 1440-rev	CCCCGAATTCAATACACACCACCTATTTTCAC
<i>sso</i> 1442-for	CCGCTAGCCCATGGATGATAAGCGGTTTCAGTT
<i>sso</i> 1442-rev	CCGGATCCTCACTCTTCTCTAATTTAAC
<i>sso</i> 1449-for	CCGCTAGCCCATGGATGGTTAGTGTAACGGAT
<i>sso</i> 1449-rev	CCGGATCCTCATAGCAAACACTAACTGAG
<i>sso</i> 1729-for	CCATATGCCATGGGTGAGGCGGGTGGGATCAA
<i>sso</i> 1729-rev	CCGGATCCTAAATTACTCCCCTCATTATTC
<i>sso</i> 1991-for	CCCCGGATCCAAATGAGTACTGATGATAATTC
<i>sso</i> 1991-rev	CCCCAAGCTTTCCTTCTCGCCACCGTAAATTA
<i>sso</i> 1997-for	CCCCCCCCCATATGCCATGGATGATAGGCGGTTTCAGGTA
<i>sso</i> 1997-rev	CCCCGAATTCACCTTTTCTTTAATTTAGCT
<i>sso</i> 1998-for	GCCATGCCATGGGCTTGATCTACTCT
	AAGGTTTTTTTAAACTTCATTGGGGT
<i>sso</i> 1998-rev	CGCCGGATCCGGGCTAAAGACAACATATTCTC

<i>sso</i> 1999-for	CGCTAGCCCATGGTTGAAGTTGAGGC TAAGGAAGCATATGAGGTTGGTGA
<i>sso</i> 1999-rev	CCCCGAATTCAATACACACCACCTATTTTCAC
s1996Duet-for	ACCGGATCCGGTGAGGGTAGCCAACTTG
s1996Duet-rev	ACCAAGCTTATTTCTTTTCACCACCTTGTTTC
s1997Duet-for	ACCCATATGATAGGCGGTTTCAGGTAG
s1997Duet-rev	ACCGGTACCCTTTTCCTTTTAATTTAGC
s1998-forDuet	ACGGGATCCGTTGATCTACTCTAA GGTTTTTTTAAAACTTCATTGGGGT
s1998-revDuet	ACCCCAAGCTTGCTAAAGACAACATATTCTC
s1999-forDuet	ACGGAAGATCTCTTGAAGTTGAGGCTAAGGA
s1999-revDuet	ACCGGGGTACCATACACACCACCTATTTTCAC
s2001-forDuet	ACCCCAATTGGTTGATCAAGCCTTGCGCT
s2001-revDuet	ACCGGGGTACCTAGAGTGGAACCTCCATT
s2002-forDuet	ACGGGATCCGATGGAGGTTCCACTC TATAATATATTTGGAGATAATTACA
s2002-revDuet	ACCCAAGCTTTCTTCTCCTCTTATTAACTCTCCA
S1450-for	CATGCCATGGGCGTGATAAGCGTGAGGACTTT
S1450-rev	GCCGAATTCGCCATCACCAACTTGAAACCCC
S2001-for	ACATATGTTGATCAAGCCTTGCGCTTA
S2001-rev	ACGAGCTCTAGAGTGGAACCTCCAT

2.1.4.2 Primers for mutation

QCD-for	CAAGAGCTTAAAAGTTCGCGATTACACTTTTGATG
QCD-rev	CATCAAAGGTGTAATCGCGAACTTTTAAGCTCTTG
QCE-Afor	CTTTCATATATCATGCCCTCGGTTTCGGCTC
QCE-Arev	GAGCCGAACCGAGGGCATGATATATGAAAG
QCH1-for	GACGTCGTTGTCCTTGCCGATATGGGAAAAGC
QCH1-rev	GCTTTTCCCATATCGGCAAGGACAACGACGTC
QCH2-for	CTCTTTCATATATGCTGAACTCGGTTTCGGC
QCH2-rev	GCCGAACCGAGTTCAGCATATATGAAAGAG
QCS-for	TCATGAACTCGGTGCGGCTCTATTTTTTCT
QCS-rev	AGAAAAATAGAGCCGCACCGAGTTCATGA

2.1.4.3 Substrates for SsoSSB

Pyrimidine-rich15	GCGTTAATCCTACCT
Purine-rich15	AGGTAGGATTAACGC
Pyrimidine-rich24	TCTTCTTCTGTGCACTCTTCTTCT
Purine-rich24	AGAAGAAGAGTGCACAGAAGAAGA

Materials and methods

Pyrimidine-rich42	AATTCTCCCTCTTCTTCTGTGCACTCTTCTTCTCCCCATCTC
Purine-rich42	TCGAGAGATGGGGAGAAGAAC TCACGTGTCAAGAAGAGGGAG
ssb-7*	CCTACCT
ssb-9*	ATCCTACCT
ssb-11*	TAATCCTACCT
ssb-13*	GTAAATCCTACCT
ssb-15*	CAGTTAATCCTACCT
ssb-17*	GGATCAAATTGTATCC
ssb-25*	TTGTGGATCAAATTGTATCCGCAA
ssb-50*	ACAGCTATGACCGAATTCCTGGGGA CTTCTTCTCAGCACAGAAGAAGAGG

* : Substrates with/without 5'-end fluorescein

2.1.4.4 Substrates for nuclease assay

50b	CCTCTTCTTCTGTGCACTCTTCTTC TCCCCAGGAATTCGGTCATAGCTGT
50u	ACAGCTATGACCGAATTCCTGGGGAG AAGAAGAGTGCACAGAAGAAGAGG
gap1mer	AAGAAGAGTGCACAGAAGAAGAGG
gap5mer	AGAGTGCACAGAAGAAGAGG
gap10mer	GCACAGAAGAAGAGG
HJ-b	TTCTGCTCGAGGGCGCCAGGGTGG GAAGAAGAGTGCACAGAAGAAGAGG
HJ-u	CTTGCATGCCTGCAGGTCGACCAGG CCACCCTGGCGCCCTCGAGCAGGAA GAATTCCTGGGGAGAAGAAGAGTGCA
mid-match	ACAGCTATGACCGAATTCCTGGGGA
nick-for	GAAGAAGAGTGCACAGAAGAAGAGG
nick-for2	CCTCTTCTTCTGTGCACTCTTCTTCTCCCC
PT-A30	GGGGAGAAGAAGAGTGCACAGAAGAAGAGG
PTB30	ACAGCTATGACCGAATTCCTGGGGA
pseudo-Y	CCTGGTTCGACCTGCAGGCATGCAAG

2.1.4.5 Substrates for Sso1450C6H

ssb-6*	CCTACC
ssb-7*	CCTACCT
ssb-9*	ATCCTACCT
ssb-11*	TAATCCTACCT
ssb-12*	CTCTTATAGAAT
ssb-13*	GTAAATCCTACCT
ssb-24*	CTTCAATTCTATAAGAGATTATC
ssb-33*	TTGCTCCGTCTTTCAATTCTATAGTAGATTAGC
ssb-36*	GATCGCTTCTTTCTTTCAATTCTATAAGAGATTATC

12nRNA*	AUUCUAUAAGAG
24nRNA*	CUUUCAAUUCUAUAAGAGAUUAUC
tRNA ^{Arg} _{E. c}	
tRNA ^{Phe} _{E. c}	Lehrstuhl Biochemie I, Universität Bayreuth

* : Substrates with/without 5'-end fluorescein

2.1.5 Buffers and solutions

AFM buffer	8 mM HEPES, 2 mM NaCl, 2 mM MgCl ₂ , pH 8.2
Binding & washing buffer	10 mM Tris-Cl, pH 7.5, 1 mM EDTA, 1 M NaCl
ATPase reaction buffer	125 mM Tris-Cl, 25 mM MgCl ₂ , 5 mM DTT, pH 7.5
Buffers for agarose and PAGE gel	
Electrophoresis buffer, 0.5x TBE	45 mM Tris 45 mM Boric acid 0.5 mM EDTA, pH8.4
Native gel loading buffer	40% (w/v) sucrose, 0.25% (w/v) BromphenolBlue, 1 mM EDTA, pH 8.0
Buffers for denaturing PAGE	
Electrophoresis buffer, 1x TBE	90 mM Tris 90 mM Boric acid 1 mM EDTA, pH8.4
Denaturing gel loading buffer	20 mM EDTA, 0.2% (w/v) Xylencyanol, 0.2% (w/v) Bromphenolblue in formamide
Buffers and solution for SDS-PAGE (Schägger-Jagow protein gel)	
S/J-gel buffer (10x)	3 mM Tris-Cl, 0.3% (w/v) SDS, pH 8.5
Cathode buffer	0.1 M Tris-Cl, 0.1 M Tricin, 0.1% SDS, pH 8.25
Anode buffer	0.2 M Tris-Cl, pH 8.9
S/J-loading buffer	100 mM Tris-Cl, 24% glycerol, 8% SDS, 4% β-mercaptoethanol, 0.02% Serva-Blue, pH 6.8
Coomassie blue staining solution	0.2 % (v/v) Coomassie Brilliant Blue R-250, 10 % (v/v) acetic acid, 30 % (v/v) ethanol
Destaining solution	10 % (v/v) acetic acid, 30 % (v/v) ethanol
Buffers for preparation of competent cells	
TfBI buffer	30 mM K-Acetate pH 5.8 (acetate acid) 100 mM RbCl 50 mM MnCl ₂ 10 mM CaCl ₂ 15 % (w/v) Glycerol
TfBII buffer	10 mM Mops pH 7.0 (NaOH) 10 mM RbCl 75 mM CaCl ₂ 15 % (w/v) Glycerol
Buffers and solution for <i>Sso</i> SSB	

Materials and methods

<i>Sso</i> SSB EMD buffer	50 mM Tris-Cl, 1mM EDTA, 1 mM DTT, pH 7.5
Ammonium sulfate solution	60% (w/v), 90% (w/v)
<i>Sso</i> SSB Sephacryl S-300HR buffer	0.1 M (NH ₄)HCO ₃
<i>Sso</i> SSB Stock buffer	50 mM sodium phosphate, 1 mM DTT, 50% glycerol, pH 7.5
<i>Sso</i> SSB binding buffer (EMSA)	20 mM Tris-Cl, 50 mM NaCl, 0.1 mg/ml BSA, pH 7.5
Buffers for refolding assays	
Buffer A	10 mM Tris-Cl, 0.1 M sodium phosphate, 0.3 M NaCl, 20% glycerol, 2 mM β-mercaptoethanol, pH 8.0
Buffer W	10 mM Tris-Cl, 0.1 M sodium phosphate, 0.3 M NaCl, 20% glycerol, 2 mM β-mercaptoethanol, pH 6.3
Buffer E	10 mM Tris-Cl, 0.1 M sodium phosphate, 0.3 M NaCl, 20% glycerol, 2 mM β-mercaptoethanol, pH 5.0
Buffers and solutions for high-through put refolding assays: see Tables 2.1 and 2.2 for details.	
Buffers for Western Blotting	
Blotting buffer	25 mM Tris-Cl, 192 mM glycine, 20% methanol, pH 8.6
TBS-T buffer	10 mM Tris-Cl, 150 mM NaCl, 0.05% tween20 pH 7.5
AP buffer	100 mM Tris-Cl, 100 mM NaCl, 5 mM MgCl ₂ pH 9.5
BCIP stock solution	5 mg/ml BCIP (5-bromo-4 chloro-3-indolyl phosphate in 100% dimethylformamide)
NBT stock solution	5% NBT (Nitro blue tetrazolium chloride in 70% dimethylformamide)
Cell lysis buffer for <i>Sso</i> 1450C6H	50 mM Tris-Cl, 150 mM NaCl, 20% glycerol 2 mM β-mercaptoethanol, 10 mM imidazole pH 7.5
Cell lysis buffer for <i>Sso</i> 2001Est	50 mM Tris-Cl, 100 mM NaCl, 10% glycerol 2 mM β-mercaptoethanol, pH 7.5
Fluorescence anisotropy buffer	20 mM Tris-Cl, 50 mM NaCl, 100 ng/ml BSA, 0.05% tween20, pH 7.5
Hybridization buffer (NEB4)	20 mM Tris/AcOH, 10 mM Mg(OAc) ₂ , 50 mM KOAc, 1 mM EDTA, pH 7.9 at 25°C
NEB3 buffer	50 mM Tris-Cl, 100 mM NaCl, 10 mM MgCl ₂ , 1 mM DTT, pH 7.9 at 25°C
Nuclease reaction buffer	50 mM Tris-Cl, 10 mM MgCl ₂ , pH 7.5
PBS	0.24 g/l KH ₂ PO ₄ , 1.44 g/l Na ₂ HPO ₄ , 0.2 g/l KCl, 8 g/l NaCl, pH 7.4
PCR buffer (high quality)	10 mM Tris-Cl, pH8.8, 50 mM KCl, 1.5 mM MgCl ₂ , 0.1 % (v/v) Triton X-100

PCR bufferII (high quantity)	20 mM Tris-Cl, pH8.55, 16 mM (NH ₄) ₂ SO ₄ , 0.01 % Tween20, 2 mM MgCl ₂
Pfu buffer	20 mM Tris-Cl, 10 mM KCl, 10 mM (NH ₄) ₂ SO ₄ , 2 mM MgSO ₄ , 0.1 mg/ml BSA, 0.1 % Triton X-100
PNK forward buffer	50 mM Tris-Cl (pH 7.6 at 25°C), 10 mM MgCl ₂ , 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA
protein stock buffer	50 mM Tris-Cl, 100 mM NaCl, 50% glycerol 1 mM DTT, pH 7.5
SAP buffer	10 mM Tris-Cl (pH 7.5 at 37°C), 10 mM MgCl ₂ 0.1 mg/ml BSA
TE (1x)	10 mM Tris-Cl, 1 mM EDTA, pH8.0
Thin layer chromatography running buffer	1 M formic acid, 0.8 M lithium chloride

2.1.6 Commercial kits

CellLytic™ B bacterial cell lysis extraction reagent	Sigma, Schnelldorf
E.Z.N.A.® Plasmid Miniprep KitII	PeqLab, Erlangen
pGEM®-T Vector System	Promega, Erlangen
Roti®-Nanoquant (Bradford Assay) solution	Roth, Karlsruhe
Wizard SV Gel and PCR Clean-Up System	Promega, Erlangen

2.1.7 Instruments and materials

96-well, 24-well Nunc-Immuno microtitre plates	International, Nunc-Immuno plate Denmark
AFM MultiMode	Veeco/Digital Instruments Inc. Santa Barbara, USA
Centrifuge Biofuge 13R	Heraeus, Hanau
Cool centrifuge RC5B	Du Pont, Bad Homburg
Dialysis membrane	Serva, Heidelberg
Environmental Incubator Shaker	New Brunswick Scientific, Edison N.J., USA
gel apparatus	Pharmacia, Freiburg
LS50B spectrofluorometer	Perkin-Elmer, USA
Microplate Spectrophotometer	BioTek Instruments, Inc., USA
Mighty Small Gelsystem	Hoefer Scientific Instr., San Francisco, USA
MP220 pH meter	Mettler-Toledo, Switzerland
Nitrocellulose Blotting Membrane	Advantec MFS, Inc., Dublin, USA
Panther™ Semi-Dry Electroloter, HEP-1	Peqlab, Erlangen
PEI-Cellulose plastic sheets	Macherey-Nagel, Düren
Petri dishes, tubes, micro tubes, semi-micro cuvettes	Sarstedt AG & Co., Nümbrecht
Quartz and glass cuvettes	Hellma Müllheim

Smart-Ladder (DNA marker)	Eurogentec, Köln
Sonifier B15	Branson, Schwäbisch Gmünd
T-personal Combi (PCR instrument)	Biometra, Gröttingen
Thermomixer 5436	Eppendorf, Hamburg
Thermostat heating block 5320	Eppendorf, Hamburg
Typ Instant Imager 2024	Canberra-Packard, Dreieich
UV-Spectral photometer DU640	Beckman, München
UV/Vis-Transilluminator with camera and computer	MWG, Ebersberg
Vortex-Genie	Bender & Holbein AG, Zürich
Water bath	Köttermann, Burgdorf

2.1.8 Chromatographic materials

Co ²⁺ -Talon	Clontech, Heidelberg
Dynabeads, Dynal MPC (<u>M</u> agnetic <u>P</u> article <u>C</u> oncentrator)	Invitrogen Dynal AS, Norway
EMD-SO ₃	Merk, Darmstadt
Ni ²⁺ -NTA-Sepharose, superflow	
Ni ²⁺ -NTA-Silicagel, spin columns	Qiagen, Hilden
Sephadex G25	
Sephacryl S-300HR	Pharmacia, Freiburg
Trifluoromethyl ketone (TFK) Sepharose CL-6B	
3-Butylsulfanyl-1,1,1-trifluoro-propan-2-one (TFK analog)	Lehrstuhl Biochemie I, Universität Bayreuth

2.1.9 Softwares

BLASTN on NCBI website (www.ncbi.nlm.nih.gov/blast/blast.cgi)	National Center for Biotechnology Information, Bethesda, USA
Corel PHOTO-PAINT version 12	Corel Co., USA
KC4 program	BioTek Instruments, Inc., USA
Origin 5.0	OriginLab Co. Northampton, MA, USA
VectorNTI Suite 7	InforMax Inc., USA
Windows XP and MS Office 2003	Microsoft Co., USA

2.2 Standard methods

2.2.1 Spectrophotometric determination

2.2.1.1 Determination of protein concentration

When nucleic acids were present, protein concentration was determined by absorbance measurements in a 1 cm quartz cuvette in a Beckmann DU-640 spectrophotometer. Protein concentration was calculated according to Ehresmann:

$$(A_{228.5}-A_{234.5}) / 3.14 = \text{mg/ml (protein)}$$

Protein concentration was also determined by Bradford protein quantification assay. This assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions stabilize the anionic form of the dye, causing a visible color change (Bradford, 1976).

In detail, BSA stock solution was diluted to a range of 1-100 $\mu\text{g/ml}$ to establish the standard curve. 50 μl from each calibration standard or from the samples was pipetted into a 96-well culture plate and was mixed with 200 μl of Roti[®]-Nanoquant (1x) reaction solution. After 5 min incubation at room temperature, OD₄₅₀ and OD₅₉₅ of the standards and samples were measured. Thereafter, the quotient OD₅₉₅/OD₄₅₀ of each sample was plotted and compared to the calibration curve from the standards (Zor and Selinger, 1996).

2.2.1.2 Determination of nucleic acid concentration

The concentration of DNA and RNA was determined by measurement of the absorbance at 260 nm. One OD₂₆₀ unit corresponds to approx. 50 μg of dsDNA, 30 μg of ssDNA or 40 μg of RNA.

2.2.1.3 Determination of bacterial cell density

The optical density (OD₆₀₀) of bacterial suspensions was measured with the Beckmann DU-640 spectrophotometer in 1cm polystyrol cuvettes. One OD₆₀₀ unit corresponds to 6×10^8 cells.

2.2.2 Gel electrophoresis

2.2.2.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used for analysis, preparation of DNA and detection of DNA-protein complexes. First, 0.4-1 g agarose was melted in 50 ml 0.5x TBE buffer

(0.8-2% (w/v)) and was then cooled down to room temperature. The samples were mixed with 1/3 volume of native gel loading buffer and were then loaded on the gel. Electrophoresis was run at 5-7 V/cm in 0.5x TBE. The bands were visualized under a long wavelength UV lamp after being stained in 0.5 mg/l ethidium bromide solution.

2.2.2.2 Native polyacrylamide gel electrophoresis

The nucleic acid-protein complexes were analyzed in a small vertical slab unit. The gel (4-8 % acrylamide/N,N'-methylenebisacrylamide (19:1)) containing 25 mM Tris-acetic acid, pH 7.0 and 10 mM MgAc₂ was polymerized by addition of 0.05% (v/v) TEMED and 0.1% (w/v) APS. The samples were mixed with 1/3 volume of native gel loading buffer and were then loaded on the gel. Electrophoresis was performed at 100 V for 1-2 h in 1x TBE.

2.2.2.3 Denaturing polyacrylamide gel electrophoresis

For typical analysis of nucleic acids, 20% polyacrylamide (acrylamide/N, N'-methylenebisacrylamide 19:1) gels of 20 × 20 × 0.075 cm containing 8.3 M urea in 1× TBE buffer were prepared. Polymerization was carried out by addition of 0.02% (v/v) TEMED and 0.1% (w/v) APS. Gels were pre-run for at least 30 min at 50 mA, 50°C, and then the pre-heated (in Thermomixer 5436 for 5 min at 95°C) samples mixed with 1/3 volume denaturing gel loading buffer were loaded onto the gel. Electrophoresis was run at 50 mA for approx. 1 h. For sequencing, thinner gels (40 × 20 × 0.02 mm) of 12% or 20% polyacrylamide containing 8.3 M urea in 1 × TBE buffer were used. Gels were pre-run for 30 min at 50°C, 30 mA, and then samples mixed with loading buffer were loaded onto the gel. Electrophoresis was run at 50°C, 30 mA for approx. 2 h. The gel was dried for 1 h at 80°C for detection of labelled nucleic acids.

2.2.2.4 SDS polyacrylamide gel electrophoresis (SDS PAGE)

The SDS PAGE was performed according to Schagger-Jagow's protocol (Schagger and von Jagow, 1987) in the Mighty Small Vertical Slab Unit (Hofer Scientific Instruments). The 12% separating gel (acrylamide/N, N'-methylenebisacrylamide, 29:1)

and the 4% stacking gel (acrylamide/N, N'-methylenebisacrylamide, 29:1) both contained 300 mM Tris-HCl pH 8.5, 0.3% (w/v) SDS. Polymerization was started by addition of 0.05% (v/v) TEMED and 0.1% (v/v) APS. Two volumes of protein samples were mixed with one volume of SDS PAGE loading buffer and then heated at 95°C for 3 min. Electrophoresis was carried out at 15 V/cm, in Schagger-Jagow double-buffer system. The protein bands were visualized by Coomassie Brilliant Blue G250.

2.2.3 Detection of radioactively labeled nucleic acids in gel

After completion of gel electrophoresis, the radioactivity of labeled nucleic acids was counted by Typ Instant Imager 2024 (Canberra-Packard, Dreieich) and was automatically transformed into a visible form.

2.2.4 Detection of unlabeled nucleic acids in gel

Nucleic acids in agarose gels or in polyacrylamide gels were visualized under a long wavelength UV lamp after being stained in 0.5 mg/l ethidium bromide solution.

2.2.5 Staining of protein gels

The protein gels were stained by 5 volumes of coomassie blue staining solution (45 % (v/v) methanol, 10 % (v/v) acetic acid and 1 % (w/v) coomassie brilliant blue G250) for approx. 30 min. The protein bands were visualized after the gel was destained (3x) in 10 volumes of destaining solution (30 % (v/v) methanol, 10 % (v/v) acetic acid).

2.3 Molecular biology methods

2.3.1 Preparation and transformation of competent cells

Preparation and transformation of competent *E.coli* cells was carried out according to modified Hanahan methods (Hanahan, 1983). In detail, an *E.coli* strain was streaked directly from a frozen stock onto the surface of a LB agar plate and was incubated overnight at 37°C. Afterwards, four or five well-separated colonies were transferred into 1 ml of LB medium and were dispersed by vortexing at moderate speed. Then the culture was diluted in 100 ml of LB medium in a 1 l flask and was incubated for 2.5-3 h at 37°C

till OD₆₀₀ reached 0.4. The culture was then cooled down to 0°C in an ice bath for 10 min. The cells were recovered by centrifugation at 2700 g in SS34 rotor (Sorvall, DuPont) for 10 min at 4°C. The cell pellets were re-suspended by gentle vortexing in approx. 20 ml of ice-cold TfBI buffer. After storage on ice for 10 min, the cells were pelleted by repeating the centrifugation step and were re-suspended gently in 4 ml of ice-cold TfBII buffer. Aliquots of 100 µl of competent cells were stored at -70°C for until use.

When transformation was performed, up to 100 ng of transforming DNA was added to 100 µl of appropriate *E.coli* competent cells. The cell suspension was kept on ice for 30 min. Following heating at 42°C for exactly 90 s in the Thermomixer 5436 (Eppendorf, Hamburg), the cell suspension was cooled down rapidly on ice bath. After mixing with 1 ml of LB medium, the cell suspension was incubated at 37°C for 45 min in the Thermomixer 5436. Then 100-500 µl of the transformed competent cells were spread onto the LB agar plate containing the appropriate antibiotic(s). The plate was incubated at 37°C overnight.

2.3.2 Culture of bacterial strains

For small-scale preparation of plasmid DNA, after transformation (2.3.1), a single colony was picked with a toothpick from LB-agar petri dishes into 5 ml LB liquid medium. *E. coli* strains were grown in LB medium supplied with the appropriate antibiotic(s). Cultures were incubated overnight at 37°C with agitation of 170 rpm in the Environmental Incubator Shaker (New Brunswick Scientific, Edison, N.J., USA).

For expression protein genes in *E.coli*, 20 ml LB medium was inoculated by a single colony of *E. coli* strain harboring the genes to be expressed, supplied with proper antibiotics and grown overnight at 37°C with agitation. 4 l LB medium supplied the same antibiotics was inoculated by a 10 ml aliquot of this culture. This culture was grown at 37°C until OD₆₀₀ reached 0.8, at which point IPTG was added to a final concentration of 0.1-1 mM to induce the expression of protein genes. The culture was grown further for 4 h. Cells were harvested by centrifugation at 5 000 g for 10 min at 4°C in SLA3000 rotor (Sorvall, DuPont).

2.3.3 Extraction of nucleic acids with phenol:chloroform

The standard way to remove proteins from nucleic acids is to extract first with phenol:chloroform and then with chloroform. In the first step, an equal volume of phenol:chloroform (organic phase) was mixed with the nucleic acid sample (aqueous phase) in a polypropylene tube with a plastic cap. Then the mixture was centrifuged at 12,000 g for 15 s in a microfuge at room temperature. The aqueous phase was later on transferred into a fresh tube by pipetting. To achieve the best recovery, the organic phase and interface were mixed well with an equal volume of TE (pH 8.0). The phases were separated by repeating the centrifugation. The aqueous phases were collected together.

2.3.4 Nucleic acid precipitation by ethanol

For precipitation of nucleic acid by ethanol, 2 volumes of ice-cold ethanol (100%, v/v) and 1 volume of nucleic acid solution were mixed with additional salt (sodium acetate, 0.3 M, pH 5.2). The mixture was stored at -20°C for 2 h and was afterwards centrifuged at 12,000 g for 15 min at 0°C. The supernatant was carefully removed and the nucleic acid pellet was washed by 70% (v/v) ethanol and centrifugation. The open tube was stored on the bench at room temperature until the last traces of fluid had evaporated. Finally, the nucleic acid pellet was dissolved in the desired volume of buffer, and was stored at -20°C or -70°C until use.

2.3.5 Dephosphorylation of nucleic acid

Phosphates can be removed from the 5'-ends of nucleic acids by phosphatases. In the present work, this reaction was catalyzed by shrimp alkaline phosphatase (SAP, Boehringer Mannheim) with SAP dephosphorylation buffer in an appropriate volume for 1 h at 37°C. The enzyme was inactivated by 20 min incubation at 70°C.

2.3.6 5'-end labeling of oligonucleotides by [γ -³²P]-ATP

Phosphates can be added back to the 5'-ends of dephosphorylated nucleic acids by bacteriophage T4 polynucleotide kinase (PNK, New England Biolabs, Frankfurt am Main). T4 polynucleotide kinase transfers the gamma-phosphate of ATP to the 5'-

hydroxyl of nucleic acid. If the gamma phosphate is radioactive (γ - ^{32}p), then the nucleic acid is labeled with the radioactive phosphate. This 5'-end labelling of DNA or RNA was performed with PNK forward buffer in 10 μl reaction solution that contained 2 μM DNA or RNA oligonucleotides, 1 μl [γ - ^{32}p]-ATP (10 $\mu\text{Ci}/\mu\text{l}$, 2 μM) and 5 U bacteriophage T4 PNK. The reaction was incubated for 30 min at 37°C. The nucleotides were extracted twice with phenol:chloroform (2.3.3). Afterwards, unincorporated ATP was removed on a sephadex G25 (<50bp) or Sephacryl HR200 (>50bp) spin column if necessary. The labeled oligonucleotide was stored at -20°C.

2.3.7 Hybridization of oligonucleotides

The double-stranded oligonucleotide hybrids (DNA/DNA, DNA/RNA, and RNA/RNA) were formed by a general hybridization method. The mixture of the complementary strands was heated in NEB4 buffer at 90°C for 1 min. Then the mixture was cooled down slowly (0.02°C/s) to 25°C. The hybrids were stored in -20°C.

2.3.8 Small scale preparation of plasmid DNA

2.3.8.1 Growth of the bacterial culture

After transformation (2.3.1), a single *E.coli* XL1-blue (Stratagene, Heidelberg) colony was transferred into 5 ml of LB liquid medium containing appropriate antibiotic(s) in a loosely capped 15 ml tube. The culture was incubated overnight at 37°C with vigorous shaking at 170 rpm in the Environmental Incubator Shaker.

2.3.8.2 Cell harvest and plasmid DNA purification

The *E.coli* XL1-blue cells were harvested by centrifugation at 10,000 g for 30 s at 4°C in a microfuge. After removing supernatant, the cells were lysed and the plasmid DNA was purified based on alkaline method by a commercial kit (E.Z.N.A.[®] Plasmid Miniprep KitII, PeqLab, Erlangen) according to the attached manual. The plasmid DNA was stored at -20°C until use.

2.3.9 Preparation of bacteriophage M13 ssDNA

2.3.9.1 Infection of bacterial *E.coli* XL1-blue cells by bacteriophage M13 ssDNA and culture of the cells

50 μ l of *E.coli* XL1-blue competent cells were infected by 1-50 ng of bacteriophage M13 ssDNA as described in 2.3.1. Afterwards the infected cells were gently mixed with 50 μ l of pre-incubated XL1-blue cells and 5 ml top agar at 47°C. The mixture was poured onto a LB agar plate and was incubated overnight at 37°C. Then one well developed bacteriophage M13 plaque on the top agar was picked and placed into 250 ml LB liquid medium. The culture was incubated at 37°C with vigorous shaking at 170 rpm in Environmental Incubator Shaker for 5-8 h.

2.3.9.2 Harvest of M13 ssDNA

After incubation, the cells were pelleted down by centrifugation at 4,000 g in a SLA1500 rotor (Sorvall, DuPont) for 15 min at 4°C. M13 ssDNA in the supernatant was moved into a beaker, and mixed with polyethylene glycol (PEG) 8000 by stirring for 1 h at room temperature. The precipitate was collected by centrifugation at 10,000 g for 20 min at 4°C. The precipitate was re-suspended in 20 ml of 10 mM Tris-Cl, pH 8.0. Afterwards, the M13 ssDNA was extracted by phenol:chloroform (2.3.3) and was dissolved in a proper volume of 10 mM Tris-Cl, pH 8.0. The final concentration of M13 ssDNA was approx. 1 mg/ml.

2.3.10 Polymerase chain reaction (PCR)

In general, the PCR was performed in a final volume of 20 μ l in the DNA thermal cycler (T-personal Combi, Biometra, Gröttingen) under the indicated conditions (DNA melting at 95°C, annealing at 56°C or 60°C according to the sequence specificity, elongation at 72°C). The reaction mixture contained approx. 5 ng of plasmid DNA or 100 ng of genomic DNA as template, 1 μ M primers, 200 μ M dNTPs and 2.4 U/0.1 U of Tap/Pfu polymerase mixture in PCR buffer.

2.3.10.1 Gene amplification from *Sulfolobus solfataricus* (Sso) P2 genomic DNA

The primers for gene amplification from *Sso* P2 genomic DNA were critically designed and checked by software, VectorNTI (InforMax, Inc., USA) for additional restriction sites and avoiding wrong anchors on the genome sequence. The PCR was performed as described in 2.3.10.

2.3.10.2 Gene amplification from plasmid vector

The primers for gene amplification from plasmid DNA were designed and checked by software, VectorNTI. The PCR was performedal way as described in 2.3.10.

