Structural and biochemical studies of cold shock domain containing proteins.

A Thesis

Submitted for the Degree of

Doctor of Philosophy

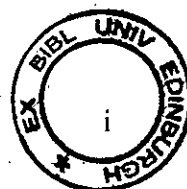## by

Hugh P Morgan, B.Sc.

Structural Biochemistry Group

The Institute of Structural and Molecular Biology

University of Edinburgh

August 2008

Abstract

The majority of transcription factors recognise target sequences in duplex form. However, single-stranded regions can be induced by torsional stress of double stranded DNA, allowing single-stranded nucleic acid binding proteins (SNABPs) access to their binding sites. SNABPs have been shown to bind with high affinity, non-specifically and specifically, to ss DNA which can regulate gene expression both positively and negatively. Gene expression can also be regulated on a translational level, by SNABPs binding to mRNA. To understand how binding of SNABPs to ss nucleic acids regulates transcription and translation, the regions of sequence specificity must be identified. Many SNABPs contain a small structural motif known as the Oligonucleotide/oligosaccharide Binding fold (OB-fold). Cold shock proteins (CSPs) are the simplest examples of OB-fold proteins, containing only a single OB-fold domain. CSPs have been shown to be directly or indirectly involved in protein transcription and translation but a clear mechanism has yet to be elucidated. This thesis describes a novel method for determining the sequence specificity of SNABPs and the structural and biochemical studies of members of the CSP family from *Salmonella typhimurium*.

Sequence specificity of single stranded DNA binding proteins: A novel DNA microarray approach.

This thesis outlines the development of a novel DNA microarray-based approach for identification of the sequence-specificity of single-stranded nucleic acid binding proteins (SNABPs). For verification, the major cold shock protein from *Bacillus subtilis* (*Bs*CspB) has been shown to bind with high-affinity to pyrimidine rich sequences, with a binding preference for the consensus sequence, 5′-GTCTTTG/T-3′. The sequence was modelled onto the known structure of *Bs*CspB and a cytosine-binding pocket was identified, which explains the strong preference for a cytosine base at position 3. This microarray method offers a rapid high-throughput approach for determining the specificity and strength of ss DNA-protein interactions. Further screening of this newly emerging family transcription factors will help provide an insight into their cellular function.

The biochemical characterisation of single stranded DNA binding cold shock proteins from Salmonella typhimurium.

In *Salmonella typhimurium*, six CSPs (*St*CspA, *St*CspB, *St*CspC, *St*CspD, *St*CspE, *St*CspH) have been identified, although their functions are yet to be clearly elucidated. With the exception of *St*CspH, these proteins have been successfully cloned, expressed and purified. The sequence specificity for each of the five CSPs were determined using the previously developed microarray assay, to reveal two different types of ss DNA binding preferences, purine or pyrimidine rich sequences. *St*CspD bound purine (guanine) rich sequences, resulting in the consensus binding sequence, 5′-ACGGGG-3′. *St*CspA, *St*CspB, *St*CspC and *St*CspE bound pyrimidine (thymine) rich sequences, with an identical consensus core binding sequence, 5′-TCTTT-3′.

The kinetics and thermodynamics of CSP/ss DNA interactions were examined for the two exemplary members, *St*CspE and *St*CspD, binding to various ss DNA sequences, using the surface plasmon resonance method and isothermal titration calorimetry. *St*CspE bound preferentially to the 5′-GTCTTTT-3′ sequence at 15°C with a 1:1 stoichiometry and high affinity (11 nM). Thermodynamic studies revealed temperature dependence of equilibrium dissociation constants (K_d), which was primarily due to a large increase in dissociation rate constant (k_d). *St*CspD bound preferentially, with a 1:1 stoichiometry and a high affinity (19 nM) to the 5′-ACGGGG-3′ sequence at 15°C, which was due to an extremely slow dissociation rate constant k_d ($2.70 \pm 0.22 \times 10^{-4}\ s^{-1}$).

Structural studies of CSPs from S. typhimurium.

The crystal structure of cold shock protein E form *Salmonella typhimurium* (*St*CspE) has been determined at 1.1 Å resolution and refined to R = 0.203. The three-dimensional structure is similar to previously determined CSPs, composed of five anti-parallel β-strands forming a classic OB-fold/five-stranded β-barrel. This first structure of a CSP from *S. typhimurium* provides a new insight into the cold shock response in *S. typhimurium* and may lead to the development of potential CSP inhibitors, vital to the food industry.

The *St*CspE/5′-GTCTTT-3′ complex proved to be difficult to crystallise and as a result the 5′-GTCTTT-3′ sequence was modelled onto the structure of *St*CspE. A cytosine specific binding pocket was identified at DNA subsite 3, explaining the preference for a cytosine base. The replacement of a thymine with a guanine at subsite 1 resulted in the gain of a hydrogen bond, explaining the preference for a guanine base at this position.

Many attempts were made to obtain crystals of *St*CspD but were unsuccessful. A computer generated model was obtained and used to generate a *St*CspD/5′-ACGGGG-3′ model, which explained the base preference seen at each of the DNA binding subsites, as observed in previous biochemical experiments.

Declaration

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and has not been submitted in whole or in part for any other degree.

Hugh P Morgan

Acknowledgements

Abbreviation list

AMPSO	3-[1,1-Dimethyl-2-hydroxyethyl)amino]-2-hydroxypropanesulfonic acid
CCP4	Collaborative Computational Project Number 4
cDNA	Complementary DNA
CSD	Cold shock domain
CSP	Cold shock protein
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
HEPES	N-(2-Hydroxyethyl)piperazine-N´-(2-ethanesulphonic acid)
IPTG	Isopropyl β-D-thiogalactopyranoside
k_a	Association rate constant
Kb	Kilobase
k_d	Dissociation rate constant
K_d	Equilibrium dissociation constant
kDa	Kilodalton
LB	Luria-Bertani broth
MALDI-Tof	Matrix Assisted Laser Desorption Ionisation Time-of-Flight
NTA	Nitrolotriacetic acid
OD_{600}	Optical density 600 nm
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
RMSD	Root mean square deviation
SA	Streptavidin
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SPR	Surface Plasmon Resonance
SSB	Single stranded DNA binding protein
ss DNA	Single-stranded deoxyribonucleic acid
SNABP	Single stranded nucleic acid binding protein
Tris	Trishydroxymethylaminomethane
T_m	Transition temperature
UV	Ultraviolet
V_m	Matthew's coefficient

Table of contents

Table of Figures

CHAPTER 1. Structure and function of cold shock domain containing proteins.

1.1 Overview: Single stranded DNA binding proteins

The famous first structure of duplex DNA was first obtained in 1953 by Watson and Crick [1]. The occurrence of base pairing, in any order, provided overwhelming evidence for sequence diversity and the role of DNA as the hereditary material of life. The importance of DNA molecules are undoubtedly reflected in the diverse set of proteins specifically evolved for care and maintenance of DNA. In a cell there are a variety of proteins with specialised functions in DNA metabolism that allow the cell to maintain or replicate its genome. In order to copy or make repairs to DNA, the double helix must be unwound to reveal two complementary strands. The need to manipulate DNA in its single stranded form has given rise to a specialised group of proteins (for reviews see [2-6]). The interactions of proteins with single stranded (ss) DNA play important roles in many cellular processes such as transcription, translation [7], DNA recombination [8], DNA repair [9], DNA replication [3], telomere maintenance [10] and cold shock response [11], as a result many types of protein folds exist to facilitate these functions (Figure 1-1).

Furthermore, DNA specifies RNA through the process of gene transcription and the mRNA molecules in turn specify all of the proteins of a cell and as a result many single-stranded RNA binding proteins also exist [12]. Various ss DNA binding proteins can also bind ss RNA molecules due to the structural similarities of both interactions [13]. As a result ss DNA binding proteins will be referred to as single-stranded nucleic acid binding proteins (SNABP).

a

OB-Fold

b

Domain II

Domain III

Active site

Domain IV

Domain I

Topoisomerase

c

Active site

E2A DNA binding
Domain

d

Zinc finger

e

Ferredoxin-like

f

KH-domain
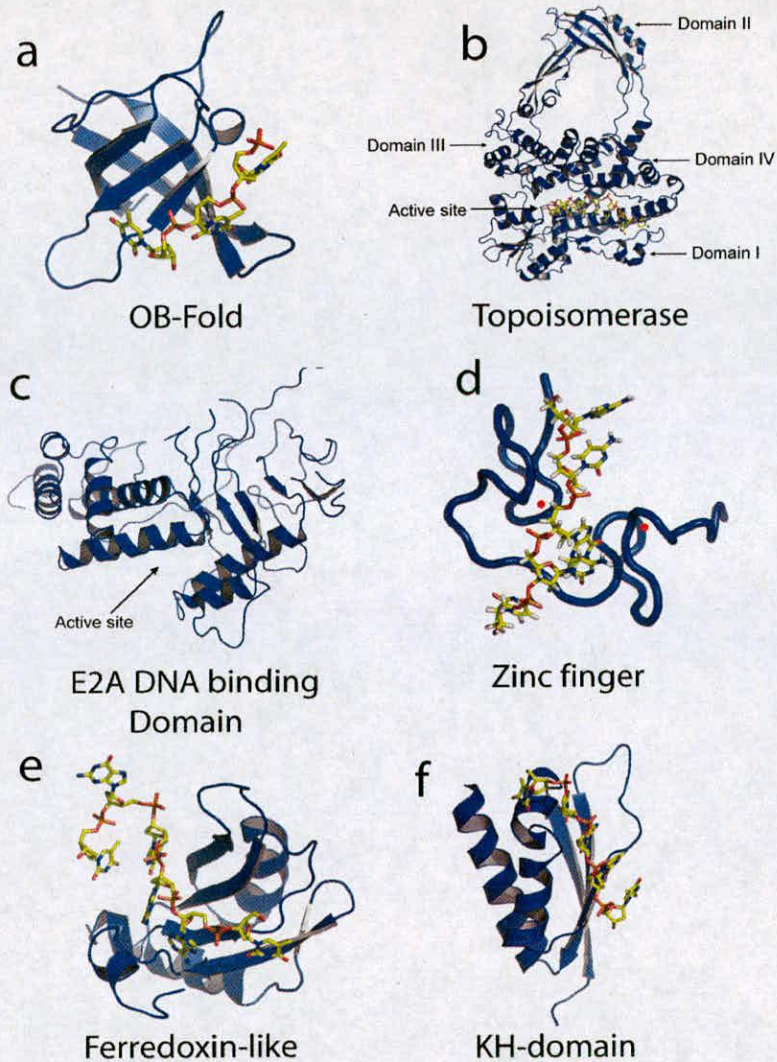
Figure 1-1. **Folds found in ss DNA binding proteins.**

(a) Structure of *Bs*CspB in complex with polythymidine [14]. This is a representation of an OB-fold, which is a five-stranded antiparallel β-sheet that is folded into a β barrel. (b) Structure of type IA DNA topoisomerase in complex with a single-stranded DNA molecule [15]. The ssDNA is bound within a groove extending from the active site at the interface of domains I, III and IV, and continuing through a region created primarily by domains I and IV. (c) Structure of the Adenovirus single-stranded DNA binding protein [16]. The active site is comprised of four helices, bundled together to form a partly opened structure to which ss DNA is thought to bind. (d) Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC) [17].This is a representation of a metal (zinc) bound fold. (e) Structure of UP1 bound to the sequence d(TTAGGG) [18]. This is a representation of a ferredoxin-like fold (comprised of ~90 amino acid residues), which is made up of a four-stranded β-sheet and two α-helices. (f) The structure of the KH3 domain of heterogeneous nuclear ribonucleoprotein K (hnRNP K) in complex with ssDNA (10-mer) [19]. This is a representation of a eukaryotic type KH-domain fold, which is made up of a single helix and three β-sheets.

1.1.1 *Properties of ss DNA and it relationship to protein binding*

The structure of ss DNA is formed by individual repeating nucleotides. Nucleotides are made up of three parts; a nucleoside base, a sugar and a phosphate (Figure 1-2e-h). ss DNA binding proteins utilise these features for binding and recognition. As shown in Figure 1-2a, the structure of ss DNA is extremely flexible due to the free rotation about the phosphodiester backbone linking adjacent nucleotides. The high degree of flexibility can result in many different structural forms for the same ss DNA ligand (Figure 1-2a-d), keeping in mind that it is energetically more favorable to bend the molecule smoothly, spreading the strain over a large distance, than to bend it sharply at isolated positions [20]. Each phosphate along the ss DNA backbone has one overall electronegative charge. Many SNABP use this to their advantage by lining the DNA binding surface with electropositive residues (for an example see Figure 1-3a [21]).

In addition to the electrostatic interactions, hydrogen bonds from amino acid side chains and/or amide and carbonyl groups of the protein have also been shown to interact in a non-sequence specific manner with the ss DNA backbone [22-24]. The electrostatic and hydrogen bonding occurring between the protein and the phosphodiester backbone of ss DNA is ever-present but constitutes only part of the energy and specificity required for ss DNA binding.
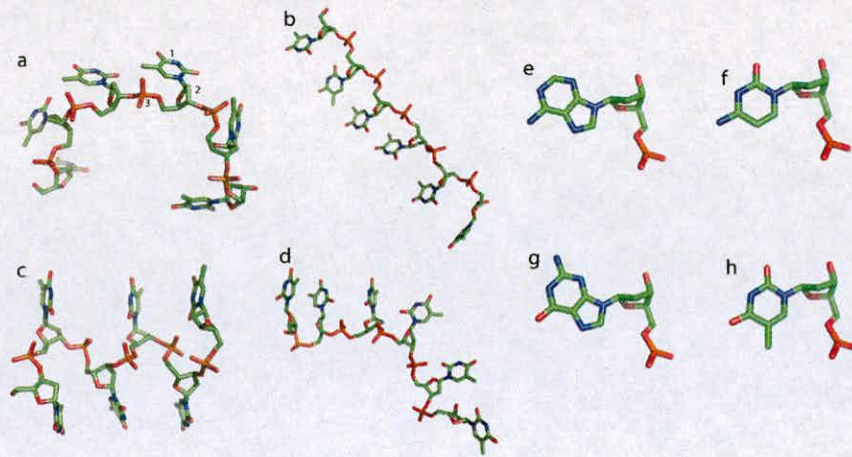
Figure 1-2. **Irregular structures of a small ss DNA ligands and nucleotides.**

(a) ss DNA taken from the *Bacillus caldolyticus* CSP structure with ss DNA bound [25]. Note the irregular structure. The nucleotide bases are in a variety of orientations and are accessible to stacking interactions with side-chain aromatics from the protein. Three parts of a nucleotide are (1) nucleoside base; (2) a sugar (3) a phosphate. **(b-c)** Examples of stacking interactions formed between bases using the same oligonucleotide. **(d)** Example of a possible binding mode; stacking interactions between three bases, while the other bases are accessible to stacking interactions with the protein. The chemical structures of the four deoxynucleoside bases **(e)** adenosine $(C_{10}H_{13}N_5O_4)$, **(f)** cytidine $(C_9H_{13}N_3O_5)$, **(g)** guanosine $(C_{10}H_{13}N_5O_5)$ and **(h)** thymidine $(C_{10}H_{14}N_2O_5)$. All structural figures were generated using Pymol [26].
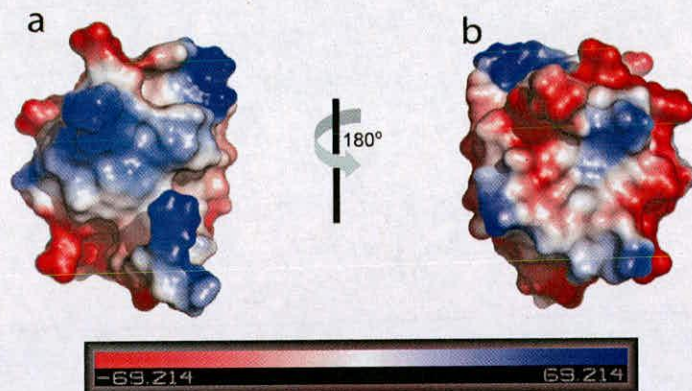
Figure 1-3. **An electrostatic surface representation of the major cold shock protein, CspB, from *Bacillus subtilis*.**

(a) The ss DNA binding face of the protein is lined with positively charged amino acids (blue) which can interact with negatively charged ss DNA. **(b)** The opposite side of the protein (180°) contains negatively charged regions on the surface (red) and are well away from the ss DNA binding interface.

4

The nucleotide base is also available for binding and recognition. The chemical structures of pyrimidine and purine bases differ significantly but they share a common aromatic ring structure at the core. These planar rings quite commonly form stacking interactions with aromatic amino acids (tyrosine, phenylalanine and tryptophan) during ss DNA and RNA binding [27]. Figure 1-4 shows an example of a stacking interaction between a tryptophan and a thymine nucleobase. The stacking interaction provides the protein with an additional binding method, which does not require specific hydrogen bond networks.
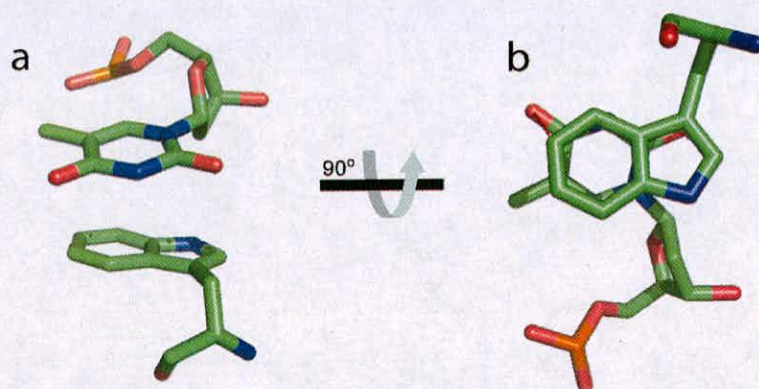
Figure 1-4. **An example of a stacking interaction.**

(a) Stick representation of a stacking interaction between a thymine base and a tryptophan amino acid and the same image rotated 90° (b). This example was taken from the *Bacillus caldolyticus* Csp /Hexathymine complex structure [25].

While the sugar moiety of ss DNA is also present for non-specific ss DNA binding it is often inaccessible to proteins. The relatively flat nature of the sugar ring allows for stacking interactions [12], although they are rarely seen. The main purpose of the sugar moiety is that it allows SNABP to distinguish between ss DNA and ss RNA, as there is a 2′ hydroxyl group present in RNA. Enzymes such as RNA and DNA polymerase must have highly selective sites, which make use of the presence or

absence of the 2′ hydroxyl group, thus preventing the mis-incorporation of incorrect nucleotides into newly synthesised nucleic acid molecules. On the other hand, it is also simple to perceive that ss DNA binding proteins can just as easily bind ss RNA in situations where the 2′ hydroxyl group does not sterically hinder binding (refer to [28, 29] for comparative examples).

1.1.2 ss DNA specificity and recognition

The interactions formed during non-specific ss DNA/protein binding (hydrogen bonding, electrostatic and stacking interactions) also form the foundation for specificity. As nucleic acids (dsDNA, dsRNA, ssRNA) share similar chemical properties they all have the potential to compete and/or bind to the protein [3]. ss DNA binding proteins have evolved key features on their surface to allow for preferentially binding of ss DNA over ds nucleic acids. Double stranded nucleic acids are not as flexible as ss DNA (it takes 50 times as much energy to bend a ds 100-mer DNA molecule into a circle than to perform the same operation with a ss DNA molecule [20]). Base pairing forms a rigid helix, which fixes the distances between phosphates, depending on the dsDNA form [30]. In contrast, ss DNA molecules have no fixed structure and nucleotide spacing can vary dramatically (Figure 1-2), which is highlighted by the mechanics of ss DNA elasticity [20, 31]. As a result ss DNA binding proteins frequently have a narrow binding cleft on the surface [32], which sterically inhibits a wide duplex molecule from binding.

The ability of ss DNA binding proteins to distinguish ss DNA from ss RNA is a major problem, as structurally they are more or less identical and quite commonly exhibit little or no difference in preference or affinity for ss DNA over ss RNA [16, 33]. For this reason in vivo concentrations of ss DNA binding proteins are tightly

6

regulated to prevent interference with RNA [34]. The 2´ hydroxyl group on the RNA can also limit sugar/aromatic side chain stacking interactions, as the 2´ hydroxyl group must be positioned away from nearby residues.

There are many examples of proteins that bind ss DNA in a sequence specific manner (Table 1-2). The basic interactions are the same but are supplemented with base-specific hydrogen bond networks (Figure 1-5) and/or small tightly fitting nucleotide binding pockets on the surface.
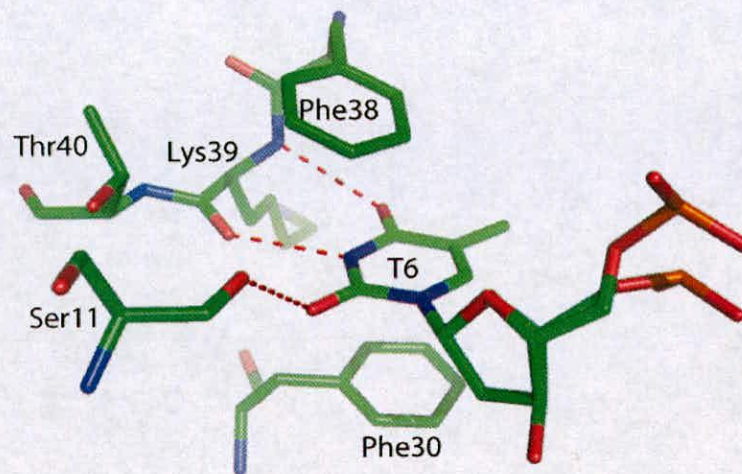
Figure 1-5. **The specific interactions made between the major cold shock protein, CspB, from *Bacillus subtilis* and a nucleotide** [14].

The hydrogen bonds (shown as dashed red lines) made between the nucleobase and the protein resemble that of Watson-Crick base pairs. Note: The degree of specificity is also enhanced by a stacking interaction between Phe30 and Phe 38.

1.1.3 *The OB-fold*

Many ss DNA-binding proteins contain a small structural motif known as the Oligonucleotide/oligosaccharide binding fold (OB-fold, for review [35]). At a general level, the OB-fold is a five-stranded antiparallel β-sheet that is folded into a β barrel (Figure 1-6). The OB-fold presents a narrow ss DNA binding cleft, which is

studded with aromatic and positively charged amino acid residues, these facilitate many of the key ss DNA/protein interactions described previously. It is the highly twisted nature of the β-sheets within the OB-fold that produces many surface exposed residues, which are freely available for ss nucleic acid binding. OB-folds are found in nine superfamilies within the SCOP (structural Classification of Proteins) database [36]. Only the nucleic acid-binding protein superfamily, which is by far the largest of the nine groups, will be described here.
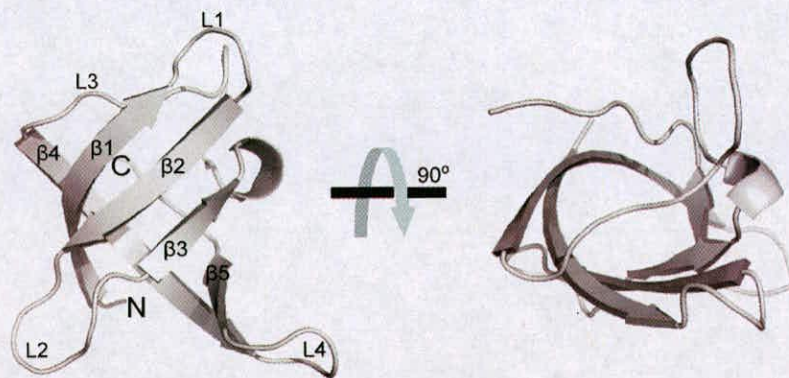
Figure 1-6. **An example of an OB-fold containing SNABP.**

The three dimensional structure of the major cold shock protein, CspB, from *Bacillus subtilis* [21]. β-strands are given as curved arrows and numbered β1- β5. Loops between β-strands are numbered L1-L4.

OB-Fold containing proteins can be classified into two major groups on the basis of ss DNA binding:

1. Proteins that bind nucleic acids without apparent or strong sequence specificity (Table 1-1 provides examples), including the *Escherichia coli* single stranded DNA-binding protein (*Eco*SSB) [23] and the gene V protein from the Filamentous bacteriophage f1 (gVp) [37].

2. Proteins that bind specifically to single-stranded regions of nucleic acids (Table 2-
1 provides examples), including the telomere end binding proteins (TEBPs),
Oxytricha nova TEBP[32], *Saccharomyces cerevisiae* Cdc13 [38] and
Schizosaccharomyces pombe protection of telomeres 1 protein (Pot1p) [39].

Table 1-1. **OB-Fold proteins which bind non-specifically to ss DNA.**

PDB ID	Name	Resol. (Å)	Method	Ref.
1EQQ/1QVC	*Eco*SSB	3.20/2.20	X-RAY	[23]
1GPC	T4 gp32	2.20	X-RAY	[40]
1JE5	g2.5	1.90	X-RAY	[41]
1L1O	*hs*RPA	2.80	X-RAY	[42]
1O7I	*Sso*SSB	1.20	X-RAY	[43]
1S30	*Hm*SSB	2.47	X-RAY	[44]
1SE8	DrSSB	1.80	X-RAY	[45]
1UE1/1UE5	MtSSB	2.50	X-RAY	[46]
1AE3	gVp	2.0	X-RAY	[37]
2CCZ	EcPP	2.70	X-RAY	[47]
1PFS	Pf3 ssDBP		NMR	[48]

Abbreviations; *EcoSSB = Escherichia coli* ssDNA-binding protein; *hsRPA = Homo sapiens*
Replication protein A; T4 gp32 = Bacteriophage T4 gene 32 protein; *SsoSSB = Sulfolobus
solfataricus* single stranded DNA binding protein; HmSSB = Human mitochondrial single stranded
DNA binding protein; *DrSSB = Deinococcus radiodurans* single-stranded DNA-binding protein;
MtSSB = Mycobacterium tuberculosis single stranded DNA binding protein; gVp = Gene V protein
from the Filamentous bacteriophage f1; EcPP = primosome protein; Pf3 ssDBP *Pseudomonas*
bacteriophage pf3 single stranded DNA binding protein.