2.3.10.3 Colony PCR

A single colony of *E.coli* cells containing recombinant DNA was picked from a LB agar plate and was placed into 0.5 ml LB liquid medium. The culture was incubated at 37°C for 4 h. Afterwards, 1 µl culture was mixed with 20 picomole of specific primers, 5 micromole dNTPs and 2 U Taq DNA polymerase (Peqlab biotechnologie GmbH, Germany) in PCR bufferII (20 mM Tris-Cl, pH8.55, 16 mM (NH₄)₂SO₄, 0.01 % Tween20, and 2 mM MgCl₂). The final volume was 20 µl. The PCR reaction was carried out as described in 2.3.10.

2.3.10.4 PCR product purification

The PCR product was isolated by agarose gel electrophoresis, and was purified by Wizard SV Gel and PCR Clean-Up System (Promega, Erlangen). In detail, the DNA fragment of interest was excised in a minimal volume of agarose using a clean scalpel. The gel slice was mixed with membrane binding solution at a ratio of 10 µl of solution per 10 mg of agarose gel slice, and the mixture was incubated at 60°C with vortexing till the gel slice was completely dissolved. The dissolved gel mixture was transferred into the SV Minicolumn assembly and was incubated for 1 min at room temperature. After centrifugation at 16,000 g for 1 min, the liquid in the collection tube was discarded. The DNA bound to the Minicolumn was then washed twice by membrane washing solution

with centrifugation (16,000 g) for 1 min after each washing step. The PCR product was released by nuclease-free water or proper buffer (e.g TE buffer) in desired volume (normally 20-60 μ l), and was then transferred into a clean microcentrifuge tube by centrifugation at 16,000 g for 1 min at room temperature. The eluted DNA was finally stored at -20°C until use. The recovery of the DNA was confirmed on an agarose gel.

2.3.10.5 Point mutation by quick change method

The PCR for generation of point mutations by the quick change method was performed in a final volume of 20 μ l in the DNA thermal cycler under the indicated conditions (30 s at 95 °C, 12 cycles of (30 s at 95 °C→1 min at 55 °C→5 min at 68 °C) for 4.5 kb Plasmid). The primers were complementary with 12 to 18 base pairs on both sides of the desired mutation site and carried the the mismatch in center. The PCR mixture contained 0.25 μ M primers, 5 ng plasmid DNA, 3 U Pfu polymerase (Promega, Mannheim) in Pfu buffer.

Following PCR, the original plasmid DNA was digested by 20 U DpnI (New England Biolab, Frankfurt am Main) for 1 h at 37°C in buffer NEB4. 1 μ l of the PCR product was transformed into XL-1 blue competent cells (2.3.1). The mutated plasmid was prepared as described in 2.3.8. The mutation was confirmed by sequencing.

2.3.11 DNA digestion by restriction endonucleases

The digestion of DNA by restriction endonucleases was performed under the conditions clarified by the manufacturer's manual. For analytical purposes, 0.1-0.2 μ g of DNA was digested for 2 h in a volume of 10 μ l, with 1 U of appropriate restriction endonuclease(s). Preparative digestion was carried out in a volume of 20-50 μ l with 2 -5 μ g of DNA and 2-10 U of appropriate restriction endonuclease(s) for over night.

2.3.12 DNA ligation

DNA fragments were ligated by T4 DNA ligase (New England Biolab, Frankfurt am Main) in T4 DNA ligase buffer. For cloning, the ligation of the PCR products was

performed using the pGEM[®]-T Vector System (Promega, Erlangen) following the instructions of the manufacturer.

2.4 Computational analysis of genes from *Sso* P2 Stain

Bioinformatic analysis in Koonin's group (Makarova et al., 2002) revealed a conserved gene context in most archaeal and some bacterial genomes suggesting a previously undetected repair function in prokaryotes. Further analysis in *Sso* P2 strain was completed in the present work including homology analysis and operon analysis.

2.4.1 Homology analysis

All the clusters of orthologous groups (COG) that contain putative DNA repair gene candidates from *Sso* P2 were scanned in NCBI (National Center for Biotechnology Information, Bethesda, USA) database. The putative genes in *Sso* P2 stain were screened out. The homologous genes in *Sso* P2 were grouped and are listed in Table 3.2. Each putative gene was then aligned using BLASTN program on NCBI website (www.ncbi.nlm.nih.gov/blast/blast.cgi) to detect the presence of functional domains.

After the denotation of *cas* genes was available on the NCBI website, all putative repair genes from *Sso* P2 were re-scanned manually in the NCBI database to identify their *cas* gene properties and their *cas* gene numbers.

2.4.2 Gene location and operon analysis in *Sso* P2

According to the putative DNA repair-related gene sequences in *Sso* P2 from NCBI gene database, these gene locations were defined. These genes were divided into few groups (Table 3.3) by comparing their location distances and their predicted functional relationships. In each group the genes are clustered, head-to-tail arranged. The overlap regions between genes were figured out and compared. The putative operons of these genes were predicted as follows: The TATA box of each gene was detected using sequence TTTTAAA around 30 bp upstream of the transcription start site as specified for *Sulfolobus* (Qureshi and Jackson, 1998). The BRE (Transcription Factor B

recognition elements of archaeal promoter) sequence was detected using weakly conserved sequence, RNWAAW (where R=purine, w=A or T, N=any nucleotide) upstream of TATA box specially for archaea (Soppa, 1999; Bell et al., 1999). The Shine-Dalgarno sequence was detected using sequence GGTGA specified for archaea (Tolstrup et al., 2000). The terminator sequence was detected using sequence TTTTTYT (where Y=pyrimidine) specified for archaea (Tolstrup et al., 2000). The tolerance values to typical TATA box, BRE, Shine-Dalgarno and terminator sequences were varied to predict the promoter and terminator regions. All the detection was performed manually in software VectorNTI.

2.5 Preparation of proteins

2.5.1 Overexpression and purification of *SsoSSB*

The plasmid pET19b-*SsoSSB* (a kindly gift from Prof. M. White, University of St. Andrews, UK) containing *ssb* gene was transformed into *E.coli* BL21-CodonPlus(DE3)-RIL (Stratagen, Heidelberg) competent cells as described in 2.3.1. After overnight incubation at 37°C, one well-grown colony was transferred into 20 ml LB medium. The culture was incubated at 37°C with vigorous shaking at 170 rpm in an Environmental Incubator Shaker overnight. Then 4 l LB medium was inoculated by 10 ml overnight culture, and was incubated at 37°C until OD₆₀₀ reached 0.6-1.0 (approx. 4 h). Afterwards, 1 mM IPTG was added to induce the target protein expression. When OD₆₀₀ was >1.6 (approx. 3 h), the cells were pelleted down by centrifugation in a Sorvall SLA3000 rotor at 4,000 g for 10 min at 4°C and were frozen at -20°C until use.

The cell pellet (~30 g) was thawed in 50 ml *SsoSSB* cell lysis buffer and was immediately sonicated (3x3 min, 1 min pause, 50% of maximum duty) with ice cooling in Sonifier B15 (Branson, Bad Lieberzell). The lysate was then centrifuged at 40,000 g for 30 min at 4°C in a Sorvall SS34 rotor. The supernatant was heated to 70°C for 30 min in a water bath (Köttermann, Burgdorf). The denatured proteins were precipitated and removed by centrifugation in a Sorvall SS34 rotor at 40,000xg for 30 min at 4°C.

Afterwards, the supernatant was applied to an EMD-SO₃ column (Merk, Darmstadt) equilibrated with *SsoSSB* EMD buffer. Bound proteins were eluted by a 500 ml linear

gradient of 0-1 M NaCl. Fractions corresponding to a distinct absorbance peak were analyzed by SDS-PAGE. The pooled eluates containing *Sso*SSB and some contaminants were then precipitated by ammonium sulfate at 60% and 95% saturation. The final precipitate was dissolved in 0.1 M (NH₄)HCO₃ and was loaded onto a Sephacryl 300HR size-exclusion column (Pharmacia, Freiburg). The eluted fractions were analyzed on SDS-PAGE. The fractions containing *Sso*SSB were lyophilized and redissolved in 50 mM sodium phosphate, pH 7.5, at a final concentration of 50 μM. Aliquots of 50 μl per tube were shock frozen in liquid nitrogen and were stored at -70°C until use. The purity of the protein was determined on SDS-PAGE.

2.5.2 Solubility analysis of Proteins from *Sso* P2 expressed in *E.coli*

2.5.2.1 Solubility analysis of single *Sso* putative repair proteins during small-scale expression

The general detection steps of *Sso* putative protein was as follows:

The plasmid containing the target *sso* protein gene was transformed into *E.coli* Rosetta(DE3)pLysS (Novagen, Darmstadt) competent cells according to 2.3.1. The *E.coli* cells were grown on plates over night at 37°C under selection of proper antibiotics. Three well-grown colonies were then picked up and were separately added into three wells containing 200 μl LB medium with proper antibiotics on 48-well plate. The plate was shaken overnight at 37°C in Environmental Incubator Shaker with agitation at 170 rpm. Then three samples, each containing 750 μl LB medium and antibiotics were separately inoculated in the three wells by 5 μl overnight culture and were incubated at 37°C with agitation of 170 rpm. 4 h later, two samples out of the three were induced by 1 mM IPTG; one was left without IPTG as control. After 3 h incubation at 37°C with shaking, the cells were transferred into 2 ml microcentrifuge tubes and were pelleted down by centrifugation at 6,000 rpm at room temperature in Biofuge 13R (Heraeus, Hanau). The cell pellet in one of the induced samples was resuspended in 50 μl water and an equal volume of SDS-PAGE loading buffer. Then the sample was heated to 95°C for 5 min immediately before being loaded on SDS-PAGE to obtain the whole cell extract. The cell pellet from the uninduced control was treated in the same way. The cell pellet of the other

induced sample was resuspended in 50 μ l of 1:1 water diluted CellLytic™ B bacterial cell lysis extraction reagent (Sigma, Schnellendorf). The mixture was vortexed for 2 min at room temperature. The cell debris and inclusion bodies were collected by centrifugation at full speed for 15 min in a Biofuge. The supernatant containing soluble proteins was carefully moved into a new microcentrifuge tube and was mixed with an equal volume of SDS-PAGE loading buffer. The precipitate containing inclusion bodies was resuspended by 50 μ l urea (8M) and was mixed with an equal volume of SDS-PAGE loading buffer. After heating at 95°C for 5 min, all the samples and the control were analyzed on SDS-PAGE as described in 2.2.5 yielding the expression levels of the whole cell extract, crude extract (soluble proteins) and insoluble form (inclusion bodies). If necessary, the Western Blot assay was performed after the SDS-PAGE (see 2.5.4).

2.5.2.2 Solubility analysis of *Sso* proteins in a small-scale coexpression system

sso1996, *sso1997*, *sso1998*, *sso1999*, *sso2001* and *sso2002* genes were amplified from *Sso* P2 genomic DNA by PCR (2.3.10.1). The genes were subcloned into duet vectors (Novagen, Darmstadt), *sso1996* and *sso1997* in pRSFDuet-1 (RSF1030 replicon, kanamycin resistance), *sso1998* and *sso1999* in pETDuet-1 (ColE1 replicon, ampicilline resistance), *sso2001* and *sso2002* in pCDFDuet-1 (CloDF13 replicon, streptomycin/spectinomycin resistance), respectively, by DNA digestion (2.3.11) and ligation (2.3.12). The replicons of the duet vectors are compatible. The duet vectors containing the target protein genes were transformed into *E.coli* BL21-CodonPlus(DE3)-RIL or Rosetta(DE3)pLysS strains in 1/2/3 vectors mode for 2/4/6 genes coexpression.

Then the expression and detection of protein genes in the coexpression vectors were performed as outlined in 2.5.2.1.

2.5.3 Protein refolding

2.5.3.1 Protein purification under denaturing conditions

The refolding of misfolded proteins in inclusion bodies was investigated by different assays in the present work. The target protein genes were amplified by PCR (2.3.10) and were subcloned into proper vectors for expression with hexahistidine tag (2.3.11) and

2.3.12). The genes were expressed in 1 l LB medium as described in 2.3.2. The cells were pelleted down by centrifugation at 4,000 g for 10 min at 4°C in Sorvall SLA3000 rotor. The cell pellet was resuspended in 20 ml PBS buffer and cell lysis was performed in Sonifier B15 (3x3 min, 1 min pause, 50% of maximum duty) with ice cooling. The cell lysate was centrifuged at 40,000 g for 30 min in SS34 rotor at 4°C. The supernatant was removed. The precipitate was resuspended in 20 ml PBS with 3 M urea and the dissolved components were removed by centrifugation at 40,000 g for 30 min in SS34 rotor at 4°C. This washing step was repeated. Then the inclusion bodies were dissolved in 20 ml Buffer A with 8 M urea. The insoluble debris was removed by centrifugation at 40,000 g for 30 min in SS34 rotor at 4°C. The supernatant was transferred onto Ni²⁺-NTA-Sepharose, superflow (Qiagen, Hilden) column (4x2 cm) which was pre-equilibrated by Buffer A with 8 M urea. The column was washed by Buffer W with 8 M urea until stable baseline was reached. The denatured target protein was eluted by Buffer E with 8 M urea. The fractions were detected on SDS-PAGE. The fraction containing the target protein was dialyzed against Buffer A over night at 4°C. This protein solution was ready for refolding investigation.

2.5.3.2 On-column refolding assay

The target protein containing a hexahistidine tag in 5 ml Buffer A with 8 M urea was immobilized on Ni²⁺-NTA-Sepharose, superflow column (2x1cm). A linear gradient of urea in buffer A (8 M to 0) was then applied within 2 h. The protein was then eluted by Buffer A containing 200 mM imidazole. Imidazole was then removed by dialysis. The fractions of all steps were analyzed by SDS-PAGE.

2.5.3.3 Rapid dilution refolding assay

The target protein in 5 ml Buffer A, 8 M urea was added drop by drop into 50 ml of Buffer A without urea, containing 5 mM GSH and 1 mM GSSG, 0.4 M arginine. Then the arginine was removed by dialysis. The dialyzed solution was centrifuged at 40,000 g for 20 min at 4°C. The supernatant and the pellet were analyzed by SDS-PAGE.

2.5.3.4 Dialysis refolding assay

The target protein in 5 ml Buffer A, 8 M urea was first dialyzed against Buffer A containing 4 M urea and GSH/GSSG (5 mM/1 mM) for 2 h at 4°C. Then the concentration of urea in the dialysis buffer was decreased to 2 M and GSH/GSSG were replaced by 0.4 M arginine. After 2 h of dialysis, the concentration of urea was further decreased to 1 M in Buffer A. Finally all urea was removed by dialysis against Buffer A. The dialyzed solution was centrifuged at 40,000 g for 20 min at 4°C. The supernatant and pellet were analyzed by SDS-PAGE.

2.5.3.5 High-throughput refolding assay

The 86 refolding mixtures were handmade from the chemicals listed in Table 2.1 and were mixed according to the recipe in Table 2.2 (Vincentelli et al., 2004). A 190 µl aliquot of each refolding solution was dispensed into 86 wells on a 96-well plate. The content of each of the 86 wells was mixed individually with 10 µl of 8 M urea denatured protein (2.5.3.1). Immediately after the mixing step, the turbidity was assessed by measuring the OD at 340 nm using a microplate reader (Microplate Spectrophotometer BioTek Instruments, Inc., USA) and the KC4 software program. The blank, that is the absorbance before adding the protein, was automatically subtracted by the computer program. The protein was considered to be soluble when OD₃₄₀ was smaller than 0.05. The plate was sealed and then stored at 4°C. Twenty-four hours later, the seal was removed and a second reading was performed.

Table 2.1: Chemicals used to make the 86 first refolding buffers.

Buffer (50 mM)	Ionic strength	Amphiphilic	Detergent (100 mM)	Reducing agent (10 mM)	Additive
NaAc, pH 4	NaCl 100 mM	Glycerol 20% (v/v)	NDSB 195	β -MSH	Arginine 800 mM
MES, pH 5	NaCl 200 mM	PEG 4000 0.05% (w/v)	NDSB 201		Glucose 500 mM
MES, pH 6	KCl 100 mM	PEG 400 0.05% (w/v)	NDSB 256		Cocktail
TRIS, pH 7					EDTA 1 mM
TRIS, pH 8					
CHES, pH 9					

The concentrations indicated are those used before adding the protein.

Cocktail consists of 50 µM of each of the following: NADH, thiamine HCl, biotine, CaCl₂, MgCl₂, CuSO₄, ZnCl₂, CoSO₄, ADP and NiCl₂.

Table 2.2 Detailed composition of each well in refolding plate.

	1	2	3	4	5	6	7	8	9	10	11
A	pH 4, β -MSH, Arg	pH4, KCl, β -MSH, NDSB256	pH 5, NaCl 200mM	pH 5, KCl, glucose	pH 6, KCl	pH 7	pH 7, NDSB201, Arg	pH 8, KCl, β -MSH, NDSB201, Arg	pH 8, β -MSH	pH 9, glycerol, NDSB201	GSH 8 mM
B	pH 4, β -MSH, NDSB256	pH 4, β -MSH, NDSB201, glucose	pH 5, NDSB256	pH 5, cocktail	pH 6, NDSB256, glucose	pH 7, KCl	pH 7, PEG4000, β -MSH	pH 8, glucose	pH 8, NDSB201	pH 9, NaCl 100mM, glucose	GSH 5mM, GSSG 2 mM
C	pH 4, NaCl 100mM, β -MSH	pH4, KCl, NDSB195	pH 5, EDTA, β -MSH, NDSB201	pH 6, PEG400, β -MSH, NDSB201	pH 6, glycerol, β -MSH	pH 7, NaCl 200mM, NDSB201	pH 7, NDSB195, glucose	pH 8	pH 9, PEG4000, β -MSH, glucose	pH 9, β -MSH, NDSB195, Arg	GSH 5mM, GSSG 5 mM
D	pH 4, glycerol	pH 4, PEG4000, glucose	pH 5, β -MSH, glycerol	pH 6, PEG400, β -MSH, NDSB195, glucose	pH 6, EDTA	pH 7, glycerol, β -MSH	pH 7, EDTA, β -MSH, NDSB195	pH 8, PEG4000, NDSB195	pH 9, KCl, β -MSH, NDSB201, glucose	pH 9, PEG400	GSH 2mM, GSSG 5 mM
E	pH 4, PEG400, glucose	pH 5, EDTA	pH 5, PEG400, β -MSH, Arg	pH 6, glycerol, NDSB256, Arg	pH 6, NaCl 100mM, β -MSH, NDSB195	pH 7, PEG4000, NDSB256, Arg	pH 7, cocktail	pH 8, β -MSH, glucose	pH 9, β -MSH	pH 9	GSSG 5 mM
F	pH 4, EDTA	pH 5, NaCl 100mM, Arg	pH 5, β -MSH, NDSB256	pH 6, NaCl 200mM, β -MSH, glucose	pH 6	pH 7, NaCl 100mM, NDSB201	pH 8, Arg	pH 8, NaCl 100mM, β -MSH, NDSB256	pH 9, NaCl 100mM, NDSB256	pH 9, EDTA, β -MSH, Arg	GSSG 10 mM
G	pH 4, NDSB201	pH 5, PEG4000, β -MSH, NDSB201	pH 5, glycerol, β -MSH, NDSB195	pH 6, β -MSH, NDSB201, Arg	pH 6, PEG4000	pH 7, β -MSH, Arg	pH 8, EDTA, NDSB256, glucose	pH 8, NaCl 200mM, β -MSH, glucose	pH 9, β -MSH, NDSB195	pH 9, NDSB195, Arg	
H	pH 4, NaCl 200mM, β -MSH, NDSB195, Arg	pH 5, NaCl 100mM, NDSB195	pH 5, β -MSH	pH 6, PEG400, NDSB256	pH 7, PEG400, β -MSH, NDSB256, glucose	pH 7, β -MSH	pH 8, glycerol, β -MSH	pH 8, PEG400, NDSB195	pH 9, NaCl 200mM, β -MSH, NDSB256	pH 9, cocktail	

*, Tris, pH 8, NaCl 150mM, EDTA 1mM

2.5.3.6 Co-refolding assay

The denatured proteins *Sso1998*, *Sso1999*, *Sso2001* and *Sso2002* in Buffer E and 8 M urea were prepared as described (2.5.3.1). These denatured proteins were mixed in a 1:1 molar ratio. Then the mixture solution was dropped into refolding buffer by rapid dilution as described in 2.5.3.3.

2.5.4 Western Blotting

Western blotting is a method to detect a specific protein in a given sample of tissue extract or fractions from protein preparation steps. The proteins that are pre-transferred onto a membrane are probed (detected) using antibodies specific for the target protein. After the unbound probes are washed away, the probe-target protein complexes are coloured by the reaction of specific substrate and reporter enzyme present on the antibody. Protein levels are evaluated through the density of insoluble dye.

The proteins to be analyzed were separated using SDS-gel electrophoresis. They were afterwards blotted from the gel onto the Nitrocellulose Blotting Membrane (Advantec MFS, Inc., Dublin, USA) in the blotting apparatus, PantherTM Semi-Dry Electroloter, HEP-1 (Peqlab, Erlangen) at a current of 1.3 mA/cm² in Blotting buffer for 2 h at room temperature. Then non-specific protein binding to the membrane was blocked by placing

the membrane in 40 ml 1x TBS-T buffer containing 2 g of non-fat dry milk (Roche, Mannheim) for 0.5 h at room temperature. After blocking, 2 µl of primary antibody (Anti-His antibody, 1:1500 diluted, Qiagen, Hilden) was incubated with the membrane in 3 ml of 1x TBS-T buffer with 0.3 % BSA for 2 h at room temperature. After rinsing the membrane to remove unbound primary antibody, the membrane was incubated with 1 µl of secondary antibody (Rabbit anti mouse IgG/AP conjugate, 1:4000 diluted, Pierce, USA) in 4 ml 1x TBS-T for 1 h at room temperature. Following washing away of the secondary antibody from the membrane, the target protein-antibody complex was coloured by 100 µl of 5 % NBT (Nitro blue tetrazolium chloride, Sigma, Schnellendorf, in 70% dimethylformamide) and 100µl50mg/ml BCIP (5-bromo-4 chloro-3-indolyl phosphate, Sigma, Schnellendorf, in 100% dimethylformamid) and 25ml Ap buffer for a few minutes at room temperature. The amount of target protein was evaluated by being compared with the positive protein marker.

2.5.5 Preparation of *Sso2001Est* fusion protein

2.5.5.1 Cloning, expression and purification of *Sso2001Est* fusion protein

sso2001 gene was amplified by PCR from *Sulfolobus solfataricus* P2 genomic DNA (2.3.10.1) using forward primer S2001-for (ACATATGTTGATCAAGCCTTGCGCTTA) and reverse primer S2001-rev (ACGAGCTCTAGAGTGGAACCTCCAT). The primers created the *NdeI* and *SacI* restriction sites (underline letters) upstream and downstream of the *sso2001* gene, respectively. The identity of the PCR fragment was confirmed by sequencing.

The *sso2001* gene was inserted into the multi-cloning sites of pIVEX2.3d-Est2.1 vector (kindly gift from Prof. Mathias Sprinzl, Universität Bayreuth) between the *NdeI* and *SacI* cleavage sites. Then the fused fragment of *sso2001* gene plus esterase gene was cut out by *NdeI* and *BamHI* from this vector and inserted into plasmid pET-28c(+) behind the hexahistidine codons to construct the expression vector.

The expression vector containing recombinant *sso2001Est* gene was transformed into *E. coli* Rosetta(DE3)pLysS strain as described (2.3.1). Then the expression conditions were optimized to enhance the soluble protein expression level. In details, after

inoculation with 10 ml of transformed, overnight cultivated cells, the LB medium was incubated at 37°C till OD₆₀₀ reached 0.8~1.0, thereafter, the medium was immediately cooled down to 30°C and the expression was induced by 0.1 mM IPTG. When OD₆₀₀ was > 1.5, the cells were harvested by centrifugation (Sorvall SLA3000 rotor, 4,000 g, 10 min, and 4°C) and kept frozen. The culture was under kanamycin and chloramphenicol selectivity pressure. The full length protein was determined both by western blot (2.5.4) for C-terminal hexahistidine and active staining for N-terminal esterase (2.5.5.2).

The cell pellet was thawed in cell lysis buffer containing 50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 10% glycerol and 2 mM β-mercaptoethanol. The resuspended cells were treated by sonication in Sonifier B15 (3x1 min, 2 min pause, 50% of maximum duty) in an ice bath. The soluble proteins were separated from cell homogenate by centrifugation (Sorvall SS34 rotor, 12,000 g, 30 min, and 4°C). The supernatant was then loaded onto the Trifluoromethyl ketone (TFK) Sepharose CL-6B affinity column which specifically binds to the esterase (Huang et al., 2007). The target protein was eluted by 20 mM of soluble inhibitor, 3-Butylsulfanyl-1,1,1-trifluoro-propan-2-one, at a flow rate of 0.2 ml/min. The pooled fractions were dialyzed against storage buffer containing 50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 50% glycerol and 1 mM DTT. The aliquots were shock frozen in liquid nitrogen and were then stored at -70°C until use. All the expression and purification steps were monitored by esterase activity detection assay (2.5.5.2).

2.5.5.2 Esterase activity detection assay

Esterase activity was assayed with β-naphthol acetate (Sigma, Schnelldorf) as substrate (Higerd and Spizizen, 1973). In detail, after SDS-PAGE (2.2.2.4), the protein gel was washed in 100 ml 0.1 M Tris, pH 7.5 for two times of 10 min. Then the gel in 100 ml 0.1 M Tris, pH 7.5 was mixed with 1 ml of β-naphthol acetate (0.1 mg/ml in ethanol) and 0.1 mg of Fast Blue BB Salt (Sigma, Schnelldorf) at room temperature. After 20 min, the gel was washed two times by 100 ml 0.1 M Tris, pH 7.5. The resulting pigment was fixed by shaking the reaction mixture in 50 ml of 30 % (v/v) methanol, 10 % (v/v) acetic acid.

2.5.6 Preparation of *Sso1450C6H* fusion protein

The *ssol450* gene was amplified by PCR from *Sulfolobus solfataricus* P2 genomic DNA (2.3.10.1) using primers S1450-for (CATGCCATGGGCGTGATAAGCGTGAGGACTTT) and S1450-rev (GCCGAATTC^uCCCATCACCAACTTGAAACCCC). The primers created the *NcoI* and *EcoRI* restriction sites (underline letters) upstream and downstream of the *ssol450* gene, respectively. The sequence of PCR fragment was proven to be correct by comparing it to the putative protein gene sequence in *Sulfolobus solfataricus* P2 genome.

The *ssol450* gene was inserted into the multi-cloning sites of plasmid pET-28c(+) behind the hexahistidine codons to construct the expression vector. The plasmid pET-28c(+) containing *ssol450* protein gene fused with C-terminal hexahistidine tag (named *ssol450C6H*) was transformed into *E.coli* Rosetta(DE3)pLysS strains as described in 2.3.1. The transformed strains were incubated over night at 37°C on LB plate. Afterwards, one well-grown colony was picked and was mixed with 20 ml LB liquid medium in a 150 ml flask. The flask was shaken over night at 37°C. 4 l LB medium was inoculated by 20 ml overnight culture and was incubated at 37°C for approx. 4 h till OD₆₀₀ reached 0.8~1.0. Then the temperature was immediately decreased down to 30°C. The protein expression was induced by addition of 0.1 mM IPTG. The culture was under kanamycin and chloramphenicol selectivity pressure. After 3 h incubation, OD₆₀₀ typically was above 1.6, and the cells were harvested by centrifugation (Sorvall SLA3000 rotor, 4,000 g, 10 min, and 4°C).

The cell pellet (~1g) was thawed in 10 ml cell lysis buffer containing 50 mM Tris-Cl, pH 7.5, 150 mM NaCl, 20% glycerol, 2 mM β-mercaptoethanol and 10 mM imidazole in ice bath and was immediately sonicated in Sonifier B15 (3x1 min, 2 min pause, 50% of maximum duty). Afterwards, insoluble proteins and cell debris were removed by centrifugation (Sorvall SS34 rotor, 12,000 g, 30 min, and 4°C). The supernatant was loaded onto a 1 ml Co²⁺-Talon metal affinity column (Clontech, Heidelberg). The column was washed by cell lysis buffer containing 50 mM imidazole until the base line was reached. Target protein was eluted by addition of 150 mM imidazole. The eluate was analyzed by SDS-PAGE. The pooled fractions were dialyzed against storage buffer

containing 50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 50% glycerol and 1 mM DTT. The aliquots were shock frozen in liquid nitrogen and were then stored at -70°C until use.

2.6 Nucleic acid-protein interaction

2.6.1 Electrophoretic mobility shift assay (EMSA)

The Electrophoretic Mobility Shift Assay is a method for studying DNA-protein interactions *in vitro*. DNA-protein complexes migrate more slowly than unbound radioactive probe and are thus visualized by discrete bands of radioactivity on a native gel image.

In the present work, after DNA-protein complex formation, the samples were mixed with one third volume of native gel loading buffer. 3-7 μ l of the mixtures were loaded on 4% polyacrylamide native gel or 1% agarose gel and then gel running was performed as described (2.2.2.2). The gels were detected as described in 2.2.2.3 and 2.2.2.4.

2.6.1 Nuclease assay

For nuclease activity detection, 10 nM of DNA or RNA oligonucleotide substrates were incubated with a proper amount of protein in buffer containing 50 mM Tris-Cl, pH 7.5, 10 mM MgCl₂ in a total volume of 10 μ l for 20 min at 50°C. The nuclease activity was quenched by denaturing gel loading buffer containing 20 mM EDTA and 0.2% SDS at 4°C. Samples were heated to 95°C for 5 min immediately before electrophoresis on 20% denaturing polyacrylamide gel. The denaturing PAGE was carried out as described in 2.2.2.3 and the 5'-end labeled oligonucleotides were detected as described in 2.2.3.

2.6.2 ATPase assay (thin layer chromatography assay)

In ATPase reactions, 2-5 μ l [γ -³²P]-ATP (approx. 40 nM) was mixed with 1 μ l ATPase candidate protein and possibly stimulating nucleic acids in ATPase reaction buffer. The volume was adjusted by water to a total of 10 μ l. The solution was incubated at 37°C or 50°C for 20 min. Then the samples (2 μ l of each sample) were spotted on PEI-Cellulose plastic sheet (Macherey-Nagel, Düren). The sheet was developed in 1 M formic

acid and 0.8 M lithium chloride. After chromatography, the sheet was dried by a blow dryer and the products were analyzed on a Typ Instant Imager 2024.

2.6.3 Fluorescence anisotropy assay

A fluorophore in solution, in the present work, fluorescein, displays a rapidly rotational motion relative to its fluorescence lifetime. Upon excitation with vertically polarized light the emitted fluorescent light is completely depolarized. Coupling of the chromophore to a larger molecule, e.g. the oligonucleotide probe reduces its rotational mobility resulting in an increased anisotropy. Upon binding of the fluorescently labeled oligonucleotide probe to a large binding protein, the anisotropy of fluorescence is further increased which serves as an indicator of complex formation. The anisotropy (r) is defined as the difference between the fluorescence intensity emitted parallel and perpendicular (I_{\parallel} and I_{\perp}) divided by the total intensity (Figure 2.1). Fluorescence anisotropies were calculated from fluorescence intensity measurements employing a vertical excitation polarizer and vertical and horizontal emission polarizers according to:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

The anisotropy values were corrected with the experimentally determined grating factor using fluorescein in the appropriate assay buffer as depolarizing sample. The fluorescence anisotropy and the fluorescence intensity of free fluorescein do not change upon addition of nucleotide binding proteins.

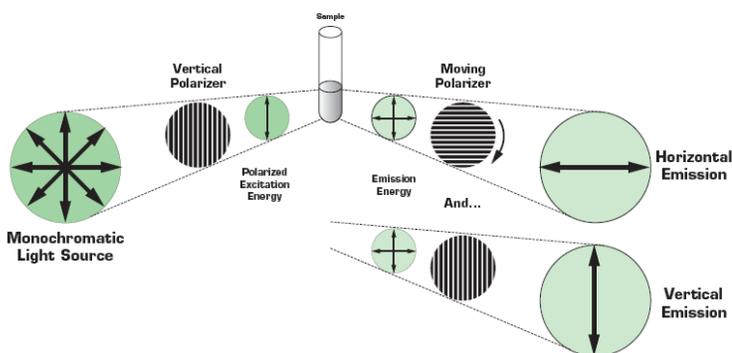


Figure 2.1: Schematic representation of fluorescence anisotropy detection. Monochromatic light passes through a vertical polarizing filter and excites fluorescent molecules in the sample cuvette. The emitted light is measured as fluorescence intensity in both the horizontal and vertical planes (PanVera Co., USA).