OB-fold containing proteins bind stretches of ss DNA and ss RNA with a binding
density of 6-7 nucleotides/monomer [49]. Due to the modular nature of the OB-fold
it is often present in more than one copy to allow a protein to bind longer ss nucleic
acids molecules [42] (e.g. Figure 1-7c). Although the OB-fold is a highly conserved
and recognised motif in single stranded nucleic acid binding proteins, there has never
been an in depth study performed to fully examine binding affinity and specificity for
ssDNA.

Table 1-2. **OB-Fold proteins which bind specifically to ss DNA.**

Protein	Recognition Sequence	Structure	Size (a.a)	Res (Å)	Ref
B.S_CspB (2ES2)	5´-d(TTTTTT)-3´	(a)	67	1.78	[14]
B.C_Csp (2HAX)	5´-d(TTTTTT)-3´	(b)	66	1.29	[25]
hnRNP A1 (2UP1)	5´-d(TTAGGG)n-3´ & 5´-d(TTAGGG)-3´	(c)	182	2.1	[51]
OnTEBP (1PA6)	5´-d(G4T4)n-3´	(d)	A=495 B=222	2.45	[32]
HmPot1 (1XJV)	5´-d(TTAGGGTTAG)-3´	(e)	174	1.73	[39]
ScCdc13 (1S40)	5´-d(GTGTGTGTGTGTG)-3´	(f)	191	NMR	[38]
HmYB-1 (1H95)	5´-d(ATTGG)-3´	(g)	79	NMR	[53]

Abbreviations; *Bs*CspB = cold shock protein B from *Bacillus subtilis*; *Bc*Csp = cold shock protein from *Bacillus caldolyticus*; *hn*RNP = heterogeneous nuclear ribonucleoprotein; *On*TEBP = *Oxytricha nova* telomere end binding protein; HmPOT1 = *Homo sapiens* protection of telomeres 1 protein; ScCdc13 = a telomere end binding protein from *Saccharomyces cerevisiae*; *Hm*YB-1 = *Homo sapiens* Y-Box binding protein. The PDB code for each structure is shown in the protein column, below the protein name. Figure 1-7 shows an enlarged diagram for each of the structures (a-g).
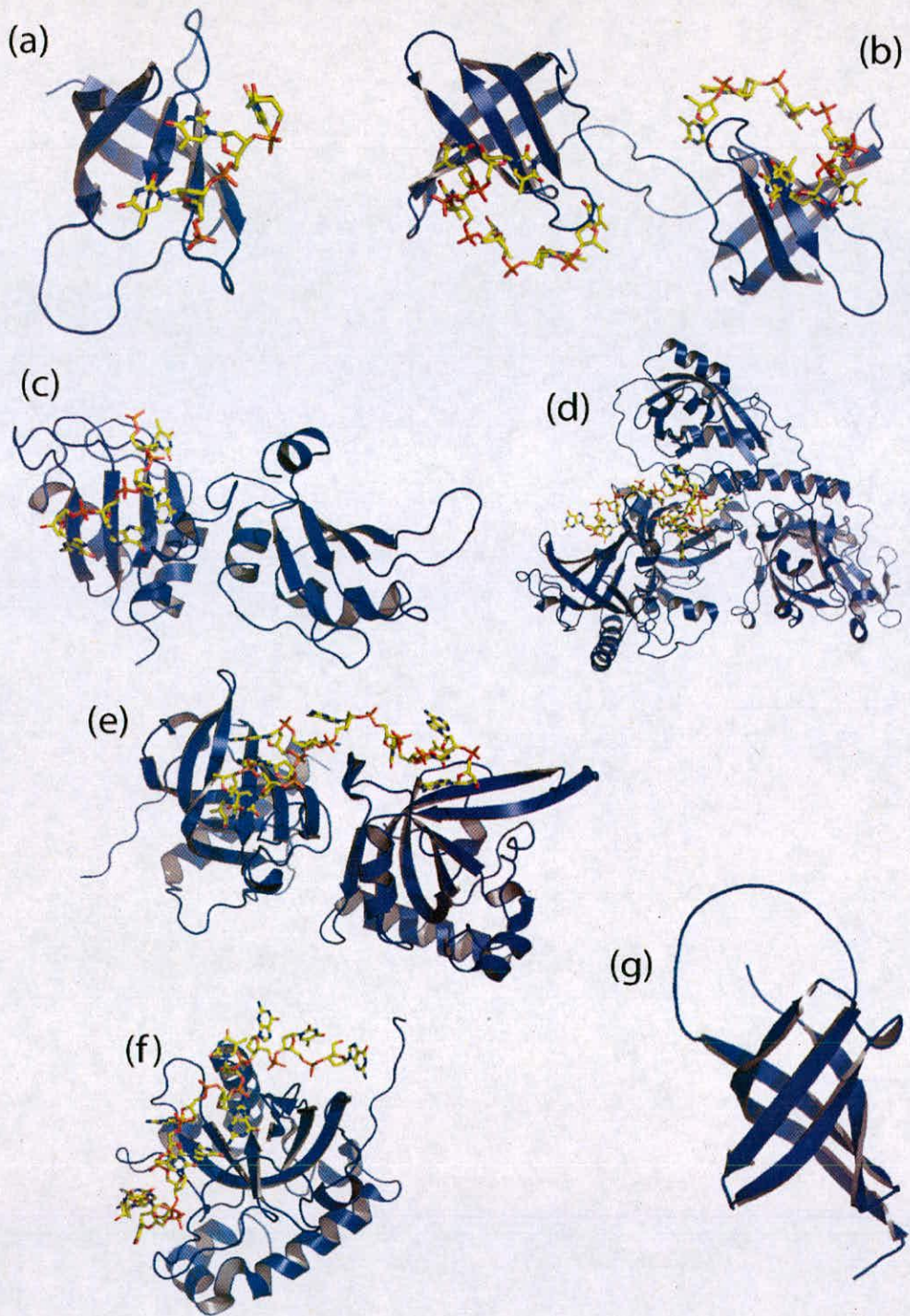
Figure 1-7. **The structures of OB-Fold proteins which bind ss DNA in a sequence specific manner.**

All the structures are described in Table 1-2. Protein structures are coloured blue and ss DNA has been coloured yellow.

1.2 A classic family of OB-fold containing proteins: Cold shock proteins

1.2.1 Overview

Cold shock proteins (CSPs) comprise a family of small proteins that are structurally highly conserved, containing a single OB-fold domain (Figure 1-6), enabling CSPs to bind single stranded nucleic acids. CSPs are induced in response to a rapid downshift in temperature but are also expressed under normal conditions, although a clear biological function has yet to be elucidated. The characteristic structural unit (OB-fold domain) is quite commonly referred to as the cold shock domain (CSD) and also occurs in other protein structures which can be found in a number of diverse organisms, from bacteria to vertebrates. Significant examples of this are the Y-Box proteins, which are structural mRNP (mRNA ribonucleoprotein particle) proteins associated with almost all mRNAs [50, 51]. Once bound to mRNA, these proteins facilitate secondary structural changes which contribute to mRNA transport, translation, processing and stabilisation [52-54].

1.2.2 Cold Shock response in bacteria

All living organisms have the remarkable ability to react and adapt to changes in their surrounding environment. Key cellular responses allow organisms to endure normally fatal conditions such as acidic shock, heat shock, cold shock, pressure and osmotic stress. Amongst these various environmental stress factors, changes in temperature are probably the most regular stress for all organisms. On a cellular level, adjustment to rapid temperature downshift (37 to 10°C) occurs by alterations in membrane composition and function, a decrease in cellular division and changes

in gene expression [55]. The cellular response in prokaryotic organisms to cold shock is defined by the production and function of a small set of cold induced proteins (CIPs) [56, 57]. Most of these proteins are directly or indirectly involved in transcription and translation (summarised in a review [11]). The most common of these are the small highly conserved family of CSPs [58]. Figure 1-8 summarises the general cellular response observed during cold shock.
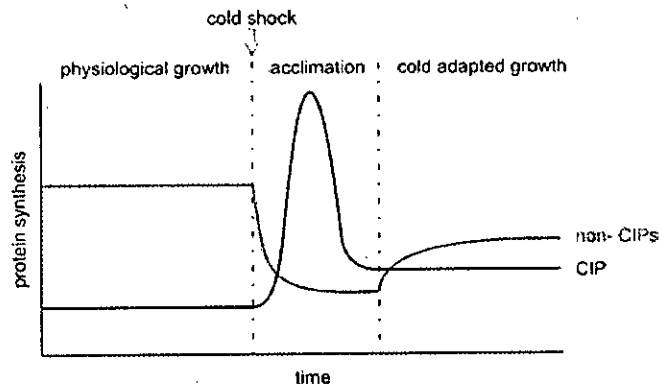
Figure 1-8. **An overview of protein expression after cold shock.**
The acclimation phase directly after cold shock is characterised by the inhibition of non-CIPs and a boost in CIPs, including CSPs. After that, the CIP expression declines and bulk protein synthesis is restored to a cold adapted level.

1.2.3 Bacterial CSPs

The CSP family are a highly conserved family of small acidic proteins with a molecular mass of approximately 7.4 kDa [59]. Two of the best-studied groups of bacterial CSPs are from *B. subtilis* and *E. coli*. To date nine CSPs have been identified in *E. coli*, *Ec*CspA-I. *Ec*CspC and *Ec*CspE are constitutively expressed at physiological temperatures [60] but increase tenfold during the cold shock response [61]. *Ec*CspD is induced under nutrient stress [62] and *Ec*CspA, *Ec*CspB, *Ec*CspG and *Ec*CspI are highly cold inducible [63-65].

*Ec*CspA is most often described as the major cold shock protein, given that cold shock of *E. coli* cells results in high cellular concentrations of *Ec*CspA, at least 100 μM [61] (13% of total cellular protein content). *Ec*CspA is also expressed at high levels (50 μM) during early exponential phase of normal physiological temperatures. The constitutive expression of *Ec*CspA, and most other CSPs [66], under non-cold shock conditions is a good indication that CSPs have additional cellular functions that have not been characterised. *Ec*CspA has been show to regulate its own expression by binding to the 5′ untranslated region (UTR, 159 bp) of its own mRNA and can also assist the translation of other CSPs [67, 68].

Other CSPs in *E. coli* are constitutively expressed at physiological temperatures and many share high sequence homology to *Ec*CspA, suggesting they could complement one another functionally, restoring cold adaptive phenotype if *Ec*CspA or other CSPs were mutated or deleted. Indeed this has been discovered; transiently induced *Ec*CspE restored the cold adaptive phenotype in a triple deletion mutant (ΔCspA ΔCspB ΔCspG) [69].

Clearly *Ec*CspA influences transcriptional processes, *Ec*CspE must also be involved in transcriptional processes as it binds the polyadenine tail at the 3′ end of mRNAs, interfering with the degradation of polynucleotide phosphorylase and RNase E at 37°C [70]. *Ec*CspE involvement in anti-termination has been shown, where purified *Ec*CspE inhibited Q-mediated transcriptional anti-termination [71, 72].

Three CSP have been discovered in *B. subtilis* [73], *Bs*CspB, *Bs*CspC and *Bs*CspD. *Bs*CspB and *Bs*CspC are cold inducible and the existence of at least one of them is essential for cell survival [74]. Like *Ec*CspA, *Bs*CspC and *Bs*CspB are also induced at the stationary growth phase [21, 75]. The major cold shock protein from *B.*

subtilis, *Bs*CspB, has similar functions to *Ec*CspA, playing the most important role during cold shock. A deletion of the *Bs*CspB gene (ΔCspB) results in a cold sensitive phenotype [76]. Only the expression of either *Bs*CspD or *Bs*CspC was able to restore the cold adaptive phenotype in the ΔCspB [74, 77], highlighting a similarity in function for all CSPs.

CSPs are found not only in the mesophilic bacteria *Esherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium* (see chapter 3 for a detailed analysis) but in most other eubacteria (Figure 1-10a), e.g. in the thermophilic *Bacillus caldolyticus* [78], in the psychrophilic *Listeria monocytogenes* [79] in hyperthermophilic *Thermotoga maritima* [80] and recently in Archaea [81].

To summarise, CSPs are induced during cold shock and their joint function is necessary for cell viability [74] but not all CSPs can complement the deletion of one or more CSP genes to restore the cold phenotype [74, 77]. One can assume that although these proteins are highly similar they may have different functions or in the case of *E. coli* they may even function together and/or have binding partners, as yet undiscovered. The constant low level expression under non-cold shock conditions also suggests the possibility of alternate functions during non-cold shock conditions. There is one function of the CSP family that is clear: their ability to bind ss nucleic acids. The different functions of CSPs may be reflected in their ability to bind specifically to different ss nucleic acid sequences. CSPs from the same family may bind different sequences, accounting for a difference in function, which has been observed during complementation experiments (and recently in *Salmonella typhimurium* – Unpublished – section 6.2.3). Further study of ss DNA specificity in CSPs may reveal potential differences in function.

1.2.4 CSPs and transcription

The majority of transcription factors recognise target sequences in duplex form. However single stranded regions can be induced by torsional stress of double stranded DNA, allowing single-stranded nucleic acid binding proteins (SNABPs) access to their binding sites [82]. In prokaryotes and eukaryotes, the non-specific and site specific binding of ss DNA binding proteins have been shown to regulate transcription both positively and negatively. There is also evidence that ss DNA binding proteins are also involved in a complex network of protein-protein interactions (for a full review see [82]). The first strong piece of evidence that CSPs were involved in transcriptional regulation was as a result of their high homology to other CSD containing proteins, namely the eukaryotic Y-Box proteins (Figure 1-10a). Y-Box proteins, which were initially characterised by their ability to bind the Y-Box motif (ATTGG) and the reverse complement (CCAAT) [83]. The Y-Box protein, YB-1, is a negative regulator of major histocompatibility complex (MHC) class II genes [83] and represses interferon-γ-induced transcription of MHC genes [84]. Y-Box proteins, MY1 and MY1a, have been shown to negatively regulate the gene encoding the nicotinic acetylcholine receptor by binding specifically to one of the DNA strands of their target site and suppress promoter activity [85]. CSPs are also characterised as having an identical binding preference for the Y-Box sequence [86] and therefore may also play key roles in transcriptional regulation.

Another key piece of evidence that CSPs were able to regulate the cold shock response on a transcriptional level is the function of the CSD containing protein, H-NS (a nucleoid-associated protein), from enterobacteria. The authors demonstrated that the cold-induced protein, H-NS, binds preferentially to ds DNA and is associated

with cold shock [87, 88]. The H-NS ratio increased fourfold during cold shock compared to physiological conditions [89], its increased expression was a direct result of CspA binding to the *H-NS* gene promoter region, thus acting as a transcriptional activator. Similar results were also observed in *E. coli*, where increased expression of the *H-NS* gene was observed during the heterologous expression of CspB from *B. subtilis* [74], demonstrating functional homology of CSPs from different bacteria.

Clearly CSPs can bind to nucleic acids and affect transcription but have also been characterised as transcriptional anti-terminators or as often called, RNA-chaperones [72]. CSPs bind single stranded nucleic acids (ss RNA and ss DNA) and prevent the formation of secondary structures, which have been shown to inhibit transcriptional processes (for an example see; [90]). Recent studies have shown that *Bs*CspB influences transcription (and translation) *in vitro* and has both negative and positive affects on the expression of many genes [91], again providing strong evidence for CSP involvement in transcriptional regulation.

1.2.5 CSPs and translation

Over the years bacterial CSPs have been shown to be involved in translational processes in a variety of different ways. They have been shown to bind to their own mRNA, preventing the formation of unfavourable secondary structures [49] (resulting from base pairing), which can mask the Shine-Dalgarno (S-D) sequence (sequence near the ribosome binding site), and thus facilitates subsequent translation initiation. *Bs*CspB has also been shown to enhance its own translation, as a downstream sequence present on its mRNA can interact with rRNA [92-94] and stabilise the binding of the ribosome to the S-D sequence.

In addition to these methods of self-enhancement of their own expression, CSPs must have other translationally related functions associated with cold shock. During cold shock major changes occur at the cellular level, reduction of bulk protein synthesis, changes in membrane fluidity and increased nucleic acid base pairing [95]. The increase in base pairing results in the formation of secondary structures (Figure 1-9a), which blocks translation initiation and elongation, thus reducing cell viability [57, 96]. CSPs have been shown to actively prevent the formation of these problematic secondary structures by binding mRNA and keeping it unfolded (Figure 1-9b), thus allowing transcriptional and translational events to proceed. Cold induced secondary structure is unlikely to be the only cause of decreased bulk protein expression during the acclimation phase (Figure 1-8). A gradual decrease in temperature does not lead to a selective block of gene expression, as one would expect with the diverse variation of 5′UTRs [97]. Therefore, the decrease in bulk protein expression observed during acclimation phase is most likely a net result of several factors; direct temperature affects on cellular kinetics and resultant protein expression [95], miscoding during elongation and increased secondary structure formation of mRNA [98].
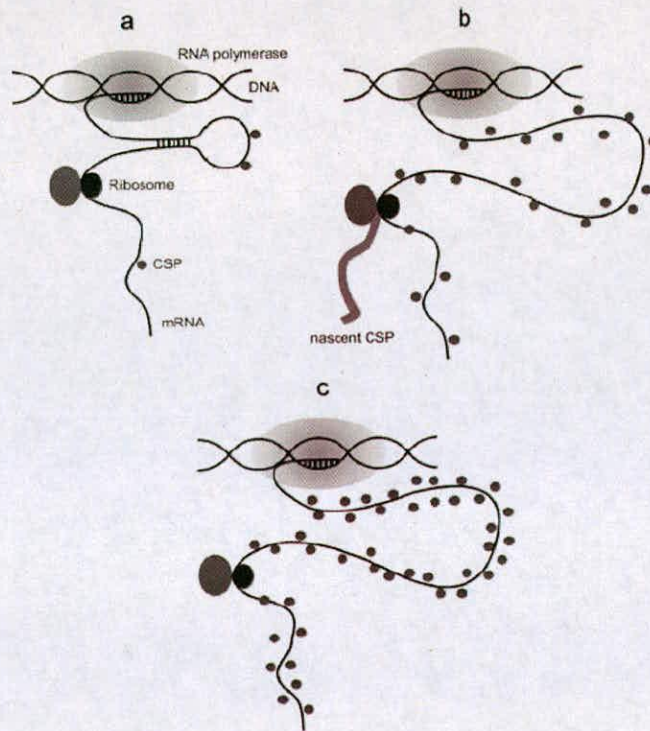
Figure 1-9. **Schematic representation of the function of CSPs during acclimation phase.**

(a) Immediately after cold shock, the translation of bulk proteins is generally reduced as a result of increased base pairing. **(b)** Intermediate stage, CSP levels increase as direct result of their ability to bind their own mRNA and enhance translation. Bulk protein synthesis is suppressed by the unspecific interaction of CSPs with mRNA. **(c)** Late acclimation, high levels of CSP suppresses translation of their own mRNA (and potentially enhances CSP transcription). CSP levels decrease within the cell and bulk protein expression resumes, although at a cold adapted level. Figure adapted from [95].

While increased expression of CSPs have regulatory effects on protein transcription and translation they also help the cell to establish protein expression at a cold adapted level [91]. As cellular concentrations of CSPs increase they have the ability to halt translation by binding non-specifically and completely covering the essential parts (for translation initiation) of the mRNA (Figure 1-9c) [91], providing an adaptation time frame or translation stop. This allows the cell to climatise itself and avoids miscoding, which results in high levels of incorrectly translated proteins [99].

During this time frame the altered properties of ribosomes [100] and the action of special proteins [101, 102] allows the cell to adapt to the reduced temperature.

Whether or not CSPs promote transcription of these specialised proteins or related proteins is still unknown. It should also be pointed out that there is no direct evidence of CSP involvement in biologically relevant RNA folding. It remains to be determined if CSPs are involved in the folding of transfer RNAs (tRNAs), ribosomal RNAs and ribozymes. Moreover, despite the fact that CSPs bind ss DNA/RNA sequences with such high-affinity, the direct relationship between the ability of CSPs to destabilise RNA secondary structure and cold-shock adaptation has not been demonstrated.

1.2.6   Expression of other CIPS

CSPs are not the only family to act as RNA chaperones, the DEAD box helicases have also been shown to unwind ds nucleic acids [103]. DEAD box helicase family are involved in many aspects of RNA metabolism [104-108] and some are also cold induced [109]. CsdA (cold-shock DEAD-box protein) was isolated as a cold induced helicase in *E. coli* [110] and is involved in RNA degradation by the degradosome [111, 112], a multienzyme complex which controls the level of transcripts in the cell by processing and degrading mRNA as the cell adapts to changing conditions [113]. CsdA has also been shown to interact directly with various components (PNPase, RNaseE and other enzymes) of the RNA degradosome under cold shock conditions [114]. CsdA is an essential CIP, although it is not clear which of the functions are essential cold-shock functions and until recently no proteins were shown to complement its function in vivo. *Ec*CspA and an exonuclease, RNase R, can complement the cold-shock function of CsdA [115].

Recently *Bs*CspB has been shown to interact with putative RNA helicases CshA and CshB [116]. This work and that of CsdA could suggest a more cooperative function between CSPs and helicases, where helicases unfold mRNA during cold shock and CSPs bind to prevent re-annealing. This seems more valid than CSPs actually breaking hydrogen bonds between complementary bases on a ds RNA molecule.

1.3 Structure and molecular properties of CSPs

1.3.1 Structural aspects of CSPs

To date the structures of four CSP structures have been solved, *Bacillus subtilis* CspB, *Thermotoga maritima* CSP, *Bacillus caldolyticus* CSP and *Escherichia coli* CspA [78, 117-119]. *Ec*CspA exists as monomer in both solution [120] and in the crystal structure [118]. *Tm*Csp exists as a monomer in solution [117], even at high temperatures [121]. *Bc*Csp exists as a dimer in the crystal structure [122]. Interestingly, *Bs*CspB exists as a dimer in the crystal structure [119] and as both a monomer and a dimer in solution depending on the concentration of phosphate ions [123]. CSPs, for which no structures are available, all form dimers in solution, e.g. *Ec*CspD [75], *Ec*CspE [124] and both CspA and CspE from *B. cereus* [125]. The CSPs form a characteristic CSD comprised of five anti-parallel β-strands forming β-barrel (Figure 1-10b, c, e and f), which interact with oligonucleotides.

1.3.2 Structural conservation of CSPs

The CSD is conserved not only in bacteria (as CSPs) but in eukaryotic organisms [126], more commonly known as the Y-Box proteins (Figure 1-10a). Y-Box proteins are involved in the regulation of transcription and translation (for review see [7, 127]) and are structurally similar to CSPs (Figure 1-10d). Both CSPs and Y-box proteins share the canonical ribonucleotide protein (RNP)1 and RNP2 [128, 129] motifs conserved on the DNA binding surfaces (Figure 1-10a). It is clear that all CSPs and CSD containing proteins belong to a large family of structurally related nucleic acid binding proteins, suggesting similar functions.

1.3.3 Nucleic acid binding activity of CSPs

CSPs have consistently been shown to bind nucleic acids, both ss DNA and ss RNA but not ds DNA [14, 25, 28, 29, 49, 130], as one would expect on further examination of the narrow ss DNA binding interface. The only exception is the eukaryotic CSP homologue, *Cladosporium herbarum* (Cla) h 8 protein, which binds both ss DNA and ds DNA with nanomolar affinity [131].

The protein-nucleic acid binding interaction occurs along the ribonucleotide protein motifs 1 and 2, which are present and conserved ($\geq$ 75%) in all CSPs (Figure 1-10a). The structures of CSPs show that the RNP1 and RNP2 motifs form the electropositive nucleic acid binding surface (Figure 1-10g and h), which are able to differentiate between various ss DNA sequences [14, 49] and show extremely strong binding preferences for particular sequences [14].

The interactions between various CSPs and ss nucleic acids have been experimentally verified using various methods. Commonly, electrophoretic mobility shift assays were used to analyse the formation of protein/ss nucleic acid complexes;

*Bs*CspB binds ss DNA [130], *Ec*CspA binds both ss RNA and ss DNA [29] and both *Tm*CspB and *Tm*CspL bind ss RNA [132]. The characteristic single tryptophan occurring at the ss DNA binding site of many CSPs also allowed for tryptophan fluorescence quenching experiments, which have been used to fully characterise the binding of *Bs*CspB to various oligonucleotides [14], these experiments have been complemented in a series of site-directed mutagenesis studies [133].

An *in vitro* selection approach (SELEX) revealed preferable consensus RNA binding sequences of UUUU, AGGGAGGGA and AAAUUU for *Ec*CspB, *Ec*CspC and *Ec*CspE, respectively [33] but was insufficient to find the preferred sequence for *Ec*CspA. In another study, *Ec*CspA has been show to bind T-, U- or C-based ss DNA templates with similar affinities [28]. In the case of *Bs*CspB it was shown to initially bind the Y-Box sequence ATTGG, as well as the reverse complement [86].