2.6.3.1 Fluorescence anisotropy measurements

All experiments were performed using a LS50B spectrofluorometer (Perkin-Elmer, USA) equipped with a polarization device and a thermostat jacket at 25°C. Excitation and emission bandwidths were adjusted to 10 and 20 nm, respectively. Fluorescence titrations using fluorescein-labeled oligonucleotides were performed at an excitation wavelength of 495 nm with a vertical polarizing filter and monitored at an emission wavelength of 525 nm using a 515 nm cutoff filter. Following sample equilibration, at least 3 data points with an integration time of 15 s were collected for each titration point.

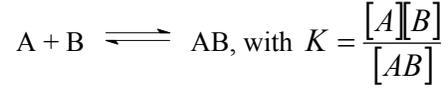
Fluorescence anisotropy measurements with nucleotide binding protein were performed in fluorescence anisotropy buffer. The salt addition was necessary to increase the dissociation constant to DNA concentrations where the signal-to-noise ratio was good enough to allow for reliable data collection (typically 2-100 nM, depending on the cuvettes used, concentrations are given in the results section). Protein was normally titrated in 10x4 mm cuvettes (1 ml) with stirring; Titrations for the determination of dissociation constant values were carried out in 10x2 mm microcuvettes in a sample volume of 150 μ l.

2.6.3.2 Competition titration

Changes in fluorescence intensity might result from interactions between protein and the fluorescein moiety which could influence the strength of complex formation. Therefore it was necessary to compete a fluorescent protein-DNA complex with the unlabeled DNA. Furthermore, some RNA substrates were not available in a fluorescein-labeled form. Competition titrations were carried out in these cases. In these experiments, the fluorescent probe alone was titrated with protein until 80% of nucleotide was bound. Increasing amounts of unlabeled probe were then added until the anisotropy remained constant. The competition titrations were performed in 10x2 mm microcuvettes in a sample volume of 150 μ l.

2.6.3.3 Data analysis

Anisotropy mean values were plotted against the protein concentration and fitted using a simple one-site binding model:



where [A] represents the concentration of free protein, [B] the free DNA concentration, and [AB] the concentration of the respective protein-DNA complex.

In case of the fluorescence intensities of free and protein bound DNA probe being the same, the anisotropy was calculated from:

$$r = \frac{[AB]}{[B_0]} r_{AB} + \frac{[B_0] - [AB]}{[B_0]} r_B$$

$$[AB] = \frac{[A_0] + [B_0] + K}{2} - \frac{\sqrt{([A_0] + [B_0] + K)^2 - 4[B_0]K}}{2}$$

with [A₀] being the total protein concentration, [B₀] the total DNA concentration, *r_B* the anisotropy of free DNA, and *r_{AB}* the anisotropy of the protein-DNA complex.

When the total fluorescence intensity changes during the titration experiment, the fractional fluorescence intensities *f_B* of the free DNA and *f_{AB}* for the protein-DNA complex have to be considered according to:

$$r = f_{AB} r_{AB} + f_B r_B$$

$$f_B = \frac{([B_0] - [AB])s}{([B_0] - [AB]) + [AB]s}$$

$$f_{AB} = \frac{[AB]s}{([B_0] - [AB]) + [AB]s}$$

with *s* representing the intensity ratio of the complex versus free probe.

Data sets were least-squares fitted with the software program Origin (OriginLab Co. Northampton, MA, USA) using the Levenberg-Marquardt algorithm (Hey et al., 2001). The curves and data points shown in the figures and tables are based on averaged fits obtained from the recalculation of triplicate measurements. The error range for each titration point is omitted from the figures to enhance clarity.

2.6.4 Atomic force microscopy (AFM)

AFM measurements were performed in the tapping mode. Thereby the cantilever is excited to the resonance oscillation with a piezoelectric driver. On the end of the cantilever an atomically sharp tip is mounted. A feedback loop controls the spatial movement of a piezoelectric positioner with Angstrom accuracy (operates in ambient environment). Oscillating tip is scanned over a surface with feedback mechanism that enables the piezoelectric scanners to keep the tip at the constant force (that is why AtomForce Microscopy) (Figure 2.2). In the present work, AFM was introduced for investigating the formation of DNA-*Sso*SSB complex.

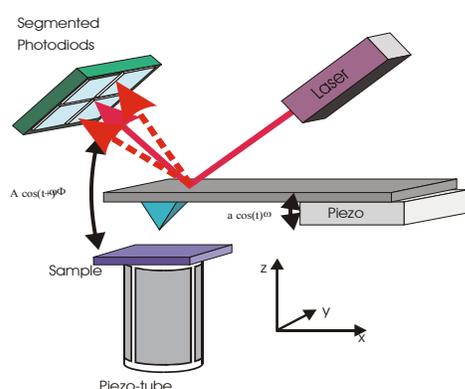


Figure 2.2: Tapping mode Atomic Force Microscopy. Tapping mode Atomic Force Microscopy Segmented photodiodes trace displacements of the cantilever due to the reflected laser beam. An atomically sharp tip is mounted on the backside of the cantilever. Due to the feedback loop cantilever with the constant amplitude (force) scans softly over the sample, achieving only intermittent tip-sample contact.

2.6.4.1 DNA substrate for AMF

A 359 bp dsDNA was obtained by PCR from the phage M13 ssDNA using primers, 5'-end biotinylated M13AFM-for (CCGGGATCCGAAGGGATATCAGCTGTTGCCCGTC) with additional *Eco*RV restriction site (underline letters) and M13AFM-rev (AGCTTCCGGCACCGCTTCTGGTGCCGGA). A smaller fragment was produced through the cleavage of the 359 bp dsDNA by the restriction enzyme, *Pvu*I (restriction site was derived from original M13 ssDNA) as described in 2.3.11. A 278 bp fragment

with biotin at one end and an 81 bp fragment without biotin were obtained in this way. The 287 bp dsDNA was isolated by agarose gel. Both 278 bp dsDNA and 359 bp dsDNA (approx. 50 pmole) were captured by streptavidin on 1 mg streptavidin coupled magnetic Dynabeads (Invitrogen DYNAL AS, Norway) in 1.5 ml microcentrifuge tubes with binding & washing buffer, respectively. The immobilization of 378 bp dsDNA was carried out by gentle rotation at room temperature. After 10 min incubation, dsDNAs were denatured by 0.1 M NaOH. Then the ssDNA (with biotin at one end)-containing beads were separated from solution by DYNAL MPC (Magnetic Particle Concentrator, Invitrogen DYNAL AS, Norway). The supernatant containing ssDNA (without biotin) was removed to new tubes. The NaOH in the supernatant was neutralized by 0.1 M HCl. The same procedure was performed on 278 bp dsDNA. The beads containing 359 nt ssDNA was then mixed with the supernatant containing 278 nt ssDNA. Hybridization of both strands was completed as described (2.3.7) and the renatured strands were cleaved from the beads by *EcoRV* in NEB3 buffer and BSA (100 µg/ml) producing a 261 bp dsDNA with an 81 nt ssDNA tail. The hybridized DNA was separated from the beads by DYNAL MPC and was purified by 2% agarose gel (2.2.2.1) and Wizard SV Gel and PCR Clean-Up System (2.3.10.4).

2.6.4.2 AFM measurements

AFM measurements were performed on a Multi Mode AFM Digital Instruments (Santa Barbara, CA) operated in tapping mode. Silicon nitride oxide sharpened tips were used. The drive frequencies ranged between 3.4 and 34 kHz. All measurements were performed in AFM buffer containing 8 mM HEPES, pH 8, 2 mM NaCl and 2 mM MgCl₂. The tailed DNA and *SsoSSB* protein were pre-incubated at a molar ratio of 1:3 in AFM buffer for 1 min. A control without protein was handled in the same way. Immobilization of the biomolecules on freshly cleaved mica was achieved via 2.5–5.0 mM NiCl₂ without further rinsing of the samples and omitting any drying. Images were kindly recorded by Dr. Marina Lysetska, Lehrstuhl Physikalische Chemie II, Universität Bayreuth.

3. Results

3.1 DNA Binding Properties of Single-stranded DNA binding protein from *Sso* P2 (*SsoSSB*)

Single stranded DNA binding proteins (SSBs) are ubiquitous in all three domains of life. These proteins participate in many cellular processes including DNA replication, DNA repair, transcription, recombination, etc. They share many common features, but differ in specific aspects. Eukaryotic and bacterial SSBs are well studied whereas the biochemical properties of archaeal SSBs are only poorly characterized. The SSB from *Sulfolobus solfataricus*, *SsoSSB*, can be considered as a representative SSB in the crenarchaea branch of the archaea domain. The *SsoSSB* gene codes for a 148 aa protein of 16 kDa. In the following section the major DNA binding properties of *SsoSSB* are described.

3.1.1 Overexpression of *ssoSSB* gene

The *ssoSSB* gene from *Sso* P2 which had been previously cloned in pET19b vector (a gift from Prof. M. White, University of St. Andrews) was overexpressed from the host strain, *E.coli* BL21(DE3)CodonPlus-RIL in LB medium. The expression level was followed by SDS-PAGE of the whole cell extract (see 2.5.1 and Figure 3.1).

3.1.2 Purification of *SsoSSB* protein

After cell harvest, the target protein was purified as follows: First, *E.coli* cells was thawed in an ice bath and immediately sonicated in Sonifier B15 (Branson, Bad Lieberzell). Following a heat shock at 70°C, most *E.coli* proteins precipitated and were then removed, together with the cell debris, by centrifugation. The supernatant was filtered and loaded on a cation exchange column (EMD-SO₃, Merk). The target protein was eluted by applying a salt gradient (0~1M NaCl). In other experiments, it was found that most of the contaminants precipitated at 60% saturation of ammonium sulfate whereas *SsoSSB* remained in solution. Therefore, the eluate from cation exchange column was precipitated by 60% ammonium sulfate. The remaining protein was

precipitated by 95% ammonium sulfate and was dissolved in 0.1 M $(\text{NH}_4)\text{HCO}_3$. Size-exclusion chromatography (Sephacryl S-300HR, Amersham Biosciences, Pharmacia) was then applied for further purification using 0.1 M $(\text{NH}_4)\text{HCO}_3$ as running buffer. Finally, the target protein was lyophilized, redissolved in 50 mM sodium phosphate, pH 7.5, and was shock frozen in liquid nitrogen. The protein aliquots were stored at $-70\text{ }^\circ\text{C}$ until use. This purification method was based on a published procedure (Wadsworth and White, 2001) with ammonium sulfate precipitation as an additional purification step. The protein recovery of purification steps was monitored by SDS-PAGE (Figure 3.1). The final concentration as determined by Bradford assay (2.2.1.1) was about $50\text{ }\mu\text{M}$. The yield was approx. 4.1 mg SsoSSB/L of cell culture.

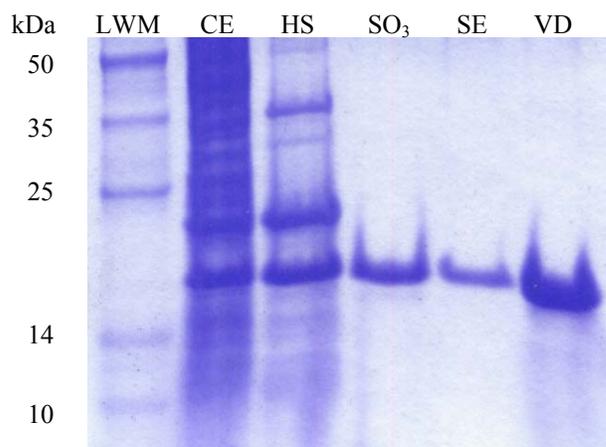


Figure 3.1: purification of *SsoSSB* protein (148aa, 16 kDa).

The fractions from purification steps were monitored by 12% SDS-PAGE. The electrophoresis was run in Schagger-Jagow buffer. kDa, kilodalton; LWM, low molecular weight protein marker; CE, crude extract; HS, supernatant of CE after heat shock and centrifugation; SO_3 , peak eluate from EMD- SO_3 cation exchange column; SE, peak eluate from size-exclusion chromatography column; VD, protein redissolved after vacuum drying.

3.1.3 Characterization of *SsoSSB*

The DNA binding specificity of *SsoSSB* was followed by native gel electrophoresis (2.2.2.2) and by fluorescence measurements (2.6.3) using plasmid DNA and synthetic oligonucleotides as binding substrates.

3.1.3.1 Gel retardation of DNA-protein complex

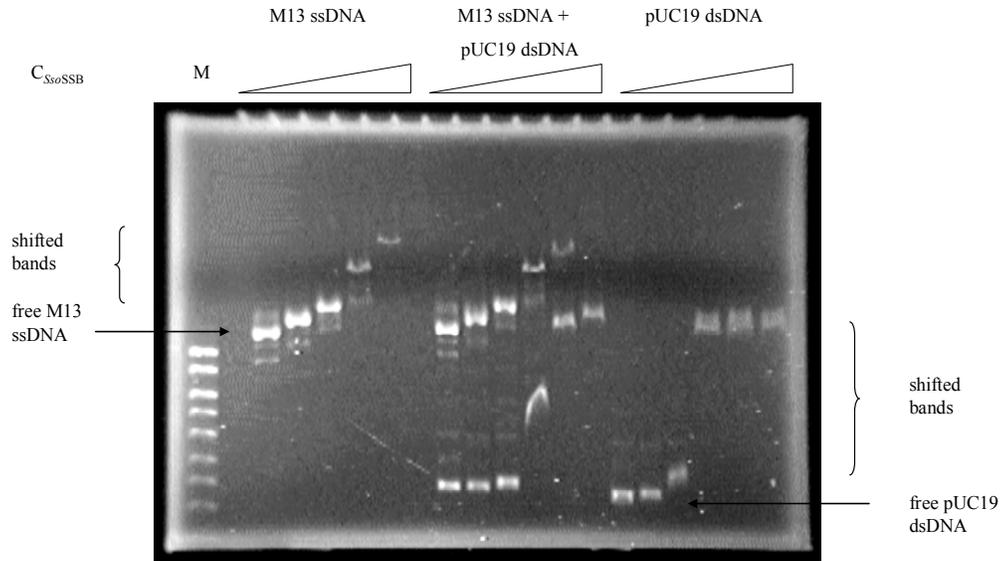


Figure 3.2: Electrophoretic mobility shift assay of DNA-*SsoSSB* complexes on 0.7% agarose gel. Concentrations of *SsoSSB* were varied as 0, 0.1, 0.2, 0.5, 1 and 4 μg . The concentrations of M13 ssDNA and of pUC19 dsDNA were fixed as 100ng. After being incubated in *SsoSSB* binding buffer at 25°C for 10 min, the samples were loaded on agarose gel with native gel loading buffer. The electrophoresis was performed in 0.5x TBE buffer. The DNA strands were stained by ethidium bromide. The free DNAs were arrowed, and the DNA-Protein complexes were indicated by brackets. M, DNA marker (smart ladder, Invitrogen, Germany)

The binding of *SsoSSB* to the large single-stranded (ss) and double-stranded (ds) DNAs was investigated by native agarose gel electrophoresis. Various concentrations of purified *SsoSSB* protein were incubated with 100 ng of circular bacteriophage M13 ssDNA molecules or 100 ng of plasmid pUC19 dsDNA in 10 μl total volume of *SsoSSB* binding buffer. After 10 min incubation at 25°C, the samples were analyzed by agarose gel electrophoresis (2.2.2.1). Protein binding to M13 ssDNA was indicated by a gradual reduction of the mobility of the M13 ssDNA with increasing SSB concentrations, until an end-point was reached where all available ssDNA binding sites were apparently saturated by protein molecules. This end-point coincided with a dramatic reduction in fluorescence intensity after staining with ethidium bromide. On the contrary, the affinity of *SsoSSB* to pUC19 dsDNA was much lower than to M13 ssDNA. The concentration of *SsoSSB*

required for a visible shift of the pUC19 dsDNA was more than 10 fold higher than required for M13 ssDNA. Within a narrow concentration range, apparent saturation of the dsDNA was observed. In these complexes, dsDNA molecules could still be efficiently stained by ethidium bromide (Figure 3.2). When both M13 ssDNA and pUC19 dsDNA were offered as binding substrates, the ssDNA was strongly preferred. This result clearly indicates that *SsoSSB* recognizes preferentially ssDNA. The binding of *SsoSSB* to the dsDNA may be explained by the presence of partially single-stranded regions on the plasmid DNA and/or by a protein-induced melting of the double-stranded DNA. By calculating the molar ratio of DNA nucleotides:*SsoSSB* molecules at the end-point of saturation, the binding site size of *SsoSSB* on M13 ssDNA is estimated as 4-6 nt per SSB monomer.

The binding of *SsoSSB* to short ssDNA oligonucleotide was followed by electrophoretic mobility shift assay (EMSA) on 4% native PAGE. Various concentrations of *SsoSSB* and 10 nM of 5'-end γ^{32} -ATP labeled single-stranded oligonucleotide substrates of 15, 24 and 42 nt were incubated for 10 min at 25°C in *SsoSSB* binding buffer. The solution was then analyzed on the native gel and DNA bands were visualized by an instant imager. The complexes formed a ladder pattern on native PAGE gels (Figure 3.3). When 15 nt ssDNA oligonucleotide was offered as a binding substrate, only one ssDNA-protein complex formed (Figure 3.3A). For a 24 nt ssDNA oligonucleotide, two complexes were observed (Figure 3.3B) whereas for 42 nt (Figure 3.3C), three distinct complexes were formed. The presence of the distinct bands with 42 nt ssDNA implies that there is little or no cooperativity in the binding event under these conditions. The shifted bands for pyrimidine-rich substrates were visible from 0.05 to 0.1 μ M, whereas for purine-rich substrates, concentrations of 0.2 to 1 μ M *SsoSSB* were needed to produce visible complexes. This indicates a higher affinity of *SsoSSB* for pyrimidine-rich ssDNA as compared to purine-rich ssDNA.

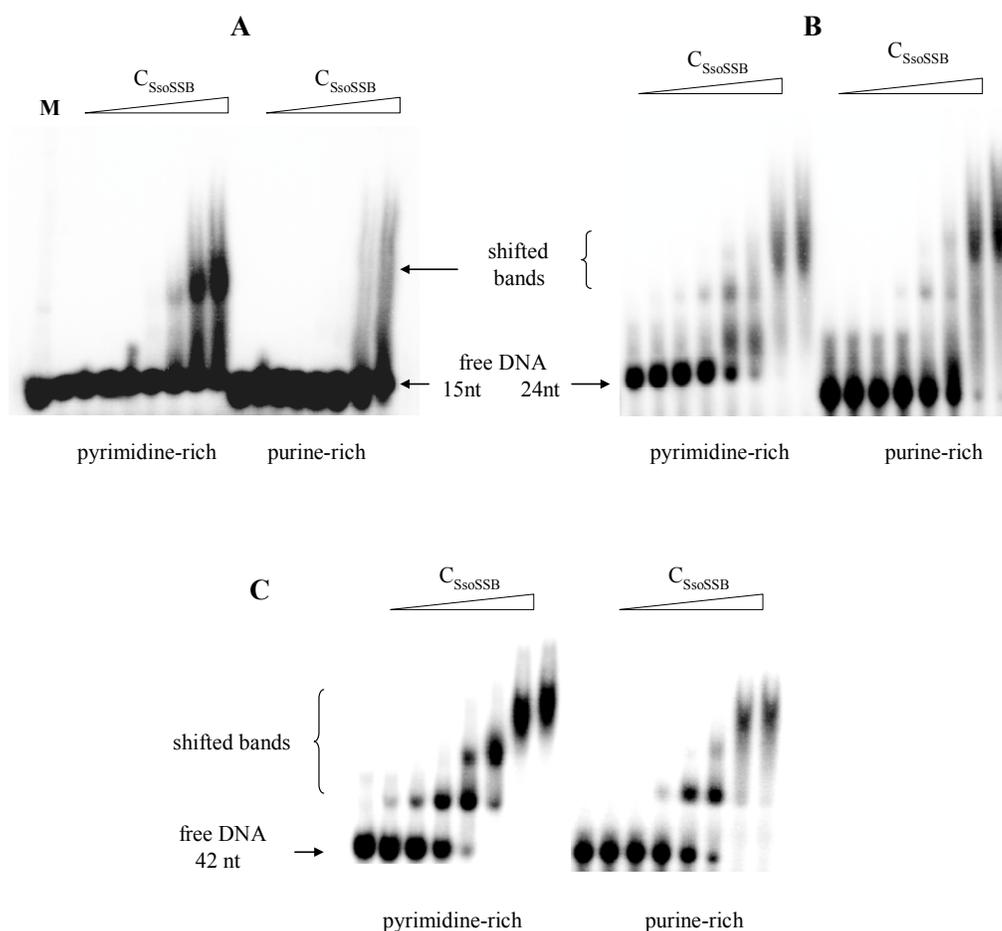


Figure 3.3: Electrophoretic mobility shift assay of *SsoSSB* - ssDNA oligonucleotide complexes on 4% native PAGE gel. The concentration of ssDNA oligonucleotide was 100 nM. The concentration of *SsoSSB* was increased from 0, 0.05, 0.1, 0.2, 0.5, 1, 4, to 8 μ M. **A**, binding of *SsoSSB* to 15 nt ssDNAs. **B**, binding of *SsoSSB* to 24 nt ssDNAs. **C**, binding of *SsoSSB* to 42 nt ssDNAs.

3.1.3.2 Binding complex detection of *SsoSSB* by AFM

Atomic force microscopy (AFM) has been successfully applied for the study of the interaction of hRPA (human replication protein A) with damaged DNA substrates (Lysetska et al., 2002). This method allows the imaging of complexes in aqueous solution and may provide information on the dimensions and in some cases on the stoichiometry of the complexes. A 261 bp dsDNA fragment with 81 nt single-stranded (ss) tail was prepared for AFM as described (2.6.4.1, Figure 3.4A).

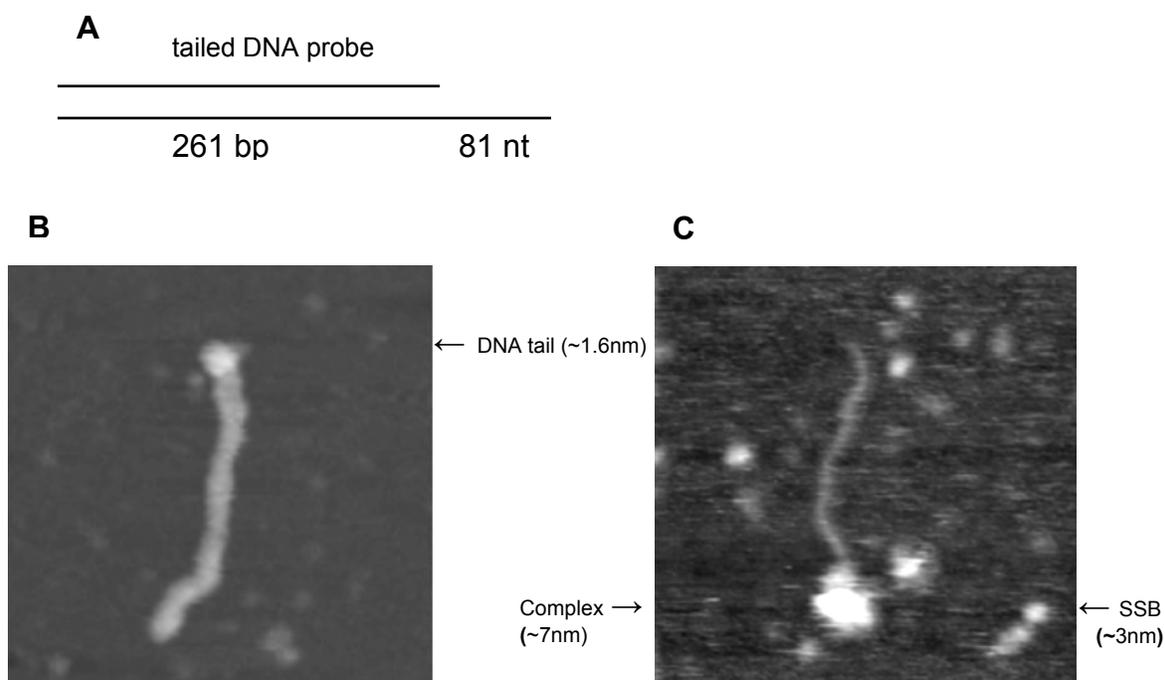


Figure 3.4: AFM images of ss-tailed dsDNA in the presence and absence of *SsoSSB*. **A**, schematic of ss-tailed dsDNA. **B**, ss-tailed DNA without protein. **C**, ss-tailed dsDNA in the presence of *SsoSSB*. The variation of brightness indicates a height of to 20 nm. The arrows indicate the ssDNA tail, the single protein molecule and the protein-ss-tailed dsDNA complex. (Images were recorded by Dr. Marina Lysetska, Lehrstuhl Physikalische Chemie II, Universität Bayreuth.)

In control experiments without *SsoSSB*, the tailed DNA probe (1-3 nM) in deposition buffer was placed on freshly cleaved mica. After 1 min for DNA equilibration on the mica surface, nickel chloride was added to the sample at a final concentration of 2.5 nM. It proved important to keep the buffer around pH 8 to achieve efficient DNA attachment to the mica surface. As Figure 3.4B shows, the coil ssDNA tail and linear dsDNA can be identified in the AFM images. The single strand part presents as a globular structure of ~1.6 nm of height which indicates a coiling and collapsing of the single-strand on the mica surface. For the ssDNA straightly stretched on the surface, one would expect a greater length and a height of ~ 1 nm.

For the AFM studies of the complexes, the ss-tailed DNA probe and *SsoSSB* were pre-incubated at a molar ratio of 1:3. The complexes were placed on the mica surface and were then immobilized by addition of nickel chloride. AFM images displayed at this low molar ratio of DNA and protein showed the formation of globular complexes on the DNA ss-tailed termini. The complexes (~7 nm in height) are obviously larger than the free

ssDNA tails (Figure 3.4C). This observation directly shows that *SsoSSB* possesses high affinity to the ssDNA region. The comparison of brightness of the ssDNA tail termini with or without protein molecules in the AFM images implies that several *SsoSSB* molecules are involved in ssDNA binding event (Figure 3.4B and C).

3.1.3.3 Stoichiometry of *SsoSSB*-ssDNA complexes as followed by fluorescence anisotropy

Fluorescence anisotropy assays the rotational diffusion of a molecule or molecule complex from the decorrelation of polarization in fluorescence, between the exciting and emitted fluorescent photons (Hey et al., 2001). The fluorescence anisotropy can be calculated from the formula:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

Where the r term indicates anisotropy, the I terms indicate intensity measurements parallel (I_{\parallel}) and perpendicular (I_{\perp}) to the incident polarization.

In the present experiments, synthetic DNA probes were used that carry a fluorescein label at the 5'-end. The experiment was performed at a concentration well above the binding constant (100 nM, 13 nt ssDNA) and therefore can be used to determine the stoichiometry of the complexes. The binding curve in Figure 3.5 shows an increase in anisotropy (r) from an initial value of 0.02 to approx. 0.14 which indicates the formation of the DNA-protein complexes. Following a linear increase of anisotropy, a sharp saturation is observed in the titration curve. By extrapolating the two linear parts of the binding curve, the stoichiometry of the complexes is obtained from the intersection point. The intersection point in Figure 3.5 corresponds to a *SsoSSB* concentration of 110 nM which is compatible with a 1:1 ratio, as one monomer of *SsoSSB* binds to one molecule of 13 nt ssDNA. The result also indicates that more than 90% of the *SsoSSB* protein is active in DNA binding. This DNA binding activity of *SsoSSB* was found to be stable for at least 6 months. Therefore in all following fluorescence anisotropy experiments, the given *SsoSSB* concentrations represent active protein concentrations.

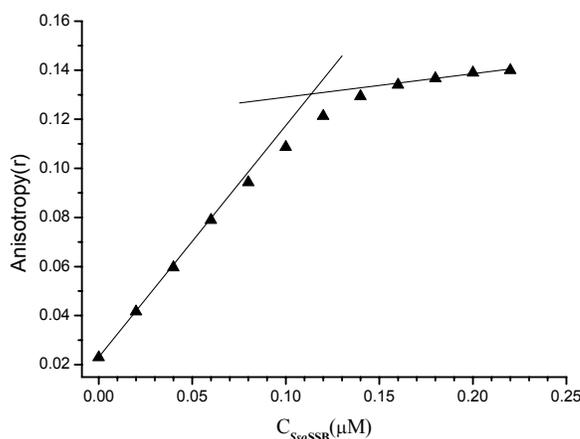


Figure 3.5: Stoichiometric titration of *SsoSSB*. Fluorescence anisotropy of a 13 nt ssDNA probe (100nM) in *SsoSSB* binding buffer without salt (50 mM NaCl) added upon titration with *SsoSSB*. The solid lines represent the linear regression curves obtained from the concentration range of 0-60 nM and of 160-220 nM of *SsoSSB*. The intersection yields the equivalence point of the titration (~110nM).

3.1.3.4 Dissociation constant of *SsoSSB*-ssDNA complexes

Given the 1:1 stoichiometry for the binding to a 13 nt oligonucleotide, it is important to know how many nucleotides are minimally required for a strong binding of *SsoSSB* to DNA. To determine the binding site size of the *SsoSSB*-DNA complexes, the binding of oligonucleotides of different lengths was investigated by fluorescence anisotropy measurements. Data analysis was greatly simplified under a 1:1 binding mode. The lengths of ssDNA probes varied from 7 to 50 nt to test the optimal length of substrate for 1:1 binding mode. Analysis of the binding curves showed that with a 13 nt ssDNA probe, the dissociation constant of 0.48 nM (Figure 3.6A) was strongest. A probe of only 7 nt was bound considerably weak. Interestingly, the dissociation constants of probes longer than 13 nt also increased strongly. The 52 nt probe was bound nearly 50 times weaker as the 13 nt probe (Table 3.1). This observation indicates that more than one *SsoSSB* molecule bind to the longer DNA probes. The binding events influence each other under these conditions and the binding mode switches to a weaker binding as compared to the binding of a single *SsoSSB* molecule.

Since the binding experiments were performed with the 5'-end fluorescein labeled oligonucleotides, control experiments were required to evaluate a potential influence of the fluorescent dye on the binding equilibrium. Therefore, a preformed complex of labeled oligonucleotide was challenged with unlabeled oligonucleotide and the displacement of the fluorescent DNA was followed by fluorescence anisotropy. The

Results

competition experiment with a 13 nt ssDNA probe revealed an approx. 3 fold weaker binding ($K_d \sim 1.6$ nM) for unlabeled ssDNA as compared to the labeled probe. Therefore, it is concluded that the fluorescent label has only a small influence on the binding equilibrium.

There was no clear saturation plateau for the longer DNAs. This points to the formation of larger complex and it indicates the weaker binding of the longer ssDNA molecules (Figure 3.6B). Together with the EMSA data, it is concluded that *SsoSSB* possesses high affinity to ssDNA in the low nanomolar range when offered a single binding site of approx. 13 nt length. With longer DNAs, binding switches to a weaker binding mode.

Table 3.1: single stranded oligonucleotide binding affinity of *SsoSSB*.