A more detailed analysis showed that *Bs*CspB is not limited to binding Y-Box sequences [49, 74, 130]. *Bs*CspB binds preferentially to polypyrimidine but not polypurine ss DNA templates. Binding to thymine rich templates occurs with high affinity (≥ 100 nM) and is independent of salt, whereas binding of poly-cytosine ss DNA templates is one order of magnitude weaker and strongly salt dependent [49]. *Bs*CspB binds to T-rich ss DNA templates with a density of 6 to 7 nucleotides/CspB monomer and is enthalpically driven, which was rationalised by the interaction of aromatic residues with the pyrimidine ring of the thymine nucleobase [130].
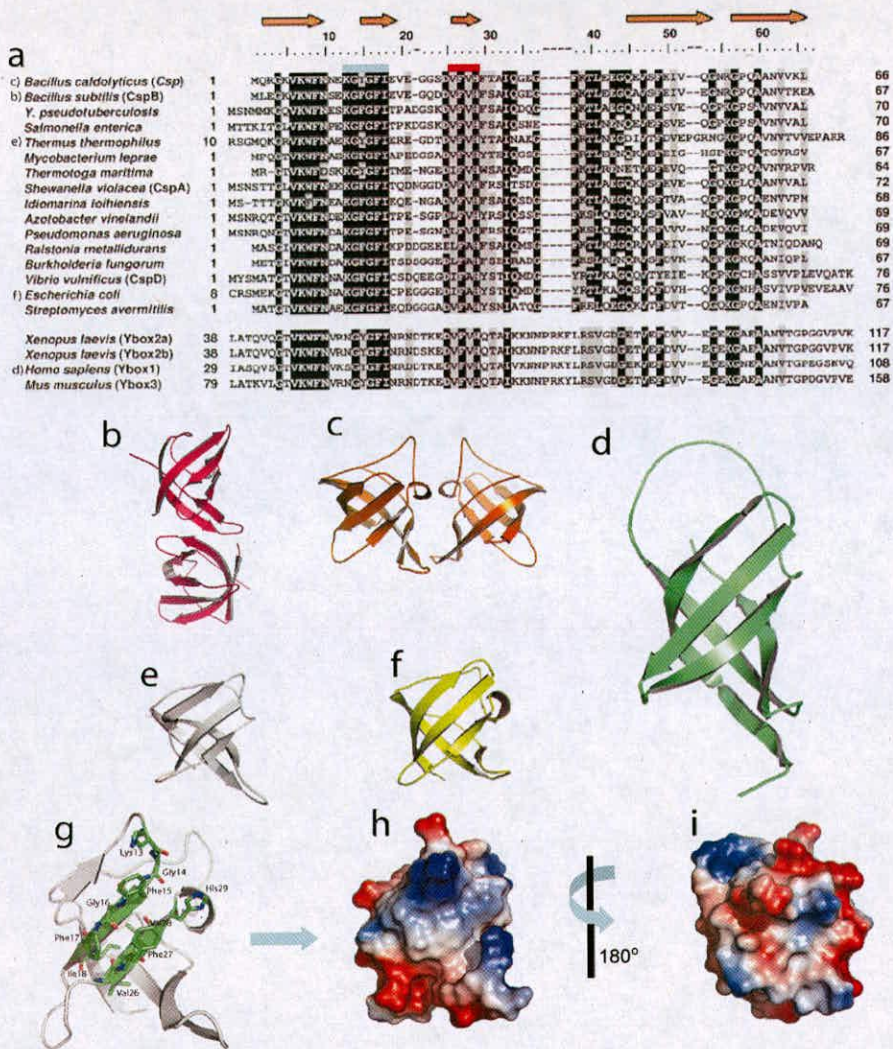
Figure 1-10. **Comparison of CSD containing protein structures.**

(a) Sequence alignment of CSPs (upper) and Y-Box proteins (lower). Residues which are conserved at a level of at least 75% identity/similarity are highlighted black or grey, respectively. The RNP1 and RNP2 motifs have been highlighted with a blue and red bar, respectively. Structures of **(b)** CspB from *B. subtilis* (1CSP), **(c)** Csp from *B. caldolyticus* (1C9O), **(d)** *H. sapiens* YB-1 (1H95), **(e)** Csp from *Thermotoga maritima* (1G6P) and **(f)** CspA from *E. coli* (1MJC). **(g)** The amino acids (shown as sticks) which comprise the highly conserved ribonucleotide protein motifs 1 ((a) – blue box) and 2 ((a) – red box) are situated along β-stands 2 and 3 on the highly positively charged (blue) ss DNA binding face **(h)** of the *B. subtilis* CspB. **(i)** The opposite side of the protein (180°), showing negatively charged regions on the surface (red).

1.3.4 The crystal structures of CSPs in complex with ss DNA

Recently two CSP/ hexathymidine (dT$_6$) complex structures have been published [14, 25], both of which show identical ss DNA-binding interfaces. In the *Bs*CspB/T6 crystal structure the dT$_6$ oligonucleotide binds to two protein molecules, which forms a long continuous chain (Figure 1-11a) but in solution the binding is assumed to occur as shown in a non-continuous fashion.

In the *Bc*Csp/T$_6$ structure the two protein chains form a domain swapped dimer with two globular functional units, made up of residues 1-35 from one protein and residues 38-65 from another protein (Figure 1-11e) [25]. Both functional units align well with monomeric crystal structures of *Bc*Csp [78], giving RMSD values of < 0.5 Å. In the *Bc*Csp/T$_6$ complex, each ss DNA molecule binds to one globular functional unit of a swapped dimer. The interactions between the protein and ss DNA molecule are common to both the *Bc*Csp/T$_6$ (Figure 1-11g) and *Bs*CspB/T$_6$ (Figure 1-11c) structures and share identical binding sites and DNA binding preferences. The authors stated that the differences in binding involving sub-sites 6 and 7 (Figure 1-11c and g), were a direct result of different crystal packing environments. Although certain interactions are only seen in one structure and not the other, the authors were able to determine the seven conserved DNA sub-sites on the surface of each CSP, which would bind the high-affinity ligand (as determined by tryptophan fluorescence experiments), TTCTTTT (Figure 1-11c and g).

The ss DNA binding results for both studies clearly show that a cytosine base at position three was key for high-affinity binding [14, 25] and the binding of ss DNA sequences was not merely dependent on the number of thymine nucleotides present but the actual positions of key bases within the sequence itself.
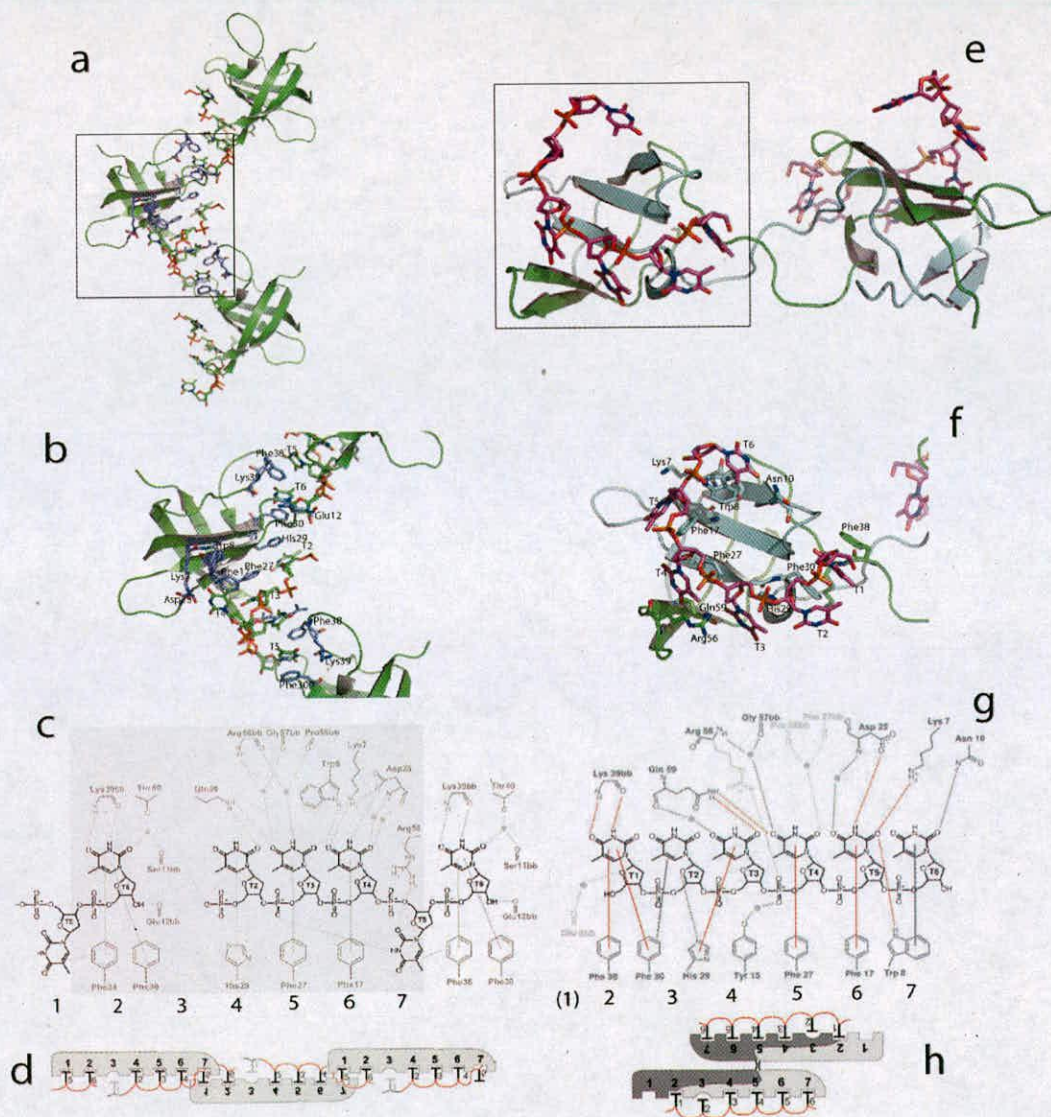
Figure 1-11. **Crystal structures of *Bs*CspB and *Bc*Csp in complex with hexathymidine.**

(a) Crystal structure of *Bs*CspB in complex with hexathymidine. (b) Enlargement of the *Bs*CspB-ss DNA contact surface (boxed area from a) showing the protein groups involved in stacking and hydrogen bonding interactions with the hexathymidine. (c) Schematic overview of intermolecular interactions: DNA (black) and protein groups (grey) interact through stacking interactions (solid lines) and hydrogen bonds (dashed lines). Some contacts are mediated by water molecules (circles). The numbers of the contact sub-sites for individual nucleobases are given at the bottom. All parts of a complete binding site from a single protein molecule are highlighted by a grey box. (d) Schematic overview of CspB/oligonucleotide assemblies. Protein molecules (grey) interact with bases (characters) from oligonucleotide molecules (strings) at distinctive binding sub-sites (cavities). In the *Bs*CspB/dT$_6$ crystal, a continuous arrangement of protein and DNA molecules is formed. A gap between the 3′ nucleotide (bound to sub-site 3) and the first structured 5′ nucleotide (bound to sub-site 4) exists, which is expected to bind the unstructured T1 nucleotide (grey). (e) Crystal structure of *Bc*Csp in complex with hexathymidine. **f** (green = chain A and blue = chain B), **g** and **h** are equivalent to b, c and d, respectively, except they represent the *Bc*Csp structure.

26

1.4   Why study cold-shock response?

A better understanding of CSPs and their function during the cold-shock response is not only interesting from a research point of view but has health and commercial implications (see Table 1-3 for examples). A better understanding of cold-shock for food-related pathogens, such as *Salmonella*, is imperative for the food industry as refrigeration is a commonly used method of food storage. Bacterial cells, which are cold-shocked before freezing, exhibit better cryotolerance, meaning that food-spoilage bacteria can survive direct freezing thus contaminating the food article.

Certain proteins are poorly produced at 37°C but could be overproduced at low temperatures using cold inducible promoters (e.g. *Ec*CspA promoter), availability of such proteins may be significant in both medicinal and structural fields of research.

Table 1-3. Commercial significance of study of cold-shock response.

Industry	Concern	Solution
Food	Spoilage of refrigerated food and infections due to food-borne pathogens	Reduce efficiency of cold- shock response of food- borne pathogens
	Instability of lactic acid bacteria starter cultures due to temperature changes during fermentations in dairy industry	Modify starter cultures to improve cold adaptation and cryotolerance
Agriculture	Reduced efficiency of 'biofertilizers' due to low temperatures	Use of cold-adapted Rhizobial cultures
	Great economic losses due to low tolerance of plants to low temperatures.	Enhance chilling resistance of plants using gene transfer from bacteria to plants
Research	Inefficient expression of proteins due to temperature sensitivity or proteolysis	Cold-inducible expression systems for production of these proteins
	Reduced efficiency of certain processes using biocatalysts at low temperatures	Use of cold-adapted enzymes

1.5 Methods used to study protein DNA/interactions

To understand how binding of SNABPs interact with ss nucleic acids, the regions of sequence specificity must be identified. Many techniques including electrophoretic mobility shift assay (EMSA) [134], nitrocellulose-binding assays [135], Southwestern blotting [136], phage display [137], UV cross-linking [138] and X-ray crystallography [139], fluorescence measurements [28], fluorescence resonance energy transfer (FRET) combined with a DNA foot-printing assay [140], surface plasmon resonance (SPR), fluorescence polarisation [141], isothermal titration calorimetry (ITC) [142], analytical ultracentrifugation [143] and systematic evolution of ligands by exponential enrichment (SELEX) [144] have all been used to effectively study specific ss DNA-protein interactions. Each method has its own strengths and weaknesses when studying protein/ss DNA interactions but the chosen method must make the most efficient use of time and resources.

1.5.1 Initial screening of protein/ss DNA interactions

To identify sequence specific ss DNA binding sites for SNABP is a somewhat daunting task. For example, if the experimental SNABP bound single stranded nucleic acids with a binding density of 7 nucleotides (nt) per monomer, then 16384 (4^7) 7-mers would be required to analyse ever possible ss DNA binding site, which would be time consuming, laborious and expensive when using the techniques described above. A high-throughput method is required which allows the user to screen for the possible binding sequence for the candidate SNABP. Chapter two of this thesis describes an innovative high-throughput assay, which provides a parallel screening system for identifying the specificity of SNABP binding. An

oligonucleotide chip was used for the identification of high-affinity 6-mer binding motifs for a SNABP. The chip contains all possible 4096 (4^6) ss hexadeoxynucleotides incorporated onto a standardised anchor, which provided a means of identifying high affinity consensus binding motifs for SNABPs.

1.5.2 *Biophysical techniques*

Once a sequence specific ss DNA binding sequence has been identified for the candidate SNABP then the molecular interactions can be studied using various biophysical techniques. For protein/ss DNA interactions, ITC and SPR (Surface Plasmon Resonance) are frequently used. Both methods have their own strengths and weaknesses (see sections 1.5.2.2 and 1.5.3.5) and are complementary to one another, providing a method of cross validating any results obtained.

1.5.2.1 Isothermal Titration Calorimetry

ITC is a thermodynamic technique for monitoring any chemical reaction initiated by the addition of a binding component, and has become the method of choice for characterising biomolecular interactions. When substances bind, heat is either generated or absorbed. Measurement of this heat allows accurate determination of binding constants (K_d), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS), thereby providing a complete thermodynamic profile of the molecular interaction in a single experiment.

In ITC, a syringe containing a "ligand" solution is titrated into a cell containing a solution of the "macromolecule" at constant temperature. When ligand is injected into the cell, the two materials interact, and heat is released or absorbed in direct proportion to the amount of binding. As the macromolecule in the cell becomes

saturated with ligand, the heat signal diminishes until only background heat of dilution is observed.

1.5.2.2 Advantages and disadvantages of using ITC for examining protein/ss DNA interactions.

ITC has typically been viewed as the "gold standard" for characterising molecular interactions. A major advantage of this method is that ITC does not require immobilisation of the reactants. The major disadvantages of using ITC is the large amounts of samples required (μM amounts of protein and ligand are required), limited throughput (although this has been improved with the development of an automated ITC machine) and provides no direct information about the kinetics of the system.

1.5.3 Surface plasmon resonance (SPR)

The core mechanics of surface plasmon resonance (SPR) are complicated (see [145] for a detailed review) but the technique itself does not require a detailed understanding of the theory. SPR instruments use optics to measure a change refractive index near the surface of a sensor chip. In Biacore instruments, a running buffer is passed under a continuous flow rate (30-100 μl.min^{-1}) over a micro-fluidic flow cell (~ 20 - 60 nl in volume). To allow for the detection of an interaction one molecule (the ligand) is immobilised onto the surface of a sensor chip. Its binding partner (the analyte) is then injected in solution (ideally with the same components and composition as the running buffer to minimise refractive index changes caused by buffer mis-match) through the flow cell, under continuous flow. As the analyte binds to the ligand the build up of protein on the sensor surface causes an increase in

refractive index. This refractive index change is measured in real time (readings are taken every 0.1 s), and the results are plotted as response units (RUs) *versus* time (termed a sensorgram). A response (background response) will also be generated if there is a difference in the refractive indices of the running and sample buffers. This background response is subtracted from the sensorgram to obtain the actual binding response. The background response is calculated by injecting the analyte into a reference flow cell (serves as a control), which has no ligand or an unrelated ligand immobilised to the sensor surface. The measurement of association and dissociation of a binding interaction allows for the calculation of association (k_a) and dissociation (k_d) rate constants and the corresponding affinity constants (k_a/k_d). One RU represents the binding of 1 pg protein/mm^2. More than 50 pg/mm^2 of analyte binding is needed to generate good reproducible responses although it is difficult to immobilise a sufficiently high density of ligand directly onto a sensor surface to achieve this level of analyte binding. To overcome this technical difficulty, Biacore sensor chips have a 100 - 200 nm thick carboxymethylated dextran matrix attached to the surface. This adds a third dimension to the surface, allowing the user to generate much higher levels of immobilisation. An added bonus is that the carboxymethylated dextran matrix is very amenable to modification for covalent immobilisation reactions by a variety of chemistries (Figure 1-12).
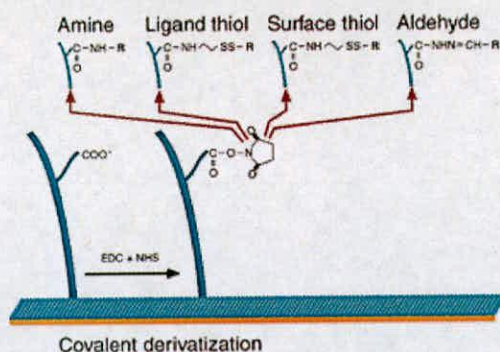
Figure 1-12. **Standard covalent immobilisation reactions used for SPR.**

Nevertheless having high levels of ligand immobilised on the surface has two important implications for the technique. High ligand density may result in a rate of analyte binding which exceeds the rate at which the analyte can be delivered to the surface (referred to as mass transport). Mass transport then becomes the rate-limiting step and as a result, the measured apparent k_a will be slower than the true k_a. Another common problem is that following dissociation of the analyte, it could rebind to an unoccupied ligand site before diffusing out of the matrix and being washed from the flow cell. Although the dextran matrix may exaggerate these kinetic artefacts (mass transport limitations and re-binding) they can affect the accuracy of the final results.

1.5.3.1 Applications of SPR

1.5.3.1.1.1 Equilibrium measurements.

Equilibrium analysis requires multiple sequential injections of analyte at different concentrations and at different temperatures. This can be very time consuming and it is only practical to perform equilibrium analysis on interactions that reach equilibrium within a reasonable time period (~5 - 10 minutes). The time it takes to reach equilibrium is determined principally by the dissociation rate constant (k_d).

High affinity interactions (K_d < 10 nM – classic antibody binding) usually have very slow k_d values and can therefore often be unsuitable for equilibrium analysis (the need to be able to regenerate the surface between runs becomes an important consideration here). Likewise, weak interactions (K_d >100 µM) are easily studied. The small sample volumes required for BIAcore injections (<20 µl) make it possible to inject the very high concentrations (>500 µM) of analyte (depending on solubility) to saturate low affinity interactions. Equilibrium affinity measurements on the BIAcore are highly reproducible. This feature and exact temperature control make it possible to estimate binding enthalpy by van't Hoff analysis. This involves measuring the change in affinity with temperature (for an example see section 4.5.4). Although not as meticulous as calorimetry, much less protein is required (for an example see [146]).

1.5.3.1.1.2 Kinetic measurements.

The fact that the BIAcore generates real-time binding data makes it well suited to the analysis of binding kinetics. However there are important limitations to kinetic analysis; due to mass transport limitations it is difficult to measure accurately k_a values faster than ~10 $\mu M^{-1} s^{-1}$. This upper limit is dependent on the size of the analyte. Faster k_a values can be measured with analytes with a greater molecular mass. This is because the larger signal produced by a large analyte allows the experiment to be performed at lower ligand densities, and lower ligand densities produce lower rates of mass transport. For different reasons measuring k_d values slower than 1 x 10^{-5} s^{-1} or faster than ~1 s^{-1} is difficult. BIAcore is easy to use and the analysis software is user-friendly, it is deceptively easy to generate kinetic data.

However obtaining accurate kinetic data can be a demanding, time-consuming task, and requires a thorough understanding of binding kinetics and the potential sources of artefact. In addition, it can be expensive, running to £1,000's of reagent to get quantitative results (for an example see [147]).

1.5.3.2 Ligand immobilisation

Immobilisation of one of the reactants onto the sensor chip surface without adversely affecting its activity can often be the most challenging step. Biacore have developed a series of sensor chips which allow for either direct covalent coupling of the ligand to the sensor surface or non-covalent immobilisation by use of an affinity tag.

1.5.3.2.1 Covalent immobilisation: A CM5 Biacore chip

CM5 chips and covalent coupling via primary amine chemistries are the standard and the simplest way to generate a surface. Primary amine coupling; one chemically modifies around 30% of the carboxylates on the matrix surface to generate succinimide esters (Figure 2-6b), which are highly reactive with primary amines (the N-terminus and the side chain of lysine residues). Then, under conditions of low ionic strength and a pH between 5.5 and 4.5 usually, where the net charge on the protein is positive and the carboxy dextran matrix is negative, the protein is weakly electrostatcially attracted close to the reactive groups, which drives the coupling reaction. The remaining activated carboxymethyl groups are then blocked by injecting high concentrations of ethanolamine over the surface.

1.5.3.2.2 Non-covalent immobilisation: A streptavidin (SA) sensor chip

SA chips are composed of a layer of carboxymethylated dextran pre-immobilised with the tetrameric protein streptavidin for immobilisation of biotinylated interaction partners.

1.5.3.2.3 Non-covalent immobilisation: A NTA Biacore chip

These function essentially like NTA-resins. You charge them with Ni^{2+} ions then immobilise a 6xhis-tagged protein onto this surface.

1.5.3.3   Activity of immobilised ligand

Levels of immobilisation should be high enough ligand to give you an adequate signal. Equation 1 illustrates the simple considerations for a 1:1 interaction. Ideally for kinetic analysis, the maximal RU signal should be ~ 100 to 150 RU for protein/protein, protein/DNA interactions and ~ 25 RU for small molecules (< ~ 1000 Da.). This kind of maximal signal maximises the ability of instrument (and the analysis software) to accurately detect and fit a 1:1 binding model to changes in the RU *versus* time to give you reliable rate constant data.

Equation 1: **RUmax = RU = (MW analyte / MW Ligand) x RU ligand immobilised**

MW = Molecular weight in daltons

However, this assumes 100% activity and this will never be the case, so higher amounts are always immobilised (~ 20 - 50% and in some cases up to 300% - 400% more) on the surface to ensure high enough activity levels. Following the generation of the surface, the activity of the surface needs to be tested. Ideally a good positive control is required, which is injected over the surface at saturating concentrations

allowing the user to assess the relative activity of the surface.

One must also consider how the analyte(s) are going to be dissociated from one another without destroying the biological properties of the ligand. If the k_d is fast, then just letting them dissociate in buffer alone may suffice, which is the best option as it is the least harsh. However, if the k_d is slow, then dissociation in just running buffer alone may make assay cycle times impractically long, in this case regeneration conditions (conditions that speed up the dissociation of complex) may need to be used.

1.5.3.4 Surface regeneration

Regeneration conditions include high salt, low or high pH, solvents, $MgCl_2$, detergents and EDTA (for NTA chip you use 350 mM EDTA, strips off Ni^{2+} and captured ligand no longer binds) etc. These regeneration conditions however, need to be stringent enough to remove bound analytes from the chip surface, but not destroy the surface. If they are too harsh, the relative activity of the sensor surface decays between cycles and produces unreliable values for affinity and rate constants.

The other important thing to remember here is that a protein captured via a His-tag interaction has its own set of kinetics (Ni and His-tag). Often these limit the experiments as the complex of protein and binding partner dissociates more slowly than the his-tag dissociates from the surface ($K_d \sim 1$ μM, with a $k_d \sim 1 \times 10^{-2} - 1 \times 10^{-3}$ s^{-1} at 25°C). This means that kinetic data obtained from this type of surface will be inaccurate and the surface will drift as a result. This becomes even more of a problem if you need to run experiments at higher temperatures (as discussed in section 4.5.2.1).

1.5.3.5 Advantages and disadvantages of using Biacore to study protein ss DNA interactions.

Biacore has several advantages and disadvantages over other biophysical techniques when measuring protein/ss DNA interactions.

Advantages: SPR provides a label free technique (No need for radioactive, fluorescent or any other labelling) for the real time analysis of molecular interactions. Rate constants for the interaction can be determined over a temperature gradient, which makes it possible to thermodynamically characterise the interaction. The application of the van't Hoff theory to SPR data allows one to derive thermodynamic constants, which complement ITC data. Impure samples can be used for Biacore experiments. The method is extremely sensitive and can accurately determine rate constants from a maximum response of 5 RU.

Disadvantages: Issues with mass transport can affect kinetic measurements. Any artificial change in response, other than that of the interaction, can also give a signal. One of the interacting molecules must be immobilised to the surface. It can sometimes be difficult to determine suitable conditions for surface regeneration. Incorrect orientation of the molecules immobilised to the surface can inhibit accessibility to the active site.

1.5.4 Structural studies of protein/DNA interactions

Examination of the structure of SNABPs in a DNA bound and unbound state, at an atomic level, can provide a detailed insight of the SNABP-ss DNA interactions, one method that allows for such an analysis is X-ray crystallography.

1.5.4.1 X-ray crystallography

There are few techniques available for the determination of three dimensional structures of protein molecules these include, nuclear magnetic resonance (NMR), Cryo-electron microscopy (Cryo-EM), computer generated homology models and X-ray crystallography. X-ray crystallography is the method used to determine the arrangement of atoms within a crystal using X-rays scattered from electrons. Using the high-resolution information the atomic structure of a molecule can be obtained. The structural model derived contains information on the protein folds and the intra- and inter-molecular contacts made between protein molecules and/or bound ligands, which has proved extremely useful in structure-based drug design [148]. X-ray crystallography has also been successfully employed to examine the specificity of protein-DNA interactions [149] and to determine associated mechanisms [150].