Length (nt) of ssDNA	K_d (nM)
7	3.6
9	0.99
11	0.61
13	0.48
15	1.7
17	3.5
25	16
50	24

K_d , dissociation constant

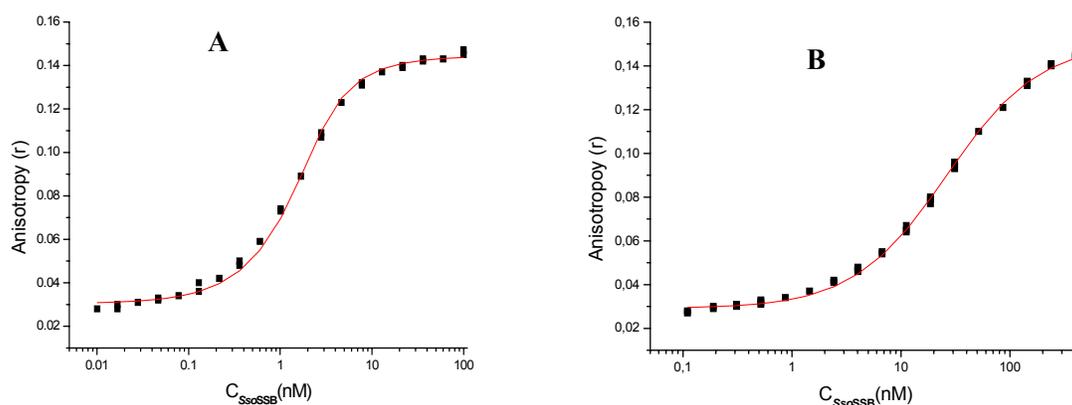


Figure 3.6: Fluorescence anisotropy measurements of *SsoSSB* binding to ssDNA probes (2 nM). **A**, binding of *SsoSSB* to 13 nt ssDNA oligonucleotide. **B**, binding of *SsoSSB* to 25 nt ssDNA oligonucleotide. The buffer contained 20 mM of Tris-Cl, pH 7.5, 50 mM of NaCl and 0.1 mg/ml BSA. The solid line is the fit according to a 1:1 binding model.

3.2 Computational analysis of the putative, DNA repair related, specifically conserved gene clusters in *Sulfolobus solfataricus* P2

At the beginning of the present work, bioinformatic analysis (Makarova et al., 2002) had revealed a conserved gene context in prokaryotic genomes suggesting the presence of previously undetected, partially conserved gene clusters that consist of more than 20 genes. These clusters are found in most archaea and some bacteria. The putative genes in these clusters were intriguingly predicted to be involved in a novel DNA repair system. The encoded proteins differed from known DNA repair factors such as RPA, XPA, XPC and *SsoSSB*. The study by Makarova et al. revealed that the gene composition and gene order in these clusters vary greatly between species. However, most genes had homologues in *Sso* P2 strain. Therefore *Sso* P2 strain was used as a model organism to investigate the properties of the genes in these gene clusters. To study the roles of these putative genes in life, a computational analysis and subsequently cloning and expression of some of the genes were performed.

3.2.1 Homology analysis

The potentially repair-related gene clusters in *Sso* P2 genome sequence were analyzed using BLASTN program on NCBI (National Center for Biotechnology Information, Bethesda, USA) website. The candidate genes are listed in Table 3.2. In *Sso* P2 genome, more than one homologous gene could be identified in each COG. The very lower E values indicate high homology between genes of the same COG. Manual analysis by VectorNTI software (InforMax.Inc.) revealed three similar groups in *Sso* P2 genome: *ssol1398~ssol1403*, *ssol1443~ssol1438*, *ssol1996~ssol2002*, one individual cluster of *ssol1449~ssol1450* and two isolated homologous genes of *ssol1729* and *ssol1991*. All of the candidate genes encoded putative proteins for which no clear biological function could be predicted. The homology search of the conserved domains of these genes in NCBI database however suggested a relationship to proteins involved in DNA replication, transcription and repair (Makarova et al., 2002). For instance, COG1203 putative genes (three homologous genes, *ssol1402*, *ssol1440* and *ssol1999*, in *Sso* P2 genome) contain a typical helicase domain. And COG1353 putative genes (two homologous genes, *ssol1729*

Results

and *ssol1991*, in *Sso* P2 genome) contain the C-terminal domain which relates to DNA or RNA polymerase family.

Table 3.2: Candidate *Sso* protein gene homologues from NCBI blast results.

	COG1468	COG1203	COG2254	COG1518	COG1688	COG1857	COG1353
<i>ssol</i> protein	<i>ssol1392</i>	<i>ssol1402</i>	<i>ssol1403</i>	<i>ssol1405</i>	<i>ssol1441</i>	<i>ssol1442</i>	<i>ssol1729</i>
genes	<i>ssol1449</i>	<i>ssol1440</i>	<i>ssol1439</i>	<i>ssol1450</i>	<i>ssol1998</i>	<i>ssol1997</i>	<i>ssol1991</i>
		<i>ssol1999</i>	<i>ssol2001</i>		<i>ssol1400</i>	<i>ssol1399</i>	

COG: Clusters of orthologous groups.

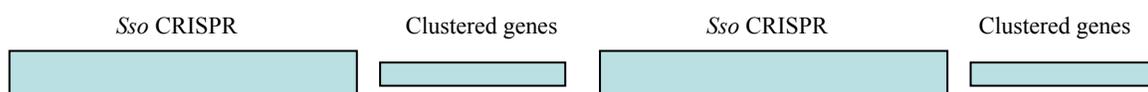


Figure 3.8: Map of locations of some *Sso* clustered candidate genes. The Clustered genes in between two CRISPRs represent genes of group *ssol1449-ssol1451*; The Clustered genes close to one CRISPR end represent genes of group *ssol1996-ssol2001*, *ssol1438-ssol1443* or *ssol1398-ssol1403*. CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats.

The further sequence analysis in the present work and in other groups (Makarova et al., 2006; Jansen et al., 2002a) interestingly showed that most of the candidate genes in *Sso* P2 or other prokaryotes located close to or in between sequence elements called originally “Long Clustered Tandem Repeats (LCTRs)”. These have been now renamed as CRISPR, (Clustered Regularly Interspaced Short Palindromic Repeats) (Figure 3.8). In further analysis the protein genes in the neighborhood of the CRISPRs were named *cas* protein genes (CRISPR-associated protein genes). The *cas* protein genes are found only in species containing CRISPR. The biological function of these Cas proteins remains unknown. During the present work, some bioinformatic researchers suggested an immunity function of Cas proteins against foreign DNA (Makarova et al., 2006; Lillestol et al., 2006). Experimental evidence for this function was however not available. The gene numbers, COG numbers, hypothetical functions and *cas* numbers of the *cas* genes screened out of *Sso* P2 are listed in Table 3.3.

Table 3.3 Clustered *cas* protein genes in *Sso* P2 with previously predicted functions.

<i>ssol</i> number	COG	Previously Predicted Functions	Gene
<i>ssol</i> 1392	1468	metal-binding, nucleic acid-binding, RecB-like nuclease	<i>cas4</i>
<i>ssol</i> 1402	1203	ATP-dependent RNA helicase homolog, DNA-helicase	<i>cas3</i>
<i>ssol</i> 1403	2254	nuclease, often fused with COG1203 protein	
<i>ssol</i> 1404	1343	small protein	<i>cas2</i>
<i>ssol</i> 1405	1518	nuclease/integrase	<i>cas1</i>
<i>ssol</i> 1439	2254	nuclease, often fused with COG1203 protein	
<i>ssol</i> 1440	1203	ATP-dependent RNA helicase homolog, DNA-helicase	<i>cas3</i>
<i>ssol</i> 1441	1688	RAMPs, possibly RNA-binding protein	<i>cas5</i>
<i>ssol</i> 1442	1857	α/β protein, possibly nuclease	<i>cas2</i>
<i>ssol</i> 1449	1468	metal-binding, nucleic acid-binding, RecB-like nuclease	<i>cas4</i>
<i>ssol</i> 8090	1343	small protein	<i>cas2</i>
<i>ssol</i> 1450	1518	nuclease/integrase	<i>cas1</i>
<i>ssol</i> 1451	3578	?	
<i>ssol</i> 1996	?	?	
<i>ssol</i> 1997	1857	α/β protein, possibly nuclease	<i>cas2</i>
<i>ssol</i> 1998	1688	RAMPs, possibly RNA-binding protein	<i>cas5</i>
<i>ssol</i> 1999	1203	ATP-dependent RNA helicase homolog, DNA-helicase	<i>cas3</i>
<i>ssol</i> 2001	2254	nuclease, often fused with COG1203 protein	

Note: *ssol*1451 and *ssol*1996 were in the gene groups with no prediction.

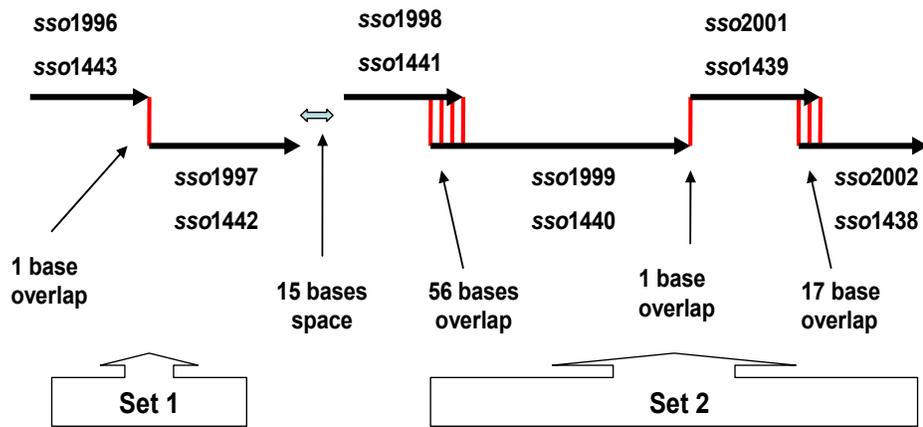
3.2.2 Operon analysis

In the present work most of the candidate genes were found in a few groups in the complete *Sso* P2 genome sequence. Each group contains a few genes in the same orientation with their homologues in other groups. This phenomenon suggests a coordinate expression and a function relationship between clustered genes.

The operon analysis focused on three similar groups in *Sso* P2 genome, the clusters of *ssol*1996-*ssol*2002, *ssol*1438-*ssol*1443 and *ssol*1398-*ssol*1403. The sequences were manually scanned for sequence elements required for transcription and translation, namely for BRE (Transcription Factor B recognition element of archaeal promotor) sequences, TATA boxes (Reeve, 2003; Bell et al., 1999; Reiter et al., 1988), Shine-Dalgarno sequences (Tolstrup et al., 2000) and terminator sequences (Reiter et al., 1988) (see 2.4.2). The three groups showed high similarity of sequence contexts. Two sets were found in each group. The genes in each set are overlapped and these two sets are spaced by a few bases or stop

codons (Figure 3.9). In each set, BRE sequences, TATA boxes, Shine-Dalgarno sequences and terminator sequences were found upstream and downstream of the open reading frames (ORFs). The ORFs in each set were found to be arranged in the same orientation with head-to-tail overlapped bases in between stop codons of one ORFs and start codons of the next ORFs implying co-regulation in gene expression (Figure 3.9).

A: Sequence contexts in *sso1996-2002* and *sso1438-1443* clusters



B: Sequence context in *sso1398-1403* cluster

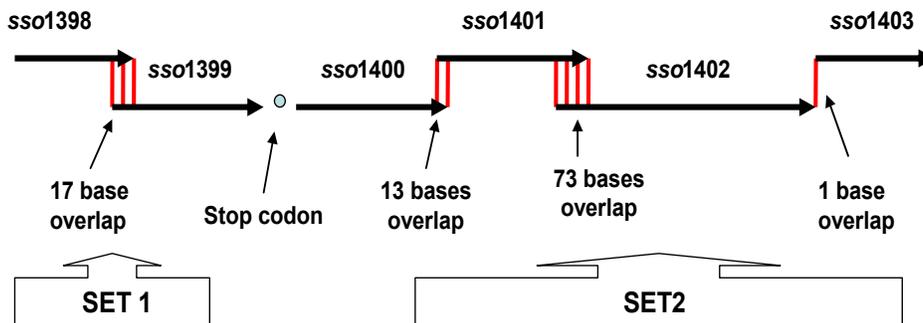


Figure 3.9 Operon analyses of three similar clusters in *Sso P2* genome. **A**, The similarity of sequence contexts between the gene clusters *sso1996-2002* and *sso1438-1443* is very high both in Set 1 and Set 2, even in the base numbers of overlaps and spacers. **B**, Genes of cluster *sso1398-1403* can be also divided into two sets which are similar to the two clusters shown in A. Sequence comparison reveals high similarity of the genes in the three clusters.

3.3 Expression screening and functional scanning of putative *Sso* proteins in *E.coli*

The products of the *sso* genes in Table 3.2 were predicted to be involved in a novel DNA repair system or genome immune system and were uncharacterized to date (Makarova et al., 2002; Makarova et al., 2006). For a characterization of their biochemical functions, it was necessary to express and purify the corresponding proteins. Until very recently, archaeal expression systems were not available, and therefore the expression was performed in *E.coli* expression systems. The strains used included *E.coli* BL21(DE3) (Stratagene), BL21-CodonPlus(DE3)-RIL (Stratagene), Rosetta(DE3)pLysS (Novagen), and Rosetta-gami(DE3) (Novagen). The vectors included pET-28c(+) (Novagen) for single protein gene expression, duet vectors (pACYCDuetTM, pETDuetTM, pCDFDuetTM and pRSFDuetTM, Novagen) carrying compatible replicons and antibiotic resistance markers for co-expression of two or more proteins.

3.3.1 Expression screening and solubility detection

In expression screening experiments, most of the candidate protein genes in Table 3.2 were amplified from *Sso* P2 genomic DNA by PCR (2.3.10.1) and were subcloned in pET-28c(+) vectors. The only exception was *sso1729* which could not be amplified. Subsequently, the vectors were transformed into *E.coli* Rosetta(DE3)pLysS, and the genes were expressed under normal conditions according to the protocols (2.5.2.1). The expression and solubility of the proteins were followed by SDS-PAGE analysis of the clear cell extracts and the insoluble fractions (2.5.2.1). Most of the products were insoluble and formed aggregates (inclusion bodies). *Sso1991* did not show expression. Only *Sso1442*, *Sso1450*, *Sso1996* and *Sso1997* were partially soluble.

For the genes that could be expressed in the first experiments only in insoluble form, various conditions were tested to increase the solubility of the expression products. Expression was tested in *E.coli* BL21(DE3), Rosetta(DE3)pLysS and Rosetta-gami(DE3) with variations of medium components, temperature, amount of inducer and cultivation time. However, all these proteins could be overexpressed only in insoluble form and no soluble gene products were observed.

3.3.2 Further soluble expression, refolding efficiency and functional screening

Computational analysis showed that in gene clusters of *ssol398~ssol403*, *ssol438~ssol443* and *ssol996~ssol2002*, all the genes in one cluster had homologous genes in the other two clusters (Table 3.2). Therefore, the expression test focused on the genes in one cluster, in this case, *ssol996~ssol2002*. Since the computational analysis had indicated a possible co-regulation *in vivo*, duet vectors were used for expression that allowed the simultaneous expression of 2/4/6 proteins in 1/2/3 vectors. The genes of cluster of *ssol996~ssol2002* were subcloned into different duet vectors which were designed specifically for gene co-expression and were afterwards expressed in *E.coli* BL21-CodonPlus (DE3)-RIL or Rosetta(DE3)pLysS in 2/4/6 genes co-expression mode (2.5.2.2). However, in solubility tests, no positive results could be achieved.

Since soluble gene products could not be obtained under various conditions, renaturation of the insoluble proteins was considered as a further step. These experiments concentrated on the proteins *Ssol999* and *Ssol2001*.

To obtain some hint on possible biochemical functions, the gene cluster of *ssol998~ssol2002* was blasted on NCBI website using BlastN. Only weak predictions were obtained for three of the proteins. *Ssol998* was predicted as an RNA binding protein, *Ssol999* as a helicase, *Ssol2001* as a nuclease whereas for *Ssol2002* there was no prediction. In this work, the sequences of these genes were manually analyzed.

The results revealed that *Ssol999* contained highly conserved walker A and walker B motifs (Figure 3.10) characteristic for helicases. These two motifs represent specific ATPase motifs which participate in ATP binding and hydrolysis, respectively. They are present in helicase superfamilies 1 and 2. Another small motif, motif III, with predicted nucleic acid unwinding function in helicase superfamily 2, was found close to motif II. Interestingly, the motif VI for RNA binding was highly conserved in *Ssol999* pointing to an ATP-dependent RNA helicase activity (Caruthers and McKay, 2002; Hall and Matson, 1999; Tuteja and Tuteja, 2004). The presence of these motifs strongly suggested that *Ssol999* could be a putative helicase.

Ssol2001 was previously predicted as a nuclease due to the presence of a HD domain. There are several examples known where helicase and nuclease are encoded within the

same operon to allow functional and regulatory cooperation (Makarova et al., 2002; Makarova et al., 2006; Komori et al., 2004; Haft et al., 2005; Yu et al., 1998). In the *sso1998-sso2002* gene cluster, *sso1999* and *sso2001* overlap head to tail suggesting a functional relationship. Therefore, these proteins were chosen for refolding and functional scanning experiments.

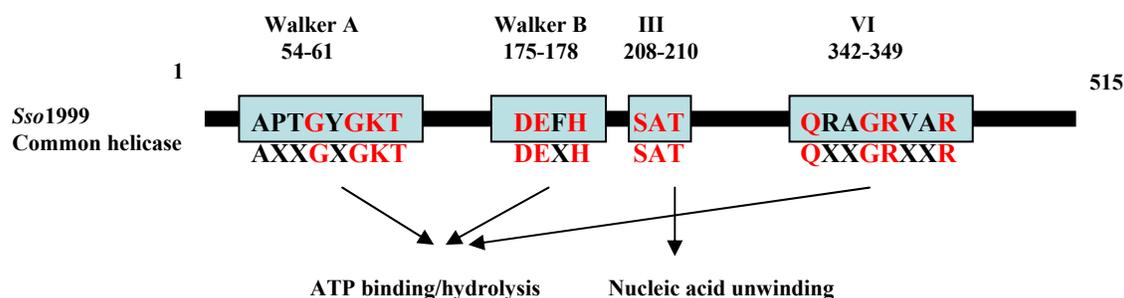


Figure 3.10: Schematic diagram of comparison between putative helicase motifs in *Sso1999* and the consensus motifs of well characterized helicases. The black bar represents the gene sequence. Amino-acid sequences of the motifs are expressed in single-letter codes in boxes for *Sso1999* and below the boxes for the conserved helicase superfamily. The most conserved residues are red labeled. The numbers above the black bar indicate the positions in the *Sso1999* sequence. The functions and names of the conserved helicase motifs are indicated by arrows.

In refolding experiments, *sso1999* and *sso2001* were overexpressed with/without N- or C-terminal hexahistidine tags in insoluble forms (inclusion bodies) from *E.coli* Rosetta(DE3)pLysS (2.5.3.1). For both proteins, a strong overexpression in inclusion bodies was observed. The inclusion bodies were solubilized in 8 M urea followed by a wash in 3 M urea (Figure 3.11). The concentrations of proteins were estimated by western blot assay using standard protein markers. Further steps to investigate the refolding efficiency included on-column refolding, rapid dilution, dialysis and high-throughput assays (Vincentelli et al., 2004).

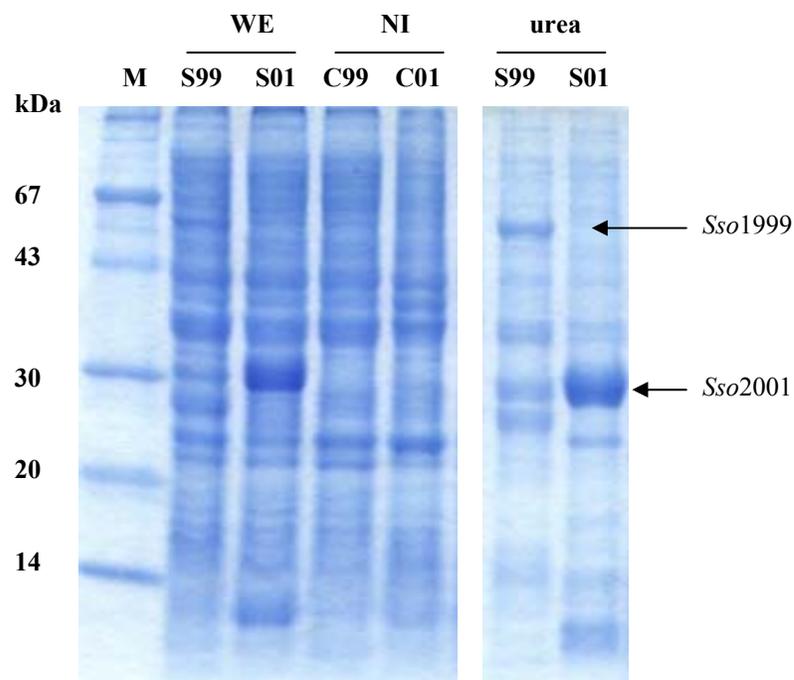


Figure 3.11: SDS-PAGE analysis of the expression of *Sso1999* (61KDa) and *Sso2001* (30KDa) proteins in insoluble forms after cell lysis and being dissolved by 8 M urea. WE, whole cell extract after induction; NI, cell culture extract without induction; urea, cell pellet proteins dissolved in 8 M urea; M, protein marker; S99, *Sso1999*; S01, *Sso2001*; C99, C01, controls without induction; kDa, kilodalton.

In on-column refolding experiments (2.5.3.2), *Sso1999* failed to get solubilized, whereas *Sso2001* could be successfully solubilized by decreasing the concentration of urea from 8 M to 0 (Figure 3.12A) during the column purification. In rapid dilution refolding experiments, after affinity column (Ni-NTA, Qiagen) or ion-exchange column (EMD-DEAE, Merk) purification under denaturing conditions (with 8 M urea), *Sso1999* and *Sso2001* proteins were rapidly diluted into 20 fold dilution buffer with 0.5 M arginine by dropping denatured protein into the renaturation solution at 4°C (2.5.3.3). Both proteins reached soluble form at low concentrations and were stable for a few days (Figure 3.12B). In dialysis experiments (2.5.3.4), the denatured proteins in 8 M urea were directly dialyzed against dialysis buffer containing 0.5 M arginine and redox reagents (GSH and GSSG). After over night dialysis at 4°C, most of the protein however precipitated and only a small fraction remained in solution (Figure 3.12C).

High-throughput renaturation (2.5.3.5) assays as described by Vincentelli et al. (Vincentelli et al., 2004) provided a chance to investigate the renaturation of denatured proteins under a broad range of conditions. These authors published a collection of refolding conditions that cover a wide range of pH-values, salt concentrations and other renaturants. According to the recipe in Table 2.2 (2.5.3.5), the components in Table 2.1

(2.5.3.5) were arranged in a 96 well plate which allowed following the renaturation of a large number of samples on a single plate. The solubilization was monitored by light absorption at 340 nM. When precipitates are present, the light is scattered resulting in an absorption increase. The data showed that low pH and arginine increased the yield of soluble protein, whereas detergents (NDSBs), ion strength and redox components (GSH, GSSG) did not.

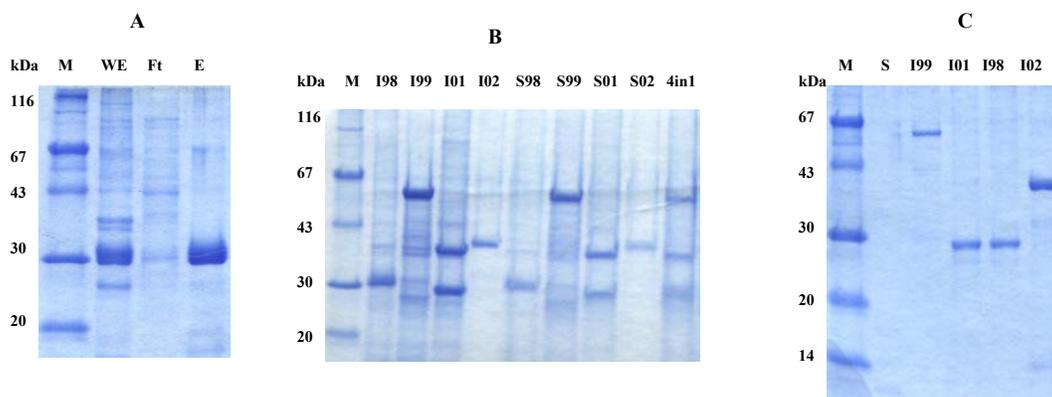


Figure 3.12: A, On-column refolding of *Sso2001* on Ni-NTA column (see 2.5.3.2 for details). The whole cell extract (WE), flow-through (Ft) and elution (E) fractions were monitored by SDS-gel. M represents protein marker. **B, Protein solubilization of *Sso 1998* (29KDa), *Sso1999*, *Sso2001* and *Sso2002* (41KDa) by rapid dilution (see 2.5.3.3 for details).** S represents soluble parts after refolding and I, insoluble parts before refolding; 4in1, co-refolding of 4 proteins (*Sso1998*~*Sso2001*). **C, Protein solubilization of *Sso1998*, *Sso1999*, *Sso2001* and *Sso2002* by dialysis (see 2.5.3.4 for details).** S represents soluble part after dialysis and I, insoluble parts after refolding; proteins in B, C: 99, *Sso1999*; 01, *Sso2001*; 98, *Sso1998*; 02, *Sso2002*. kDa. kilodalton

Co-refolding assay, based on interactions between different protein molecules, supplies a chance that the refolding of one protein may enhance the refolding efficiency of the others (Trivedi et al., 1997). In the present work, the co-refolding test of *Sso1998*, *Sso1999*, *Sso2001* and *Sso2002* was carried out by rapid dilution assays (2.5.3.3) in refolding buffer containing 0.1 M sodium phosphate, pH 8.0, 5 mM of GSH, 0.5 mM of GSSG, 0.5 M of arginine (Figure 3.12B, lane 4in1). The soluble forms of all four proteins were observed.

Activity tests of the solubilized proteins were performed after different refolding assays. The soluble forms of all four proteins were assayed for ATPase activity, DNA/RNA binding ability, and nuclease activity in a large series of assays. However, none of these activities could be unambiguously identified. There was no ATPase activity

Results

detectable for *Sso1999* which was previously predicted as a helicase. The negative results were obtained for *Sso1998* and *Sso2002*. The only protein for which a positive result could be obtained was *Sso2001* which showed a nuclease activity after refolding in buffers with 0.8 M arginine in pH 7, 8 and 9 in high-throughput assay (see 2.5.3.5 and Figure 3.13).

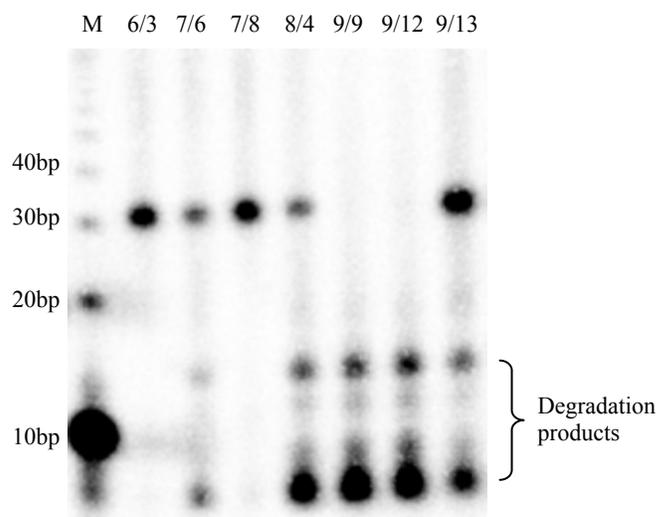


Figure 3.13: Nuclease assay of renatured *Sso2001* after high-throughput refolding. 10nM of 5'-end labeled 30 bp dsDNA (PT-AB30) was incubated with renatured *Sso2001* in buffer containing 20 mM tris-Cl, pH 7.5, 50 mM NaCl, 10 mM MgCl₂ at 50°C for 60 min. The samples were then analyzed by 20% denaturing PAGE. *Sso2001* refolded under some conditions showed nuclease activity (lane 7/6, 8/4, 9/9, 9/12 and 9/13). The numbering of lanes 6/3~9/13 represents the well numbers on the refolding plate (see Table 3.4).

Table 3.4: Refolding buffer composition of soluble fractions from Figure 3.13.

Well No.	6/3	7/6	7/8	8/4	9/9	9/12	9/13
pH	MES, 6	Tris, 7	Tris, 7	Tris, 8	CHES, 9	CHES, 9	CHES, 9
Refolding buffer components	Glycerol, NDSB256, Arg	PEG4000, NDSB256, Arg	β-MSH, Arg	KCl, β-MSH, NDSB201, Arg	β-MSH, NDSB195, Arg	EDTA, β-MSH, Arg	NDSB195, Arg

Concentrations: glycerol, 20% (v/v); PEG4000, 0.05% (w/v); NDSB, 100 mM; β-MSH, 10mM; Arg, 800 mM; EDTA, 1 mM

3.4 Expression, purification and characterization of *Sso2001* protein from *Sso* P2 with esterase as a fusion partner

Sso2001 is located in COG2254. The proteins belonging to this COG had been first predicted as nucleases in a DNA repair system (Makarova et al., 2002), and afterwards were denoted as Cas proteins involved in a genome immune system (Makarova et al., 2006). Some of COG2254 proteins were found to be fused with COG1203 proteins to form helicase-nuclease bifunctional proteins (Makarova et al., 2006). However, none of the proteins has been biochemically characterized to date. In *Sso* P2, the computational analysis in the present work revealed separate genes for the two activities. The potential helicase gene *sso1999* (COG1203) overlaps head-to-tail with the potential nuclease gene *sso2001* (COG2254). In the following, the expression, purification and characterization of *sso2001* is described.

3.4.1 Gene amplification from *Sulfolobus solfataricus* P2

sso2001 gene was amplified by PCR from *Sso* P2 genomic DNA (2.3.10.1) using forward primer S2001-for (ACATATGTTGATCAAGCCTTGCGCTTA) and reverse primer S2001-rev (ACGAGCTCTAGAGTGGAACCTCCAT). The primers created the *Nde*I and *Sac*I restriction sites (underline letters) upstream and downstream of the *sso2001* gene, respectively. The sequence of the PCR fragment was shown to be correct when compared to the predicted gene sequence in *Sso* P2 genome.

3.4.2 *Sso2001* fusion protein gene expression

The data in 3.3 show that the *sso2001* gene expression product in *E.coli* is insoluble and its refolding does not work well. Therefore, expression as a fusion protein was considered as an alternate way. In these experiments, a number of fusion proteins were studied. The *sso2001* gene was fused with various prokaryotic folding helper sequences (DsbA, Disulfide bond Formation protein A; SlyD, N-terminal prolyl isomerase domain as *E.coli* cytosol folding helper protein; Fd, phage fd gene-3 protein for disulfide folding, kindly gifts from Prof. Franz Xaver Schmid, Universität Bayreuth) or esterase gene from thermophilic bacteria, *Alicyclobacillus acidocaldarius* (kindly gift from Prof. Mathias

Results

Sprinzi, Universität Bayreuth) (Agafonov et al., 2005) which serves at the same time as an affinity tag and as a reporter to monitor the extent of expression. The expression of the fusion protein genes was then followed in *E.coli*. The SDS-PAGE and western blot results indicated that only the fusion of *sso2001* gene and esterase gene (*sso2001Est*, construction detailed in Figure 3.14) could be expressed partially soluble in *E. coli* Rosetta(DE3)pLysS strain. Expression of fusion proteins with DsbA, SlyD and Fd did not yield soluble proteins. The expression conditions for the *Sso2001*-esterase fusion protein were then optimized to enhance the soluble protein expression level. In detail, after inoculation with 10 ml of transformed, overnight cultivated cells, the LB medium was cultivated at 37°C till OD₆₀₀ reached 0.8~1.0, then the medium was immediately cooled down to 30°C and the expression was induced with 0.1 mM IPTG. When OD₆₀₀ increased over 1.5, the cells were harvested and pelleted down by centrifugation. The presence of the full length protein was determined both by western blot against C-terminal hexahistidine (2.5.4) and active staining for N-terminal esterase (2.5.5). The amount of soluble full length protein was obviously enhanced when *sso2001Est* gene was expressed at lower temperature and at a lower induction level suggesting that a slower growth speed was beneficial for this fusion protein expression.