Bonded atoms are about 1.5 Å apart and cannot be visualised under visible light magnification as they require a much shorter wavelength (λ) than that of visible light (4000-7000 Å). X-ray radiation has a λ in the range of 0.02 to 100 Å and can be used to visualise the atoms which make up the molecule of study. X-rays are photons of energy with wave-like aspects, once these waves meet with electrons they are quickly absorbed by the electron cloud, forcing the electron to vibrate at the same frequency as the incoming X-ray. The electron then emits X-ray waves, at the same frequency, in all directions. In some of these directions the scattered beams are in phase and reinforce each other to give diffracted beams i.e. constructive interference. The resulting X-ray waves are then visualised using a detection system (X-ray film or a detector), which exhibits spots where X-ray waves have impinged, known as Bragg reflections. These two dimensional images of the reflections observed are

recorded and can be used to obtain the corresponding three dimensional coordinates. Protein crystals are 3-demensional structures composed of repeating units, known as unit cells. Unit cells are related through a series of parallel planes, known as Bragg planes and can be considered as a series of mirrors which reflect the incoming X-rays. The position of each reflection spot is defined by the size and shape of the unit cell and the inherent crystal symmetry. The geometry of diffraction from a crystal lattice can be determined using Bragg's law (Figure 1-13);
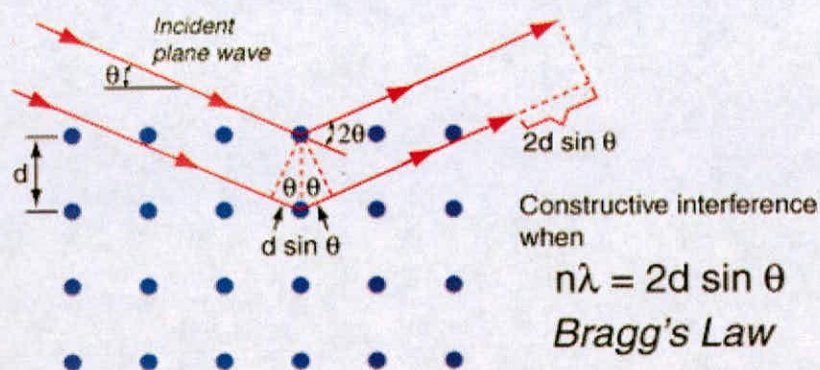
Figure 1-13. **Bragg' law.**

The x-ray scattering of objects in a crystal lattice. Figure adapted form [151].

where n is an integer, λ is the wavelength of X-rays, θ is the angle between the incident and reflected X-rays scattered by the parallel planes and d is the interplanar spacing [151].

To summarize, X-ray crystallography consists of three steps. The first step is to obtain adequate diffracting crystals of the material being studied. Obtaining adequate diffracting crystals is often the toughest and most crucial step in the crystallographic process. Various methods and techniques are available [152], although obtaining

crystals can be a long, time consuming step which may take years. In the second step, the crystals are exposed to an intense beam of single wavelength X-rays (depending on the experiment) to produce a series of reflections. The crystal is gradually rotated and each set of reflections produced are recorded until a complete set of reflections, covering a full rotation, are obtained. In the third and final step, these data are combined computationally to produce a model of the arrangement of atoms within the crystal. The model is then refined to produce the final crystal structure.

1.5.4.2 Crystallisation of protein-ligand complexes

If optimal crystallisation conditions exist for the native protein, obtaining protein-ligand complexes can still be extremely difficult. Two methods currently exist for successfully obtaining protein-ligand complexes, co-crystallisation or soaking the ligand into the preformed crystal.

The soaking method requires the production of apo protein crystals in the absence of ligands. Protein crystal can then either be transferred into a similar crystallisation solution containing the ligand at an appropriate concentration (10 times higher than the affinity [153]) or a highly concentrated solution of ligand can be added to the crystallisation solution containing the fully grown protein crystals. This method has several problems such as, damaging of protein crystals, resulting in diminished diffraction and/or obtaining an alternate space group [154]. Another consideration when using this method is the spatial limitation of the crystal packing. If the ligand binding site of the protein molecule is blocked by adjacent molecules then larger ligands maybe unable to reach their target site.

The co-crystallisation method involves the crystallisation of the purified protein-ligand complex. The major drawback of this method is that the crystallisation conditions obtained for apo crystal growth often do not produce crystals of protein-ligand complexes as the ionization state of functional groups (a direct result of pH) involved in ligand binding may determine whether ligand binding will occur, for this reason the screening procedure must be repeated to obtain favorable conditions for complex crystallisation.

1.5.4.3 Data collection

There are three common X-ray sources available, X-ray tubes, rotating anode tubes and particle storage rings, which produce synchrotron radiation in the X-ray region. For standard laboratory use, rotating anode sources are commonly used to obtain diffraction data and are ten times more powerful than X-ray tubes. Particle storage rings produce the most powerful X-ray sources. In these giant rings, charged particles are forced to circulate at near the speed of light. When accelerated and forced into a curved motion, these charged particles then emit energy as synchrotron radiation (X-rays). Synchrotron sources are available only at storage rings and have several advantages over rotating anode sources. Data sets can be collected in hours at a synchrotron source versus days for a single data set collection from a rotating anode source. Synchrotron sources also allow for data collection from extremely small crystals [155].

1.6 Project outlines

1.6.1 Examining the sequence specificity of single stranded DNA binding proteins: A novel DNA microarray approach

Clearly one of the major functions of single stranded nucleic acid binding proteins (SNABPs) is their ability to regulate both transcription [82] and translation [156] by binding to regions of ss DNA. To understand how binding of SNABPs to ss nucleic acids regulates transcription and translation, the regions of sequence specificity must first be identified. Although, many techniques (e.g. electrophoretic mobility shift assay (EMSA), nitrocellulose-binding assays, Southwestern blotting, phage display, UV cross-linking and X-ray crystallography) were developed to study sequence specific ss DNA-protein interactions. These techniques made use of non-immobilised ss DNA in liquid phase to probe ss DNA interactions with other molecules (e.g. proteins, drugs and ligands) and all suffered from being time consuming, laborious, expensive and incapable of high-throughput screening. To resolve these problems, a novel DNA microarray-based approach has been developed for the identification of the sequence-specificity of SNABPs, which will help provide an insight into their cellular function. Work from this project resulted in a publication –

Morgan, H.P., *et al.*, Sequence specificity of single-stranded DNA-binding proteins: a novel DNA microarray approach. **Nucleic Acids Res**, 2007. 35(10): p. e75.

1.6.2 Examination of a family of CSPs

The differences reported for the function of CSPs and their homologous eukaryotic equivalents lead to the hypothesis that there may be a different function between CSPs from the same family. These differences in function may therefore be reflected in their ability to recognise different ss DNA sequences. In *Salmonella typhimurium*,

six CSP (*St*CspA, *St*CspB, *St*CspC, *St*CspD, *St*CspE, *St*CspH) have been identified and the cold inducibility of *St*CspA , *St*CspB and *St*CspH has been reported [157-160], although their functions are yet to be clearly elucidated. To determine differences in ss DNA binding affinity and sequence recognition all six *S. typhimurium* CSPs were cloned, expressed (with the exception of *St*CspH) and purified. Each *S. typhimurium* CSP was then characterised, regarding ss DNA sequence specifity, using the previously developed microarray assay. Two types of recognition sequences were identified, pyrimidine (T) and purine rich (G) binders (*St*CspE = pyrimidine rich binder and *St*CspD = purine rich binder). Further biochemical studies of the *St*CSP/ss DNA interactions were performed using advanced biophysical techniques.

Work from both this project formed the basis of a second publication:

Morgan, H.P., *et al.* Structural and thermodynamic studies of a single-stranded DNA binding cold shock protein from *Salmonella typhimurium*. To be published.

1.6.3 *Structural studies of S. typhimurium CSPs*

Crystallisation of both *St*CspE and *St*CspD, with and without the consensus ss DNA sequence, would allow for the characterisation of both types of ss DNA binders at a molecular level. The results would explain the differences seen in sequence recognition and may help to determine the mechanism of CSP/ss DNA binding and even their cellular function. Despite many attempts only crystals of *St*CspE were obtained, from which a structure was determined. Models for *St*CspD, *St*CspD/ss DNA and *St*CspE/ss DNA were created to explain differences in ss DNA recognition. Work from this project provided the basis of a third publication:

Morgan, H.P., *et al.* The crystal structure of cold shock protein E from *Salmonella typhimurium*. To be published.

43

CHAPTER 2. Sequence specificity of single stranded DNA binding proteins: A novel DNA microarray approach.

2.1 Aim

To examine the use of a generic oligonucleotide chip as the basis for the development of a high-throughput assay for the identification of sequence-specific ss DNA-protein interactions.

2.2 Summary

Oligonucleotides similar in sequence to those found on the surface of the array were synthesised and used in a series of gel shift experiments to determine if sequence specific ss DNA binding proteins (e.g. *Bs*CspB) would successfully bind to the chip surface. Proteins were then labelled appropriately (either directly labelled with Cy5 or detected using Cy3-penta-His antibodies). The subsequent binding to the chip could then be visualised using a standard microarray scanner. This work demonstrates that microarray technology can provide rapid high-throughput assays for the identification of sequence specific ss DNA-protein interactions.

2.3 Introduction

Genome sequencing has allowed SNABPs to be identified and characterised for a range of eukaryotic and prokaryotic organisms. To understand how binding of SNABPs to ss nucleic acids regulates transcription and translation, the regions of sequence specificity must be identified. Many techniques including electrophoretic mobility shift assay (EMSA) [134], nitrocellulose-binding assays [135], Southwestern blotting [136], phage display [137], UV cross-linking [138] and X-ray crystallography [139] were developed to study sequence specific ss DNA-protein

interactions. Available techniques including, fluorescence measurements [28], polymerase chain reaction (PCR), fluorescence resonance energy transfer (FRET) combined with a DNA foot-printing assay [140], surface plasmon resonance (SPR) and fluorescence polarisation [141] have all been used to effectively study specific ss DNA-protein interactions. The most frequent approach used to study the sequence specificity of DNA binding molecules is by systematic evolution of ligands by exponential enrichment (SELEX). This method allows for the identification of sequences which bind with high affinity to the molecule of interest [144]. This method has been used mostly to select for double stranded DNA molecules that bind to the target but it has been also been used to screen ss DNA molecules [161] [162]. SELEX has advantages compared to the previously used methods but still is lacking in its suitability for high throughput analysis as numerous microarray experiments can be completed in a single day, thus providing a detailed analysis of binding site recognition at an unparalleled rate.

These techniques described above make use of non-immobilised ss DNA in liquid phase to probe ss DNA interactions with other molecules such as proteins, drugs and ligands; all suffered from being time consuming, laborious, expensive and incapable of high-throughput. Oligonucleotides immobilised to solid supports overcome many of these problems, providing an important tool for the rapid high-throughput examination of sequence-specific DNA protein interactions.

Two current studies [163, 164] have used microarrays displaying all possible 8-mer and 10-mer DNA duplexes to effectively study the sequence-recognition of both transcription factors and small molecules. These methods illustrate the potential

high-throughput use of x-mer arrays in examining the DNA binding properties of duplex binding molecules but leave the area of SNABPs specificity unexplored.

The innovative high-throughput assay described here provides a parallel screening system for identifying the specificity of SNABP binding. The major cold shock protein from *Bacillus subtilis,* CspB, was used to develop this microarray based assay. This protein influences transcription and translation *in vitro* [91] by binding to stretches of 6-7 nucleotides [130] of ss DNA with a high degree of specificity [86, 130].

We have used an oligonucleotide chip, for the identification of high-affinity 6-mer binding motifs for CspB. The chip contains all possible 4096 ss hexadeoxynucleotides incorporated onto a standardised anchor. The oligonucleotides on the array were originally designed to hybridise to folded mRNA [165] which requires a significant spacer between the array surface and the recognition hexamer (Figure 2-1).

The use of a competitor protein in this assay, allowed the identification of high affinity DNA binding sites. The binding affinity of a competitor protein will limit the amount of ss DNA binding sites available to the CspB. The competitor protein chosen was a single-stranded DNA-binding protein from *Sulfolobus solfataricus* (*Sso*SSB). *Sso*SSB has a molecular weight of 16 kDa and binds non-specifically to ss DNA with a binding density of 5 nt per monomer and an apparent dissociation constant (K_d) of ~90 nM [43]. Thus, the competitive binding of the *Sso*SSB protein provided a means of identifying high affinity consensus binding motifs for CspB by reducing non-specific and weak CspB-ss DNA binding.

2.4   Materials and methods

2.4.1   Microchip manufacture

Oligonucleotide chips were supplied by Nyrion Ltd and contained all possible 4096 ss hexadeoxynucleotides incorporated into a general structure (Figure 2-1), 5′-NH2-C12-Spacer-AAAAAAAAAA-NNNNNNNNN-XXXXXX-3′, where N was any random base and X was a specific hexadeoxynucleotide. Each chip is made up of a 4 x 4 meta-grid and each of these sub grids contains 18 columns x 15 rows of spots. The spots are 135 $\pm$ 20 μm in diameter.
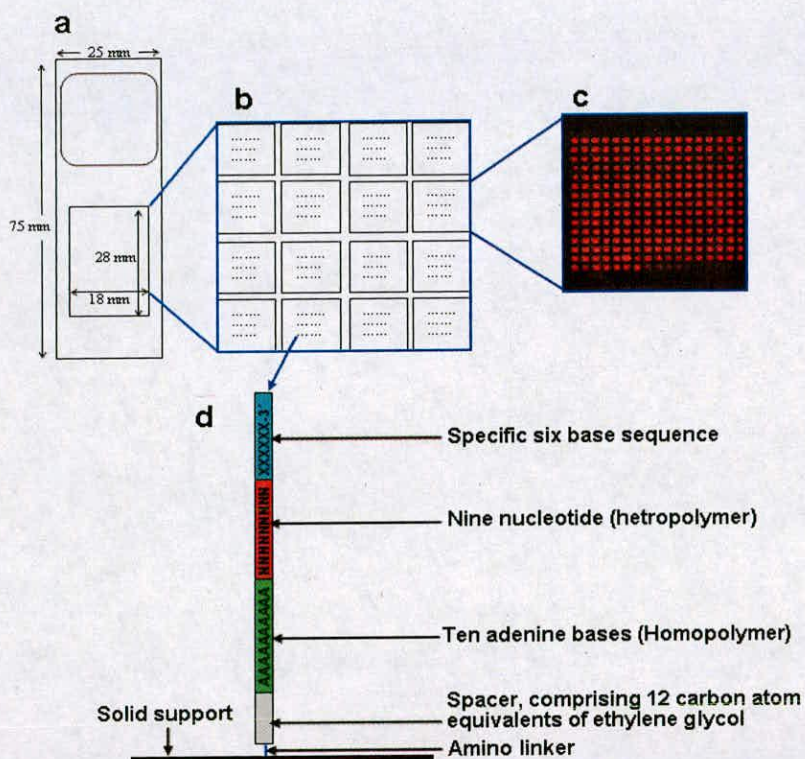
Figure 2-1. **The design of the microarray.**

(a) Oligonucleotide chip. (b) Grid pattern: A 4 X 4 meta-grid each sub-array having 18 columns x 15 rows of spots, each spot is 135 $\pm$ 20 μm.
(c) Magnification of a portion of the microarray bound by *Sso*SSB-Cy5.
(c) A schematic representation of a hexadeoxynucleotides on the oligonucleotide chip surface.

Oligonucleotides were immobilised to the chip surface using standard Exiqon amino-link chemistry. All arrays were manufactured by pin spotting, according to completely standard commercial practices (contact [165] for details of array design). This was all done under contract by MWG custom arrays. For control purposes arrays are batch tested using a standard mRNA template and a standard QC procedure expected to give a standard signal. This standard signal serves as a positive and negative control for all arrays. MWG spot biotin on the surface of the array, which also serves as a negative standard control (generates zero signal).

2.4.2 Expression and purification of recombinant SsoSSB

A mutant version of the SsoSSB protein from was constructed by changing the C-terminal glutamate residue to a cysteine (E145C mutant) [166], allowing for the incorporation of spin labels and fluorescent probes on the C-terminal tail. This mutation minimises the affect of labelling on DNA binding activity as the C-terminal glutamate is not involved in ss DNA binding. The E145C mutant was constructed only as a precaution if the amine reactive labelling methods were unsuccessful. Protein expression was induced by addition of 0.2 mM IPTG at 37°C for 3 hours, after which cells were pelleted and frozen until required. Cell lysis, centrifugation and chromatography steps were carried out at 4°C. Cells (20 g) were thawed in 50 ml lysis buffer (50 mM Tris–HCl pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM DTT) and immediately sonicated for 5 × 1 min with cooling. The lysate was centrifuged at 40 000 g for 45 min. DNase I [40 µg/ml] and RNase A [10 µg/ml]) were then added to the cell lysate and incubated at room temperature for 30 minutes with gentle agitation. The supernatant was heated to 70°C for 30 min in a water bath, and denatured proteins were precipitated by centrifugation at 40 000 g at 4°C. The

supernatant was analysed by SDS–PAGE, and shown to contain recombinant SSB, which migrated as a band of ~16 kDa as expected. The supernatant was diluted 5-fold with buffer A (50 mM Tris–HCl pH 7.5, 1 mM EDTA, 1 mM DTT) and applied to a Heparin-Sepharose (Amersham) column equilibrated with buffer A. *Sso*SSB was eluted over a linear gradient comprising 0–1 M NaCl. Fractions corresponding to a distinct absorbance peak were analysed by SDS–PAGE, pooled and concentrated. A subsequent gel filtration step (HR 10/30 Superdex-200) in a buffer containing, 10 mM Tris/HCl pH 7.5 , 150 mM NaCl, 1 mM EDTA and 1 mM DTT, resulted in essentially homogeneous *Sso*SSB, as determined by SDS-PAGE analysis (Figure 2-2b). This method is an adaptation of the previously published method [167]. *Sso*SSB was concentrated using a Viva Spin column (MWCO = 5 kDa) and quantified using both the Bradford method and the theoretical extinction coefficient, $\varepsilon 280nm = 12660\ M^{-1}.\ cm^{-1}$.
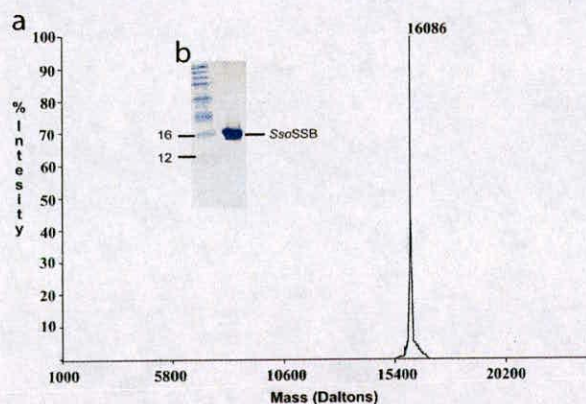
Figure 2-2. **Purified recombinant single stranded DNA binding protein from the crenarchaeote *Sulfolobus solfataricus*.**

(**a**) MALDI-Tof mass spectrum, observed 16086 Daltons, calculated 16111 Daltons. (**b**) Purified *Sso*SSB protein.

2.4.3 Cloning His₆-CspB

Primers *Bs*CSP fwd (5′-dAGC**CATATG** TTA GAA GGT AAA GTA AAA TAA - 3′) and *Bs*CSP rev (5′-dCG**GATCC** TAA CGC TTC TTT AGT AAC GTT AGC-3′) were used in a PCR with plasmid DNA (pET11-CspB vector provided by Michael Wunderlich (University of Bayreuth) containing the CspB gene. Bases were added to the primers to introduce the *Nde*I and *Bam*HI restriction sites (underlined). These sites were used to clone the PCR product in *Nde*I-*Bam*HI-digested pET28a vector, resulting in the plasmid pET28a_*Bs*CspB (Figure 2-3).
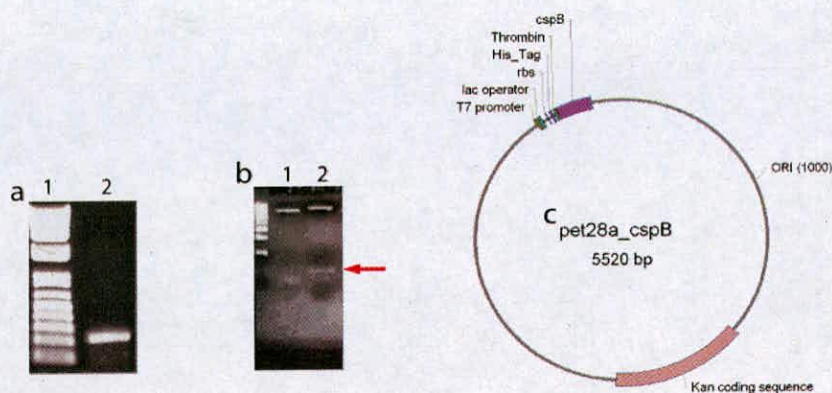
Figure 2-3. Cloning of *Bs*CspB.

(a) PCR of *Bs*CspB (lane 2). **(b)** Restriction digest of cloned PCR product. Lane 1: uncut pET28a_*Bs*CspB plasmid DNA, lane 2 digested pET28a_*Bs*CspB plasmid DNA showing a band of appropriate size (Red marker ~215 bp). 1 Kb marker used in gel b and 100 bp ladder used in gel a. **(c)** Plasmid map of pET28a_*Bs*CspB construct.

2.4.4 Expression and purification of recombinant His₆-CspB

BL21 (DE3)pLysS *Escherichia coli* (E. coli) was transformed with pET28_B.S._CspB and transformants were grown in Luria-Bertani medium, containing 50 µg/ml kanamycin at 37°C with agitation. One-litre cultures were grown to an OD_{600} of 0.5-0.7 and IPTG (isopropyl-β-D-thiogalactosidase) was then added to a final concentration of 1 mM. Incubation was then continued for an

additional 5-6 hours and the cells were harvested at 8,000 rpm in a JLA-9.1000 rotor for 12 minutes at 10°C. Pellets were then frozen in liquid nitrogen and stored at -80°C. Cell pellets were resuspended in lysis buffer (20 mM Tris HCl pH 8.0, 500 mM NaCl, 0.1% Triton X-100, 0.1 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM EDTA and protease inhibitors) to a final volume of 30 ml/5 g of cells and then lysed using a French-press. DNase I [40 µg/ml] and RNase A [10 µg/ml]) were then added to the cell lysate and incubated at room temperature for 30 minutes with gentle agitation. As an initial step, the fusion protein was purified using a Ni-NTA resin affinity column, as per manufacturer's instructions and then purified to homogeneity as described previously [168]. Briefly, to remove minor contaminants the fractions containing His_6-CspB were pooled and dialysed overnight into a buffer containing 20 mM Tris/HCl pH 6.8, 1 mM DTT. The solution was applied to an HR 5/5 Mono-Q (1 ml) anion exchange column. Bound protein was eluted with a NaCl-gradient ranging from 0-1 M. CspB eluted at a concentration of 250 mM NaCl. A subsequent gel filtration step (HR 10/30 Superdex-75) in a buffer containing 10 mM Tris/HCl pH 7.5 and 100 mM NaCl) resulted in visually pure His_6-CspB, as determined by SDS-PAGE analysis (Figure 2-4). His_6-CspB was concentrated using a Viva Spin column (MWCO = 3.5 kDa) and quantified using both the Bradford method and the theoretical extinction coefficient, $\varepsilon280nm = 5690 \, M^{-1} . cm^{-1}$.
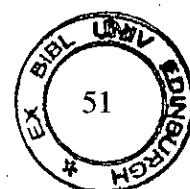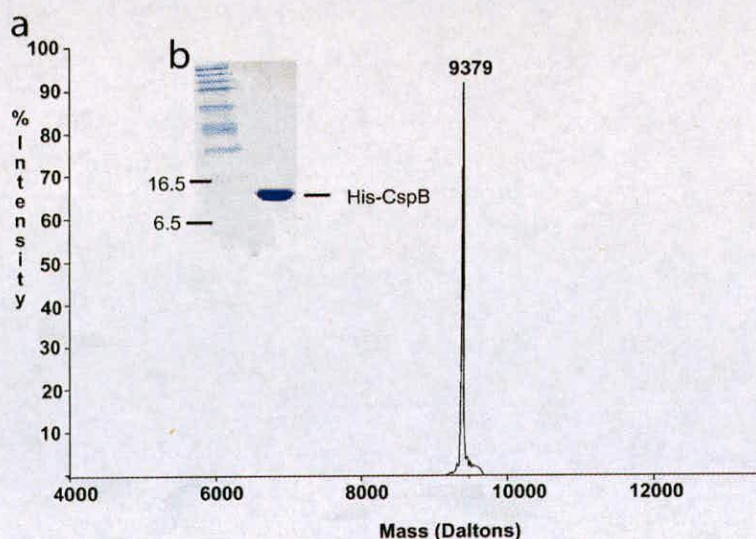
Figure 2-4. **Purified recombinant cold shock protein B from** *Bacillus subtilis.*

(**a**) MALDI-Tof mass spectrum, observed 9379 Daltons, calculated 9528 Daltons (the difference in mass is possibly a result of the loss of a methionine residue from the N-terminal). (**b**) Purified His_6-CspB

2.4.5 *Electrophoretic mobility shift assay (EMSA)*

300 pmol of each ss DNA template were 5′-end labelled by incubating templates with 0.03 mCi of $[\gamma-^{32}P]ATP$, T4 polynucleotide kinase and T4 polynucleotide kinase buffer in a total volume of 60 µl at 37°C for 2.5 hours. The reaction was stopped by heat inactivation (30 minutes at 65°C). Unincorporated $[\gamma-^{32}P]ATP$ was removed with QIAquick Nucleotide Removal Kit (Qiagen).

For a standard EMSA, 20 pmol of labelled ss DNA was mixed with increasing amounts of protein (total volume, 18 µl) at 4°C for 20 minutes in binding buffer (50 mM Tris, pH 8.0, 100 mM NaCl) unless stated otherwise. 2 µl of dye solution (20% glycerol, 0.034% bromophenol blue) was added to the samples prior to gel electrophoresis.

Electrophoresis was performed in TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0) buffer through a non-denaturing acrylamide gel (a 10% or 20% gel

was used depending on the protein; e.g. for 50 ml of 20% gel; 25 ml of 40% acrylamide, 2.5 ml of 10x TBE, 0.5 ml APS and 50 µl Temed) at 75 volts until the samples had entered the gel and then at 100 volts at 4°C (overnight for a 20% gel and 12 hours for a 10% gel). Autoradiographs were obtained by exposing gels to Kodak BioMax MS film for 1-3 hours at room temp.