3.4.3 Purification of *Sso2001Est* fusion protein

The cell pellet was thawed in cell lysis buffer and the resuspended cells were treated by sonication. The soluble proteins were separated from cell homogenate by centrifugation and the supernatant was then loaded onto the Trifluoromethyl ketone (TFK) Sepharose CL-6B (kindly gift from Prof. Mathias Sprinzi, Universität Bayreuth, see structure detail in Figure 3.15B) affinity column which specifically binds to the esterase (2.5.5.1) (Huang et al., 2007). The target protein was eluted by 20 mM of soluble inhibitor (Figure 3.15B). The pooled fractions were dialysed and aliquots were shock frozen in liquid nitrogen and stored at -70°C until use. All the expression and purification steps were monitored by esterase activity assay (2.5.5.2). The activity staining of esterase with Fast Blue BB Salt and β -naphthyl-acetate (Higerd and Spizizen, 1973) indicated that the full length fusion protein was efficiently and rapidly purified by TFK affinity column (Figure 3.15A).

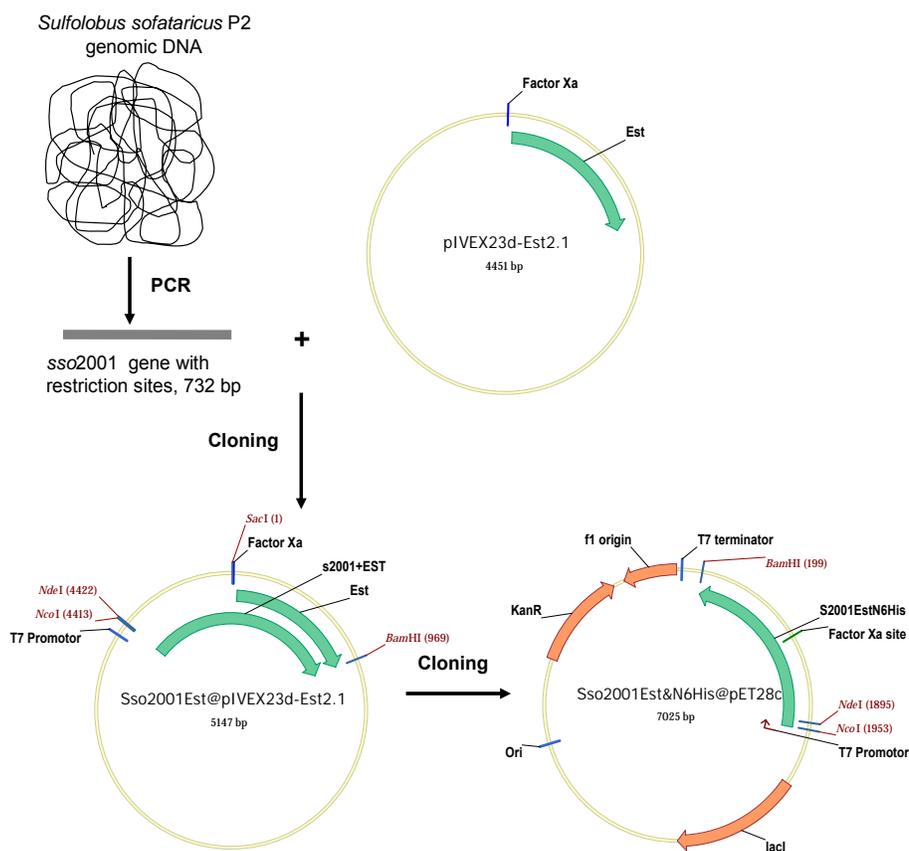


Figure 3.14: Construction of *sso2001Est* fusion protein gene expression vector. The *sso2001* gene was inserted into the multicloning site of pIVEX2.3d-Est2.1 vector (kindly gift from Prof. Mathias Sprinzl, Universität Bayreuth) between the NdeI and SacI cleavage sites. Then the fragment of *sso2001* gene plus esterase gene was cut out by NdeI and BamHI from this vector and inserted into plasmid pET-28c(+) behind its hexa-histidine codons which allows the expression of a N-terminal His₆-*Sso2001*-esterase fusion protein.

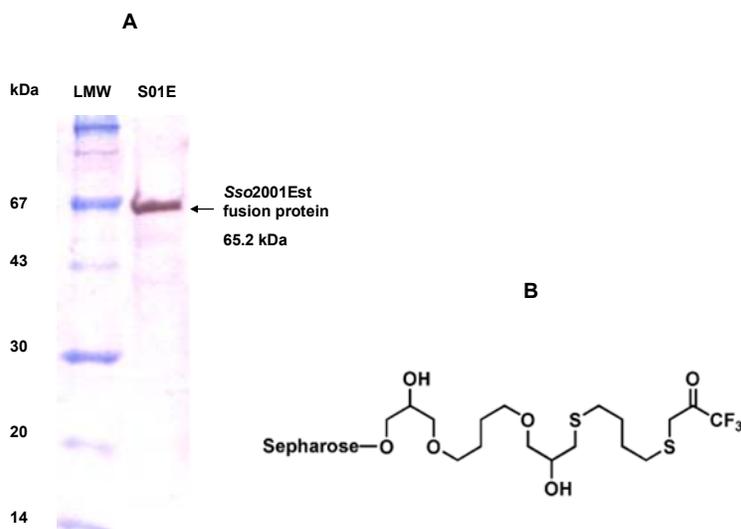


Figure 3.15: Purification of full length *Sso2001Est* fusion protein. **A**, *Sso2001Est* (S01E, 65.2 kDa) protein was purified by TFK (Trifluoromethyl ketone) sepharose CL-6B affinity column and was active stained after SDS-PAGE gel, LMW, low molecule weight marker; kDa, kilodalton. **B**, The structure of TFK modified Sepharose CL-6B.

3.4.4 Nuclease activity detection

In this experiment, the 5'-labeled ssDNA or dsDNA oligonucleotides were prepared as described (2.3.6 and 2.3.7). The nuclease assay was carried out with 10 nM of DNA substrates and 5 nM of *Sso2001Est* fusion protein in nuclease reaction buffer for 20 min at 50°C. Samples were then analyzed by electrophoresis on 20% denaturing PAGE. The gel showed a degradation of the dsDNA molecules into smaller oligomer fragments, whereas only minor degradation of ssDNA was observed. Controls with esterase alone showed no degradation of both ssDNA and dsDNA. The degradation of radioactively labeled DNA oligonucleotide substrates indicated that native *Sso2001Est* fusion protein possesses nuclease activity and prefers dsDNA oligonucleotide as substrate (Figure 3.16).

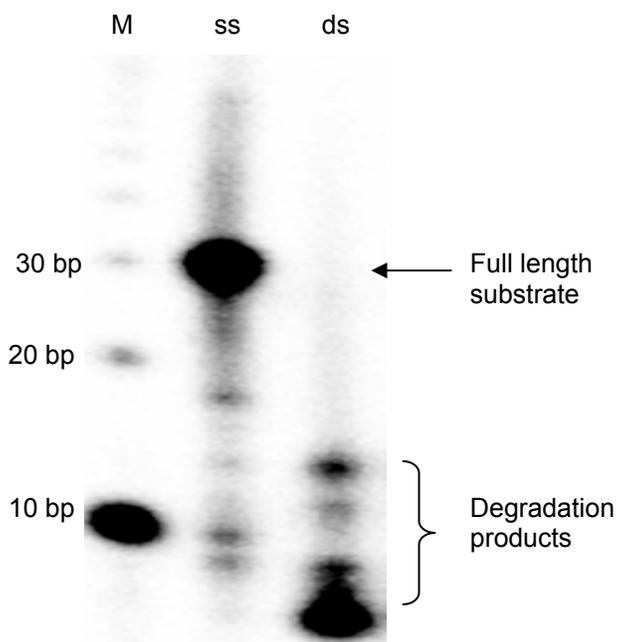


Figure 3.16: Nuclease assay of *Sso2001Est* protein. The substrates were single-stranded PT-A30 in lane ss and, hybridized dsDNA (PT-AB30) in lane ds. 10 bp DNA marker was loaded in lane M. All the substrates were 5'-end radioactively labeled. The electrophoresis was carried out on 20% denaturing PAGE.

3.4.5 Reaction condition investigation

The *Sso2001Est* nuclease reaction conditions including pH, temperature and the effect of bivalent metal ions were varied to obtain the optimal nuclease activity.

3.4.5.1 Temperature

The temperature dependence of *Sso2001Est* nuclease activity was investigated in the range from 25°C to 75°C. The other parameters were fixed as described (2.6.1). The optimal reaction temperature was 50°C which might be due to the balance of the thermostability between the thermophilic enzyme *Sso2001* and the fusion partner, mesophilic esterase (Figure 3.17A and B).

3.4.5.2 Optimal pH

The pH range was varied from 2 to 11 by using different buffer components (H₃PO₄, pH 2; glycine, pH 3; NaAc, pH 4; MES, pH 5 and 6; Tris-Cl, pH 7 and 8; CHES, pH 9; CAPS, pH 10 and 11). The other parameters were fixed as described (2.6.1). High nuclease activity of *Sso2001Est* was observed in the neutral pH range and pH 3. The pH-

Results

dependence showed an unusual behaviour with steep increases and decreases on both sides of the optimal pH values (Figure 3.17C and D).

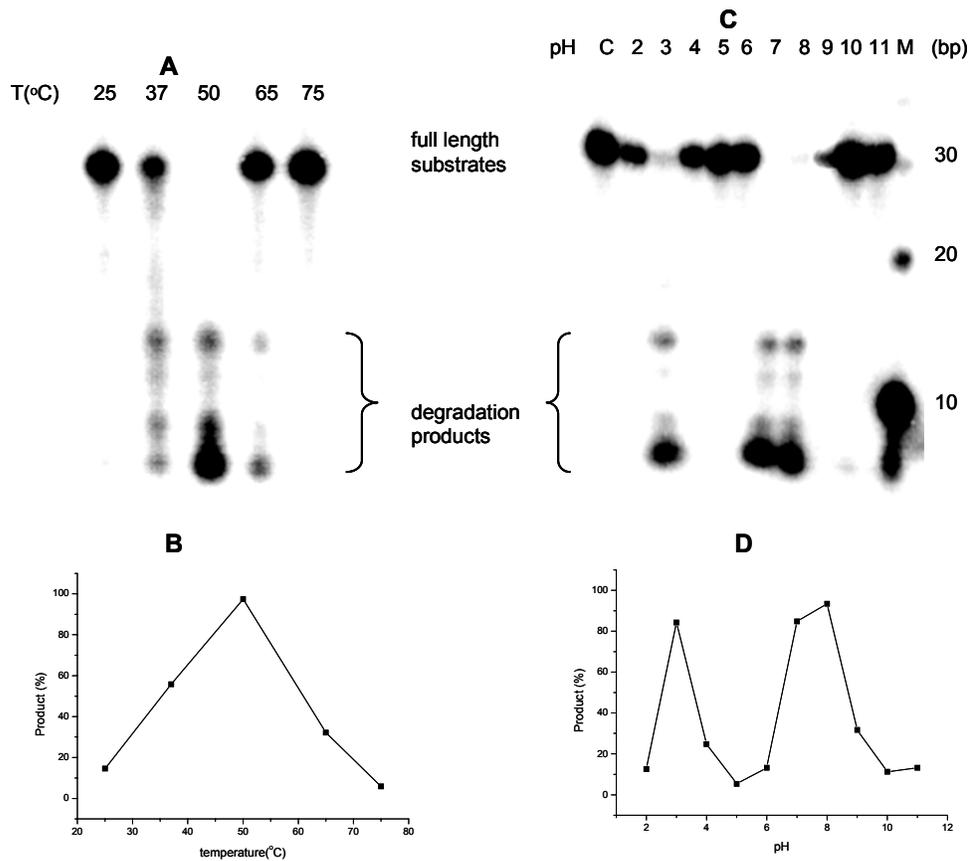


Figure 3.17: pH and temperature dependence of the nuclease activity of *Sso2001Est*. The optimal temperature was observed at 50°C (A and B). The protein shows activity in the neutral pH range and at pH 3 (C and D). In C, Lane C represents a control sample without protein, and M, the 10 bp DNA marker. The 5'-end labeled dsDNA oligonucleotide PT-AB30 was used as a substrate.

3.4.5.3 Effect of bivalent metal ions

The effect of presence/absence of Ca^{2+} , Mg^{2+} , Mn^{2+} , Ni^{2+} , or Zn^{2+} ions on nuclease activity was investigated. Cleavage occurred in the presence of 10 mM MgCl_2 , but not within Ca^{2+} , Mn^{2+} , Ni^{2+} , or Zn^{2+} ions. EDTA inhibited the nuclease activity by chelating Mg^{2+} ions (Figure 3.18).



Figure 3.18: Divalent metal ion effect on nuclease activity of *Sso2001Est*. The dsDNA probe was 5'-end radioactively labeled ds DNA (PT-AB30). Lane 1, MgCl₂ (10 mM); Lane 2, NiSO₄ (2 mM); Lane3, MnCl₂ (5 mM); Lane 4, ZnCl₂ (2mM); Lane 5, MgCl₂ (10 mM) with EDTA (20mM); Lane C, control without protein. With Magnesium ions, the dsDNA substrate was totally degraded in 20 min.

3.4.6 Protein active site determination by mutants

To verify the intrinsic nuclease activity of *Sso2001* and to exclude contaminating nuclease activity, it was necessary to create inactive point mutations of *Sso2001*. A multi-alignment of *Sso2001* sequence was performed by BlastP program on NCBI website. All the proteins hit by the program are hypothetical and are from Archaea. They all belong to COG2254 and are predicted as Cas proteins (Haft et al., 2005). Three highly conserved motifs could be identified by the sequence alignment, a HD motif, a HE motif and a conserved Serine residue (Figure 3.19). The presence of a highly conserved HD domain of these proteins suggests that they belong to the HD hydrolase superfamily. The highly conserved HE domain is without any prediction, neither is the conserved serine residue. The conserved sequence motifs offered an opportunity to experimentally investigate the relationship between the conserved residues of *Sso2001* protein and the nuclease activity.

Results

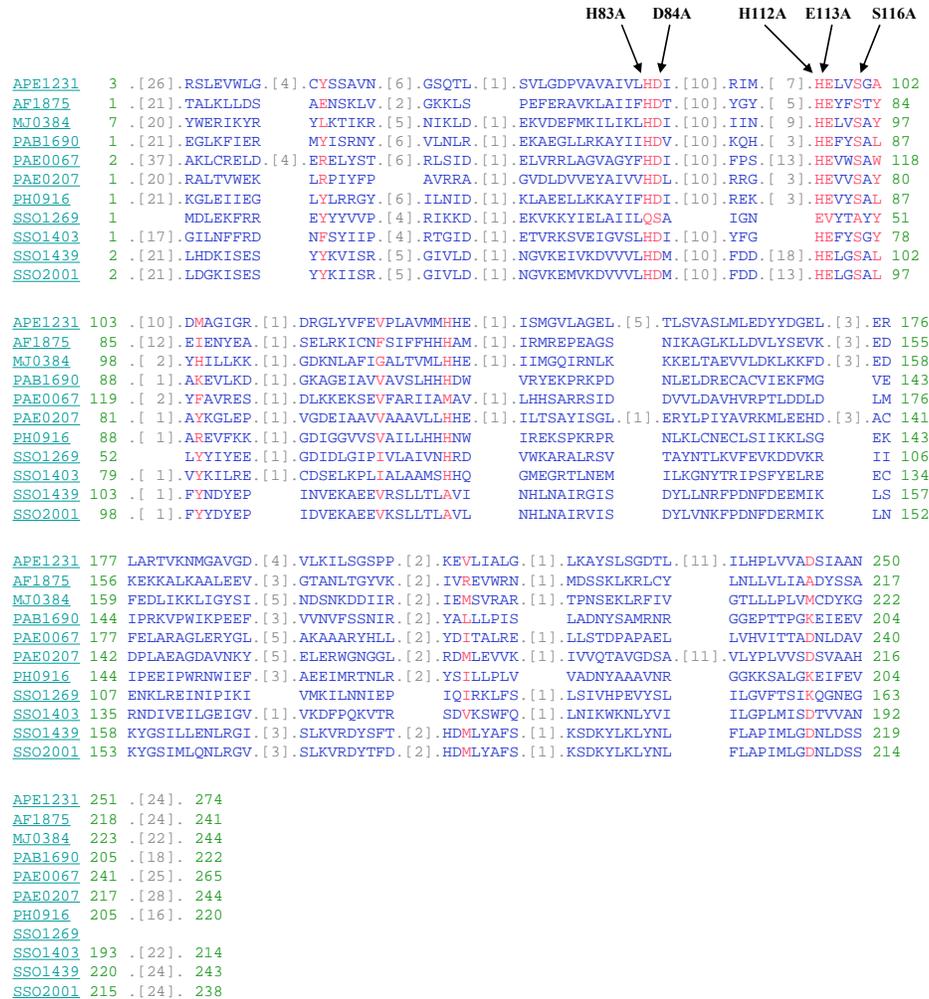


Figure 3.19 Multi-sequence alignment of *Sso2001* homologs by BLASTP. The highlighted red letters represent the conserved residues. All putative proteins are from Archaea and possess the highly conserved HD domains, HE domains and serine residues. The residues mutated in the present work are indicated by arrows. The point mutants are named as H83A, D84A, H112A, E113A and S116A. The double mutants are named as HH with H83 and H112 both to A and DE with D84 and E113 both to A. Numbers represent the residue positions in *sso2001*Est gene sequence. Single letters in sequence represent amino acid residues.

Point mutations of D84, E113, S116 and double mutations of H83 plus H112 and D84 plus E113 in *Sso2001*Est sequence were created by the quick change method (2.3.10.5). The point mutants are H83A, D84A, H112A, E113A and S116A (Figure 3.19). The double mutant of H83 and H112 both to A is HH, that of D84 and E113 both to A is DE, namely. The nuclease activity of the mutants was followed and compared with that of the wild type. The degradation patterns indicated that the D84 residue was most important for nuclease activity. The activity was highly decreased in the E113A mutant, but was not

totally inhibited. The mutation of serine residue did not disturb the nuclease activity at all. The nuclease activity of the double mutant DE was completely inhibited which shows that these two residues are crucial for nuclease activity. The double-mutant HH shows only slightly decreased nuclease activity (Figure 3.20).

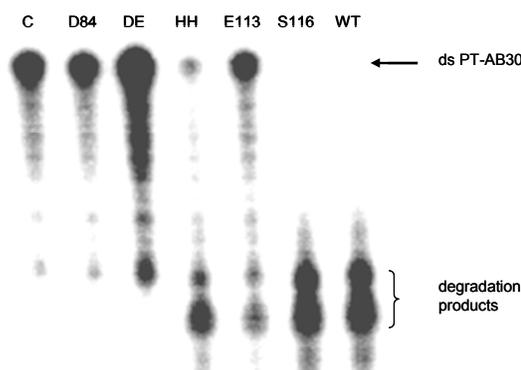


Figure 3.20: Nuclease activity detection of mutants of *Sso2001Est*. C, control without protein; D84A, the point mutant of D84 to Ala residue; DE, the double mutant of D84 plus E113 both to Ala residues; HH, the double mutant of H83 plus H112 both to Ala residues; E113, point mutant of E113 to Ala residue; S116, point mutant of S113 to Ala residue; WT, wild type of *Sso2001Est*. The substrate was 5'-end radioactively labeled PT-AB30 dsDNA.

3.4.7 Substrate specificity of nuclease activity

The next point to be investigated was the structure specificity and the sequence specificity of the nuclease activity. Various DNA oligonucleotides including single-strands, double-strands, pseudo fork-strands and double-strands with sticky ends were prepared as substrates for nuclease detection (see the schematic map of the substrate structures in Figure 3.21). The substrates and *Sso2001Est* fusion protein were mixed and incubated under optimal conditions for 10 min. The reactions were quenched immediately in denaturing gel loading buffer with additional 50 mM EDTA and 0.2% SDS to obtain partially degraded products. The degradation of 5'-end radioactively labeled upper or bottom strands in dsDNA oligonucleotides (Figure 3.21, Lane 6, 7, 8, 14, 15 and 16) indicated an endonuclease activity that starts from the 3'-end. The cleavage stopped at approx. 4 base pairs from 5'-end implying that dsDNA strands shorter than 4 base pairs are not optimal substrates for nuclease activity. The cleavage of double strands with sticky ends (Figure 3.21, Lane 9 to Lane 17) showed the double strand priority of the enzyme activity. The sequence (Table 3.5) analysis showed that darker spots

Results

(embraced by red double brackets in Figure 3.21) produced by enzyme cleavage mostly represented G residues suggesting the enzyme cleaved the dsDNA at G residues preferably. No structure specificity of the enzyme activity was observed from this experiment.

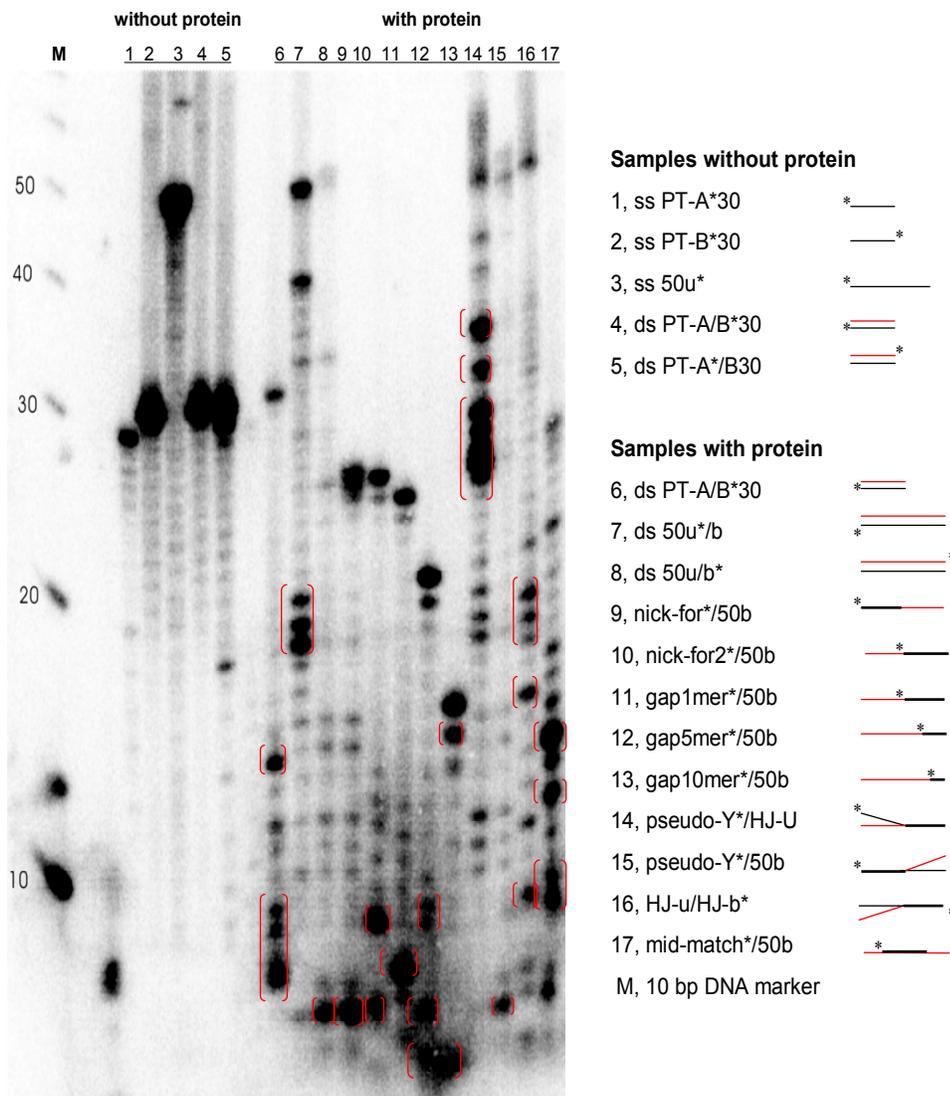


Figure 3.21: Digestion of various 5'-end radioactively labeled oligonucleotide substrates by *Sso2001Est* nuclease. After nuclease activity was quenched, samples were analyzed on a 20% sequencing gel. The structure of the oligonucleotides is indicated schematically on the left. The radioactively labeled 5'-ends are marked with black stars. The spots embraced by the double brackets represent G residues in the oligonucleotide sequences. The sequences of the oligonucleotides are collected in Table 3.5.

Table 3.5: The sequences of 5'-end radioactively labeled DNA oligonucleotide substrates for site specific nuclease detection.

Name	Sequence
ss PT-A*30	5'-*CCTCTTCTTCTGTGCACTCTTCTTCTCCCC-3'
ss PTB*30	5'-*GGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3'
ss50U*	5'-*ACAGCTATGACCGAATTCCTGGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3'
ds PT-AB*30	5'-CCTCTTCTTCTGTGCACTCTTCTTCTCCCC-3' 3'-GGAGAAGAAGACACGTGAGAAGAAGAGGGG*-5'
ds PT-A*B30	5'-*CCTCTTCTTCTGTGCACTCTTCTTCTCCCC-3' 3'-GGAGAAGAAGACACGTGAGAAGAAGAGGGG-5'
ds 50u/b*	5'-*ACAGCTATGACCGAATTCCTGGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC*-5'
ds 50u*/b	5'-*ACAGCTATGACCGAATTCCTGGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
nick-for*/50b	5'-*ACAGCTATGACCGAATTCCTGGGGA-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
nick-for2*/50b	5'-*GAAGAAGAGTGCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
gap1mer*/50b	5'-*AAGAAGAGTGCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
gap5mer*/50b	5'-*AGAGTGCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
gap10mer*/50b	5'-*GCACAGAAGAAGAGG-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'
pseudo-Y*/HJ-U	5'-*ACAGCTATGACCGAATTCCTGGGGA CCTGGTGCACCTGCAGGCATGCAAG-3' 3'-AAGGACGAGCTCCCGGGTCCCAAC GGACCAGCTGGACGTCCGTACGTTC-5'
pseudo-Y*/50b	5'-*ACAGCTATGACCGAATTCCTGGGGA CCTGGTGCACCTGCAGGCATGCAAG-3' 3'-TGTCGATACTGGCTTAAGGACCCCT CTTCTTCTCACGTGTCTTCTTCTCC-5'
HJ-u/HJ-b*	5'-CTTGCATGCCTGCAGGTGACACAGG CCACCCTGGCGCCCTCGAGCAGGAA-3' 3'-GGAGAAGAAGACACGTGAGAAGAAG GGTGGGACCGCGGGAGCTCGTCCTT*-5'
mid-match*/50b	5'-*GAATTCCTGGGGAGAAGAAGAGTGA-3' 3'-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5'

Note: stars represent radioactively labeled 5'-end of oligonucleotides; red letters the paired bases.

3.4.8 Steady state kinetics of nuclease activity

The Michaelis constant (K_m) is an important parameter that can describe the apparent affinity of an enzyme for the substrate under steady-state conditions. The K_m value is determined by Michaelis-Menten equation:

$$V_0 = V_{\max} * [S] / (K_m + [S])$$

Where V_0 is the initial velocity of the enzyme in enzymatic reaction, V_{\max} is the maximum velocity, $[S]$ is the substrate's concentration and K_m is the Michaelis constant. It refers to the substrate concentration where the velocity V is half of the maximal value V_{\max} .

In this study, the K_m value was determined as follows:

Double-stranded 5'-end radioactively labeled PT-AB30 (5'-CCTCTTCTTCTGTGCACTCTTCTTCTCCCC-3') DNA substrate at eight different concentrations was digested by *Sso2001Est* fusion protein in a proper concentration in a series of time course experiment under optimal conditions. Nuclease activity in the aliquots in every time point was quenched immediately with additional 50 mM EDTA and 0.2% SDS on ice bath. After gel electrophoresis, the isotope intensity was counted in Typhoon Instant Imager. The percentages of the digestion products were calculated automatically by software of the Imager and, plotted against time points to generate the initial velocities of the enzyme in different substrate concentrations. Finally, the substrate concentrations and the initial velocities were analyzed to obtain the Michaelis-Menten curve by software, Origin (Origin Lab, USA). From the curve, the K_m value was read out as 22.0885 ± 3.22519 nM (Figure 3.22). This value indicated a rather high affinity of *Sso2001* to the dsDNA substrate.

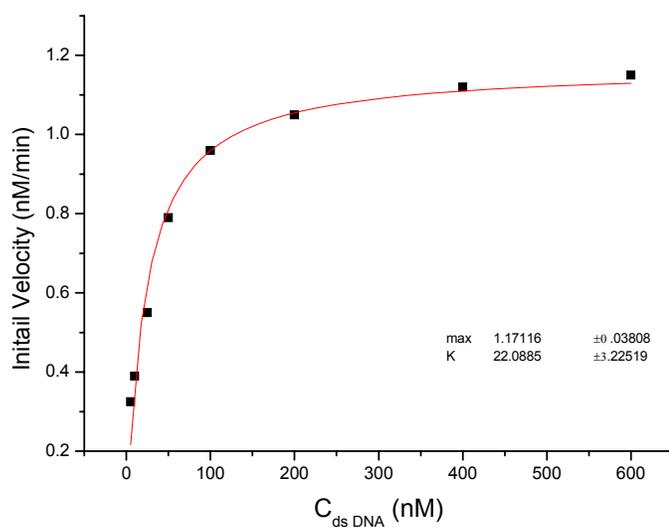


Figure 3.22: K_m determination for dsDNA in the nuclease assays with *Sso2001Est*. The substrate concentrations were varied from 5 nM to 600 nM. The initial velocities were obtained from the time course of a series of digestion experiments performed at 50°C. The data fitting to Michaelis-Menten equation was performed by software, Origin (OriginLab Corporation). The K_m value from the curve was approx. 22 nM and V_{max} was 1.17 nM/min.

3.5 Expression, purification and characterization of *Sso1450C6H* fusion protein from *Sso* P2

The *Sso1450* protein belongs to COG1518. The proteins encoded in this COG are ubiquitous in archaeal and bacterial species that harbor CRISPR in their genomes. The COG1518 proteins are denoted as Cas1 proteins with possible integrase/recombinase function. In some bioinformatic analysis (Makarova et al., 2002) the Cas1 proteins have been also predicted being repair proteins. It was therefore hoped that the biochemical characterization of *Sso1450* could help to reveal the roles of these proteins *in vivo*.

3.5.1 Protein gene expression

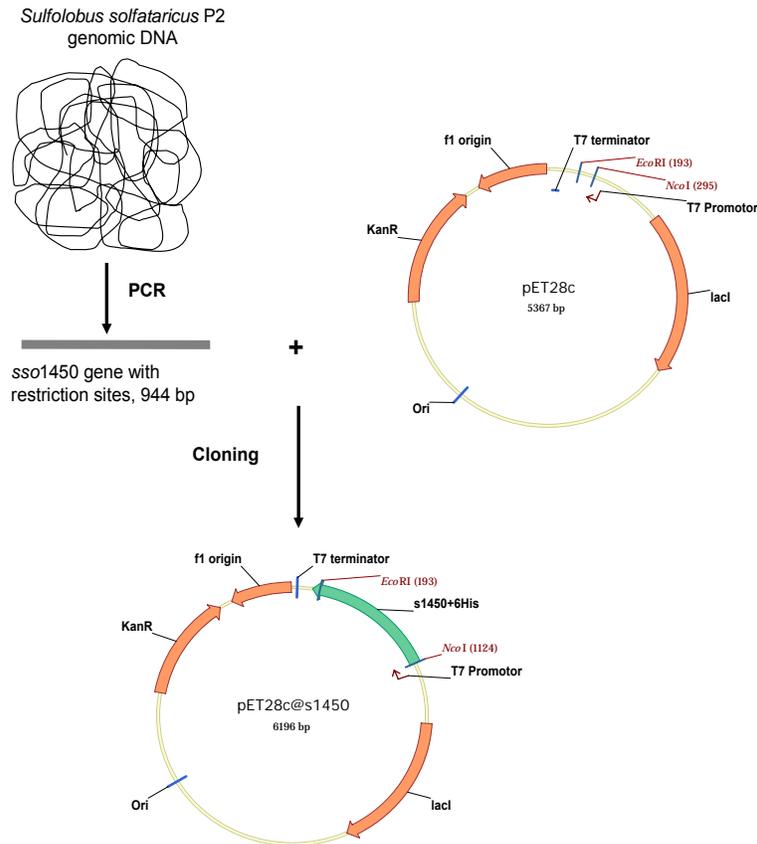


Figure 3.23: Construction of *sso1450C6H* protein gene expression vector. The *sso1450* gene was amplified from *Sulfolobus solfataricus* P2 genome by PCR and was inserted into the multi-cloning sites of plasmid pET-28c(+) behind the hexahistidine codons to construct the expression vector.