2.4.6 *Labelling of ss DNA binding proteins with Cy5 dye*

In the standard procedure, the content of 1 vial ("to label 1 mg of protein") of cyanine 5 (Cy5) mono-functional dye (Amersham) was dissolved in 50 µl of anhydrous DMSO. Proteins were dialysed into buffer B (150 mM, Na_2CO_3 (pH 9.3, pH was adjusted with H_3PO_4) and concentrated using a viva spin column (molecular weight cut off (MWCO) = 5 kDa). Typical working concentrations of proteins were 1 mg/ml, unless stated otherwise. 10 µl of Cy5 dye solution was pipetted into 200 µl protein solution under slow vortexing. After 30 minute incubation at 25°C in the dark, the reaction was terminated by the addition of 300 µl of 100 mM NaH_2PO_4 (to suppress further labelling) to the sample. To separate the unbound dye, the sample was loaded onto a PD-10 column (10 ml bed of Sephadex G-25M), which had been pre-equilibrated in a buffer A (100 mM NaCl, 50 mM NaH_2PO_4 and 1 mM EDTA [pH 7.5, pH adjusted with NaOH]). The column was then washed with buffer A (2 × 1 ml) and the labelled protein was then eluted by adding 2 ml of water to the column. The extent of the modification was assessed using MALDI-TOF mass-spectrometry. Protein concentration was determined before and after labelling, Cy5-protein concentration was calculated as per manufactures instructions.

2.4.7 Protein Hybridization

Microarrays were pre-wet with phosphate-buffered saline (PBS) and 0.01% Triton X-100 and then blocked in 2% non-fat dried milk for 1 hour. Blocked microarray slides were washed once with PBS and 0.1% Tween 20, and once with PBS and 0.01% Triton X-100. Protein (Cy5 labelled and unlabelled) binding to ss 25-mers (containing all possible [4096] 6-mer sequences) on a generic microchip was carried out in a hybridization chamber (Camlab, RTP/7870). Protein binding was performed in a humid chamber at 4°C with 80 µl of protein binding reaction mix containing: 50 mM KCl, 20 mM Tris (pH 8.0), 2% (w/v) non-fat dried milk, 0.2 µg/µl bovine serum albumin and 40 µM of test protein. Slides were covered with a siliconised cover-slip (BDH cover glass 22x50 mm, Cat. No. 406/0188/42, Borosilicate Glass) and incubated for 1 hour at 4°C. The cover-slip was removed and the slide was washed (3x) in a slide chamber filled with PBS and 0.05% Tween-20, with PBS and 0.01% Triton X-100 (3x) and once with PBS for 3 minutes each. Excess water was removed from the slide surface (by flicking), which was allowed to dry before scanning. This method is an adaptation of the previously published method [169]. Various methods (denaturing conditions; including high temperatures, various concentrations of detergents and pH range in combination with high NaCl concentration) were used to remove bound protein from the chip surface in order to reuse the array but were all unsuccessful as they either had detrimental affects on the oligonucleotides bound to the surface of the array or were unable to remove bound protein.

2.4.8 Competitive assay

The method was essentially the same as above except that both His_6-CspB and *Sso*SSB proteins were added to the binding reaction mix at the specified molar ratio.

The binding reaction was carried out as before. The array was then incubated for 1 hour in a humid chamber at 4°C with 100 μl of diluted (1:100 in blocking buffer) Alexa 532-conjugated polyclonal antibody to His_5 (Molecular Probes). After incubation the array was washed (3x) with PBS and 0.05% Tween-20 and once with PBS for 3 minutes each. Excess water was removed from the slide surface (by flicking), which was allowed to dry before scanning.

2.4.9   Microarray analysis: Data collection

All microarray slides were scanned using an ArrayWorx microarray scanner at a range of laser settings, the highest of which produced a saturated signal for the majority of spots. The Alexa-532 (Cy3 equivalent) fluorophore was excited at 532 nm and the emission was recorded at 570 nm. The Cy5 fluorophore was excited at 633 nm and the emission was recorded at 675 nm. The data was filtered initially using a series of quality control criteria so that only high quality spots were used in our analysis. For each array we removed any flagged spots, these were spots that had dust flakes, scratches and irregular spots (spots that were larger than the expected average size). The average size of a spot is 135 ± 20 μm in diameter; any spot which did not correspond to this size constraint was excluded from the data, meaning only spots with similar DNA concentrations were collected. All microarray TIF images were quantified using Imagene Version 5.0 software.

2.4.10 Microarray analysis: Data processing

The extent of background fluorescence was initially determined from an array experiment using bovine serum albumin (BSA). The level of background

fluorescence from the spots and array surface was found to be similar. Therefore, the average fluorescence between spots on the array surface was used as the background value throughout the experiments, which was minimal in comparison to the average signal intensity. Background subtracted median intensities were calculated for each spot on the microarray and the data was normalised according to the total signal intensity, so that the average spot intensity was the same for each replicate slide (x3). The normalised data of each competitive array experiment was used to generate a list of the high intensity sequences/spots (the highest to lowest intensity), i.e. spots which were above a threshold level of intensity (Figure 2-5a). High intensity spots/sequences which occurred in all three replicates were carried forward for further data analysis. This procedure minimised the occurrence of any false positives[*] or negatives[*] in the overall data collection. The average intensity was calculated and the sequences were ranked accordingly. The lists of sequences generated were condensed to include only the best binding sequences. These were spots that had intensity above 55% normalised fluorescence and at least 6 standard deviations away from the global mean intensity (Figure 2-5b). The final list contained a total of 50 high affinity binding sequences for His_6-CspB.

[*]False positives = spots which fluoresce highly on one array but not on all three arrays. Total = 6.5%; False negatives = spots which did not fluoresce on one array but fluoresced highly on two arrays. Total = 2%

Competitive-Normalised Fluorescence

SD-for high affinity binding sites
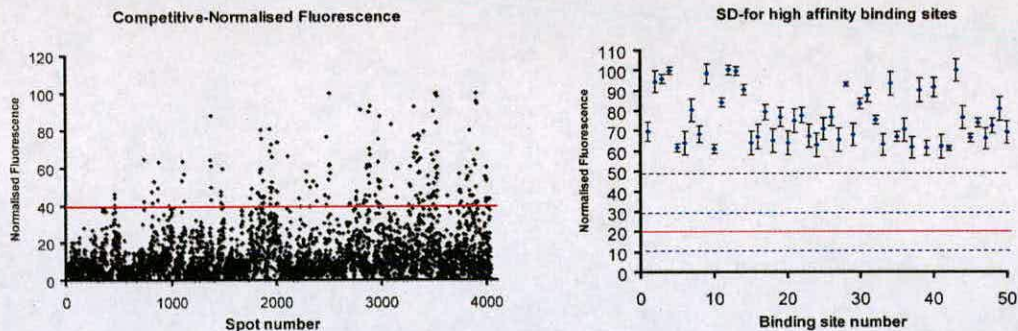
Figure 2-5. **The determination and statistical analysis of high-affinity binding sites (highly fluorescent spots).**

(a) The normalised fluorescence intensities for each spot from the competitive array experiment. The normalised intensity for His$_6$-CspB and *Sso*SSB bound to the chip surface and detected using anti-His-Alexa 532. (b) The standard deviation for each of the high-affinity binding sites. The red line represents the global mean intensity for non-specific binding, the dashed blue lines represents two standard deviations above and below the global mean and the dashed black line represents 6 standard deviations above the global mean intensity.

2.4.11 Isothermal Titration Calorimetry (ITC)

ITC experiments were carried out as described previously [130] with minor modifications. Four oligonucleotides were used for the experiments (Figure 2-16b, pg 77): both possibilities of the consensus binding sequence (ITC1 and ITC2) and a positive (ITC3[14]) and negative control (ITC$_{control}$). Each exothermic heat pulse (Figure 2-16a, upper panels) corresponds to an injection of 5 µl of each oligonucleotide (100 µM) into the cell containing 5 µM CspB at 28°C. Integrated heat data (Figure 2-16a, lower panels) is represented by a differential binding curve, which was fitted to a single-site binding model to give the stoichiometry of binding (N), binding affinity (K$_d$) and enthalpy of binding (ΔH) for each heptanucleotide.

2.4.12 Isolation of the Cy5-labelled SsoSSB/ss DNA complex from a native polyacrylamide gel

The non-radioactive Cy5-*Sso*SSB/ss DNA complex was excised from a native polyacrylamide gel using adjacent autoradiogram results as a guide. The gel slice was cut into small pieces and placed into a small piece of pre-wet dialysis membrane (MWCO = 3500 Da) clamped at one end. 1 ml of 1 x SDS-sample buffer (containing no bromophenol blue; 2 x SDS-sample buffer; 1 ml of 1.5 M Tris, pH 8.6, 600 μl of 20% β-mercaptoethanol, 3 ml of glycerol and volume adjusted to 10 ml with ddH$_2$O) was added to the gel pieces and the membrane was then sealed, ensuring that all air bubbles have been removed. The membrane was placed in a SDS-PAGE gel rig and completely filled with running buffer. This was run at 70 V for 1 hour. The contents from the dialysis membrane were added to a microcentrifuge tube and centrifuged at 13,000 rpm for 10 minutes. 20 μl of the supernatant was added to a 0.5 ml fluorescence (Quartz) cuvette and adjusted to a final volume of 5 ml using SDS-sample buffer. A spectrofluorometer (slit sizes set at; emission = 1 and excitation = 1) was used to record the emission spectrum from 655-750 nm, while exciting at 647 nm for Cy5.

2.4.13 MALDI-TOF (MS) of intact proteins.

Proteins were resuspended in either 0.1 or 1.0% trifluoroacetic acid (TFA) to give 1 mg/ml. A saturated solution of sinapinic acid was prepared in a solution containing 30% acetonitrile and 0.1% TFA for the matrix solution. The sample for MALDI-TOF analysis was then prepared by diluting 0.5 μl of protein solution to 0.5 μl with matrix solution and was spotted onto a flat, stainless steel plate from PerSeptive Biosystems. MALDI-TOF data were collected on a PerSeptive Voyager Elite

MALDI-TOF instrument. Standard instrument operating settings for sinapinic acid

were used at all times. Data were analysed using voyager software.

2.5 Results and discussion

2.5.1 Cy5 labelled BsCspB no longer binds ss DNA

In order to visualise binding of proteins to single stranded oligonucleotides (ON) immobilised on the generic chip the protein must first be labelled with an appropriate fluorescent dye. For this procedure to work the resulting fluorescently labeled protein should have high fluorescence quantum yields and retain the biological activities of the unlabeled protein. A number of suitable amine-reactive reagents are available for this purpose (Molecular Probes). Amine-reactive dyes react with non-protonated aliphatic amine groups, including the terminal amine of proteins and the ε-amino group of lysines (Figure 2-6). Protein conjugations are run in carbonate buffers with a pH ranging from 8.5–9.5, which is an optimal pH for modifying lysine residues.
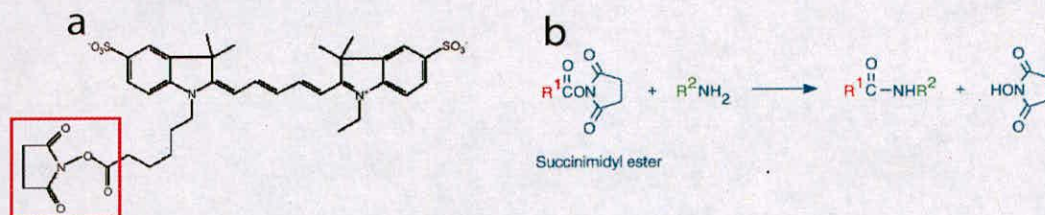
Figure 2-6. **The Cy5 covalent labelling reaction using direct amine coupling.** **(a)** 2-D representation of the chemical structure of the Cy5 succinimidyl ester (MW = 791.99 Da), the reactive group (141 Da) is highlighted red. Cy5 has a MW of 650.99 after standard amine coupling reaction **(b)**. This reaction results in a labelled protein with MW of; Protein + 650.99/Cy5 incorporated.

BsCspB was covalently labelled with the mono-functional dye Cyanine 5 (Cy5; Amersham), as described (Materials and Methods - 2.4.6). Analysis of labelled

*Bs*CspB-Cy5 using MALDI-Tof mass spectrometry indicated that the majority contained one dye moiety per protein molecule (see section A6). Electrophoretic mobility shift assays (EMSA) were conducted with a single-stranded 25-mer (ONc:5′–dATCCTACTGATTGGCCAAGGTGCTG-3′), labelled at the 5′-end with γ-^{32}P. ONc is a characterised binding sequence used in previous experiments [86]. The gel-shift results for unlabelled and labelled *Bs*CspB indicated that labelling with Cy5 significantly affected the binding affinity and specificity as the labelled protein, even at high concentrations, no longer bound ss DNA (Figure 2-7). The elimination of ss DNA binding activity is most likely a result of coupling Cy5 to a lysine near or within the ss DNA binding site.
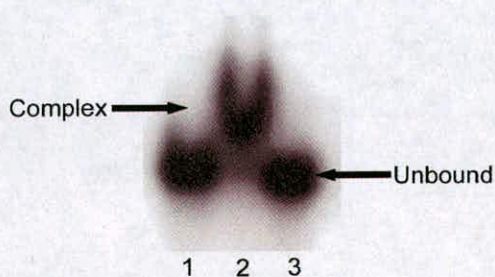
Figure 2-7. **Electrophoretic mobility shift assay of BsCspB and BsCspB-Cy5 and its complex with ONc.**

15-20 pmol of 5′-end γ-^{32}P-labelled ONc was mixed with 500 pmol of Cy5-labelled (lane 3) and unlabeled (lane 2) BsCspB and incubated for 10 minutes at 4°C. Lane 1 was a control and contained 15-20 pmol of 5′-end γ-^{32}P-labelled ONc only. Protein-DNA complexes were separated on a 7% native polyacrylamide gel.

2.5.2   *CspB and His$_6$-CspB have similar affinity for ss DNA.*

As direct labelling of *Bs*CspB with Cy5 rendered it inactive, a new method of detection was devised using an anti-his Cy5-conjugated antibody to detect His-tagged CspB once bound to the surface of the array. An N-terminal 6xHis-tag was added to *Bs*CspB by cloning the sequence into the pET28 vector (Novagen). A series

61

of EMSA were conducted (see below) to ensure that the addition of a His-tag had not affected the ss DNA binding properties of *Bs*CspB.

A ss 25-mer [133] (ONc: 5′–dATCCTACTGATTGGCCAAGGTGCTG-3′), labelled at the 5′-end with γ-^{32}P was used to compare the binding affinities and specificities of His$_6$-CspB and CspB. Figure 2-8 shows a gel-shift experiment performed for ONc in the presence of decreasing amounts of His$_6$-CspB (lanes 2-6) and non-His-tagged CspB (lane 7). Lanes 3, 4 and 5 show decreasing migration patterns for the His$_6$-CspB ss DNA complex as less protein molecules bind. The data show that the addition of the His-tag does not significantly affect the binding affinity or specificity of the protein.
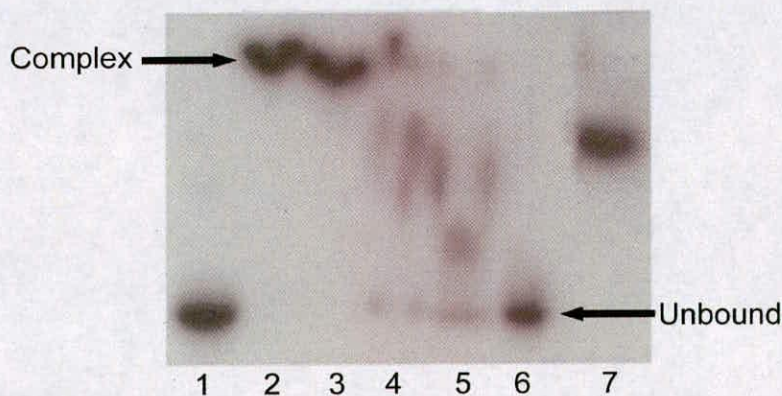
Figure 2-8. **Electrophoretic mobility shift assay of His$_6$-CspB and its complex with ONc.**

15-20 pmol of 5′-end γ-^{32}P-labelled ONc was mixed with 0, 500, 300, 100, 50 and 25 pmol (lanes 1-6) of His$_6$-Csp. Lane 7 contains 500 pmol of CspB (untagged) as a positive control. Protein-DNA complexes were separated on a 20% native polyacrylamide gel.

2.5.3 Determining the number of flanking bases required for ss DNA-CspB complex formation

The effect of flanking DNA at the 3′ end of the oligonucleotides, on CspB binding, was also examined by EMSA using a series of oligonucleotides that were structurally similar to the oligonucleotides found on the array; the only difference was that a varying numbers of bases were added to the 3′ end.

The incorporation of unknown bases (N) or wobbles within the oligonucleotides (ON's), while creating a more viable experiment, was thought to be inappropriate as we would be unable to predict any secondary structure formed by the ss DNA. The N-bases were therefore changed to A-based templates because CspB does not bind to homogeneous A-based templates [130]. CspB binds to the core Y-Box binding sequence (ATTGG) with a proposed binding density of 6-7 nt per monomer [130]. However, the Y-Box binding sequence (ATTGG) may require a certain quantity of flanking bases for CspB binding. In order to determine the optimum number of flanking bases required for CspB binding the following ON's were used (Table 1). This experiment will provide valuable information for the optimising the sequence of the ON's on the array (i.e. the required amount of flanking bases added to the 3′ end). Gel shift assays were conducted with four ss 25-mers, ON1, ON2, ON3 & ONc (Table 1), labelled at their 5′-ends with ^{32}P. The ON's have been flanked successively with adenine bases at the 3′ end.

Table 1. Oligonucleotides used for CspB binding experiments

Oligonucleotide	Sequence	Shift
ON1	5′-dAAAAAAAAAA-AAAAAAAAA-GATTGG-3′	slight
ON2	5′-dAAAAAAAAAA-AAAAAAAAA-GATTGG-A-3′	slight
ON3	5′-dAAAAAAAAAA-AAAAAAAAA-GATTGG-AAAA-3′	slight
ONc	5′-dATCCTACTGATTGGCCAAGGTGCTG-3′	YES

The Y-Box binding sequence, **ATTGG**.
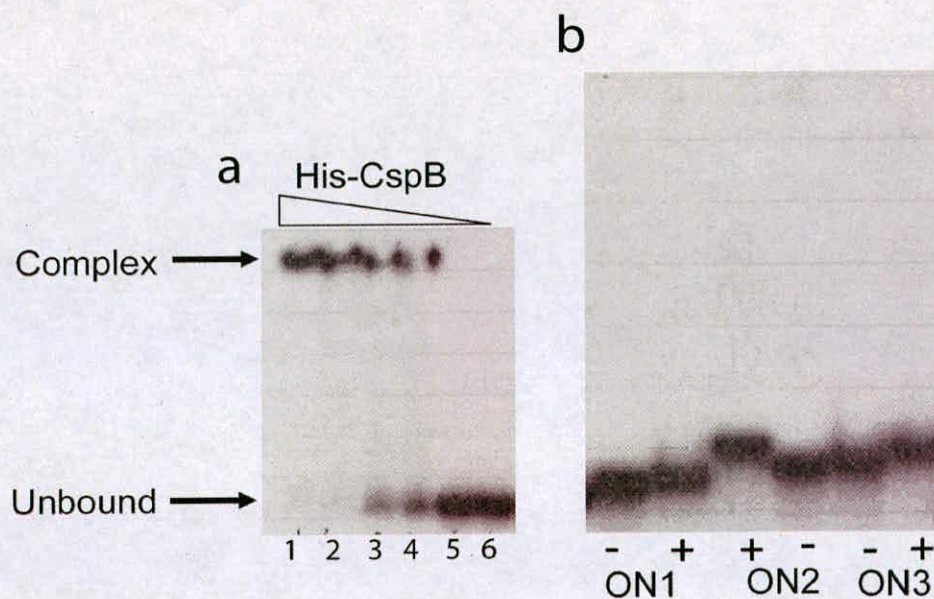
Figure 2-9. **Electrophoretic mobility shift assay of His-CspB and its complex with ON1, ON2, ON3 and ONc.**

(**a**) 15-20 pmol of 5′-end γ-^{32}P-labelled ONc was incubated with 500, 250, 100, 50, 20 and 0 pmol (lanes 1-6) of His-CspB. (**b**) 15-20 pmol of 5'-end γ-^{32}P-labelled ON1, ON2 and ON3 was incubated with (+) and without (-) 250 pmol of His$_6$-CspB. All reaction mixtures were incubated for 10 min at 4°C. Protein-DNA complexes were separated on a 20% native polyacrylamide gel. Gel b has been enlarged 2x to highlight the shifted bands.

Figure 2-9b shows a gel-shift experiment performed for ON1, ON2 and ON3 in the presence (+) and absence (-) of His$_6$-CspB. For ON1, ON2 and ON3 a slight shift can be seen when comparing unbound ss DNA (-) to the relevant protein-ss DNA complexes formed (+). The shifts are minute in comparison to the shift seen for the control oligonucleotide, ONc (Figure 2-9, lanes 1-6), as only a single monomer binds to ON's 1, 2 and 3 (bind the ATTGG site only) and three to four monomers of protein have been shown to bind to ONc [130]. The binding of a single monomer of CspB to a ON, which is similar in size to ONc, did not produce a shifted band one third in size (Figure 2-9a lane 1 *versus* Figure 2-9b lane ON1+). A possible reasons

for the difference in the size of the shift seen is that CspB is binding to the 3′ end of ON's 1, 2 and 3 and this is affecting how the complex migrates through the gel. The similarity of the shifts seen for ON1, ON2 and ON3 would suggest that flanking bases at the 3′ end are not required and a lack of them does not seem to have an effect on the binding of His$_6$-CspB.

2.5.4 *Cy5 labelled SsoSSB binds ss DNA.*

Concern developed during the course of this work concerning weak and non-specific binding of *Bs*CspB to ss DNA sequences on the surface of the array, which was later confirmed (compare Figure 2-13b with c). To eliminate weak binding of *Bs*CspB a competitor protein was used. The candidate competitor protein chosen was a single-stranded DNA-binding protein from *Sulfolobus solfataricus* (*Sso*SSB). Further experiments confirmed (sections 2.5.4-2.5.7) *Sso*SSB use as a suitable competitor protein.

*Sso*SSB was covalently labelled with the mono-functional dye Cyanine 5 (Cy5; Amersham), as described (Materials and Methods - 2.4.6). Analysis of labelled *Sso*SSB-Cy5 using MALDI-Tof mass spectrometry indicated that the majority contained two dye moieties per protein molecule (*Sso*SSB = 16184 Da, 2xCy5 labelled *Sso*SSB = 17462 Da). EMSA were conducted to verify that *Sso*SSB retained its DNA binding activity subsequent to labelling with Cy5. A single stranded 25-mer (ONc:5′–dATCCTACTGATTGGCCAAGGTGCTG-3′), labelled at the 5′-end with γ-^{32}P, was used to compare the binding affinity of unlabelled and Cy5-labelled *Sso*SSB. Figure 2-10 shows a gel-shift experiment performed with ONc in the presence of increasing amounts of unlabelled (lanes 1-4) or labelled (lanes 5-8)

protein. The similarity in the EMSA for both unlabelled and Cy5 labelled protein suggests that labelling did not significantly affect the binding affinity of the protein. The double banding seen (Figure 2-10) may be a result of the gradual dissociation of labelled DNA from SsoSSB. The fact that previous SsoSSB-ss DNA binding studies using ITC have shown that SsoSSB bound to ss DNA with a ratio of 5 nt/monomer [43] suggests that the double banding seen in the gel is probably a result of the EMSA technique used and the on- off-rates of SsoSSB-ss DNA complex formation.
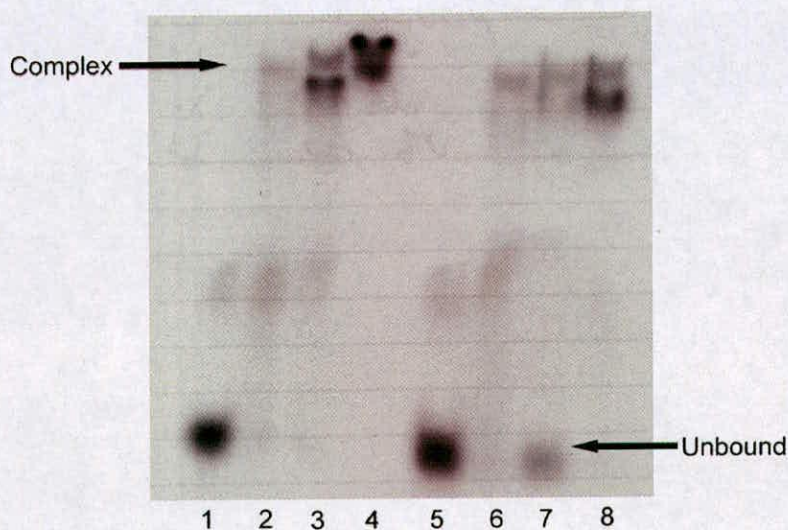
Figure 2-10. **Electrophoretic mobility shift assay of SsoSSB and its complexes with ONc.**

15-20 pmol of 5´-end γ-^{32}P-labelled ONc was mixed with increasing amounts of Cy5-labelled (lanes; 6, 7 & 8) and unlabeled (lanes; 2, 3 & 4) SsoSSB and incubated for 10 minutes at 4°C. Amounts of protein: lanes 2 & 6, 60 pmol; lanes 3 & 7, 120 pmol; lanes 4 & 8, 300 pmol. Lanes 1 and 5 were controls and contained 15-20 pmol of 5´-end γ-^{32}P-labelled ONc only. Protein-DNA complexes were separated on a 10% native polyacrylamide gel.

66

2.5.5 Complex formation monitored by Cy5 fluorescence

A general concern was that the shifted bands (Figure 2-10, lanes 6-8) for the labelled protein may be due to a small percentage (~10%) of unlabelled protein present in the sample, for that reason the gel-shift assay was repeated with γ-^{32}P-labeled ONc and non-radioactive ONc. The radioactive complex serves as a visible marker for the excision of the non-radioactive complex in the adjacent lane. The non-radioactive SsoSSB-ssDNA (ONc) complex was excised from the native polyacrylamide gel and isolated as described (Materials and Methods section 2.4.12). The fluorescence intensity of the isolated protein-ssDNA complex was then measured using a spectrofluorometer, which allowed for the detection of SsoSSB-Cy5. Figure 2-11 shows a peak in the spectrum at 665 nm when the protein solution was excited at 647 nm, demonstrating that Cy5 is present and consequently SsoSSB-Cy5 must be binding ss DNA. The SsoSSB-Cy5 protein had clearly retained its capability to bind ss DNA and was carried forward for use in microarray experiments.
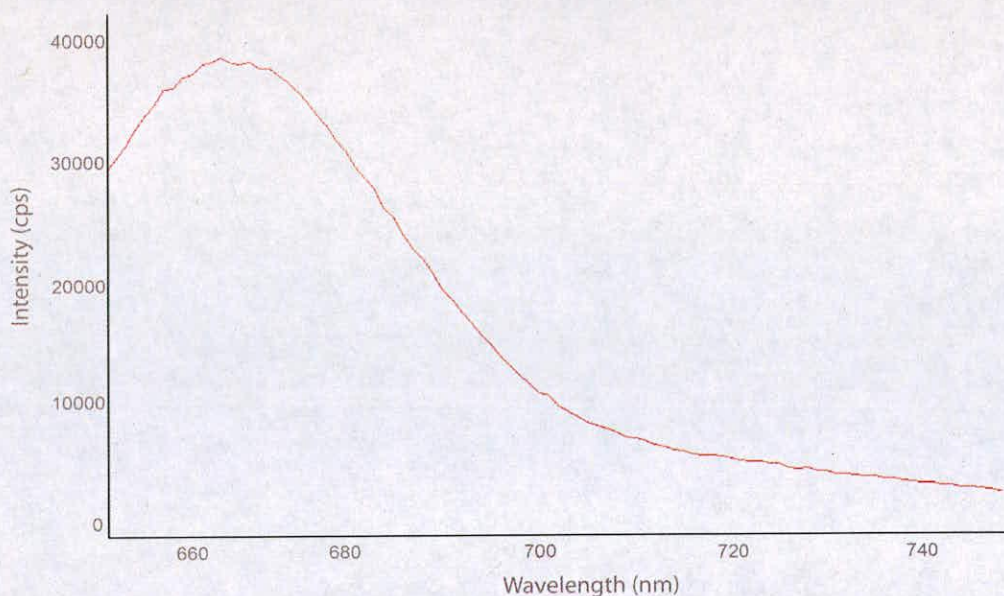
Figure 2-11. **The emission spectrum recorded for *Sso*SSB-Cy5.**

The excitation (slit size =1) and emission (slit size = 1) wavelengths were 647 and 665 nm, respectively.

2.5.6 *SsoSSB-Cy5 binds non-specifically to all ss DNA sequences.*

A generic chip (Figure 2-1) was constructed, which contained all possible 4096 ss hexadeoxynucleotide sequences found in DNA and incorporated into a general construct, 5′-NH-A_{10}-N_9-X_6-3′ (X=G, C, A, or T and the stretch of 9 Ns is composed of random bases). The binding of *Sso*SSB-Cy5 to the generic chip was analysed and all the spots on the array fluoresced with similar intensities (Figure 2-1c), consistent with non-specific binding of *Sso*SSB-Cy5 [43].

2.5.7 *CspB is competitive with SsoSSB.*

A competitive EMSA was used to show that His_6-CspB binds more strongly than *Sso*SSB to the previously reported high affinity Y-box binding motif [119] (ATTGG). A 25-mer, ON1 (5′ –dA_{19}-**GATTGG**-3′), which contains the Y-box binding motif, was labelled at the 5′-end with γ-^{32}P. ON1 is similar in composition to

68

the oligonucleotides found on the generic chip. Figure 2-12 shows the result of a gel-shift experiment performed with ON1 in the presence of varying amounts of His_6-CspB and SsoSSB proteins. An intermediate band can be seen in lane 5, corresponding to the formation of the His_6-CspB-SsoSSB-ON1 complex. This band occurs only when the SsoSSB/CspB proteins are in a molar ratio of approximately 1:1. The degree of migration through the gel of protein ss DNA complexes relates not only to molecular weight but also the combined charge of His_6-CspB (theoretical pI 5.84; negatively charged during gel running conditions of pH 8.0), SsoSSB (theoretical pI 9.16; positively charged) and the additional negative charge of the ss DNA. The more negatively charged the ss DNA/protein complex the faster it will migrate to the cathode and the resulting band will move further into the gel. The intermediate band seen in lane 5 must therefore be the result the combined charge of both His_6-CspB and SsoSSB molecules binding ON1.
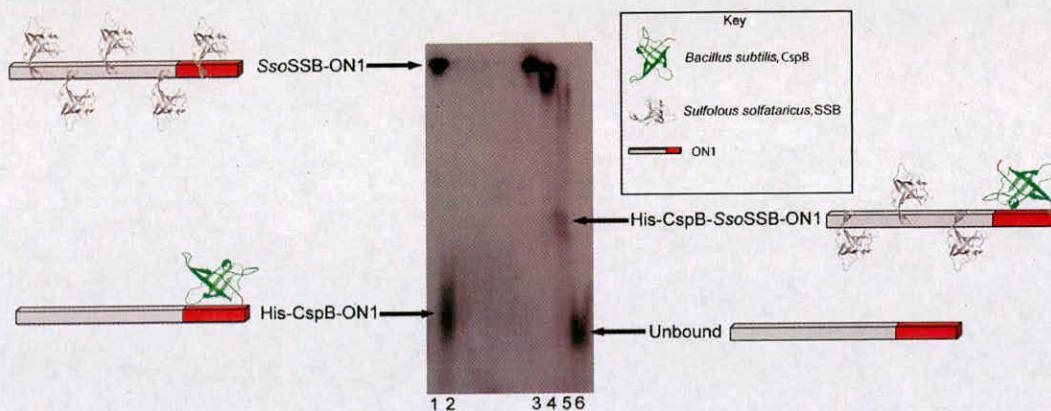
Figure 2-12. **Electrophoretic mobility shift assay of *Sso*SSB, CspB and the complexes formed with γ-^{32}P-labelled ON1.**

15-20 pmol of 5′-end γ-^{32}P-labelled ON1 was used at all times. Lanes; (1) 300 pmol *Sso*SSB; (2) 500 pmol CspB; (3) 300 pmol *Sso*SSB & 100 pmol CspB; (4) 300 pmol *Sso*SSB & 200 pmol CspB; (5) 300 pmol *Sso*SSB & 300 pmol CspB; (6) Unbound ON1 probe. Protein-DNA complexes were separated on a 10% native polyacrylamide gel. Diffuse bands are a common result of low affinity interactions ($K_d > \sim1$ μM). ON1: 5′-dAAAAAAAAAA-AAAAAAAAA-GATTGG-3′

2.5.8 *The competitive binding assay can be transferred to microarray format.*

The binding of His$_6$-CspB to the generic chip was analysed. For explanation purposes only a schematic representation of the assay is shown in Figure 2-13a. The fluorescent signal from His$_6$-CspB bound to the chip was recorded (Figure 2-13b). About 20% of the spots had signals greater than threshold level (which was set at 40% of the maximum fluorescent signal). To eliminate weakly bound CspB, an equimolar mixture (as determined by gel-shift, Figure 2-12) of a competitor *Sso*SSB, and His$_6$-CspB was incubated with an oligonucleotide chip and bound His$_6$-CspB was detected (Figure 2-13b). High intensity fluorescence spots were observed in repeatable patterns on the arrays, which is indicative of selective binding affinity [169].
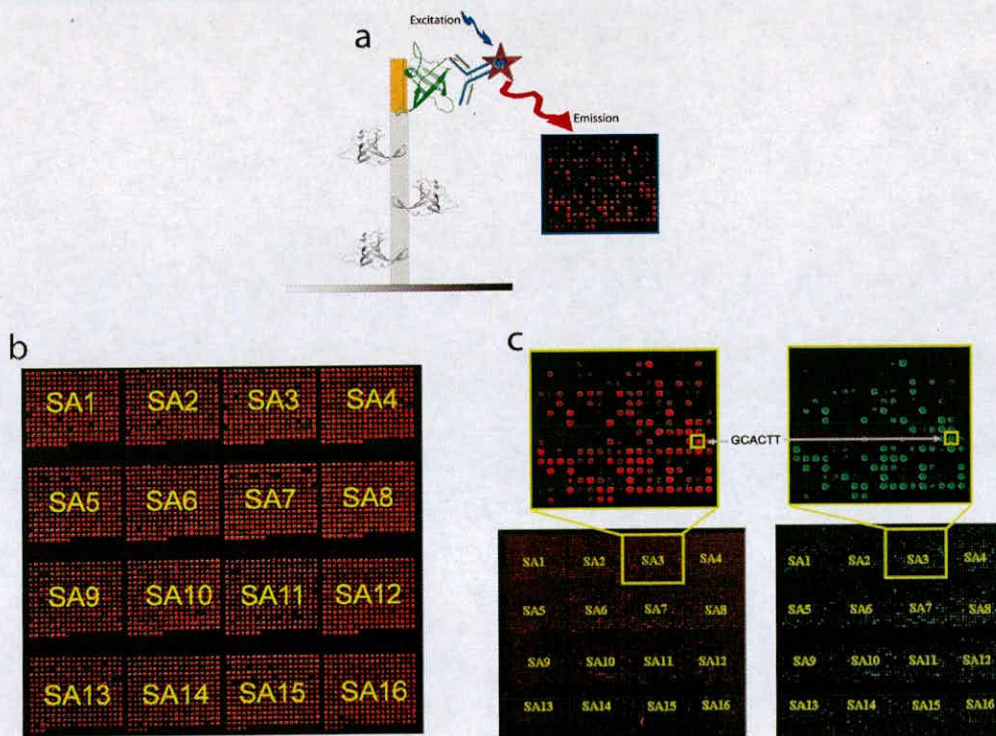
Figure 2-13. **Competitive microarray assay.**

(a) A schematic representation of the competitive protein binding assay. A high intensity spot (red) on the array surface corresponds to several monomers of SsoSSB binding to the large spacer region on the array (Hetero/Homopolymer) and 1/2 monomers of His_6-CspB binding with high affinity to a specific 6-mer sequence. A low intensity or black spot on the array surface corresponds to several monomers of SsoSSB binding the full length of the 25-mer oligonucleotide (ON: 5′-AAAAAAAAAA-NNNNNNNNN-XXXXXX). SsoSSB protein is shown in grey, His_6-CspB is shown in green and the alexa 532-conjugated polyclonal antibody to His_5 is shown in blue. **(b) His_6-CspB bound to microarray (sub-arrays (SA) 1-16).** The fluorescence intensities of the spots are shown in red and the brighter the red colouring for a spot, the higher the intensity. Black spots are indicative of no signal intensity. An analysis of intensities showed that about 20% of the spots had signals greater than threshold level (which was arbitrarily set at 40% of the maximum fluorescent signal). **(c)** Competitive microarray assay. Two individual microarray results for the competitive protein binding assay between His_6-CspB (specific binding protein) and SsoSSB (non-specific binding protein). The fluorescent signal is a result of His_6-CspB being detected on the surface using alexa 532-conjugated polyclonal antibody to His_5 (Fig. 2-12a). Microarray images shown here were results from two different batches of arrays and proteins. Sub-array three from each competitive protein binding experiments has been enlarged to highlight the repeated pattern seen on all arrays. The oligonucleotide, GCACTT, appears as a bright spot on both arrays. The microarrays have been artificially coloured as a visual aid only. *Abbreviations, SA = Sub-array*

2.5.9   High affinity motifs identified by microarray analysis, are validated by electrophoretic mobility shift assay.

An EMSA was carried out to confirm the binding-site data generated from the oligonucleotide chip analysis. The high affinity binding motif, GCACTT, was chosen from the data (Figure 2-13c) to examine if the His_6-CspB could compete successfully with the SsoSSB protein for this binding site. A 25-mer, ON-Microarray test (ON-Mt:5′–dAAAAAAAAAA-GCACTT-AAAAAAAAA-3′),   containing   the   high affinity binding motif was labelled at the 5′-end with γ-^{32}P.

Figure 2-14 shows a gel-shift experiment performed with ON-Mt in the presence of varying amounts of His_6-CspB and SsoSSB proteins. An intermediate band (His-Csp-SsoSSB-ONc complex) can be seen in lanes 3 to 11 for ON-MT, indicating that the CspB protein competed with SsoSSB for binding to the (microarray determined) high affinity binding site, GCACTT.
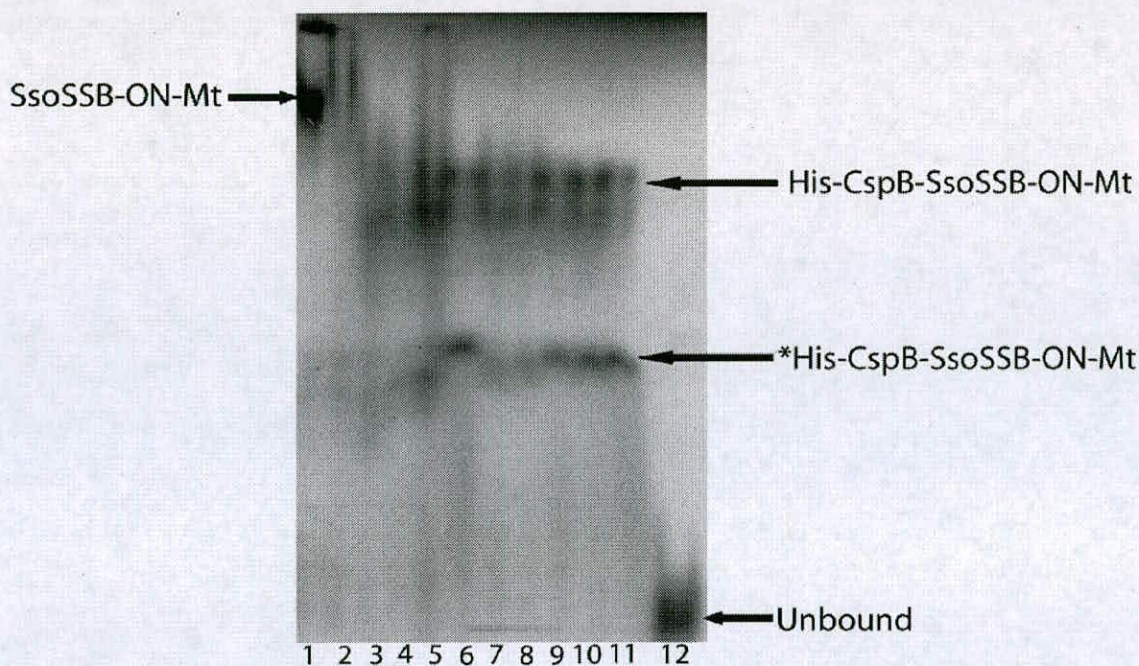
Figure 2-14. **Electrophoretic mobility shift assay of *Sso*SSB, CspB and the complexes formed with ON-Microarray test (ON-Mt).**

15-20 pmol of 5'-end γ-^{32}P-labelled ON-Mt was used at all times. Lanes 1-11 contain a fixed amount of *Sso*SSB (200 pmol). Lanes; (1) 100 pmol His$_6$-CspB; (2) 200 pmol His$_6$-CspB; (3) 300 pmol His$_6$-CspB; (4) 400 pmol His$_6$-CspB; (5) 500 pmol His$_6$-CspB; (6) 600 pmol His$_6$-CspB; (7) 700 pmol His$_6$-CspB; (8) 800 pmol His$_6$-CspB; (9) 900 pmol His$_6$-CspB; (10) 1000 pmol His$_6$-CspB; (11) 1100 pmol His$_6$-CspB; (12) Unbound ON-Mt probe. All 20 µl reaction mixtures were incubated for 20 min at 4°C. Protein-DNA complexes were separated on a 10% native polyacrylamide gel at 4°C.

*The band is seen as a result of increasing concentrations of His$_6$-CspB and is likely due to additional His$_6$-CspB molecules binding to ON-Mt which results in more *Sso*SSB molecules being displaced.

2.5.10 *Identification of the consensus high affinity binding sequence, 5´-GTCTTTG/T-3´, for CspB.*

The highest intensity spots identified from the microarray analysis indicate that the strongest CspB binding sites are pyrimidine rich (Figure 2-15a). The high incidence of thymine bases within the high-affinity 6-mer sequences agrees with previous reports that CspB has a preference for T rich stretches of ss DNA [49]. The presence of a stretch of ten adenines in the linker region of each oligo (Figure 2-1) is likely to

lead to hairpin formation for T-rich sequences and may be expected to down weight

the occurrence of poly T sequences. There is indeed a low intensity for TTTTTN

sequences, which may in part be caused by the formation of such hairpins. Despite

this effect, the averaging procedure still generates a T-rich consensus sequence. The

standard motif alignment method (Genedoc) was used to align the resulting top fifty

(as described in materials and methods - 2.4.10) high-affinity CspB binding

sequences (Figure 2-15b). A sequence alignment window ten bases in length was

used; only seven (coloured columns) out of those ten positions are significantly

(>40%) populated. Analysis of the relative distribution of each base within this

proposed heptanucleotide binding site gives a CspB consensus binding sequence of

5′-GTCTTTG/T-3′ (Figure 2-15e & f). Analysis of the microarray binding results

indicates that CspB can accommodate the binding of a heptanucleotide with a strong

binding preference for cytosine at position 3 and thymine at positions 2, 4 and 6. This

is in agreement with another recent study [14], where the sequence-specific binding

of heptapyrimidines to CspB was analysed by tryptophan fluorescence quenching

experiments. Interestingly the microarray results show that neither the Y-Box

recognition motif [86], 5′-ATTGG-3′, nor its reverse complement 5′-CCAAT-3′,

bind strongly to CspB. The ATTGG sequence has recently been shown to bind with a

low affinity for CspB at 15°C (K_d = 5.3 μM [170]), which is similar to the results

described here for CspB binding to sequences containing ATTGG at 4°C.

Figure 2-15. **Analysis of CspB binding sites.**

(**a**) Fifty highest affinity 6-mer binding sequences for His$_6$-CspB. The chosen sequences are ranked into four groups, according to fluorescence intensity, and coloured accordingly; decreasing intensity from red>pink>blue>green. (**b**) Multiple sequence alignment of the top fifty 6-mer binding sequences for His-CspB. The overall length of the sequence alignment is numbered 1-10 and positions 3-9 are coloured, as they are occupied by a meaningful number of bases (>40%). (**c**) The absolute base distribution for His$_6$-CspB obtained from alignment of all 50 sequences shown in Figure 9B. (**d**) The relative base distribution for His$_6$-CspB. (**e**) 3-D bar chart showing the relative base distribution for each base within the proposed heptanucleotide binding site. (**f**) The sequence logo (generated using the EnoLOGOS program), corresponding to the frequency of each base within the heptanucleotide binding site.

2.5.11 Verification of DNA microarray results using Isothermal Titration Calorimetry (ITC)

The interactions of His_6-CspB with both variants of the preferential binding sequences (ITC-1 and ITC-2), a positive (ITC3 [14]) and negative control ($ITC_{control}$ [130]) (see Figure 2-16b for details of sequences) were assayed by ITC to determine solution-based equilibrium constants. All ITC assays were repeated twice for all oligonucleotides examined. ITC directly measures changes in heat during complex formation. The calorimetric analysis of His_6-CspE binding to ss DNA is shown in Figure 2-16. The interactions of His_6-CspB with all oligonucleotides were exothermic, releasing heat during complex formation. The ITC assay for the His_6-CspB/ss DNA interactions yielded equilibrium dissociation constants of 30 ± 2 nM for ITC1, 44 ± 4 nM for ITC2 and 43 ± 2 nM for ITC3 and all with stoichiometries of approximately 1:1 (Figure 2-16b).

There is an order of magnitude difference in the K_ds for the oligonucleotides described here and the K_ds described previously [14], for similar/identical oligonucleotides. This is likely due to a difference in temperature, buffer conditions and the method used (tryptophan fluorescence quenching). The control sequence (TTCTTTT-ITC3), used here and in the previous study allows us to compare and scale the results from the two methods.

Thus, both the ITC and EMSA results confirm that the microarray assay does indeed select and identify tight binding sequences. This screening procedure could therefore be used as a general method for the rapid identification of high-affinity binding sites for SNABPs
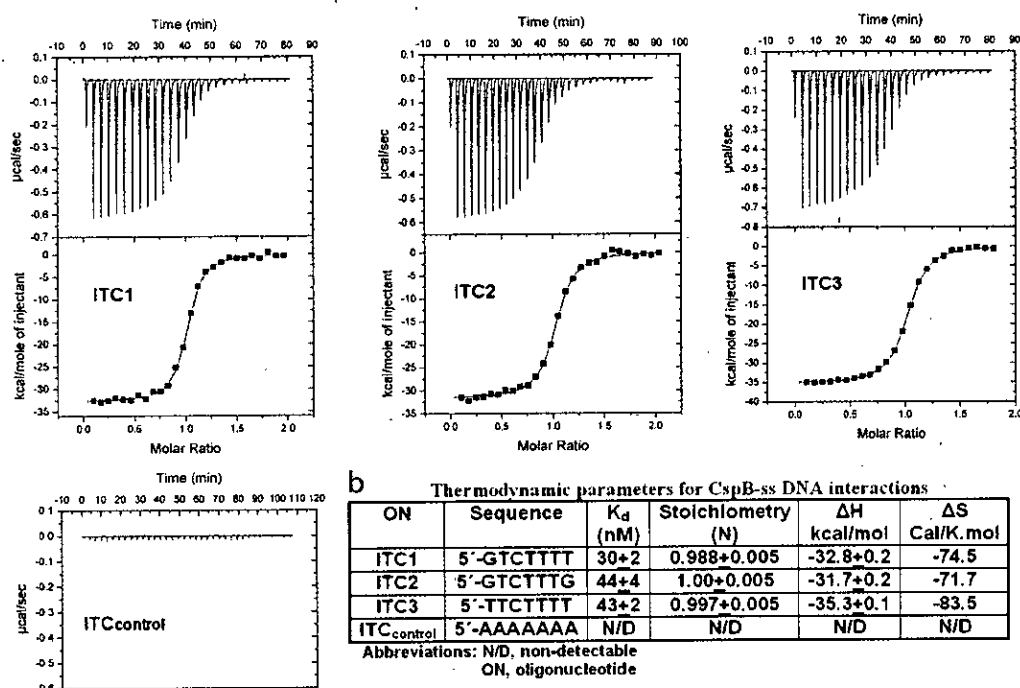
Figure 2-16. **Isothermal titration calorimetry for ss DNA binding to recombinant His₆-CspB.**

Each exothermic heat pulse (upper panel) corresponds to an injection of 5 µl of oligonucleotide (100 µM) into the cell containing 5 µM His₆-CspB at 28°C.

Integrated heat data (lower panels) constitutes a differential binding curve, which was fitted to a single-site binding model to give, the stoichiometry of binding (N), binding affinity (K_d) and enthalpy of binding (ΔH) for each oligonucleotide (Table b). The heat of the reaction was obtained by integrating the peak after each injection of ss DNA oligonucleotide and correcting for the heats of dilution using the Origin® V.7 software. The resulting curve (lower panel) was fitted to a single class binding site model by non-linear least square analysis. All data have been corrected for heats of dilution. **(b)** Thermodynamic parameters for CspB-ss DNA interactions are shown in the table. Non-detectable = $K_d \geq 5$ µM.

2.5.12 CspB-cytosine model explains the preference for a cytosine at position 3.

The X-ray structure for CspB has been reported previously [168]. Figure 2-17 shows the positively-charged face of CspB, highlighting amino acid residues known to be involved in DNA binding [133]. Molecular modelling of CspB with the consensus

oligonucleotide ITC1, using the programme WITNOTP , identified a pocket in the centre of the DNA binding face, which provides an ideal shape and hydrogen-bonding complementarity for binding cytosine. Three hydrogen bonds are formed between the docked cytosine base and amino acid side chains Ser31, His29 and the backbone oxygen of Phe27 (Figure 2-17), providing specificity for cytosine over the other bases.

The recent published structure of CspB in complex with hexathymidine [14] shows that the ligand binds to two protein molecules. The nucleobase of T2, T3 and T4 make contact with one protein molecule, T5 bridges between protein molecules and T6 binds to the next protein molecule. The contacts made by hexathymidine provide the necessary scaffold for the complex to crystallize but the sequence fails to bind across the face of the CspB protein, as this sequence lacks the key cytosine nucleobase at position 3 [14], which here we have shown is required for optimum ss DNA docking (Figure 2-16 and Figure 2-17).
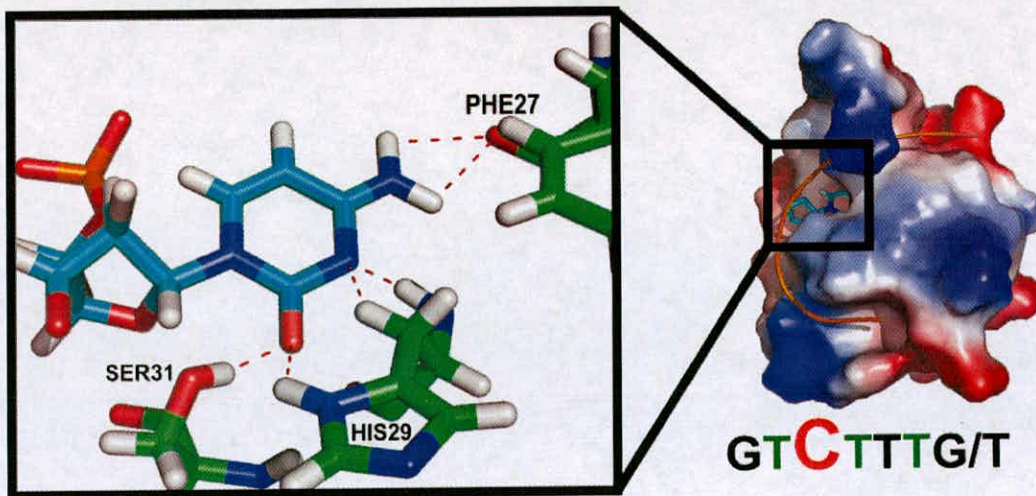
Figure 2-17. **A molecular model of a potential cytosine-binding pocket on CspB.** The electrostatic surface potential of CspB and the backbone trace of the bound consensus heptanucleotide and the hydrogen bonds (depicted as dotted lines) between CspB and a cytosine nucleobase in the proposed binding pocket.