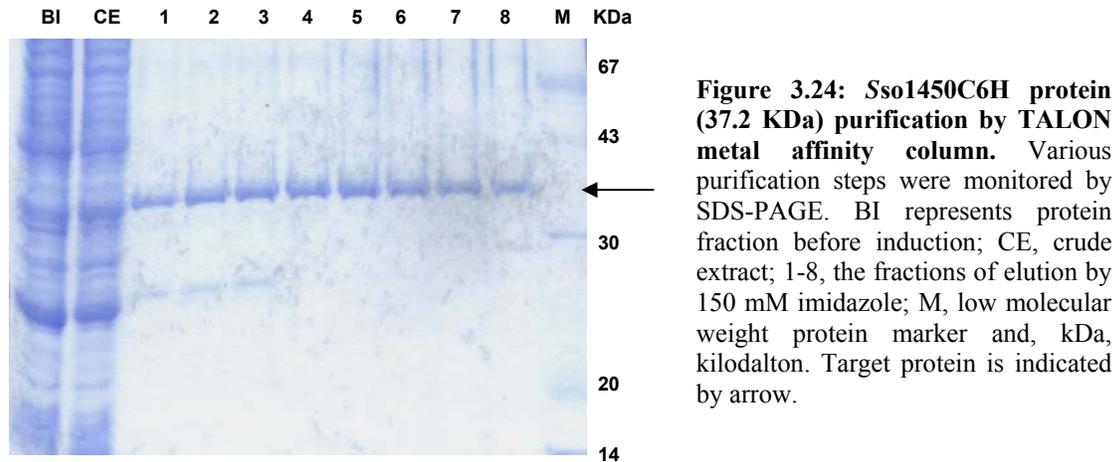
ssol1450 gene was amplified by PCR from *Sso* P2 genomic DNA (2.2.10.1) using primers S1450-for (CATGCCATGGGCGTGATAAGCGTGAGGACTTT) and S1450-rev (GCCGAATTCCCCATCACCAACTTGAAACCCC). The primers created the *Nco*I and *Eco*RI restriction sites (underline letters) upstream and downstream of the *ssol1450* gene, respectively. The sequence of PCR fragment was shown to be correct by comparison to the predicted protein gene sequence in *Sulfolobus solfataricus* P2 genome.

The plasmid pET-28c(+) containing *ssol1450* protein gene fused with C-terminal hexahistidine tag (namely *ssol1450C6H*) was transformed into *E.coli* Rosetta(DE3)pLysS strains. The transformed strains were incubated over night at 37°C on LB plate. After that, one well-grown colony was picked up and was mixed with 20 ml LB liquid medium in a 150 ml flask. The flask was shaken over night at 37°C. 4 l LB medium was inoculated by 20 ml over night culture and was incubated at 37°C for approx. 4 h till OD₆₀₀ reached 0.8~1.0. Then the temperature was immediately decreased down to 30°C. The protein expression was induced by addition of 0.1 mM IPTG. After 3 h incubation, OD₆₀₀ normally was > 1.6, and the cells were harvested by centrifugation. All the culture procedures were under kanamycin selection pressure. As mentioned in *Sso2001Est* protein preparation, the lower temperature and lower amount of inducer enhanced the expression level of *ssol1450* gene in *E.coli* system.

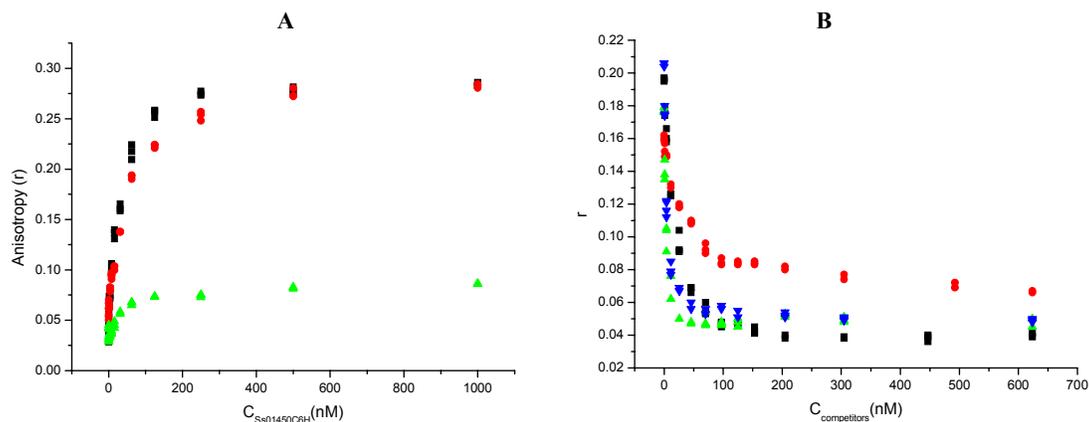
3.5.2 Protein purification

The cell pellet (~1g) was thawed in 10 ml cell lysis buffer containing 2 mM β-mercaptoethanol and 10 mM imidazole on ice and was immediately sonicated in Sonifier B15. Afterwards, most of the *E.coli* proteins and cell debris were precipitated and were removed by centrifugation. The supernatant was loaded onto a 1 ml TALON metal affinity column (Clontech Laboratories, Inc.). The column was washed by cell lysis buffer with additional 50 mM imidazole. Target protein was eluted by addition of 150 mM imidazole. The molecular weight of the eluted protein was determined by SDS-PAGE (Figure 3.24) corresponded to the size of *Sso1450*.

Results



3.5.3 Binding of *Sso1450C6H* on nucleic acid substrates



Sso1450 protein had not yet been characterized biochemically. Therefore its DNA/RNA binding properties were tested by fluorescence anisotropy assays (2.6.3) using various single-stranded and double-stranded oligonucleotides that were 5'-end

labeled with fluorescein. The anisotropy value of the labeled DNA probe increased from 0.04 to 0.27 with increasing protein concentration indicating the complex formation between *Sso1450C6H* and the DNA probes. Similar binding curves were obtained for single-stranded and double-stranded DNA (Figure 3.25A). In addition to dsDNA and ssDNA, *Sso1450C6H* also binds to ssRNA and dsRNA. Since fluorescently labelled RNA probes were not available, the RNA binding was analyzed by a series of competition titrations (Figure 3.25B). In these experiments, a complex between *Sso1450C6H* and a fluorescein-labeled 13 nt ssDNA was preformed and then challenged with increasing amounts of unlabelled ssRNA or dsRNA. As shown by the decrease in anisotropy with increasing concentration of RNA, the RNA was able to compete efficiently with the labeled DNA for binding to *Sso1450C6H*. From these competition titrations the binding constants for the RNA could be determined. The data show that *Sso1450* binds to ssDNA, dsDNA and RNA with approximately the same affinity (Figure 3.25B). Furthermore, the length-dependence for the binding of various single-stranded DNAs was followed. The data reveal an optimal length of the DNA substrates of 13 nt (Table 3.6).

Table 3.6: Dissociation constant values (Kd) for the binding of *Sso1450C6H* to DNAs or ssRNAs of different lengths.

Length(nt)	Kd(nM)
6	165
7	127
9	56.2
11	23
12	20
12 (ssRNA)	29
13	19.6
24	34.7
24 (ssRNA)	51
33	37
36	47.1

3.5.4 Binding mode of *Sso1450C6H* on nucleic acid substrates

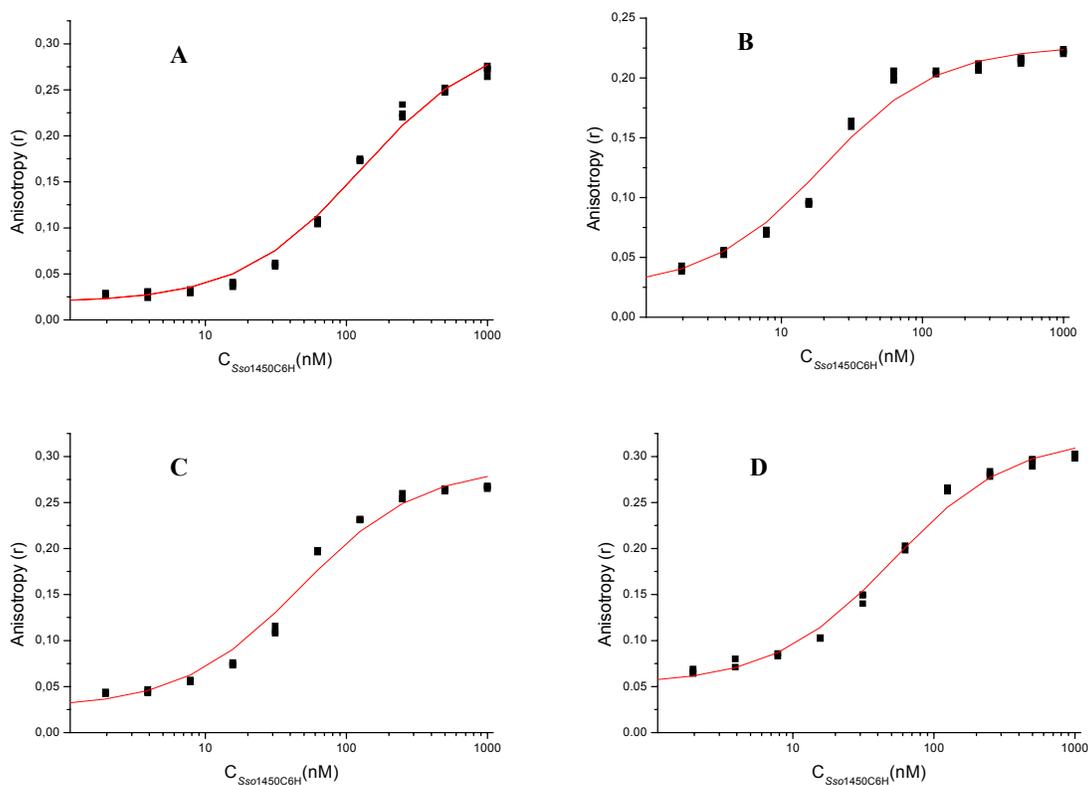


Figure 3.26: One-site model fitting curves of *Sso1450C6H* -ssDNA complexes. A, titration of 7 nt ssDNA; **B**, titration of 12 nt ssDNA; **C**, titration of 24 nt ssDNA probe; and **D**, titration of 36 nt DNA probe.

To investigate the binding model further, the anisotropy variations of fluorescein-labeled ssDNA substrates of different lengths in the present of *Sso1450C6H* were analyzed. The raw data could not be fitted well to one-site binding model (Figure 3.26). No obvious saturation plateau was reached even at 200-fold excess of protein over DNA implying that protein multimers form on the ssDNA substrate, possibly in a cooperative manner. An exact evaluation of the binding curves was therefore not possible. The formation of large complexes on ssDNA oligonucleotides was also shown by EMSA experiments. When protein concentration was 10-fold excess of ssDNA substrate, the formation of large DNA-protein complex was visible on the native gel indicating that in low salt condition, protein molecules aggregate on the ssDNA substrate (Figure 3.27).

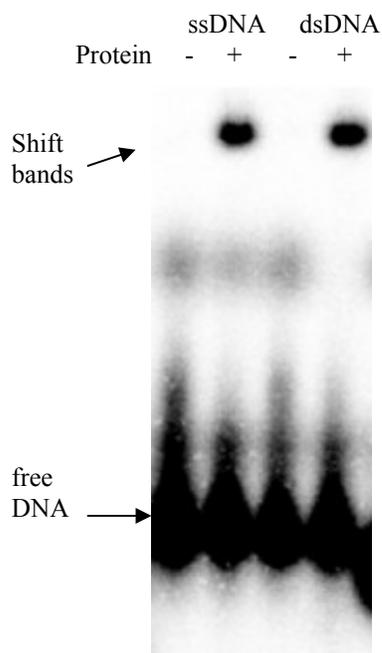


Figure 3.27: Formation of large complex of *Sso1450C6H* with DNA in EMSA experiment. Concentration of protein was 1 μ M, concentration of ssDNA (ssb-36) was 100nM. The strands were 5'-end radioactively labeled. The samples were incubated at 25°C for 30 min, and then were loaded on 4 % native gel.

3.5.5 Renaturation activity of *Sso1450C6H* for denatured dsDNA

The titration data show that *Sso1450C6H* binds with comparable affinity to ssDNA, dsDNA and also RNA substrates. Furthermore, large complexes are observed that are probably due to the aggregation of protein molecules on DNA substrate. These properties suggested that *Sso1450* might participate in DNA and/or RNA renaturation and strand annealing. To investigate this point, *Sso1450C6H* was incubated with separated complementary DNA strands (287 bp) and the protein-induced renaturation of the strands was then followed by native gel analysis. The re-formed dsDNAs (Figure 3.28, lane 3-5) indicates that *Sso1450C6H* is able to promote DNA renaturation even in a one to one molecular ratio model in 20 min at 30°C, suggesting a highly efficient catalysis of DNA annealing. The dsDNA background (Figure 3.28, Lane 2) might be due to the automatic zipping occurrence between completely paired DNA strands even at low temperature.

Results

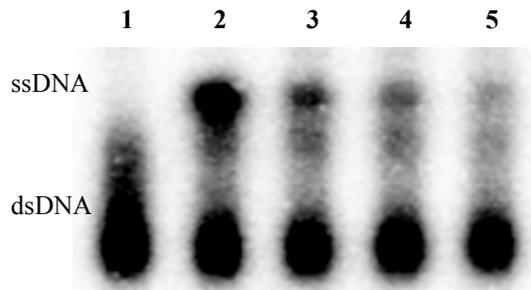


Figure 3.28: Renaturation of DNA strands promoted by *Sso1450C6H*. A 5'-end labeled 287 bp dsDNA was obtained by PCR and was then denatured at 95°C for 5 min. The renaturation was performed with 5 nM of DNA probe at increasing concentrations of *Sso1450C6H* in 20 min at 30°C. The reaction was then stopped by addition of 0.5 % SDS and 20mM EDTA. The samples were then analyzed on a 4 % native PAGE gel. Lane 1, native dsDNA; Lane 2, denatured DNA strands; Lane 3 to 5, denatured DNA strands with 10 nM, 40 nM and 400 nM of *Sso1450C6H* protein.

4. Discussion

4.1 Characterization of *Sso*SSB protein

SSBs are indispensable elements in all living organisms. Most SSBs bind ssDNA non-specifically, recruiting a variety of enzymes involved in the biological processes such as replication, recombination and repair of DNA. Eukaryotic and bacterial SSBs are well studied in the past decades (Wold, 1997; Raghunathan et al., 2000; Webster et al., 1997; Kur et al., 2005), whereas the biological properties of archaeal SSBs are poorly understood. In the present work the DNA binding properties of *Sso*SSB, as a representative SSB in crenarchaea, was characterized.

4.1.1 *Sso*SSB protein binds single-stranded DNA with high affinity

The DNA binding affinity of *Sso*SSB was investigated by fluorescence anisotropy and gel retardation assays using oligonucleotides as model binding substrates. The fluorescence titrations indicate a high affinity of *Sso*SSB to ssDNA oligonucleotides. The competition titration data with dsDNA oligonucleotides as competitors indicate over 100 fold higher binding affinity to ssDNA as compared to dsDNA. This result is consistent with previous reports from other groups (Haseltine and Kowalczykowski, 2002; Wadsworth and White, 2001) on the binding of *Sso*SSB to DNA. To quantify the binding affinity to ssDNA and dsDNA substrates, the fluorescence anisotropy assays were evaluated using a 1:1 binding model. This model allowed a good fit of the experimental data. The optimal binding dissociation constant is in subnanomolar range. The binding was also followed on long circular ss or dsDNA molecules using gel retardation assays. The results show binding of *Sso*SSB to dsDNA is much weaker than that to ssDNA. This is well proved by an experiment (Figure 3.2) where both ssDNA and dsDNA were offered as binding substrates. In this case, *Sso*SSB only bound ssDNA. This completely consists with the direct observation of *Sso*-DNA complexes by AFM. The binding of *Sso*SSB to the plasmid dsDNA is explained by the breathing of the dsDNA and the binding of *Sso*SSB to transiently formed double-stranded region. Bubbles formed during DNA breathing at high temperature may be captured by the *Sso*SSB molecules. Such a

behaviour of *Sso*SSB has been suggested by White's group (Cubeddu and White, 2005). Although SSB proteins are abundant in the cell, it is improbable that this process often occurs *in vivo*.

4.1.2 *Sso*SSB protein is a monomer in solution

The subunit structure of crenarchaeal SSB proteins has been a matter of debate (Wadsworth and White, 2001; Haseltine and Kowalczykowski, 2002). Multimeric forms (primarily tetrameric, less dimeric and monomeric form) have been reported (Haseltine and Kowalczykowski, 2002), whereas one group suggested a monomeric structure only (Wadsworth and White, 2001).

In this study, when the *Sso*SSB protein was purified and analyzed by gel filtration method on FPLC, only one peak was observed corresponding to the predicted size of *Sso*SSB. This strongly indicates that *Sso*SSB is a monomer in solution. The heating step (heat shock at 70°C for 20 min) before gel filtration had been included in the purification of *Sso*SSB. This step was convenient for the further purification steps, but it did not induce protein multimerization. This experimental result reveals that under physical salt concentration and neutral pH range, *Sso*SSB proteins in solution are mostly present as simple monomers. This form of *Sso*SSB is different from that of *Eco*SSB which is a homotetramer in solution (Chedin et al., 1998; Kelly et al., 1998) as well as that of RPA which has a heterotrimeric structure providing a total of four ssDNA-binding domains (OB folds) (Brill and Bastin-Shanower, 1998). It is even different from that of euryarchaeal SSB protein which resembles eukaryotic RPA with multi DNA-binding domains (Brill and Bastin-Shanower, 1998; Barns et al., 1996).

4.1.3 Binding mode of *Sso*SSB to ssDNA

Structural analysis of *Eco*SSB protein confirms the presence of four OB folds forming the functional homotetramer in ssDNA binding with “unlimited” inter-tetramer cooperativity at low salt concentration or with “limited” inter-tetramer cooperativity at high salt concentration (Raghunathan et al., 1997; Lohman et al., 1986; Lohman and Overman, 1985). In contrast, human RPA binds ssDNA in two very different modes. In a major binding mode, the four DNA-binding domains with OB folds and the zinc-ribbon

motif interact on the DNA substrate allowing the RPA heterotrimer to cover 30 nt in total (Kim et al., 1992a). This binding mode exhibits high affinity and low cooperativity. The second mode, which is less stable and may be a precursor of the 30 nt mode, has an 8-10 nt binding site with a lower affinity and a higher cooperativity (Blackwell and Borowiec, 1994).

The question of the binding mode of *SsoSSB* to ssDNA was followed by using various ssDNAs including phage M13 ssDNA and ssDNA oligonucleotides of different lengths as binding substrates. The results (Figure 3.2) show a gradual shift of the ssDNA mobility with increasing protein concentration and the DNA strands can be less stained by ethidium bromide. At saturation, the protein-bound DNA strands are difficult to be visualized under UV light. Distinct intermediate bands that can be well stained are observed at lower SSB concentration implying that the double-stranded regions in large ssDNA molecules can still be well stained. Along with increasing concentration of *SsoSSB*, these double-stranded regions are melted by the event of *SsoSSB* binding to single-stranded regions. However it is difficult to evaluate the cooperativity between SSB molecules.

There were some technical obstacles for ssDNA oligonucleotide retardation by *SsoSSB* on PAGE gel in a previous research (Wadsworth and White, 2001) probably due to the unfavorable conditions of electrophoresis. By overcoming these difficulties, the shifted DNA bands could be well resolved on the native PAGE gel in this study (Figure 3.3). The ssDNA oligonucleotides were separately retarded by the *SsoSSB* protein molecules according to the increase of protein concentration till saturation plateau. The complexes are interpreted as the successive binding of one, two and three *SsoSSB* molecules to the 42 nt ssDNA with increasing *SsoSSB* concentrations. This result indicates low or the absence of cooperativity with these substrates. For comparison, *EcoSSB* and RPA both show cooperativity on DNA binding, (Lohman and Ferrari, 1994; Bochkareva et al., 2002; Bochkareva et al., 2001; Lohman et al., 1988). The fluorescence anisotropy titrations show that the binding of the small oligonucleotides can be described by 1:1 complex with a *SsoSSB* monomer. Under these conditions, the dissociation constant strongly increased when ssDNA substrates were larger than 20 nt indicating a switch in the mode a *SsoSSB* molecule interacts with the DNA. Possibly conformational

changes take place on *Sso*SSB when longer oligonucleotides are bound. A recent work in our group showed that protein-DNA binding enthalpy was temperature independent (Kernchen and Lipps, 2006). The results also indicate that at lower salt concentration (50 mM NaCl) and a wide range of temperature *Sso*SSB binds small ssDNA as a monomer.

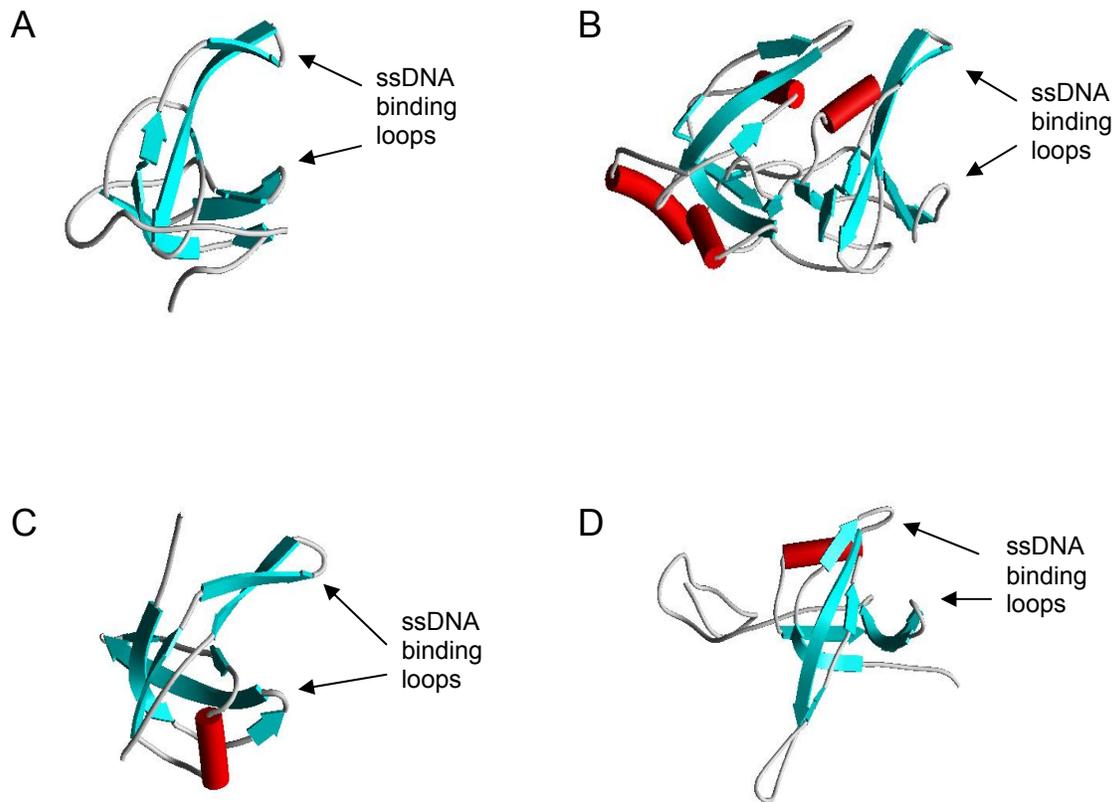


Figure 4.1: SSB proteins from three domains of life. **A**, the crystal structure of *Sso*SSB monomer; **B**, that of human RPA DBD-A and DBD-B domains; **C**, that of human mitochondrial SSB; **D**, that of *Eco*SSB. The ssDNA binding loops are indicated by black arrows. The important aromatic residues for ssDNA binding, Phe and Trp, are in the binding loops or in the nearby ribbons.

4.1.4 The binding site size and DNA-protein interaction

By calculating the molar ratios of phage M13 ssDNA nucleotides versus *Sso*SSB when saturation point was reached, the binding site size of *Sso*SSB was estimated to be

approx. 4-6 nt per protein monomer. This result is completely consistent with that of other groups (Haseltine and Kowalczykowski, 2002; Kerr et al., 2001). The staining of ethidium bromide on phage M13 ssDNA strands was gradually reduced along with the increase of protein concentration suggesting the protection of DNA bases by *SsoSSB*.

It is generally assumed that SSBs bind to ssDNA via an intercalation of aromatic amino acids into the stacked bases of the ssDNA (Figure 4.1). Such a binding mode will interfere with the binding of ethidium bromide to the DNA that also occurs by intercalation (Gago, 1998). The finding in the present work of a stronger binding of *SsoSSB* to pyrimidine-rich DNA as compared to purine-rich DNA is in line with this interpretation. The stronger self stacking of purine bases (Boon and Barton, 2002) will hinder the intercalation of aromatic residues from *SsoSSB*. A similar tendency of stronger binding to pyrimidine-rich nucleotides was also found for human RPA (Kim et al., 1992a) and many other ssDNA binding proteins (Kim and Wold, 1995). Previous research on human RPA has confirmed that the aromatic ring stacking, not the hydrogen bond, contributes strongly to complex formation between the protein molecule and the DNA strand (Figure 4.1B) (Bochkarev et al., 1997; Brill and Bastin-Shanower, 1998). In this interaction, Trp and Phe play major roles. Interestingly, recent research reported that these two residues contact the DNA in the crystal structure of *SsoSSB*-ssDNA complex (Kerr et al., 2003) implying that the stacking of aromatic amino acid residues on the individual bases of ssDNA is a major binding mode (Figure 4.1A). For human RPA, the pyrimidine binding priority was suggested to be due to the hyperphosphorylation of human RPA induced by DNA damage (Patrick et al., 2005). However this point is still unclear.

4.1.5 Does *SsoSSB* represent the ancestral SSB of three domains?

To our knowledge, both bacterial and eukaryotic SSBs are oligomers when binding to ssDNA, as well as euryarchaeal and mitochondrial SSBs (Webster et al., 1997; Lohman and Ferrari, 1994; Kelly et al., 1998; Wold, 1997; Komori and Ishino, 2001; Yang et al., 1997). These proteins share similar OB folds which play the main roles for ssDNA binding (Murzin, 1993). They bind ssDNA substrates with higher or lower cooperativity.

Furthermore, they contain c-terminal sequences or additional subunits for recruitment of other proteins.

*Sso*SSB expresses differences from the hitherto characterized SSBs. It is the first SSB protein characterized in Archaea. It shows only low cooperativity in binding ssDNA substrate. It binds ssDNA as a monomer. It contains a single OB fold which is structurally closer to eukaryotic SSBs (Figure 4.1), whereas its C-terminus shares similarity with bacterial SSBs. *Sso*SSB lacks the zinc ribbon motif which has been found to stabilize the heterotrimer of human RPA (Bochkareva et al., 2000). Recent data have shown that the C-terminus of *Sso*SSB plays a role for protein-protein interaction with other proteins and, interestingly, all these proteins contain zinc finger(s) (Napoli et al., 2005; Carpentieri et al., 2002; Richard et al., 2004) that are diverse both in structures and functions (Laity et al., 2001).

All together, *Sso*SSB is assumed to mimic a simplest ssDNA binding protein that consists of the minimum components for ssDNA binding and protein-protein interaction. It might be the closest one to the ancestral SSB in phylogenesis.

4.2 Computational analysis and expression screening of the putative protein genes from *Sso* P2

In the past years in our group, on aspect of the study was focused on DNA repair proteins, for instance, human RPA, XPA, and XPC. Besides the investigation of newly discovered *Sso*SSB, the screening of novel repair proteins was continuously performed to expect further understanding of repair mechanisms in thermophilic archaea. Therefore, when a prediction of novel, undetected repair system from Koonin's group (Makarova et al., 2002) appeared, it holded our interest. *Sso* is the most widely studied organism of the crenarchaeal branch of the Archaea. Its genome had been completely sequenced (She et al., 2001) just before the starting of this project. Therefore, to study these putative repair genes in more detail, *Sso* P2 was chosen as a model strain.

4.2.1 Scanning and collecting putative protein genes in *Sso* P2

Previous bioinformatic analysis discovered a partially conserved cluster of coding sequences consisting of more than 20 genes in most archaea and some bacterial

hyperthermophiles (Makarova et al., 2002). These putative protein genes were predicted to code for DNA helicases, HD-superfamily hydrolases, RecB family exonucleases, DNA polymerases. Furthermore, other completely uncharacterized protein genes were found in these clusters. However there were no detailed computational analyses for one specific genome. To well study homologous genes of this cluster in *Sso* P2, the genome of this organism was analyzed and the identified potential genes are listed in Table 3.2 according to their COG numbers. Normally, there were more than one homologous genes in each COG. These genes were divided into a few groups according to gene location, gene orientation, and gene order. Five highly conserved core genes, COG1857, COG1688, COG1203, COG1468 and COG1518, are included in Table 3.3. These analyses provide an integrated model system for further investigation of the putative genes and their relations.

4.2.2 Do these genes represent a novel repair system?

Earlier, the putative genes had been predicted to code for a novel repair system in archaea (Makarova et al., 2002). However, the sequence analysis of the present work revealed that the putative genes in *Sso* P2 are related to CRISPR.

In many prokaryotes, CRISPR loci are flanked by *cas* protein genes that have been defined by homology analysis (Godde and Bickerton, 2006; Jansen et al., 2002a). The presence of the *cas* genes was invariably associated with the presence of CRISPR loci in the genomes suggesting a functional relationship between the *cas* genes and the CRISPRs. The proteins encoded by *cas* genes share homology with proteins involved in DNA recombination and repair. This is the reason that many Cas proteins had been predicted as repair proteins before CRISPRs was observed. During the present work, Koonin's group also corrected the prediction of these proteins to Cas proteins (Makarova et al., 2006). Although the biological function of CRISPRs and Cas proteins remains obscure, the newer bioinformatic analyses strongly suggest the presence of a CRISPR mediated immune system in archaea and some bacteria. This immune system appears to share some features with RNAi in eukaryotes (Mojica et al., 2005; Makarova et al., 2006; Peng et al., 2003; Barrangou et al., 2007; Haft et al., 2005). In a recent work it has been shown that CRISPR provides, together with associated *cas* genes, resistance against bacteriophages.

The experimental finding of non-coding RNAs and the key protein Argonaute in archaea (Tang et al., 2005; Tang et al., 2002; Song et al., 2004) also lends support to this hypothesis. However, none of the Cas proteins appears to be related to Argonaute proteins.

The sequence analysis in *Sso* P2 genome in the present work reveals that the putative genes collected in Table 3.2 are located very close to CRISPRs or in between two CRISPRs. Furthermore, the sequence alignment shows that most of these genes are predicted as *cas* protein genes implying the relationship between these genes and CRISPRs. This result is compatible with other CRISPR and *cas* gene analyses and suggests that the putative protein genes are *cas* genes which are functionally related to CRISPRs. Together with CRISPRs, the *cas* genes might be involved in a novel immune system in *Sso* P2. The earlier hypothesis (Makarova et al., 2002) that these COG genes might fight function in DNA repair appears to be no longer valid. These COG genes seem to function rather as *cas* genes providing resistance against foreign infections.

4.2.3 Operon analysis of putative genes in *Sso* P2 genome

In the last decade, great progress has been made with the understanding of archaeal transcription and translation. The archaeal DNA-dependent RNA polymerase has been characterized, archaeal cell-free transcription systems have been developed (Soppa, 1999; Bell and Jackson, 2001; Bell et al., 2001; Reeve, 2003), and the complete sequencing of a number of archaeal genomes revealed the transcriptional control sequences. These studies discovered not only the factors involving in transcription and translation processes, but also the specific sequences in archaeal genome which were recognized by these factors, such as A+T rich TATA-box like element (Qureshi and Jackson, 1998), purine-rich BRE region (Soppa, 2001; Bell et al., 1999), T-rich terminator (Reiter et al., 1988) and Shine-Dalgarno sequence (Tolstrup et al., 2000).