2.6 Conclusions

SNABPs have been reported to activate transcription by binding to a specific recognition sequence upstream [171] or within [172] a promoter, resulting in activation or repression of transcription. Here the major cold shock protein, CspB, of *Bacillus subtilis,* which is capable of binding single stranded nucleic acids [130] and affects expression of over 100 genes under cold shock conditions [173], was used as the test protein for the development of a novel DNA microarray binding assay [174]. Database analysis of the *B. subtilis* genome reveals that 89 copies of the consensus binding sequence (5′-GTCTTTT-3′) exist within potential SNABP promoter regions (up to 100 bases upstream of the ATG start for all genes), of which only 24 have an assigned function. The use of an unbiased genomic assay to identify optimal binding sequences for SNABPs *in vitro,* may provide insight into their role in regulating

cellular functions. The full implications of these sequences on gene expression and binding of CspB remains to be resolved.

CHAPTER 3. Cloning, expression and initial characterisation of a family of six CSP paralogues from *Salmonella typhimurium*.

3.1 Aim

To clone, express and purify each of the six CSPs from *S. typhimurium* and to determine the sequence specificity of ss DNA recognition using the previously developed microarray assay. The similarities or differences observed in ss DNA recognition sequences for each CSP will serve to highlight potential functions for each member of this CSP family.

3.2 Summary

This chapter details the cloning, expression, purification and initial ss DNA binding characterisation of each member of the *S. typhimurium* CSP family. Large amounts of pure protein (mg/ml) were produced using bacterial expression systems. Purified proteins were then used for ss DNA binding experiments, biophysical characterisation (Chapter 4) and structural studies (Chapter 5). Results are summarised in Table 3-1.

Table 3-1. Stages of results completed for each of the *S. typhimurium* CSPs.

Protein	Cloned	Expressed	Purified	ss DNA specificity	Consensus binding Seq.	Crystallisation Trails	Crystals obtained
*St*CspA					GT-T/C-T/C-TTC		
*St*CspB					G-T/C-T-C/G-TTT		
*St*CspC					TCTTT-T/C		
*St*CspD					ACGGGG		
*St*CspE					GTCTTTT		
*St*CspH							

Completed stages are coloured green and uncompleted are coloured red. Sequence specificity data were generated from the microarray assay described in chapter 2.

3.3 Introduction

As discussed in detail in chapter one, CSPs have been shown to bind to stretches of ss DNA (6-7 nucleotides) [28, 49, 130] with high affinity and a high degree of specificity [14, 25, 86, 130, 174] but also in a sequence unspecific manner [29]. This ability to bind nucleic acids led to the conclusion that CSPs stabilise mRNA secondary structures [29] induced by low temperatures [175] and resulted in their classification as RNA chaperones. Controversially, CSPs have been shown to influence transcription and translation *in vitro* [91] and act on bulk protein expression not as RNA chaperones but by inhibiting their transcription and translation by unspecific binding to nucleic acids.

The CSD containing proteins (Y-Box factors) found in eukaryotic organisms share high sequence similarity to CSPs [126], and show even greater structural resemblance [176], with the distinctive DNA binding framework (RNP1 and RNP2) [128] present. However, the Y-Box proteins have been allocated a diverse set of functions compared to CSPs, which includes their ability to regulate transcription and translation, both positively and negatively [7]. The differences reported between the function of CSPs and their homologous eukaryotic equivalents led to the hypothesis that there may be different functions amongst CSPs from the same family, which have yet to be identified.

In *Salmonella typhimurium*, six CSPs (*St*CspA, *St*CspB, *St*CspC, *St*CspD, *St*CspE, *St*CspH) have been identified and the cold inducibility of *St*CspA, *St*CspB and *St*CspH has been reported [157-160], although their functions are yet to be clearly elucidated. The complete ss DNA binding characterisation of this family may highlight differences/similarities in DNA binding, corresponding to the potential

functions of each CSP. Before the characterisation of this family of proteins can be preformed, pure samples of each CSP must be obtained.

3.3.1 Protein production and purification strategy

Recombinant proteins were expressed using a standard expression system, which can produce many copies of the desired protein within the host cell. This is accomplished by creating and inserting an expression vector into the host cell. Expression vectors contain all the necessary genetic coding to produce the protein, including a promoter, a transcription terminator sequence (e.g. TAG), a ribosome binding site and an antibiotic resistance gene to allow for plasmid maintenance.

The pET expression system was used in this work to overproduce each of the *S. typhimurium* CSP. This system makes use of an extremely strong bacteriophage T7 promoter, which is highly selective for T7 RNA polymerase. pET expression vectors contain a T7 promoter upstream of the multiple cloning site (protein expression site), which is tightly controlled by the lac operator. The *CSP* gene was cloned into the expression vector and inserted into a genetically modified bacterial host cell, which expresses T7 RNA polymerase (e.g. *E. coli* BL21(DE3)). Expression of T7 polymerase in the host cell is then induced by the addition of a chemical, isopropyl-β-D-thiogalactoside (IPTG). T7 polymerase binds to the T7 promoter on the expression vector, resulting in high levels of expression of the *CSP* gene downstream.

The pET expression vector contains a hexa-histidine (His$_6$) fusion tag which is added to the N-terminal of the CSP to facilitate purification of the recombinant protein. His-tagged CSPs were expressed and the harvested cells were lysed to release the proteins from the cytosol. The affinity of histidine residues for nickel [177] allowed

for the selective purification of His$_6$CSPs using a Ni-NTA agarose column. The procedure for the purification of His$_6$CSPs consisted of four general steps, cell lysis, binding to nickel, washing (removes unbound proteins) and elution of the bound His$_6$CSP using an elevated concentration of imidazole at a constant pH.

It can be difficult to completely remove contaminating nucleic acids during the purification of ss DNA binding proteins and as a result elaborate purification protocols were developed [178, 179] and expensive affinity resins have been used [180, 181]. An advantage of using a histidine fusion tag is the ability to purify proteins under denaturing conditions, which allows for the purification of fully denatured His-tagged proteins. In this work His$_6$CSPs were purified under denaturing conditions, which completely removed all contaminating nucleic acids. The development of refolding techniques for His$_6$CSPs, based on the available methods [182], allowed for the production of active His$_6$CSPs.

3.3.2 Overview: Refolding proteins

Initially proteins are incubated in a buffer containing a strong denaturant (typically 2-8 M urea or 3-8 M guanidine hydrochloride) and a reducing agent (dependant on the presence of cysteine residues), which unfolds the protein. Refolding of the proteins is then initiated by the removal of the denaturant. The efficiency of protein refolding depends on the competition between correct folding and aggregation. Aggregation results when unfolded protein molecules have their hydrophobic core exposed to solvent, and an aggregate would result from interactions between hydrophobic regions of different polypeptide chains. Aggregates can be minimised by refolding proteins at a low concentration range (10-100 µg/ml) [183] but refolding conditions need to be optimised for each individual protein (buffer conditions (pH and ionic

strength), temperature and many different additives in various mixes). Many different methods for refolding proteins have been described [182], which include dialysis, slow dilution, rapid dilution, pulse renaturation and size-exclusion chromatography. Out of all these methods size-exclusion chromatography has been shown to produce some of the highest levels of refolded active proteins in a single step [184]. During protein refolding, using size-exclusion chromatography, the denaturant is removed while the protein slowly migrates through the column. Protein molecules passing through the size-exclusion chromatography column are partitioned in the gel media and are dynamically trapped and separated from each other, greatly reducing the contacts made between protein molecules. This separation of protein molecules from each other has been observed to greatly suppress aggregation during refolding [185].

3.4 Materials and Methods

3.4.1 Cloning

cDNA corresponding to all six (*St*CspA-H) full length CSPs were generated by PCR using *S. typhimurium* strain 1344 chromosomal DNA as a template. Sequences were amplified using high-fidelity Taq polymerase (Roche) and the corresponding forward and reverse primers found in Table 3-1. The resulting PCR products were cloned into a pCR2.1 TOPO vector using the TOPO-TA Cloning® kit (Invitrogen) and digested with *NdeI* and *HindIII* (New England Biolabs). The digested inserts were ligated into a *NdeI-HindIII*-digested pET28a vector (Novagen) and verified by DNA sequencing.

Table 3-1 **Primers used to created S.T_CSP clones**

Clone	Primer position	Primer sequence 5´ to 3´
pET28a_S.T _CSPA	Forward Reverse	ACACA *CAT ATG* TCC GGT AAA ATG ACT GGT ATC CA *AAGCTT* TTA CAG GCT GGT TAC GTT GCC AGC
pET28a_S.T _CSPB	Forward Reverse	ACGCA *CAT ATG* ACG ACG AAA ATC ACT GGT TTA CGA *AAGCTT* TTA AAG CGC CAC AAC GTT GAC CGC
pET28a_S.T _CSPC	Forward Reverse	ACGCA *CAT ATG* GCA AAG ATT AAA GGT CAA GTT CGA *AAGCTT* TCA GAT AGC TGT TAC GTT AAC AGC
pET28a_S.T _CSPD	Forward Reverse	ACACA *CAT ATG* GAA ACG GGT ACT GTA AAG TGG CA *AAGCTT* CTA TGC AAC GGC CTC TGC TTC GAT
pET28a_S.T _CSPE	Forward Reverse	AGGCA *CAT ATG* TCT AAG ATT AAA GGT AAC GTT CGA *AAGCTT* TTA CAG AGC AGT TAC GTT TGC AGC
pET28a_S.T _CSPH	Forward Reverse	AGGCA *CAT ATG* TCT CGT AAA ATG ACA GGA ATT CGA *AAGCTT* TTA TGA AAG ATA AAC GTT GGC GGC

HindIII (shown in italics) and *NdeI* (shown in bold italics) restriction sites were added to primers to allow for directional cloning.

3.4.2 Expression of S. typhimurium CSP

Chemically competent Rosetta BL21(DE3)pLysS cells were transformed with the verified constructs and expressed by inoculating 50 ml of Luria-Bertani medium (containing 50 µg/ml kanamycin) and incubated over night at 37°C with agitation. One-litre cultures were grown to an OD_{600} of 0.5-0.7 and expression was induced at

37°C by adding IPTG (isopropyl-β-D-thiogalactosidase) to a final concentration of 1 mM. The incubation was then continued for 4 hours and the cells were harvested at 8,000 rpm in a JLA-9.1000 rotor for 12 minutes at 10°C. Pellets were then frozen in liquid nitrogen and stored at -80°C.

3.4.3   Preparation of cell lysate

Cell pellets were resuspended in lysis buffer (20 mM Tris HCl pH 8.0, 500 mM NaCl, 0.1 mM phenylmethylsulphonyl fluoride (PMSF) and 1 mM EDTA-free protease inhibitors) to a final volume of 30 ml/5 g of cells and lysed by pulsed sonication on ice for 5 minutes. The lysate was centrifuged at 23000 rpm at 10°C for 45 minutes using a JA 25.50 rotor. The supernatant was filtered through a 0.2 μm filter and DNase I [40 μg/ml] and RNase A [10 μg/ml]) were then added and incubated at room temperature for 30 minutes with gentle agitation.

3.4.4   Native purification of recombinant His₆CspE

The supernatant was loaded onto a 12 ml Ni-NTA agarose column at 1 ml.min^{-1}, pre-equilibrated in buffer A (50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole, pH 8.0). The column was washed with eight column volumes of buffer A. To further remove nucleic acids, a gradient (0-30% over five column volumes) of buffer B (10 mM Tris and 1.5 M NaCl, pH 7.6) was used and held at 30% until A_{280} reading returned to baseline. His₆CspE was eluted over a ten-column volume elution gradient with buffer C (50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole, pH 8.0). Fractions containing recombinant protein were pooled and concentrated to ≤1 ml using a Viva Spin column (MWCO = 3.5 kDa) and loaded onto a HiLoad 26/60

Superdex 75 column (Amersham Biosciences), pre-equilibrated in a buffer containing 20 mM Tris-HCl (pH 7.5).

3.4.5 Purification of CSPs (StCspA, StCspB, StCspC, StCspE) by denaturation and refolding using size-exclusion chromatography

The supernatant was loaded onto a 12 ml Ni-NTA agarose column at 1 ml.min^{-1}. The column was washed with eight column volumes of buffer A and the protein was eluted over an eight-column volume elution gradient with buffer C. Eluted fractions were pooled and concentrated using a Viva Spin column (MWCO = 3.5 kDa). Protein purity was excellent but DNA contamination was high (as determined by A_{280}:A_{260} = approx. 1:2). To remove bound nucleic acids 1 ml of His$_6$Csp [5-10 mg/ml] was added to 9 ml of buffer D (8 M urea, 300 mM NaCl and 50 mM Tris, pH 8.0) and incubated at room temperature (RT) overnight. The sample was loaded onto a self-packed 2 ml Ni-NTA gravity column at RT and washed with 10 column volumes of buffer D. Protein was eluted with buffer E (8 M urea, 300 mM NaCl and 50 mM Tris, pH 5.0; made immediately prior to use) by manually collecting 2 ml fractions. Fractions containing high concentrations of His$_6$CSP (fractions 5 and 6 - determined by Bradford assay) were either refolded immediately or stored at RT (free from any light sources) for up to two months before refolding.

The protein was refolded by loading 200 µl of concentrated protein sample onto a HiLoad 26/60 Superdex 75 column (Amersham Biosciences), pre-equilibrated in a buffer E (20 mM HEPES, 150 mM NaCl and 3 mM EDTA (pH 7.4)). The column flow rate was 0.5 ml.min^{-1} for the first 2 ml (to quickly dilute the urea) and then reduced to 0.2 ml.min^{-1} (to allow the protein to refold).

3.4.6 Purification of recombinant His₆CspD

The supernatant was loaded onto a 12 ml Ni-NTA agarose column at 1 ml.min^{-1}, pre-equilibrated in buffer A. The column was washed with eight column volumes of buffer A to remove unbound proteins. To remove nucleic acids, His₆CspD was denatured on the column using a 0-100% gradient of buffer F (6 M Guanidine, 300 mM NaCl and 50 mM Tris pH 8.0) over 10 column volumes. His₆CspD was refolded on the column by running a gradient of buffer G (20 mM HEPES pH8.0, 150 mM NaCl) against buffer F over 10 column volumes. The protein was eluted using an elevated concentration of imidazole (buffer C). Fractions containing recombinant protein were pooled and concentrated to ≤1 ml using a viva spin column (MWCO = 3.5 kDa) and loaded onto a HiLoad 26/60 Superdex 75 column (Amersham Biosciences), pre-equilibrated in a buffer containing 20 mM Tris-HCl (pH 7.6).

3.5   Results and discussion

Section 1: Cloning, expression and purification of *S. typhimurium* cold shock proteins

3.5.1   The identification and sequence analysis of the S. typhimurium cold shock genes

The *B. subtilis* genome sequence has been published [73]. The genome encodes the

well characterised major cold shock gene, *CspB* [86], which encodes a 67 amino acid

product. The *S. typhimurium* (Strain 1344) genome sequence has recently been

completed (Sanger Institute). A BLAST search of the *S. typhimurium* genome

sequence was carried out using CspB as a homology model to identify any possible

cold shock genes. Six cold shock genes were identified, which encode 69-73 amino

acid-products. The CSP sequences share a high degree of homology (Figure 3-1 and

Table 3-2), the highest being 84 % identical between *St*CspE and *St*CspC, the lowest

is 26.3 % between *St*CspH and *St*CspD. *St*CspH has such low sequence identity

(46%) to the other members of the family that its function as a CSP was doubted but

has been confirmed elsewhere [160].

Figure 3-1. Alignment of the amino acid sequences of cold shock proteins from *Salmonella typhimurium*.

Identical residues are shown in black, similar residues in grey and non-identical in white.

Table 3-2. **Sequence identity matrix for *S. typhimurium* CSP**

Seq->	CSPA	CSPB	CSPC	CSPD	CSPE	CSPH
CSPA		64.2 %	67.1 %	43.4 %	68.5 %	48.5 %
CSPB	64.2 %		68.5 %	46.0 %	70.0 %	45.7 %
CSPC	67.1 %	68.5 %		42.6 %	84.0 %	42.8 %
CSPD	43.4 %	46.0 %	42.6 %		42.6 %	26.3 %
CSPE	68.5 %	70.0 %	84.0 %	42.6 %		47.1 %
CSPH	48.5 %	45.7 %	42.8 %	26.3 %	47.1 %	

3.5.2 *Amplification and cloning of the S. typhimurium cold shock genes*

Specific oligonucleotide primers (Material and Methods section 3.4.1) containing *NdeI* and *HindIII* restriction sites were used to amplify the open reading frames of the six *Csp* genes from *S. typhimurium* strain 1344 chromosomal DNA. The amplification products are all approximately 0.2 Kb in size (Figure 3-2), with *CspD* visibly larger, as expected (Figure 3-1).
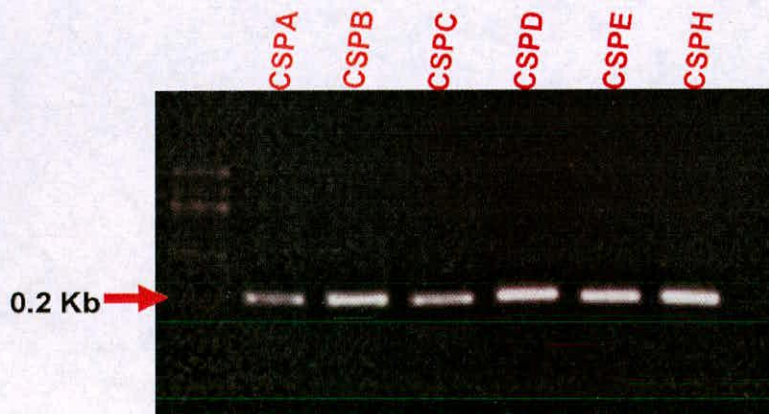
Figure 3-2. **PCR amplification of the *Csp* genes from *S. typhimurium* (strain 1344) chromosomal DNA.**

100-bp DNA ladder used.

The amplification products were ligated into the plasmid pCR®2.1-TOPO® and the resulting plasmids were transformed into chemically competent *E. coli* TOP10 cells. Transformants were isolated on LB agar plates containing X-gal (40 µg/ml) and kanamycin (50 µg/ml). Positive transformants were isolated using standard blue-

white screening. Plasmid DNA was isolated from several transformants and digested with *NdeI* and *HindIII* to confirm the presence of each *csp* gene. The resulting 0.2 Kb restriction fragments were purified from the agarose gel and ligated into *NdeI-HindIII*-digested pET28a vector (Novagen), downstream of the 6xHistidine coding region, and transformed into chemically competent *E. coli* DH5-α cells. Gel electrophoresis showing a restriction digest of plasmid DNA (pET28a_*St*Csp) isolated from the transformants can be seen in Figure 3-3. The correct sequences were verified by sequencing (A1.2) and are now referred to as; pET28a_*St*CspA, pET28a_*St*CspB, pET28a_*St*CspC, pET28a_*St*CspD, pET28a_*St*CspE and pET28a_*St*CspH.
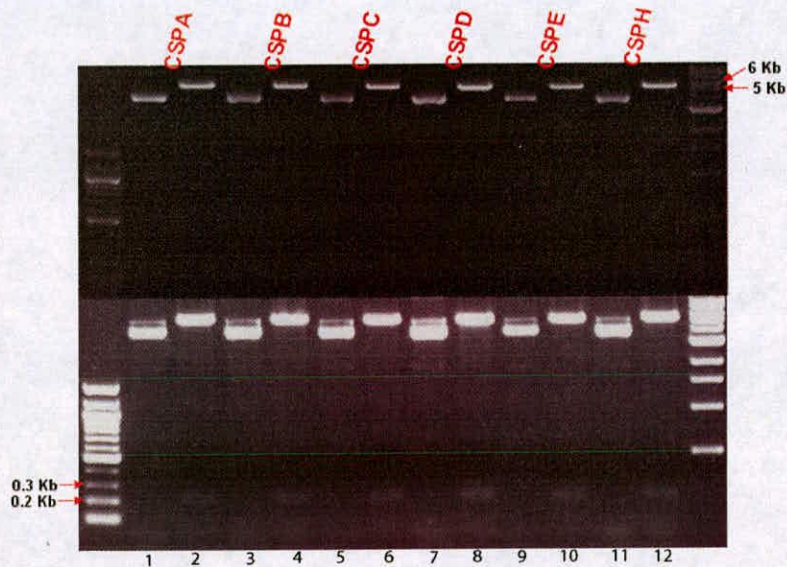
Figure 3-3. **Agarose gel electrophoresis of *NdeI* and *HindIII* digested and undigested plasmid DNA isolated from *E. coli* DH5-α carrying the plasmids pET28a_CSPA-H.**

Lanes: (1, 3, 5, 7, 9 & 11) plasmid DNA isolated from *E. coli* DH5-α containing pET28a_CSPA-H; (2, 4, 6, 8, 10 & 12) The same plasmid DNA digested with *NdeI* and *HindIII*. The resulting 5-Kb and 0.2-Kb indicates that the *StCsp* genes are present. The same gel is shown at two different gain levels to allow for visibility of the 0.2-Kb fragment. 100 bp and 1 Kb MW ladders were used.

3.5.3 Expression of the S. typhimurium cold shock proteins

Rosetta BL21(DE3)pLysS *E. coli* cells were transformed with pET28a_CspA-H and expression of the *Csp* genes was achieved by the T7 promoter-driven expression system in *E. coli,* induced by adding IPTG (1 mM). Cell where grown for a further six hours at 37°C and expression of CSPs was analysed by SDS-PAGE (Figure 3-4).
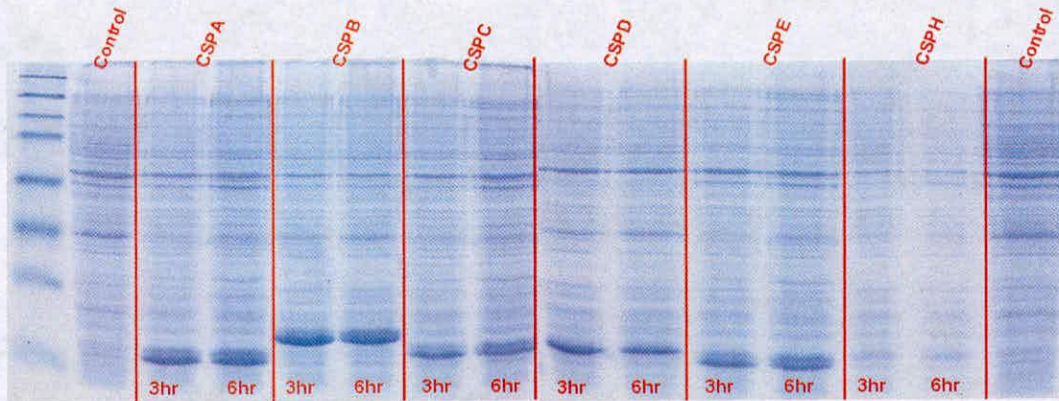
Figure 3-4. Expression of *S. typhimurium* His$_6$CSP at 37°C.

Cell extracts from Rosetta BL21(DE3)pLysS harbouring pET28_CSPA-H. Samples were analysed by 15% SDS-PAGE and visualised by Coomassie staining. Lanes show soluble cell extract after induction at 37°C for 3 and 6 hours for each of the cold shock proteins. His$_6$CspH expression is visible but low. Controls; Left - Rosetta BL21(DE3)pLysS cell extract after 6 hrs of growth. Right - BL21(DE3)pLysS harbouring pET28_CSPA – No induction.

All constructs analysed gave very high levels of soluble protein expression in *E. coli* BL21(DE3) pLysS strain, with the exception of *St*CspH. Poor expression of *St*CspH was a direct result of protein induction by the addition of IPTG. The induction of *St*CspH caused *E. coli* to stop growing. Various expression strains, growth conditions and induction levels were used to try and express *St*CspH but the final result was the same. Although all the His$_6$CSPs are of similar molecular weights (between 9.5-10 kDa) *St*CspB runs at a higher molecular weight on the SDS-PAGE gel (Figure 3-4). All *St*CSPs contain three proline residues (Figure 3-1), except *St*CspB, which has an

additional proline present in its amino acid composition, this minimises the spacing between each of the proline residues. Prolines introduce natural "kinks" in polypeptide chains which are not straightened out by denaturing. The additional proline induced bends in the polypeptide chain of *St*CspB are likely causing it to run slower through SDS-PAGE (for example, ventricular myosin light chain [186]).

3.5.4 Overproduction and native purification of His$_6$CspE

Cell lysate containing His$_6$*St*CspE was bound to a Ni-NTA agarose column and unbound proteins were washed away. The column was washed with a buffer containing 500 mM NaCl, to remove any electrostatically bound nucleic acids (Figure 3-5aI) and bound His$_6$*St*CspE was eluted using an elevated concentration of imidazole at a constant pH to displace the histidine tag from the nickel (Figure 3-5aII). The purified His$_6$*St*CspE migrated as a single protein band on a SDS-15% polyacrylamide gel (Figure 3-5c). Gel filtration with a HiLoad 26/60 Superdex 75 column was performed. Recombinant His$_6$CspE eluted as two peaks and was more than 95% pure as judged by SDS-PAGE (Figure 3-5b and c), with an experimentally determined molecular weight of 9.6 kDa (Figure 3-5d). Peaks 1 and 2 had retention voloumes of 16.7 ml and 18.5 ml, which corresponds to molecual weight standards of ~8 kDa and ~1.5 kDa. His$_6$*St*CspE has a calculated molecular weight of 9.6 kDa, suggesting that the structure of His$_6$*St*CspE from peak 2 (His$_6$*St*CspE in complex with ss DNA) must be elongated (a long and thin cylinder) resulting in a smaller hydrodynamic radius. Only His$_6$CspE from peak 1 could be used for further experiments as protein from peak two was highly contaminated with nucleic acids (Table 3-2). Samples from peaks 1 and 2 were further analysed by native gel electrophoresis (Figure 3-5e). His$_6$CspE from peak 1 (uncontaminated) fails to fully

enter the gel (Figure 3-5e – blue arrow). This is a direct result of the protein's high pI (9.23); the overall positive charge of the protein prevents it from moving towards the positively charged cathode. His$_6$$St$CspE/ss DNA complex has a negative charge caused by the bound DNA and migrates easily towards the cathode (Figure 3-5e). His$_6$$St$CspE from peak 2 (contaminated) migrates into the gel, as a result of bound negatively charged nucleic acids. This confirms that His$_6$$St$CspE from peak 1 was pure and free of nucleic acids.