In the present work, these data were used to predict gene location, gene arrangement and gene relation of the putative *cas* (COG) genes from *Sso* P2 strain to provide information on gene expression and functional relations. Since more than twenty *cas* genes exist in *Sso* P2 genome, the analysis concentrated on three regions. The potential protein genes in these regions are shown in Table 3.3 and Figure 3.9. The finding of

operons indicates that three regions in genome can be identified that contain three groups of gene strings organized in operons. In each group, the BRE/TATA-box regions, Shine-Dalgarno sequences and terminator regions are extremely similar to that in the other groups. The sequences of BRE/TATA-box are conserved and the distance to the start codon varies from 13 nucleotides to 37 nucleotides. However, in front of some putative genes, a different sequence, which is not as same as Shine-Dalgarno sequence, is observed that it is not obviously conserved but G-rich. This is consistent with the former analysis (Torarinsson et al., 2005; Tolstrup et al., 2000) that showed *Sso* to use two different translation initiation mechanisms, one with and the other without Shine-Dalgarno sequence. The leaderless (without Shine-Dalgarno sequence) translation could be a molecular fossil from LUCA (Condo et al., 1999).

In each of the gene strings, the putative ORFs are clustered and overlapped sharing one or more nucleotides with adjacent genes. This implies that these genes should be highly conserved during evolution (Krakauer, 2000). They might be functionally related and co-expressed in transcription and translation (Shinkai et al., 2007; Viswanathan et al., 2007). Together with function prediction of these genes, it appears that the clustered, overlapped genes could be co-expressed in an efficiently controlled pathway to complete a certain task *in vivo* that might be immunity.

4.2.4 *cas* gene expression, Cas protein refolding and enzymatic activity detection

In the present work, over ten putative *cas* protein genes and their homologues had been tried to be expressed in soluble active form in various *E.coli* expression systems. Most of them failed to be expressed in soluble form, except *ss01442*, *ss01450*, *ss01997* and *ss01998*. The formation of insoluble gene products indicates an incorrect folding of the proteins. There are a number of factors that can contribute to the expression of recombinant proteins as insoluble aggregates, including high expression rate, unfavorable folding conditions (temperature, pH), or the absence of suitable chaperone proteins that help to avoid incorrectly folded states (Baneyx and Mujacic, 2004). Other factors specific for *ss0* gene expression could be the toxicity to *E.coli* cells of Cas proteins harbouring DNA binding activity or nuclease activity and a requirement of high temperature for

correct folding (Ward et al., 2000; Andreotti et al., 1995). When expressed at low temperatures, incorrect folding of thermophilic proteins may occur (Ogasahara et al., 1998). It is also possible that an unknown trigger might be necessary for initiating *cas* gene expression, like a certain response to DNA damage in *Sso* (Salerno et al., 2003). This trigger could be the expression of the neighbour gene(s), and the invading of foreign genetic elements. In the present study, a large number of different expression conditions for *Sso1999* and *Sso2001* had been tested. Co-expression had been tried as well as the expression as fusion with proteins such as SlyD (Scholz et al., 2006), DsbA (Wulfiging and Rappuoli, 1997) and fd (Martin and Schmid, 2003). The latter proteins function as chaperonins in bacteria and have been shown to improve the soluble expression of variety of proteins. However, all these experiments failed to improve the soluble expression of these *cas* genes. Probably *Sso*-specific chaperonins are required to improve the expression. A few *Sso* chaperonins have been described (Guagliardi et al., 1994; Ruggero et al., 1998; Guagliardi et al., 2004). But these proteins were not available for the present work. The overlapping ORFs from computational analysis imply not only a functional relation but also the possibility of co-expression of these genes. The co-expression of *sso1998~sso2002* in 2/4/6 mode failed in *E.coli* which points to a different expression pathway in *Sso*. In the *E.coli* expression system, the genes are arranged in sequential order; they are expressed one by one separately, whereas in *Sso* P2 the genes overlap implying that these genes could be transcriptionally coupled (McCarthy, 1990), or the overlap per se might function in the regulation of gene expression (Johnson and Chisholm, 2004) *Sso*-based expression systems have been developed only recently and were not available for the present work.

Due to the insoluble form of the *Sso* Cas proteins expressed in *E.coli* system, experiments on the renaturation of the proteins were performed. This effort focused on *Sso1999* and *Sso2001* proteins. The refolding screenings were based on present methods with successful practice (Middelberg, 2002; Tsumoto et al., 2003; Vincentelli et al., 2004) including on-column refolding, rapid dilution, step-wise dialysis, high-throughput assay and co-refolding. Soluble proteins were obtained from some of the screenings in the presence of arginine indicating its important role in refolding. This is consistent with other reports (Ishibashi et al., 2005; Tsumoto et al., 2005). For primary screening, high-

throughput assay (Vincentelli et al., 2004) has advantages of investigation of a number of conditions including pH, buffer components, detergents and other additives. In the present work *Sso1999* and *Sso2001* were both successfully re-solubilized by this assay.

Following protein refolding, the enzymatic activities of soluble *Sso1999* and *Sso2001* were scanned. *Sso1999* was predicted to be a helicase. Typically, helicases catalyze the separation of duplex oligonucleotides into single stands in an ATP-dependent reaction. Stimulation of ATPase activity by DNA or RNA is a corollary activity to duplex unwinding, and is often taken as evidence to indicate that a protein may be a helicase. *Sso1999* has well conserved walker A and walker B motifs for ATP-binding and hydrolysis, conserved motif VI for RNA binding, besides motif III for nucleic acid unwinding. Unfortunately, in the present work, no ATPase activity was observed in presence/absence of DNA (plasmid dsDNA, ss/ds DNA oligonucleotides) or RNA (tRNA, dsSiRNA, ssRNA oligonucleotides). Repeat sequences and spacer sequences from CRISPR also failed to stimulate the ATPase activity of *Sso1999*. This may be due to several reasons: first, the protein is misfolded in inactive form although being soluble; second, a specific DNA or RNA may be required for ATPase activity; and third, a protein co-factor may be necessary. In a very recent research, infection of foreign phage sequence activated the Cas protein associated immune system of the host strain (Barrangou et al., 2007) indicating the functional relationship between Cas protein, CRISPR and invading sequence, whereas the individual role of Cas protein remains unknown.

4.3 Expression and characterization of *Sso2001*Est protein

4.3.1 Soluble expression of *Sso2001* with esterase

Sso2001 is one of the putative core proteins in the Cas protein group (Makarova et al., 2002; Makarova et al., 2006). In the first expression screening experiments when this project was initiated, *sso2001* failed to be expressed in soluble form. Efforts to induce soluble expressions included variations of parameters such as culture media composition, growth temperature, inducer dosage and codon usage, use of different *E.coli* strains and plasmids, adding folding helper sequences, chaperone co-expression and refolding. Finally, it was found that a fusion protein gene of *sso2001* with esterase gene from

thermophilic bacterial *Alicyclobacillus acidocaldorius* could be expressed partially in soluble form. The 34 kDa esterase is a thermostable, single chain protein that folds into a one domain structure with one active center that possess a lipase-like Ser-His-Asp catalytic triad. The overall fold is typical for α/β hydrolases. The N- and C- terminal ends of the protein expose on the esterase surface and are not involved in forming the catalytic center of the enzyme. The esterase itself is thermostable and active over a broad pH range. This provides the possibility to fuse the esterase with other polypeptides without affecting the enzymatic activity (Agafonov et al., 2005).

In the present work, *sso2001* gene was fused with esterase at the C- terminus to construct the fusion protein gene, *sso2001Est*. The soluble expression of the fusion protein was successful. The advantages of fusion with esterase were firstly, the soluble full length fusion protein was monitored in real time by esterase activity assay (Manco et al., 1998), and secondly, that the fusion protein could to be purified in a single step by TFK affinity column which specifically binds to esterase (Huang et al., 2007). The fact that the fusion protein can be obtained in a soluble form indicates that the folding of *Sso2001* is facilitated by the presence of the well soluble esterase that apparently performs a chaperon-like function. This finding, together with computational analysis, suggests that in *Sso* strain, the soluble expression of *sso2001* gene might depend on the expression of the overlapped neighbours, *sso1998*, *sso1999* and *sso2002*. Together with one or more of them, the expressed proteins might interact to form a structural and functional complex. Another possibility is some other genes out of *cas* gene clusters could be co-expressed with *cas* genes (Viswanathan et al., 2007). However, all the experiments to achieve co-expression of the *sso2001* gene with its overlapping neighbours had failed in the present work.

4.3.2 *Sso2001* is as a endonuclease

sso2001 belongs to a new protein gene group, COG2254 in computational analysis (Makarova et al., 2002; Makarova et al., 2006; Haft et al., 2005). All the homologous protein genes in this group are present in archaea and are uncharacterized with hypothetical nuclease activity. In *Sso* P2, there are four genes from COG2254, namely *sso1269*, *sso1403*, *sso1439*, and *sso2001* and all have a conserved HD domain. To test

the nuclease hypothesis, *sso2001* gene was chosen for further studies and was expressed with a fusion partner, the esterase.

In primary experiments, *sso2001* was found to be an endonuclease. It cleaved the DNA at internal sequences and produces short oligonucleotide fragments. It prefers dsDNA over ssDNA.

4.3.3 Characterization of *Sso2001* nuclease

Sso2001 was characterized with respect to pH, temperature, site specificity, metal ion dependency, active site determination and Michaelis constants.

The cleavage reaction was highly dependent on the pH of the buffer. The optimal pH of *Sso2001*Est fusion protein nuclease activity was in the range of neutral pH (7~8). However, an approx. 10% lower activity was found at pH 3. This phenomenon of two pH optima is rare for enzyme activity, although some enzymes show different pH optima for different activities or for different substrates (Desai and Shankar, 2003). One exclusive example is sialidase with different forms at corresponding pHs (Venerando et al., 2003). Therefore, it seems that the *Sso2001*Est fusion protein is active both in acidic and neutral pH in different enzyme forms.

The optimal temperature of *Sso2001*Est fusion protein has been found to be 50°C. For an enzyme from a thermophilic organism with an optimal growth temperature of around 75°C, one would expect a higher temperature optimum. The lower temperature optimum might be due to the balance of the thermostability between the thermophilic enzyme *Sso2001* and the esterase which optimal temperature is 55°C. Since the esterase is important for the solubility of *Sso2001*, this suggests that the temperature optimum of the fusion protein is not the real optimum of *Sso2001 in vivo*. Another explanation for the rather low temperature optimum could be the melting of the dsDNA oligonucleotides at high temperature so that the nuclease loses the specific substrates *in vitro*. The cleavage occurred more efficiently in the vicinity of G bases. Other examples are the nucleases from *Streptomyces antibioticus*, *Neurospora crassa* and *Chlamydomonas reinhardtii* that showed a preference for dG bases in dsDNA or ssDNA oligonucleotide (Burton et al., 1977; Linn and Lehman, 1965; Cal et al., 1996), although the majority of nucleases reported so far are base non-specific.

The activity tests showed furthermore that the divalent magnesium ion is essential for the nuclease activity that is completely inhibited by metal chelater, EDTA. However, the replacement of Mg^{2+} by Ca^{2+} , Mn^{2+} or Zn^{2+} leads to a loss of enzyme activity. That means *Sso2001* might be a metal-requiring protein and, magnesium plays an important role in protein functional conformation stability of the active site(s) and in the catalytic processes of the enzyme (Barondeau and Getzoff, 2004). This is common to nucleases (Desai and Shankar, 2003; Kevin N.Dalby, 1993) that the cleavage of phosphodiester bonds require divalent one or more metal ions. Most of the metal binding ligands of nuclease are provided by aspartic acid, glutamic acid, histidine, serine, and cysteine amino acid side chains (Chakrabarti, 1990; Glusker, 1991). In the present work, the importance of some of the highly conserved residues was investigated by point mutation. The activity assay of the point mutants shows that aspartic acid and glutamic acid residues (D84 and E103), neither two histidine residues (H83 and H104), nor serine residue (S106), are crucial for nuclease activity. This is consistent with the finding that Mg^{2+} is essential for catalyzing DNA cleavage by *Sso2001*. It is also consistent with data on protein structures of magnesium-requiring enzymes, where the amino acids that coordinate Mg^{2+} are aspartic acid and glutamic acid residues (Sutera, Jr. et al., 1999; Jockovich and Myers, 2001; Glusker, 1991; Shatilla et al., 2005; Gite and Shankar, 1995).

4.3.4 Is *Sso2001* a HD-superfamily enzyme?

Homology search predicted *Sso2001* to be a HD-like nuclease (Makarova et al., 2002; Makarova et al., 2006). In HD-superfamily proteins, histidine and aspartic acid residues form the active domain (namely HD from these two conserved residues). They both cooperate with metal ions and activate the enzyme activity (Yakunin et al., 2004; Aravind and Koonin, 1998). The alignment on *Sso2001* with its homologues identifies, in addition to HD sequence, a conserved HE motif and a conserved serine residue. The HE motif is not found in HD-superfamily enzymes and these also do not contain a conserved serine residue. By point mutation, the importance of all the conserved residues for the nuclease activity was studied. The mutations of H83 in HD motif and H104 in HE motif only slightly interfered with the nuclease activity, and the mutation of serine residues had no effect. On the contrary, by the mutation of D84 in HD motif and E103 in HE motif, the

nuclease activities were strongly reduced implying that not HD domain, but rather conserved aspartic acid and glutamic acid residues are essential for the activity. Possibly both cooperate with magnesium ion binding. Therefore, the mutational studies indicate that *Sso2001* does not belong to the HD-superfamily of nuclease. It should be a newly defined nuclease in COG2254.

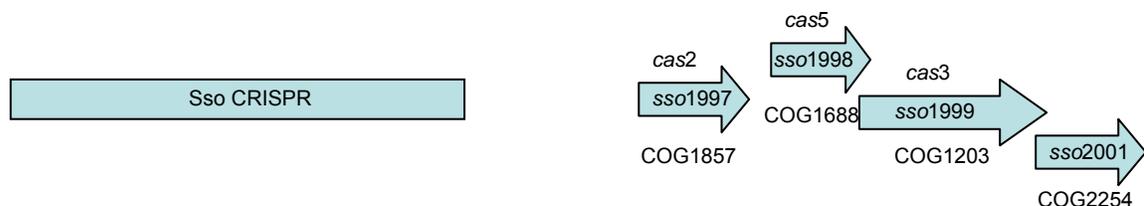


Figure 4.2: Some *cas* protein genes' locations close to one of *Sso* CRISPR sequences. In sequence alignment, some *Sso* P2 genes (numbers are noted in the blue arrows) were found belonging to various COGs (numbers are noted below the blue arrows). All these genes above were predicted as *cas* genes (numbers were noted above the blue arrows) except *sso2001* which co-exists with *sso1999*, like other COG2254 genes with *cas3* genes. *sso1998*, *sso1999* and *sso2001* are head-to-tail overlapped implying function relationship.

The neighborhood of the genes for a helicase and a nuclease is common in DNA repair systems (Aravind et al., 1999; Marti and Fleck, 2004). Furthermore, fusion proteins are sometimes found that harbour both helicase and nuclease activities (Komori et al., 2004; Yu et al., 1998). The sequence analysis shows a head-to-tail arrangement and overlapping of the genes for *sso2001* and *sso1999*, similar to most of the COG2254 and COG1203 hypothetical protein genes (Figure 4.2). *Sso1999* is predicted to be a helicase candidate with well defined motifs and *Sso2001* contains a HD domain in its sequence. Recent analysis reveals that some COG2254 protein genes are fused within COG1203 protein genes in *cas3* cluster (Makarova et al., 2006). Both genes are always located close to CRISPR sequence or in between two CRISPRs. Their functions were predicted to be related to the metabolism of CRISPRs suggesting that they could be part of a novel immune system in archaea for dsRNA unwinding and cleavage (Makarova et al., 2006; Haft et al., 2005). The study in the present work certainly defined the nuclease activity of *Sso2001* although its biological role remains to be characterized. Up to now, no sequence-specific cleavage of CRISPR sequences could be observed for *Sso2001*.

4.4 Characterization of *Sso1450C6H*

4.4.1 Nonspecific binding of *Sso1450C6H* protein to nucleic acid

Homology searches did not provide a prediction of the biochemical properties of *Sso1450* protein. In the present work, *ssol450* could be expressed in a soluble form as a fusion protein with a C-terminal hexahistidine tag. *Sso1450C6H* protein could be shown to be a nucleic acid binding protein. It appears to bind in sequence-independent way to ssDNA, dsDNA and RNA. In fluorescence anisotropy experiments, *Sso1450C6H* bound various kinds of oligonucleotides with similar affinity with binding constant in nanomolar range. Interestingly, no clear nucleic acid binding motif (Wei Yang and Gregory D Van Duyne, 2004) was found in the *Sso1450C6H* sequence using different computational algorithms. Therefore, mutational studies of the protein could not be performed in a rational way.

The binding mode of *Sso1450C6H* was investigated by fluorescence anisotropy assay using oligonucleotides as substrates. The data could not be fitted well with a one-site binding model. A better fit was obtained with a model based on the binding of dimers indicating that more than one protein molecules participate in the binding of the oligonucleotide substrate. This was confirmed on native PAGE experiments where a large complex was observed with excessive amount of protein. This might be due to that the nucleic acid molecules were embedded by protein molecules to form large aggregate which was consistent with the previous results from other proteins containing nucleic acid binding ability, such as RecA, Rad51 and Dmc1 (Weinstock et al., 1979a; Tsang et al., 1985; Sauvageau et al., 2005; Wei Yang and Gregory D Van Duyne, 2004).

4.4.2 *Sso1450C6H* promotes annealing of complementary DNA strands

Since *Sso1450* bound ssDNA, dsDNA and RNA with comparable high affinity, it was hypothesized that *Sso1450* could be involved in the annealing of complementary DNA or RNA strands. Such an activity is required for recombination events and has been shown for a number of proteins such as RecA and p53 (Bakalkin et al., 1994; Bryant and Lehman, 1986; Weinstock et al., 1979b). Therefore, experiments were performed to investigate a potential annealing activity of *Sso1450*. Besides nucleic acid binding

activity, another function of *Sso1450C6H* was screened out namely that it promotes annealing of complementary DNA strands in the presence of magnesium. An annealing activity could be indeed observed at excess concentration of protein over that of DNA suggesting that the aggregation of protein molecules and DNA strands is required to bring about the annealing of complementary DNA strands in the presence of Mg^{2+} . This is consistent with the report on RecA catalyzed DNA annealing (Weinstock et al., 1979a; Sauvageau et al., 2005; Cole and Kmiec, 1994; Wabiko et al., 1983; Tsang et al., 1985). Apparently, the formation of protein-ssDNA aggregates serves to provide a highly effective concentration of DNA strands, which would enhance the rate of annealing of complementary strands. This would also imply that *Sso1450* is able to bind two single-strands and therefore has two DNA binding sites. This point has to be investigated in the further experiments. In addition, *Sso1450* might not be the only enzyme catalyzing annealing reaction, as it was not very efficient at lower concentration (Figure 3.28, Lane 3 and Lane 4). It perhaps requires the presence of other factors (Shinohara et al., 1998; Kantake et al., 2002).

4.4.3 What might be the role of *Sso1450 in vivo*?

A CRISPR associated system (CASS) is newly discovered in all archaeal and some bacterial genomes with putative immune function of defense against invading phage and plasmid DNAs (Makarova et al., 2006; Haft et al., 2005). CASS is often eliminated from genomes during evolution. This plasticity of CASS operon organization suggests that the CRISPR-associated genomic regions are “hot spots” of recombination leading to genome rearrangement. Until now the functional characterization of this region is only based on computational analysis. *Sso1450* belongs to COG 1518. All the proteins in COG 1518 were predicted to be involved in DNA repair before CASS had been discovered (Makarova et al., 2002). Further computational analysis classified them as Cas1 proteins and generally highly essential and conserved. Cas1 proteins were the only Cas proteins found consistently in all species that contain CRISPR loci. Therefore, Cas1 protein is a best universal marker of the CASS. The finding of the annealing activity and a high affinity binding to ssDNA, dsDNA and RNA in the present work suggests that *Sso1450*

might be involved in annealing events required during the excision and insertion of the CRISPR sequences.

In former sequence analysis (Makarova et al., 2002), the COG1518 proteins had been predicted to harbour a nuclease activity. However, in the present work, no nuclease activity could be detected with various kinds of nucleic acid substrates including repeat and spacer sequences derived from the CRISPRs that are located close to the *ss01450* gene. On the other hand, the finding of a nucleic acid binding activity and DNA annealing activity strongly suggests that *Sso1450* plays a role in DNA rearrangements. This agrees with a model in which COG 1518 protein (to which *Sso1450* belongs) is assumed to mediate the insertion of repeated sequences into foreign dsDNA derived from retro-transcribed phage or plasmid RNA by its integrase/recombinase activity (Figure 1.5, (Makarova et al., 2006). Since both nucleic acid binding and annealing are commonly essential in integrase/recombinase event (Sauvageau et al., 2005; Heyer et al., 1988; Makhov and Griffith, 2006). The apparent sequence-independence of *Sso1450C6H* might be due to either the lack of enzyme cofactor/partner or the incorrect conditions.

5. Summary

Following the complete sequencing of the genome of *Sulfolobus solfataricus* (*Sso*) P2, this organism has been widely used as a model strain for crenarchaea. The present work concentrated on the characterization of a newly discovered single-stranded binding (SSB) protein from *Sso* P2 and the computational and experimental analyses of clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (*cas*) genes, respectively. The DNA-binding properties of *Sso*SSB, the organization of the CRISPR loci and the biochemical properties of some of the *cas* gene products of *Sso* P2 were studied.

Size-exclusion chromatography indicated that *Sso*SSB exists as a monomer in solution. Using fluorescence anisotropy as a method to study the interaction of *Sso*SSB to DNA, it can be shown that *Sso*SSB binds as a monomer to small oligonucleotides. The approximate binding site size is 4- 6 nt per protein molecule as determined by native electrophoresis. *Sso*SSB shows a more than 10 fold higher binding affinity to single-stranded (ss) DNA as compared to double-stranded (ds) DNA, which is consistent with former reports. The dissociation constant could be determined to be in a low nanomolar range. Furthermore, *Sso*SSB preferentially binds to pyrimidine-rich ssDNA as compared to purine-rich DNA. This property is similar to that observed for human replication protein A (RPA).

The clustering of repeat sequences in CRISPR loci and the associated *cas* genes have emerged recently as a new genomic feature of Archaea and of some Bacteria. Formerly, the *cas* genes have been predicted to encode repair proteins. In the present work, five CRISPR loci and their *cas* genes in *Sso* P2 were analyzed with respect to their genomic organization. The repeats of the CRISPR loci show highly conserved sequences at regular intervals, separated by spacer sequences of similar size. Most of the *cas* gene groups flanking a CRISPR locus contain homologous genes that were also found at other CRISPR loci. The *cas* genes could be grouped by gene location and gene order. Mostly, in each group they were head-to-tail arranged implying a functional relation. The operons of the *cas* genes *ssol1996-2002*, *ssol1438-1443* and *ssol1398-1403* could be shown to contain Transcription Factor B recognition element (BRE), TATA-box, Shine-Dalgarno

Summary

and terminator sequences specific for *Sulfolobus*. Most *cas* genes could not be expressed in a soluble form in *E.coli*, even when the expression conditions were widely varied. Refolding of the insoluble proteins was then undertaken. *ssol442*, *ssol996* and *ssol997* could be expressed in partially soluble form in *E.coli*, however catalytical activities could not be identified for these proteins. Refolding of *Ssol999*, a putative helicase, yielded a soluble protein. However no helicase and ATPase activity could be detected in the renatured *Ssol999*.

Defined biochemical activities could be only assigned to the proteins *Ssol450* and *Ssol2001*. In the latter case, the *ssol2001* gene was fused with an esterase gene from *Alicyclobacillus acidocaldarius*, and was co-expressed in a soluble form. The enzymatic screening indicated that *Ssol2001* harbored a nuclease activity. Further experiments showed that *Ssol2001* was an endonuclease with specificity for cleavage near G residues. The nuclease activity was optimal at the neutral pH range with another activity peak at pH 3. Specific point mutations introduced in *Ssol2001* indicate that this protein was not a HD-family nuclease as previously predicted.

The protein *Ssol450* (COG1518), which is considered to be a marker protein of the CRISPR and Cas system, bound nucleic acids, including ssDNA, dsDNA and RNA, with high affinity. The dissociation constant of binding to DNA oligonucleotides was in the nanomolar range. EMSA experiments indicated an aggregation of *Ssol450* on the DNA substrates. Interestingly, *Ssol450* promoted the annealing of complementary ssDNAs. This finding supported a role of *Ssol450* in the recombination of repeat sequences of the CRISPR system as suggested by Koonin's group (Makarova et al., 2006). The CRISPRs were thought to play a major role in a newly discovered genome immune system in prokaryotes.

6. Zusammenfassung

Die Sequenzierung des kompletten Genoms von *Sulfolobus solfataricus* (*Sso*) P2 macht ihn zu einem der Modellorganismen für Crenarchaea. Diese Arbeit befasst sich mit der Charakterisierung eines neu entdeckten SSB Proteins von *Sso* P2, sowie der computergestützten Analyse von *clustered regularly interspaced short palindromic repeats* (CRISPR) und der experimentellen Untersuchung von *CRISPR-associated* (*cas*) Genen. Die DNS-Bindungseigenschaften von *Sso*SSB, die Organisation der CRISPR Loki und die biochemischen Eigenschaften einiger *cas* Genprodukte von *Sso* P2 wurden untersucht.

Durch Größenausschlusschromatographie konnte gezeigt werden, dass *Sso*SSB in Lösung als Monomer vorliegt. Mittels Fluoreszenzanisotropie konnte die Wechselwirkung von *Sso*SSB in seiner monomeren Form mit kurzen Oligonukleotiden nachgewiesen werden. Mit Hilfe der nativen Gelelektrophorese konnte die Größe der Bindungsstelle mit 4-6 Nukleotide pro Proteinmolekül bestimmt werden. *Sso*SSB zeigt eine mehr als zehnmal so hohe Bindungsaffinität für Einzelstrang-DNS (ssDNS) im Vergleich zu Doppelstrang-DNS (dsDNS), was mit bereits veröffentlichten Daten übereinstimmt. Die Dissoziationskonstante ist im nieder-nanomolaren Bereich angesiedelt. Darüber hinaus bindet *Sso*SSB, ähnlich wie humanes Replikationsprotein A (RPA), bevorzugt an pyrimidinreiche ssDNS.

Die Sequenzwiederholungen im CRISPR Locus und die dazugehörigen *cas* Gene haben sich jüngst zu einer neuen genomischen Eigenschaft der Archaea und einiger Bakterien herausgestellt. Ursprünglich wurde angenommen, dass die *cas* Gene für Reparaturproteine codieren. In der vorliegenden Arbeit wurden fünf CRISPR Loki und die zugehörigen *cas* Gene in *Sso* P2 auf ihre genomische Organisation hin untersucht. Darüber hinaus wurde die Expression der *cas* Gene analysiert. Die Sequenzwiederholungen der CRISPR-Loki weisen hoch konservierte Sequenzabschnitte in regelmäßigen Intervallen auf, die durch Platzhaltersequenzen ähnlicher Größe getrennt sind. Die meisten der *cas* Gengruppen eines benachbarten CRISPR Locus enthalten homologe Gene, die auch in anderen CRISPR-Loki gefunden werden. Die *cas* Gene können anhand ihres Genlokus und ihrer Genfolge gruppiert werden. In einer Gengruppe

sind sie meistens „Kopf an Schwanz“ angeordnet, was eine funktionelle Beziehung andeutet. Die *cas* Gene *sso1996-2002*, *sso1438-1443* und *sso1398-1403* besitzen jeweils eine für *Sulfolobus* spezifische *transcription factor B recognition element*, TATA-Box, Shine-Dalgarno- und Terminatorsequenz. Die meisten *cas* Gene konnten trotz unterschiedlich getesteter Bedingungen nicht in löslicher Form in *E.coli* exprimiert werden. Die unlöslichen Proteine wurden rückgefaltet. *Sso1442*, *Sso1996* und *Sso1997* konnten in teilweise löslicher Form in *E.coli* exprimiert werden. Dennoch konnte für sie keine biochemische Funktion gezeigt werden. Aus der Rückfaltung von *Sso1999*, einer vermeintlichen Helikase, ging ein lösliches Protein hervor. Es konnten jedoch weder ATPase- noch Helikaseaktivität nachgewiesen werden.

Definierte biochemische Aktivitäten konnten nur für *Sso1450* und *Sso2001* gezeigt werden. *Sso2001* wurde als Fusionsprotein mit einer Esterase aus *Alicyclobacillus acidocaldarius* in löslicher Form coexprimiert. Die enzymatische Untersuchung wies darauf hin, dass *Sso2001* Nukleaseaktivität besitzt. Weitere Experimente zeigten, dass es sich bei *Sso2001* um eine Endonuklease handelt, die spezifisch in der Nachbarschaft von Guaninen schneidet. Das pH-Optimum dieser Nuklease liegt im neutralen Bereich und zusätzlich bei pH 3. Spezifische Punktmutationen weisen darauf hin, dass es sich bei *Sso2001* nicht um ein Mitglied der HD-Nuklease-Familie handelt wie ursprünglich angenommen.

Das Protein *Sso1450* (COG1518), welches als Markerprotein des CRISPR- und Cas-Systems angesehen wird, bindet Nukleinsäuren wie ssDNS, dsDNS und RNS mit hoher Affinität. Die Dissoziationskonstante für die Bindung von DNS liegt im nanomolaren Bereich. EMSA-Experimente deuten auf eine Aggregation von *Sso1450* auf den DNS-Substraten hin. Interessanterweise fördert *Sso1450* die Anlagerung von komplementären ssDNS. Diese Tatsache unterstützt die Annahme, dass *Sso1450* eine Rolle in der Rekombination der Wiederholungssequenzen im CRISPR-System spielt, wie die Gruppe um Koonin (Makarova et al., 2006) bereits postuliert. Das CRISPR-System steht im Verdacht eine zentrale Rolle in einem neu gefundenen genomischen Immunsystem von Prokaryoten zu spielen.

7. References

- Agafonov,D.E., Rabe,K.S., Grote,M., Huang,Y., and Sprinzl,M. (2005). The esterase from *Alicyclobacillus acidocaldarius* as a reporter enzyme and affinity tag for protein biosynthesis. *FEBS Lett.* *579*, 2082-2086.
- Andreotti,G., Cubellis,M.V., Nitti,G., Sannia,G., Mai,X., Adams,M.W., and Marino,G. (1995). An extremely thermostable aromatic aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochim. Biophys. Acta* *1247*, 90-96.
- Aravind,L. and Koonin,E.V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.* *27*, 4658-4670.
- Aravind,L. and Koonin,E.V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.* *23*, 469-472.
- Aravind,L., Walker,D.R., and Koonin,E.V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* *27*, 1223-1242.
- Bakalkin,G., Yakovleva,T., Selivanova,G., Magnusson,K.P., Szekely,L., Kiseleva,E., Klein,G., Terenius,L., and Wiman,K.G. (1994). p53 binds single-stranded DNA ends and catalyzes DNA renaturation and strand transfer. *Proc. Natl. Acad. Sci. U. S. A* *91*, 413-417.
- Baneyx,F. and Mujacic,M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* *22*, 1399-1408.
- Barns,S.M., Delwiche,C.F., Palmer,J.D., and Pace,N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. U. S. A* *93*, 9188-9193.
- Barondeau,D.P. and Getzoff,E.D. (2004). Structural insights into protein-metal ion partnerships. *Curr. Opin. Struct. Biol.* *14*, 765-774.
- Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A., and Horvath,P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* *315*, 1709-1712.
- Bastin-Shanower,S.A. and Brill,S.J. (2001). Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. *J. Biol. Chem.* *276*, 36446-36453.
- Bell,S.D. (2005). Archaeal transcriptional regulation--variation on a bacterial theme? *Trends Microbiol.* *13*, 262-265.
- Bell,S.D. and Jackson,S.P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* *6*, 222-228.

References

- Bell, S.D. and Jackson, S.P. (2001). Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.* *4*, 208-213.
- Bell, S.D., Kosa, P.L., Sigler, P.B., and Jackson, S.P. (1999). Orientation of the transcription preinitiation complex in archaea. *Proc. Natl. Acad. Sci. U. S. A* *96*, 13662-13667.
- Bell, S.D., Magill, C.P., and Jackson, S.P. (2001). Basal and regulated transcription in Archaea. *Biochem. Soc. Trans.* *29*, 392-395.
- Bernstein, D.A., Eggington, J.M., Killoran, M.P., Mistic, A.M., Cox, M.M., and Keck, J.L. (2004). Crystal structure of the *Deinococcus radiodurans* single-stranded DNA-binding protein suggests a mechanism for coping with DNA damage. *Proc. Natl. Acad. Sci. U. S. A* *101*, 8575-8580.
- Blackwell, L.J. and Borowiec, J.A. (1994). Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol. Cell Biol.* *14*, 3993-4001.
- 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.
- Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. (2001). Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. *EMBO J.* *20*, 612-618.
- 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*, 3932-3936.
- Bochkareva, E., Korolev, S., and Bochkarev, A. (2000). The role for zinc in replication protein A. *J. Biol. Chem.* *275*, 27332-27338.
- Bochkareva, E., Korolev, S., Lees-Miller, S.P., and Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* *21*, 1855-1863.
- Bolotin, A., Quinkis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* *151*, 2551-2561.
- Boon, E.M. and Barton, J.K. (2002). Charge transport in DNA. *Curr. Opin. Struct. Biol.* *12*, 320-329.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* *72*, 248-254.

- Brill,S.J. and Bastin-Shanower,S. (1998). Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A. *Mol. Cell Biol.* *18*, 7225-7234.
- Brock,T.D. (1978). *Thermophilic Microorganisms and Life at High Temperatures*. Springer, New York).
- Bryant,F.R. and Lehman,I.R. (1986). ATP-independent renaturation of complementary DNA strands by the mutant recA1 protein from *Escherichia coli*. *J. Biol. Chem.* *261*, 12988-12993.
- Burton,W.G., Roberts,R.J., Myers,P.A., and Sager,R. (1977). A site-specific single-strand endonuclease from the eukaryote *Chlamydomonas*. *Proc. Natl. Acad. Sci. U. S. A* *74*, 2687-2691.
- Caetano-Anolles,G. and Caetano-Anolles,D. (2003). An evolutionarily structured universe of protein architecture. *Genome Res.* *13*, 1563-1571.
- Cal,S., Nicieza,R.G., Connolly,B.A., and Sanchez,J. (1996). Interaction of the periplasmic dG-selective *Streptomyces antibioticus* nuclease with oligodeoxynucleotide substrates. *Biochemistry* *35*, 10828-10836.
- Carpentieri,F., De,F.M., De,F.M., Rossi,M., and Pisani,F.M. (2002). Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. *J. Biol. Chem.* *277*, 12118-12127.
- Caruthers,J.M. and McKay,D.B. (2002). Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* *12*, 123-133.
- Chakrabarti,P. (1990). Interaction of metal ions with carboxylic and carboxamide groups in protein structures. *Protein Eng* *4*, 49-56.
- Chedin,F., Seitz,E.M., and Kowalczykowski,S.C. (1998). Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. *Trends Biochem. Sci.* *23*, 273-277.
- Chen,L., Brugger,K., Skovgaard,M., Redder,P., She,Q., Torarinsson,E., Greve,B., Awayez,M., Zibat,A., Klenk,H.P., and Garrett,R.A. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J. Bacteriol.* *187*, 4992-4999.
- Ciaramella,M., Pisani,F.M., and Rossi,M. (2002). Molecular biology of extremophiles: recent progress on the hyperthermophilic archaeon *Sulfolobus*. *Antonie Van Leeuwenhoek* *81*, 85-97.
- Cole,A.D. and Kmiec,E.B. (1994). ATP-independent DNA renaturation catalyzed by a protein from *Ustilago maydis*. *Eur. J. Biochem.* *220*, 75-82.

References

- Condo, I., Ciammaruconi, A., Benelli, D., Ruggero, D., and Londei, P. (1999). Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* *34*, 377-384.
- Cubeddu, L. and White, M.F. (2005). DNA damage detection by an archaeal single-stranded DNA-binding protein. *J. Mol. Biol.* *353*, 507-516.
- Dabrowski, S., Olszewski, M., Piatek, R., and Kur, J. (2002). Novel thermostable ssDNA-binding proteins from *Thermus thermophilus* and *T. aquaticus*-expression and purification. *Protein Expr. Purif.* *26*, 131-138.
- De, F.M., Esposito, L., Pucci, B., De, F.M., Manco, G., Rossi, M., and Pisani, F.M. (2004). Modular organization of a Cdc6-like protein from the crenarchaeon *Sulfolobus solfataricus*. *Biochem. J.* *381*, 645-653.
- DeBoy, R.T., Mongodin, E.F., Emerson, J.B., and Nelson, K.E. (2006). Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. *J. Bacteriol.* *188*, 2364-2374.
- Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N., and Heinemann, U. (2000). Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* *303*, 35-47.
- DeLong, E.F. and Pace, N.R. (2001). Environmental diversity of bacteria and archaea. *Syst. Biol.* *50*, 470-478.
- Desai, N.A. and Shankar, V. (2003). Single-strand-specific nucleases. *FEMS Microbiol. Rev.* *26*, 457-491.
- Dionne, I., Robinson, N.P., McGeoch, A.T., Marsh, V.L., Reddish, A., and Bell, S.D. (2003). DNA replication in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* *31*, 674-676.
- Durand, P., Mahe, F., Valin, A.S., and Nicolas, J. (2006). Browsing repeats in genomes: Pygram and an application to non-coding region analysis. *BMC. Bioinformatics.* *7*, 477.
- Edgar, R.C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC. Bioinformatics.* *8*, 18.
- Eggington, J.M., Haruta, N., Wood, E.A., and Cox, M.M. (2004). The single-stranded DNA-binding protein of *Deinococcus radiodurans*. *BMC. Microbiol.* *4*, 2.
- Filipkowski, P., Duraj-Thatte, A., and Kur, J. (2006). Novel thermostable single-stranded DNA-binding protein (SSB) from *Deinococcus geothermalis*. *Arch. Microbiol.* *186*, 129-137.

- Filipkowski,P., Duraj-Thatte,A., and Kur,J. (2007). Identification, cloning, expression, and characterization of a highly thermostable single-stranded-DNA-binding protein (SSB) from *Deinococcus murrayi*. *Protein Expr. Purif.* 53, 201-208.
- Gago,F. (1998). Stacking interactions and intercalative DNA binding. *Methods* 14, 277-292.
- Geiduschek,E.P. and Ouhammouch,M. (2005). Archaeal transcription and its regulators. *Mol. Microbiol.* 56, 1397-1407.
- Gite,S.U. and Shankar,V. (1995). Single-strand-specific nucleases. *Crit Rev. Microbiol.* 21, 101-122.
- Glusker,J.P. (1991). Structural aspects of metal liganding to functional groups in proteins. *Adv. Protein Chem.* 42, 1-76.
- Godde,J.S. and Bickerton,A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62, 718-729.
- Grissa,I., Vergnaud,G., and Pourcel,C. (2007a). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52-W57.
- Grissa,I., Vergnaud,G., and Pourcel,C. (2007b). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics.* 8, 172.
- Grogan,D.W. (2000). The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol.* 8, 180-185.
- Guagliardi,A., Cerchia,L., Bartolucci,S., and Rossi,M. (1994). The chaperonin from the archaeon *Sulfolobus solfataricus* promotes correct refolding and prevents thermal denaturation in vitro. *Protein Sci.* 3, 1436-1443.
- Guagliardi,A., Mancusi,L., and Rossi,M. (2004). Reversion of protein aggregation mediated by Sso7d in cell extracts of *Sulfolobus solfataricus*. *Biochem. J.* 381, 249-255.
- Gutell,R.R., Weiser,B., Woese,C.R., and Noller,H.F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 32, 155-216.
- Haft,D.H., Selengut,J., Mongodin,E.F., and Nelson,K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS. Comput. Biol.* 1, e60.
- Hall,M.C. and Matson,S.W. (1999). Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* 34, 867-877.

References

- Hanahan,D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* *166*, 557-580.
- Haseltine,C.A. and Kowalczykowski,S.C. (2002). A distinctive single-strand DNA-binding protein from the Archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* *43*, 1505-1515.
- Hey,T., Lipps,G., and Krauss,G. (2001). Binding of XPA and RPA to damaged DNA investigated by fluorescence anisotropy. *Biochemistry* *40*, 2901-2910.
- Heyer,W.D., Evans,D.H., and Kolodner,R.D. (1988). Renaturation of DNA by a *Saccharomyces cerevisiae* protein that catalyzes homologous pairing and strand exchange. *J. Biol. Chem.* *263*, 15189-15195.
- Higerd,T.B. and Spizizen,J. (1973). Isolation of two acetyl esterases from extracts of *Bacillus subtilis*. *J. Bacteriol.* *114*, 1184-1192.
- Howland, John L. *The Surprising Archaea: Discovering Another domain of Life*. [ISBN 0-19-511183-4], 67-71. 2000. Oxford University Press.
Ref Type: Generic
- Huang,Y., Humenik,M., and Sprinzl,M. (2007). Esterase 2 from *Alicyclobacillus acidocaldarius* as a reporter and affinity tag for expression and single step purification of polypeptides. *Protein Expr. Purif.* *54*, 94-100.
- Iftode,C., Daniely,Y., and Borowiec,J.A. (1999). Replication protein A (RPA): the eukaryotic SSB. *Crit Rev. Biochem. Mol. Biol.* *34*, 141-180.
- Ishibashi,M., Tsumoto,K., Tokunaga,M., Ejima,D., Kita,Y., and Arakawa,T. (2005). Is arginine a protein-denaturant? *Protein Expr. Purif.* *42*, 1-6.
- Ishino,Y., Shinagawa,H., Makino,K., Amemura,M., and Nakata,A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* *169*, 5429-5433.
- Jansen,R., Embden,J.D., Gaastra,W., and Schouls,L.M. (2002a). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* *43*, 1565-1575.
- Jansen,R., van Embden,J.D., Gaastra,W., and Schouls,L.M. (2002b). Identification of a novel family of sequence repeats among prokaryotes. *OMICS.* *6*, 23-33.
- Jockovich,M.E. and Myers,R.S. (2001). Nuclease activity is essential for RecBCD recombination in *Escherichia coli*. *Mol. Microbiol.* *41*, 949-962.
- Johnson,Z.I. and Chisholm,S.W. (2004). Properties of overlapping genes are conserved across microbial genomes. *Genome Res.* *14*, 2268-2272.

- Kantake,N., Madiraju,M.V., Sugiyama,T., and Kowalczykowski,S.C. (2002). Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. *Proc. Natl. Acad. Sci. U. S. A* 99, 15327-15332.
- Kawarabayasi,Y., Hino,Y., Horikawa,H., Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H., Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Otsuka,R., Nakazawa,H., Takamiya,M., Kato,Y., Yoshizawa,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Masuda,S., Yanagii,M., Nishimura,M., Yamagishi,A., Oshima,T., and Kikuchi,H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res.* 8, 123-140.
- Kelly,T.J., Simancek,P., and Brush,G.S. (1998). Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci. U. S. A* 95, 14634-14639.
- Kernchen,U. and Lipps,G. (2006). Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from *Sulfolobus solfataricus*. *Biochemistry* 45, 594-603.
- Kerr,I.D., Wadsworth,R.I., Blankenfeldt,W., Staines,A.G., White,M.F., and Naismith,J.H. (2001). Overexpression, purification, crystallization and data collection of a single-stranded DNA-binding protein from *Sulfolobus solfataricus*. *Acta Crystallogr. D. Biol. Crystallogr.* 57, 1290-1292.
- Kerr,I.D., Wadsworth,R.I., Cubeddu,L., Blankenfeldt,W., Naismith,J.H., and White,M.F. (2003). Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *EMBO J.* 22, 2561-2570.
- Kevin N.Dalby. Models for Nuclease Catalysis: Mechanisms for General Acid Catalysis of the Rapid Intramolecular Displacement of Methoxide from a Phosphate Diester. Anthony J.Kirby. 1993. *J. CHEM. SOC. PERKIN TRANS.* Florian Hollfelder. Ref Type: Generic
- Kim,C., Snyder,R.O., and Wold,M.S. (1992a). Binding properties of replication protein A from human and yeast cells. *Mol. Cell Biol.* 12, 3050-3059.
- Kim,C., Snyder,R.O., and Wold,M.S. (1992b). Binding properties of replication protein A from human and yeast cells. *Mol. Cell Biol.* 12, 3050-3059.
- Kim,C. and Wold,M.S. (1995). Recombinant human replication protein A binds to polynucleotides with low cooperativity. *Biochemistry* 34, 2058-2064.
- Komori,K., Hidaka,M., Horiuchi,T., Fujikane,R., Shinagawa,H., and Ishino,Y. (2004). Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *J. Biol. Chem.* 279, 53175-53185.

References

- Komori,K. and Ishino,Y. (2001). Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J. Biol. Chem.* *276*, 25654-25660.
- Krakauer,D.C. (2000). Stability and evolution of overlapping genes. *Evolution Int. J. Org. Evolution* *54*, 731-739.
- Kur,J., Olszewski,M., Dlugolecka,A., and Filipkowski,P. (2005). Single-stranded DNA-binding proteins (SSBs) -- sources and applications in molecular biology. *Acta Biochim. Pol.* *52*, 569-574.
- Laity,J.H., Lee,B.M., and Wright,P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* *11*, 39-46.
- Li,T., Sun,F., Ji,X., Feng,Y., and Rao,Z. (2003). Structure based hyperthermostability of archaeal histone HPhA from *Pyrococcus horikoshii*. *J. Mol. Biol.* *325*, 1031-1037.
- Lillestol,R.K., Redder,P., Garrett,R.A., and Brugger,K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea.* *2*, 59-72.
- Lin,Y.L., Chen,C., Keshav,K.F., Winchester,E., and Dutta,A. (1996). Dissection of functional domains of the human DNA replication protein complex replication protein A. *J. Biol. Chem.* *271*, 17190-17198.
- Linn,S. and Lehman,I.R. (1965). AN ENDONUCLEASE FROM NEUROSPORA CRASSA SPECIFIC FOR POLYNUCLEOTIDES LACKING AN ORDERED STRUCTURE. II. STUDIES OF ENZYME SPECIFICITY. *J. Biol. Chem.* *240*, 1294-1304.
- Lohman,T.M., Bujalowski,W., Overman,L.B., and Wei,T.F. (1988). Interactions of the *E. coli* single strand binding (SSB) protein with ss nucleic acids. Binding mode transitions and equilibrium binding studies. *Biochem. Pharmacol.* *37*, 1781-1782.
- Lohman,T.M. and Ferrari,M.E. (1994). *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* *63*, 527-570.
- Lohman,T.M. and Overman,L.B. (1985). Two binding modes in *Escherichia coli* single strand binding protein-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 co-operativity of *Escherichia coli* single strand binding protein. *J. Mol. Biol.* *187*, 603-615.
- Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251-260.

- Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997b). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-260.
- Lysetska,M., Knoll,A., Boehringer,D., Hey,T., Krauss,G., and Krausch,G. (2002). UV light-damaged DNA and its interaction with human replication protein A: an atomic force microscopy study. *Nucleic Acids Res.* 30, 2686-2691.
- Makarova,K.S., Aravind,L., Grishin,N.V., Rogozin,I.B., and Koonin,E.V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* 30, 482-496.
- Makarova,K.S., Grishin,N.V., Shabalina,S.A., Wolf,Y.I., and Koonin,E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* 1, 7.
- Makhov,A.M. and Griffith,J.D. (2006). Visualization of the annealing of complementary single-stranded DNA catalyzed by the herpes simplex virus type 1 ICP8 SSB/recombinase. *J. Mol. Biol.* 355, 911-922.
- Manco,G., Adinolfi,E., Pisani,F.M., Ottolina,G., Carrea,G., and Rossi,M. (1998). Overexpression and properties of a new thermophilic and thermostable esterase from *Bacillus acidocaldarius* with sequence similarity to hormone-sensitive lipase subfamily. *Biochem. J.* 332 (Pt 1), 203-212.
- Marc,F., Sandman,K., Lurz,R., and Reeve,J.N. (2002). Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J. Biol. Chem.* 277, 30879-30886.
- Marsh,V.L., McGeoch,A.T., and Bell,S.D. (2006). Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM. *J. Mol. Biol.* 357, 1345-1350.
- Marti,T.M. and Fleck,O. (2004). DNA repair nucleases. *Cell Mol. Life Sci.* 61, 336-354.
- Martin,A. and Schmid,F.X. (2003). Evolutionary stabilization of the gene-3-protein of phage fd reveals the principles that govern the thermodynamic stability of two-domain proteins. *J. Mol. Biol.* 328, 863-875.
- Marx,J. (2007). Microbiology. New bacterial defense against phage invaders identified. *Science* 315, 1650-1651.
- 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. (Tokyo)* 127, 329-335.
- McCarthy,J.E. (1990). Post-transcriptional control in the polycistronic operon environment: studies of the atp operon of *Escherichia coli*. *Mol. Microbiol.* 4, 1233-1240.

References

- Middelberg,A.P. (2002). Preparative protein refolding. *Trends Biotechnol.* *20*, 437-443.
- Mojica,F.J., ez-Villasenor,C., Garcia-Martinez,J., and Soria,E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* *60*, 174-182.
- Mojica,F.J., ez-Villasenor,C., Soria,E., and Juez,G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* *36*, 244-246.
- Mojica,F.J., Ferrer,C., Juez,G., and Rodriguez-Valera,F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* *17*, 85-93.
- Mongodin,E.F., Hance,I.R., DeBoy,R.T., Gill,S.R., Daugherty,S., Huber,R., Fraser,C.M., Stetter,K., and Nelson,K.E. (2005). Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. *J. Bacteriol.* *187*, 4935-4944.
- Murzin,A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* *12*, 861-867.
- Myllykallio,H., Lopez,P., Lopez-Garcia,P., Heilig,R., Saurin,W., Zivanovic,Y., Philippe,H., and Forterre,P. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* *288*, 2212-2215.
- Napoli,A., Valenti,A., Salerno,V., Nadal,M., Garnier,F., Rossi,M., and Ciaramella,M. (2005). Functional interaction of reverse gyrase with single-strand binding protein of the archaeon *Sulfolobus*. *Nucleic Acids Res.* *33*, 564-576.
- Ogasahara,K., Nakamura,M., Nakura,S., Tsunasawa,S., Kato,I., Yoshimoto,T., and Yutani,K. (1998). The unusually slow unfolding rate causes the high stability of pyrrolidone carboxyl peptidase from a hyperthermophile, *Pyrococcus furiosus*: equilibrium and kinetic studies of guanidine hydrochloride-induced unfolding and refolding. *Biochemistry* *37*, 17537-17544.
- Otterlei,M., Warbrick,E., Nagelhus,T.A., Haug,T., Slupphaug,G., Akbari,M., Aas,P.A., Steinsbekk,K., Bakke,O., and Krokan,H.E. (1999). Post-replicative base excision repair in replication foci. *EMBO J.* *18*, 3834-3844.
- Patrick,S.M., Oakley,G.G., Dixon,K., and Turchi,J.J. (2005). DNA damage induced hyperphosphorylation of replication protein A. 2. Characterization of DNA binding activity, protein interactions, and activity in DNA replication and repair. *Biochemistry* *44*, 8438-8448.

- Peng,X., Brugger,K., Shen,B., Chen,L., She,Q., and Garrett,R.A. (2003). Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes. *J. Bacteriol.* *185*, 2410-2417.
- Perales,C., Cava,F., Meijer,W.J., and Berenguer,J. (2003). Enhancement of DNA, cDNA synthesis and fidelity at high temperatures by a dimeric single-stranded DNA-binding protein. *Nucleic Acids Res.* *31*, 6473-6480.
- Pfuetzner,R.A., Bochkarev,A., Frappier,L., and Edwards,A.M. (1997). Replication protein A. Characterization and crystallization of the DNA binding domain. *J. Biol. Chem.* *272*, 430-434.
- Pourcel,C., Salvignol,G., and Vergnaud,G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* *151*, 653-663.
- Qureshi,S.A. and Jackson,S.P. (1998). Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* *1*, 389-400.
- Raghunathan,S., Kozlov,A.G., Lohman,T.M., and Waksman,G. (2000). Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat. Struct. Biol.* *7*, 648-652.
- Raghunathan,S., Ricard,C.S., Lohman,T.M., and Waksman,G. (1997). Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution. *Proc. Natl. Acad. Sci. U. S. A* *94*, 6652-6657.
- Reeve,J.N. (2003). Archaeal chromatin and transcription. *Mol. Microbiol.* *48*, 587-598.
- Reiter,W.D., Palm,P., and Zillig,W. (1988). Transcription termination in the archaeobacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res.* *16*, 2445-2459.
- Richard,D.J., Bell,S.D., and White,M.F. (2004). Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Res.* *32*, 1065-1074.
- Rother,M. and Metcalf,W.W. (2005). Genetic technologies for Archaea. *Curr. Opin. Microbiol.* *8*, 745-751.
- Ruggero,D., Ciammaruconi,A., and Londei,P. (1998). The chaperonin of the archaeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing. *EMBO J.* *17*, 3471-3477.
- Salerno,V., Napoli,A., White,M.F., Rossi,M., and Ciaramella,M. (2003). Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* *31*, 6127-6138.

References

Sandman,K. and Reeve,J.N. (2005). Archaeal chromatin proteins: different structures but common function? *Curr. Opin. Microbiol.* 8, 656-661.

Sauvageau,S., Stasiak,A.Z., Banville,I., Ploquin,M., Stasiak,A., and Masson,J.Y. (2005). Fission yeast rad51 and dmc1, two efficient DNA recombinases forming helical nucleoprotein filaments. *Mol. Cell Biol.* 25, 4377-4387.

Schagger,H. and von Jagow,G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368-379.

Scholz,C., Eckert,B., Hagn,F., Schaarschmidt,P., Balbach,J., and Schmid,F.X. (2006). SlyD proteins from different species exhibit high prolyl isomerase and chaperone activities. *Biochemistry* 45, 20-33.

Shamoo,Y., Friedman,A.M., Parsons,M.R., Konigsberg,W.H., and Steitz,T.A. (1995). Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature* 376, 362-366.

Shatilla,A., Ishchenko,A.A., Saparbaev,M., and Ramotar,D. (2005). Characterization of *Caenorhabditis elegans* exonuclease-3 and evidence that a Mg²⁺-dependent variant exhibits a distinct mode of action on damaged DNA. *Biochemistry* 44, 12835-12848.

She,Q., Singh,R.K., Confalonieri,F., Zivanovic,Y., Allard,G., Awayez,M.J., Chan-Weiher,C.C., Clausen,I.G., Curtis,B.A., De,M.A., Erauso,G., Fletcher,C., Gordon,P.M., Heikamp-de,J., I, Jeffries,A.C., Kozera,C.J., Medina,N., Peng,X., Thi-Ngoc,H.P., Redder,P., Schenk,M.E., Theriault,C., Tolstrup,N., Charlebois,R.L., Doolittle,W.F., Duguet,M., Gaasterland,T., Garrett,R.A., Ragan,M.A., Sensen,C.W., and Van der,O.J. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U. S. A* 98, 7835-7840.

Shinkai,A., Kira,S., Nakagawa,N., Kashihara,A., Kuramitsu,S., and Yokoyama,S. (2007). Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.* 189, 3891-3901.

Shinohara,A., Shinohara,M., Ohta,T., Matsuda,S., and Ogawa,T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* 3, 145-156.

Song,J.J., Smith,S.K., Hannon,G.J., and Joshua-Tor,L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434-1437.

Soppa,J. (2001). Basal and regulated transcription in archaea. *Adv. Appl. Microbiol.* 50, 171-217.

Soppa,J. (1999). Transcription initiation in Archaea: facts, factors and future aspects. *Mol. Microbiol.* 31, 1295-1305.

- Sullivan,S., Sink,D.W., Trout,K.L., Makalowska,I., Taylor,P.M., Baxevanis,A.D., and Landsman,D. (2002). The Histone Database. *Nucleic Acids Res.* *30*, 341-342.
- Sutera,V.A., Jr., Han,E.S., Rajman,L.A., and Lovett,S.T. (1999). Mutational analysis of the RecJ exonuclease of *Escherichia coli*: identification of phosphoesterase motifs. *J. Bacteriol.* *181*, 6098-6102.
- Tang,T.H., Bachellerie,J.P., Rozhdestvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J., and Huttenhofer,A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U. S. A* *99*, 7536-7541.
- Tang,T.H., Polacek,N., Zywicki,M., Huber,H., Brugger,K., Garrett,R., Bachellerie,J.P., and Huttenhofer,A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* *55*, 469-481.
- Tolstrup,N., Sensen,C.W., Garrett,R.A., and Clausen,I.G. (2000). Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles.* *4*, 175-179.
- Torarinsson,E., Klenk,H.P., and Garrett,R.A. (2005). Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* *7*, 47-54.
- Trivedi,V.D., Raman,B., Rao,C.M., and Ramakrishna,T. (1997). Co-refolding denatured-reduced hen egg white lysozyme with acidic and basic proteins. *FEBS Lett.* *418*, 363-366.
- Tsang,S.S., Chow,S.A., and Radding,C.M. (1985). Networks of DNA and RecA protein are intermediates in homologous pairing. *Biochemistry* *24*, 3226-3232.
- Tsumoto,K., Ejima,D., Kita,Y., and Arakawa,T. (2005). Review: Why is arginine effective in suppressing aggregation? *Protein Pept. Lett.* *12*, 613-619.
- Tsumoto,K., Ejima,D., Kumagai,I., and Arakawa,T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* *28*, 1-8.
- Tuteja,N. and Tuteja,R. (2004). Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur. J. Biochem.* *271*, 1849-1863.
- Tyson,G.W. and Banfield,J.F. (2007). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.*
- van Embden,J.D., van,G.T., Kremer,K., Jansen,R., van Der Zeijst,B.A., and Schouls,L.M. (2000). Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J. Bacteriol.* *182*, 2393-2401.
- van,d., V, Driessen,A.J., and Konings,W.N. (1998). The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles.* *2*, 163-170.

References

- Venerando,B., Fiorilli,A., Croci,G., Tringali,C., Goi,G., Mazzanti,L., Curatola,G., Segalini,G., Massaccesi,L., Lombardo,A., and Tettamanti,G. (2003). Acidic and neutral sialidase in the erythrocytes of patients with type 2 diabetes: an answer to comments by Richard et al. *Blood* *101*, 2071.
- Vincentelli,R., Canaan,S., Campanacci,V., Valencia,C., Maurin,D., Frassinetti,F., Scappucini-Calvo,L., Bourne,Y., Cambillau,C., and Bignon,C. (2004). High-throughput automated refolding screening of inclusion bodies. *Protein Sci.* *13*, 2782-2792.
- Viswanathan,P., Murphy,K., Julien,B., Garza,A.G., and Kroos,L. (2007). Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. *J. Bacteriol.* *189*, 3738-3750.
- Wabiko,H., Ogawa,T., and Ogawa,H. (1983). Defective complex formation between single-stranded DNA and mutant recA430 or recA1 protein of Escherichia coli. *Eur. J. Biochem.* *137*, 263-267.
- Wadsworth,R.I. and White,M.F. (2001). Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. *Nucleic Acids Res.* *29*, 914-920.
- Walther,A.P., Gomes,X.V., Lao,Y., Lee,C.G., and Wold,M.S. (1999). Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. *Biochemistry* *38*, 3963-3973.
- Ward,D.E., Kengen,S.W., van Der,O.J., and de Vos,W.M. (2000). Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon pyrococcus furiosus and its role in alanine production. *J. Bacteriol.* *182*, 2559-2566.
- 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 (SSB) from E. coli and human mitochondria. *FEBS Lett.* *411*, 313-316.
- Wei Yang and Gregory D Van Duyne (2004). Protein-nucleic acid interactions. *Current Opinion in Structural Biology* *2004*, *14*: 1-3.
- Weinstock,G.M., McEntee,K., and Lehman,I.R. (1979b). ATP-dependent renaturation of DNA catalyzed by the recA protein of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A* *76*, 126-130.
- Weinstock,G.M., McEntee,K., and Lehman,I.R. (1979a). ATP-dependent renaturation of DNA catalyzed by the recA protein of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A* *76*, 126-130.
- White,D. (1995). *The Physiology and Biochemistry of Prokaryotes*. Oxford University Press).

- Williams, K.R. and Konigsberg, W. (1978). Structural changes in the T4 gene 32 protein induced by DNA polynucleotides. *J. Biol. Chem.* *253*, 2463-2470.
- Woese, C.R. and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A* *74*, 5088-5090.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A* *87*, 4576-4579.
- Wold, M.S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* *66*, 61-92.
- Wulfing, C. and Rappuoli, R. (1997). Efficient production of heat-labile enterotoxin mutant proteins by overexpression of *dsbA* in a *degP*-deficient *Escherichia coli* strain. *Arch. Microbiol.* *167*, 280-283.
- Yakunin, A.F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C.H., and Edwards, A.M. (2004). The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2',3'-cyclic phosphodiesterase, 2'-nucleotidase, and phosphatase activities. *J. Biol. Chem.* *279*, 36819-36827.
- Yang, C., Curth, U., Urbanke, C., and Kang, C. (1997). Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 Å resolution. *Nat. Struct. Biol.* *4*, 153-157.
- Yu, M., Souaya, J., and Julin, D.A. (1998). The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A* *95*, 981-986.
- Zor, T. and Selinger, Z. (1996). Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Anal. Biochem.* *236*, 302-308.

8. Acknowledgement

First of all, I would like to express my deepest appreciation to my graduate supervisor, Prof. Dr. Gerhard Krauss, who not only offered me the opportunity to work in his group, but also continuously supports me with his insightful direction, patient teaching and critical discussion. From his instructions, I have acquired more than the academic knowledge and that will benefit my life in the future. My deep appreciation is also to his wife, Mrs. Silvia Krauss, for her delicious food, wonderful hospitality and the friendly invitation to the university's international club.

I gratefully thank

Dr. Georg Lipps, whose helpful discussion and precise analyses always bring me the development of the ideas.

Dr. Yiwei Huang, whose lots of smart skills and kindly advice both in lab and in life provide me invaluable support and help.

Dr. Silvia Berkner, who prepared Sso P2 genomic DNA for my present work and shared her experiences in archaea research.

Dr. Irina Dieser, whose friendly care always spreads from tiny things.

Ms. Kirsten Beck, for her technical tips and delicious cakes.

Mr. Martin Sanchez, whose enthusiastic, friendly help let me strongly recognize his warm heart behind a sunny face.

Ms. Linda Wenzel, who provided me support on protein expression as well as the pleasure on the parties she have organized.

Ms. Kristina Kufner, for her valuable discussion about my thesis.

I want to extend my thanks to:

Prof. Mathias Sprinzl and his group, for providing plasmids, TKF resin and many other helpful hands.

Prof. Dr. Franz-Xaver Schmid, Dr. Barbara Eckert, Mr. Roman Jakob for their help in protein expression and refolding.

Dr. Holger Dobbek and his group, for their kindly presents of co-expression vectors and *E.coli* strains.

Dr. Marina Lysetska, for her help in AFM measurements.

Dr. Sheng Li and Dr. Yiran Wang, for their encouragement and help.

I lack the space here to acknowledge all of the other individuals and groups that their numerous supports and help let me overcome the challengeable work in lab and enjoy the wonderful time in Germany.

I am grateful to my parents, who have been an inspiration throughout my life, no matter wherever I am, or whatever I have. Finally, a real debt is owed to my wife and daughter. The happiness they offer me is far beyond the words of my appreciation.

Erklärung

Hermit erkläre ich, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ferner erkläre ich, dass ich anderweitig mit oder ohne Erfolg nicht versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Bayreuth, December 08, 2007

Dong Han