Table 3-2. **DNA contamination found in eluted peaks**

	A_{260}	A_{280}	A_{595} Bradford	Concentration by A_{280}
PEAK 1	0.25	0.24	0.64 = 0.9 mg/ml	0.8 mg/ml
PEAK 2	1.96	1.2	0.54 = 0.75 mg/ml	2 mg/ml

Extinction coefficient = 5500 M^{-1} cm^{-1}
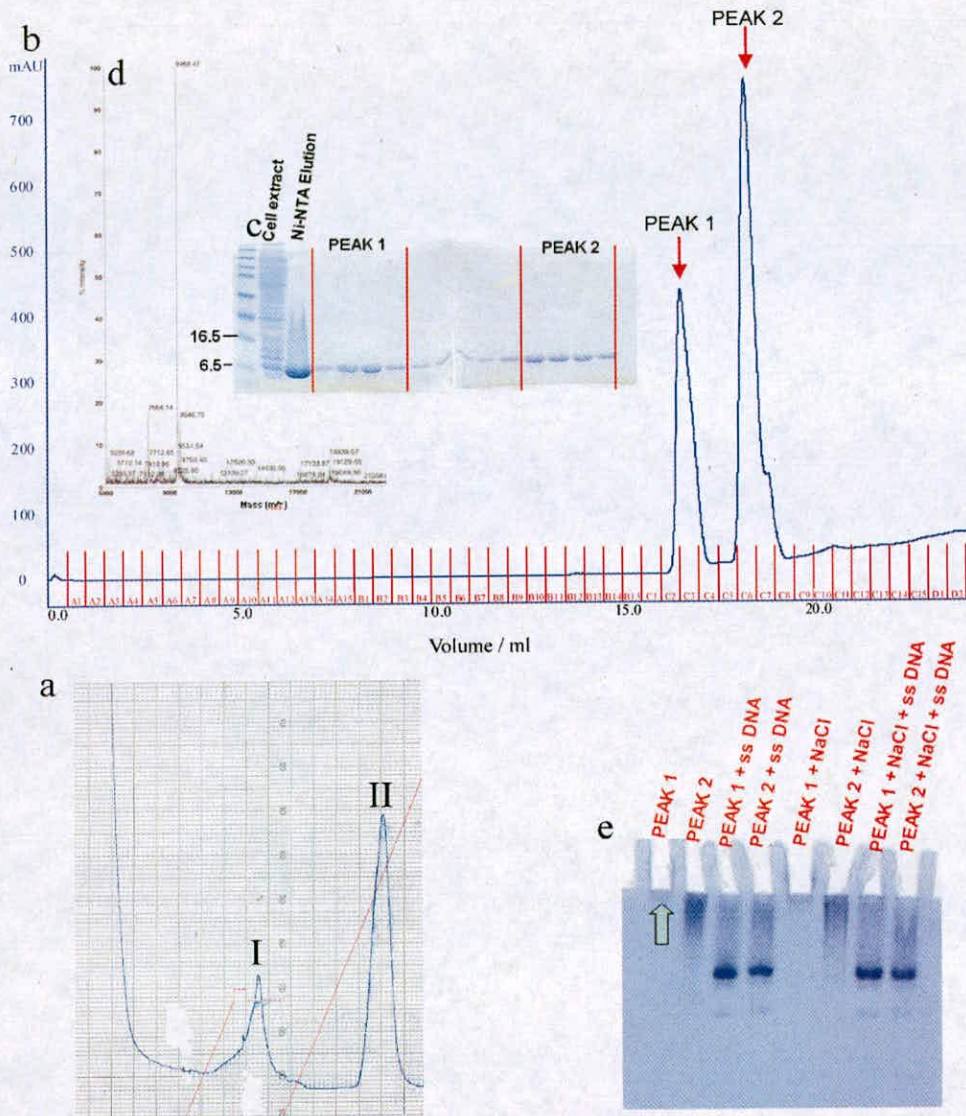
Figure 3-5. **Native purification of His₆*St*CspE.**

(a) A_{280} trace from a single Ni-NTA purification. (I) A_{280} peak corresponding to the removal of electrostatically bound nucleic acids, a result of running a 500 mM NaCl gradient. (II) A_{280} peak corresponding to imidazole-eluted His₆*St*CspE. (b) Gel-filtration with Superdex 75. Two peaks can be seen on the elution profile and the corresponding fractions were analysed by SDS-PAGE (c). (d) MALDI-Tof mass spectrum, observed 9668 Daltons, calculated 9614 Daltons. (e) 5 µg samples (determined by Bradford method-Table 3-2) from peaks 1 & 2 ran on a 5% native gel with and without 200 mM NaCl and/or excess ss DNA binding sequence, 5´-GTCTTTT-3´, as determined for *Bs*CspB (Chapter 2 - Figure 2-15).

96

3.5.5 Purification of CSPs (StCspA, StCspB, StCspC, StCspE) by denaturation and refolding using size-exclusion chromatography

The native purification method used for purifying His$_6$StCspE was used for the purification of the other StCSPs. A typical gel filtration elution profile (Superdex 75) of purified His$_6$StCSPs (Figure 3-6a) eluted as two major peaks, one corresponding to the void volume of the column (7.5 ml) and another broad peak corresponding to the His$_6$StCSP. Although His$_6$StCSPs were more than 95% pure, as judged by SDS-PAGE (Figure 3-6b) they were all highly contaminated with nucleic acids (>80%, as determined by A$_{280}$:A$_{260}$). Many attempts were made to remove bound nucleic acids by adjusting the native purification protocol but were either unsuccessful or produced extremely low amounts of His$_6$StCSPs.
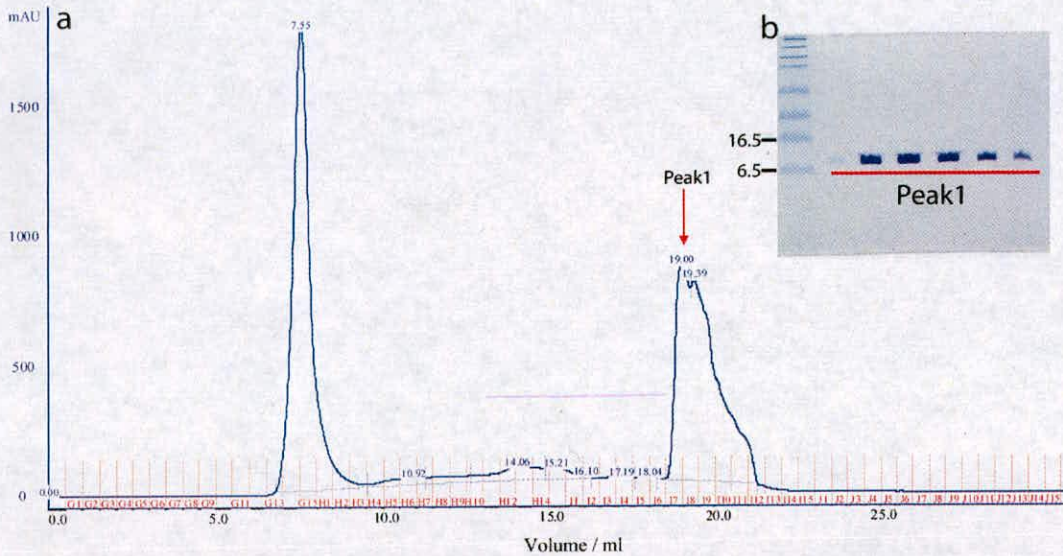
Figure 3-6. **Native purification of StCSPs.**

(a) An example of an elution profile from gel-filtration chromatography using a Superdex-75 HR column for His$_6$StCspA. In addition to the void peak, a single peak can be seen on the elution profile (Peak 1). (b) Fractions from peak 1 analysed by SDS-PAGE.

One method of removing bound nucleic acids was to denature the purified His$_6$CSPs, thus releasing the bound nucleic acids, and then refolding the His$_6$CSP.

3.5.6 Refolding CSPs

A high concentration of fully denatured His$_6$StCSP (~5-10 mg/ml) was bound to a Ni-NTA agarose column and unbound nucleic acids were washed away. Bound His$_6$Csp was eluted using a low pH buffer (pH 4.5). At a low pH histidines are protonated and do not interact with the nickel. Fractions containing high concentrations of His$_6$StCSP were either refolded immediately or could be stored at RT free from any light sources for up to 2 months before refolding. Fully denatured His$_6$StCSP (free of nucleic acid contamination), at a concentration of 1-2 mg/ml, was applied to a HiLoad 26/60 Superdex 75 column, pre-equilibrated with refolding buffer (20 mM HEPES, 150 mM NaCl and 3 mM EDTA (pH 7.4). The column was eluted with refolding buffer and the resulting profiles for each His$_6$StCSP purified using this technique are shown in Figure 3-7, along with the corresponding SDS-PAGE gels. The average protein recovery was ~94% and their molecular weights were confirmed by MALDI-Tof mass spectrometry. For each run a highly symmetrical single peak was observed for His$_6$StCspA, His$_6$StCspB, His$_6$StCspC and His$_6$StCspE. The single symmetrical peaks observed are good indicators that the proteins have correctly refolded. The broad peak observed for His$_6$CspD (Figure 3-7d) is suggestive that the refolding conditions used here did not allow the protein to refold correctly. The similar retention volumes of approximately 13.5 ml (Table 3-3) for His$_6$StCspA, His$_6$StCspB, His$_6$StCspC and His$_6$StCspE, corresponds to a molecular weight standard of ~12.0 kDa (see Figure 5.13 for molecular weight standards), suggesting that all the His$_6$StCSPs are monomeric.
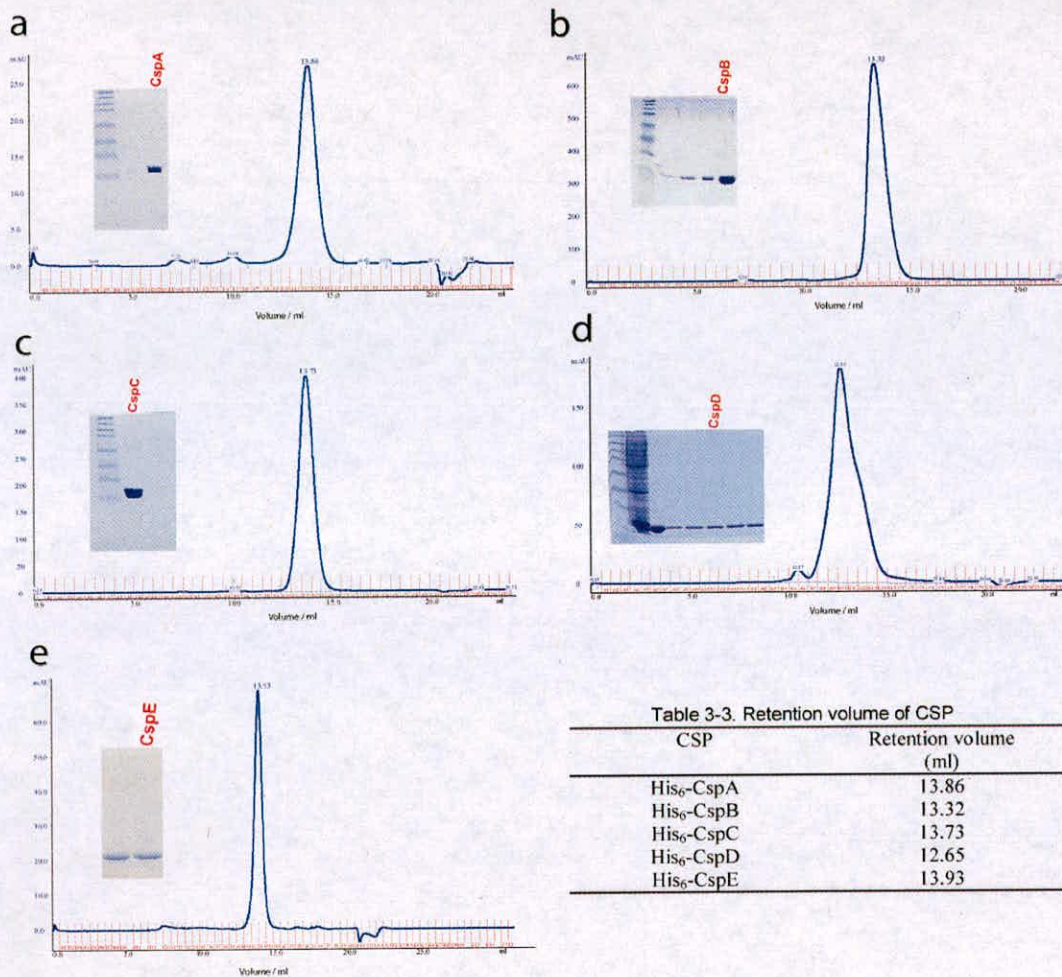
Figure 3-7. Elution profiles for refolding of His₆CSPs.

Denatured His₆CSPs (His₆CspA, His₆CspB, His₆CspC, His₆CspD and His₆CspE) were applied onto a Superdex-75 HR column. Refolding buffer was used to elute the protein (0.2 μl.min⁻¹) and the resulting fractions were analysed for protein content by SDS-PAGE analysis (inset).

3.5.7 Purification of recombinant His₆CspD

Despite their everyday use in protein folding studies the exact mechanism of action

of guanidine hydrochloride (GdnHCl) and urea on proteins is not understood [187].

The variations in stabilising and destabilizing effects of the different denaturants

must arise from their influence on protein conformations. Unfortunately, a clear

understanding of their mode of action is lacking but there are advantages of using one denaturant over another [188] when devising a refolding protocol.

Cell lysate containing His_6CspD was bound to a Ni-NTA agarose column and unbound proteins were washed away. To unfold His_6CspD and remove nucleic bound nucleic acids, a denaturant buffer containing 6 M GdnHCl (buffer F) was used. His_6CspD was unfolded on the column using a gradient (0-100%) of buffer F. His_6CspD was refolded on Ni-NTA agarose column by running a gradient (0-100%) of buffer G (20 mM HEPES pH8.0, 150 mM NaCl) against buffer F. Bound His_6CspD was eluted using an elevated concentration of imidazole at a constant pH to displace the histidine tag from the nickel particles. The purified His_6CspD migrated as a single protein band on a SDS-PAGE (15%) polyacrylamide gel (Figure 3-8b). Gel filtration with a HiLoad 26/60 Superdex 75 column was performed to analyse the percentage of correctly refolded His_6CspD and aggregate. Approximately 10% of incorrectly refolded recombinant His_6CspD eluted in the void peak (7.5 ml), while the majority of His_6CspD eluted as two major symmetrical peaks (Figure 3-8a) with retention volumes of 11.40 and 13.5 ml, corresponding to molecular weight standards of 30 and 14 kDa. This suggests the formation of dimeric structures under these conditions. Analysis of the corresponding fractions confirmed that His_6CspD was greater than 90% pure, with an experimentally determined molecular weight of 10 kDa (Figure 3-8).
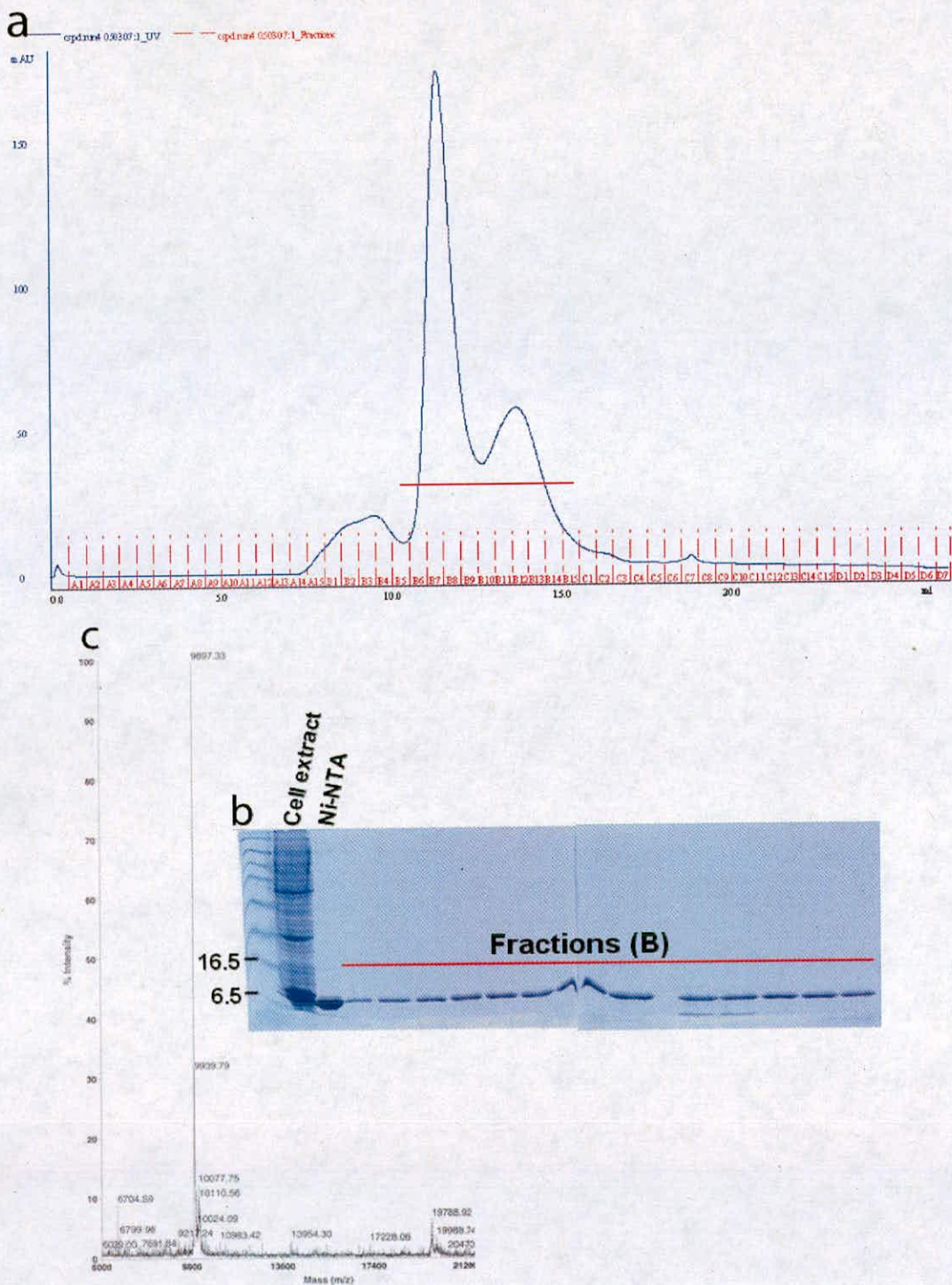
Figure 3-8. **Purification of His$_6$CspD.**

(a) Gel-filtration with Superdex 75. Two peaks can be seen on the elution profile and the corresponding fractions were analysed by SDS-PAGE (b). (c) MALDI-Tof mass spectrum of His$_6$CspD, observed 9996 Daltons, calculated 10048 Daltons.

3.6 Section 2: The examination of single stranded DNA binding of *S. typhimurium* CSPs

3.6.1 *Refolded CSPs bind ss DNA*

Electrophoretic mobility shift assays (EMSA) were conducted to verify that refolded *S. typhimurium* cold shock proteins have retained their DNA binding activity. A single stranded 25-mer (ONc:5′–dATCCTACTG<u>ATTGG</u>CCAAGGTGCTG-3′) substrate containing the Y-Box binding motif (underlined) [86], labelled at the 5′-end with γ-^{32}P, was used to examine the DNA binding activity of refolded CSPs. Figure 3-9 shows a typical gel-shift experiment performed with ONc in the presence of increasing amounts of His$_6$*St*CspE. All refolded proteins bound ss DNA (also confirmed by microarray experiments section 3.6.3) indicating that they have all been correctly refolded during the purification process.
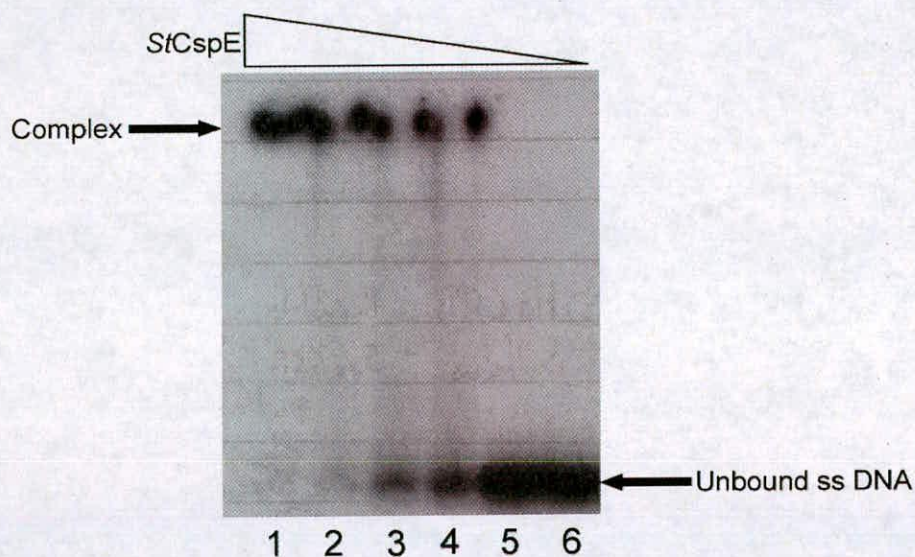
Figure 3-9. An example of a typical electrophoretic mobility shift assay for *S. typhimurium* CspE and its complexes formed with ss DNA.

15-20 pmol of 5′-end γ-^{32}P-labelled ONc was mixed with increasing amounts of CSP and incubated for 10 minutes at 4°C. Amounts of protein: lane 1, 500 pmol; lane 2, 250 pmol; lane 3, 125 pmol; lane 4, 62.5 pmol: lane 5, 31.25 pmol. Lane 6 was a control and contained 15-20 pmol of 5′-end γ-^{32}P-labelled ONc only. Protein-DNA complexes were separated on a 20% native polyacrylamide gel.

3.6.2 Natively purified and refolded CSPs show identical characteristics

All refolded CSPs bind ss DNA but confirmation that refolded CSPs have equivalent ss DNA binding activity to natively purified CSPs was required. To verify that there was no difference between natively purified and refolded CSPs, the interaction between His_6CspE and an oligonucleotide, 5′-GTCTTTT-3′, was analysed by ITC to determine solution-based equilibrium constants. The ss DNA ligand used here was the high affinity binding sequence determined for the *B. subtilis* CspB (Figure 2-15). The ITC assay for the $His_6StCspE$/5′-GTCTTTT-3′ interaction yielded apparent equilibrium dissociation constants of 36 ± 4 nM for natively purified and 41.3 ± 3.4 nM for refolded $His_6StCspE$ (Figure 3-10 and Table 3-3). These affinity values are not significantly different and demonstrate that both refolded and natively purified $His_6StCspE$ exhibit the same ss DNA binding characteristics. $His_6StCspE$ serves as an exemplary test member from the group of refolded *S. typhimurium* CSPs and suggests that the use of similar refolding techniques would work for all CSPs, thus maintaining ss DNA binding characteristics.
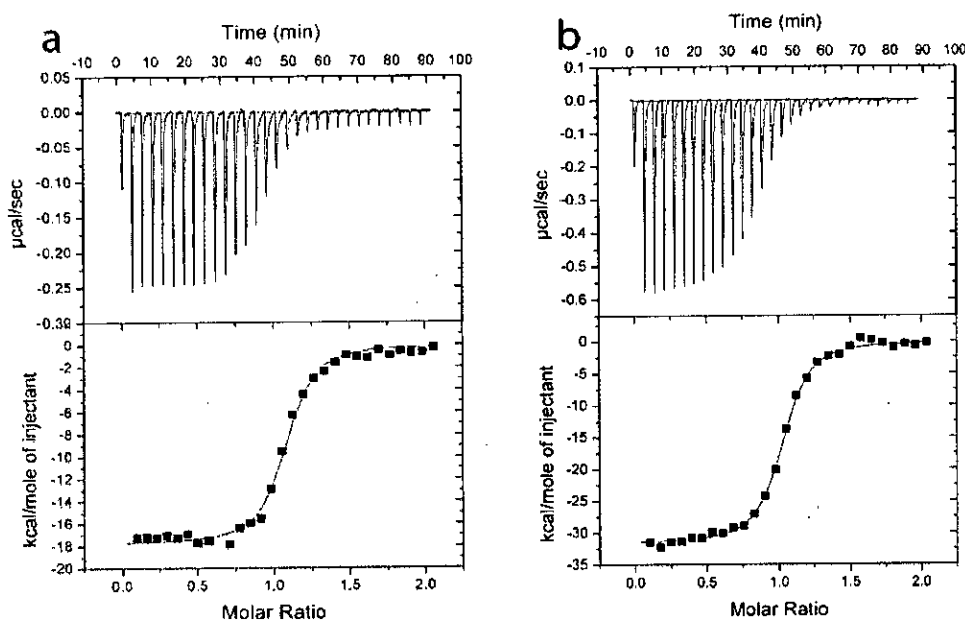
Table 3-3. K_d for refolded and native His_6-CspE.

Protein	Sequence	$K_{d\ 25°C}$ (nM)	Stoichiometry (n)
(a) His_6-CspE (Native)	5′-GTCTTTT-3′	35.9±4.0	1.1±0.01
(b) His_6-CspE (Refolded)	5′-GTCTTTT-3′	41.3±3.4	1.0±0.01

Figure 3-10. **Isothermal titration calorimetry for ss DNA binding to natively purified and refolded $His_6StCspE$.**

Each exothermic heat pulse (upper panel) corresponds to an injection of 5 μl of oligonucleotide (100 μM) into the cell containing 5 μM of either natively purified (**a**) or refolded $His_6StCspE$ (**b**) at 25°C. Thermodynamic parameters for $His_6StCspE$/ss DNA interactions are shown in table 3-3.

3.6.3 The identification of consensus high affinity binding sequences for S. typhimurium CSPs

The microarray assay described in chapter two was used to examine the DNA sequence specificity of the family of CSPs from *S. typhimurium*, excluding *St*CspH. The results show that *St*CspA, *St*CspB, *St*CspC and *St*CspE all bind to very similar pyrimidine rich consensus sequences, while CspD binds to purine (guanine) rich sequences (Figure 3-11). Indeed, we could expect *St*CspA, *St*CspB, *St*CspC and

*St*CspE to behave very similarly as they all share a high level of sequence identity

(**Table 3-2**).
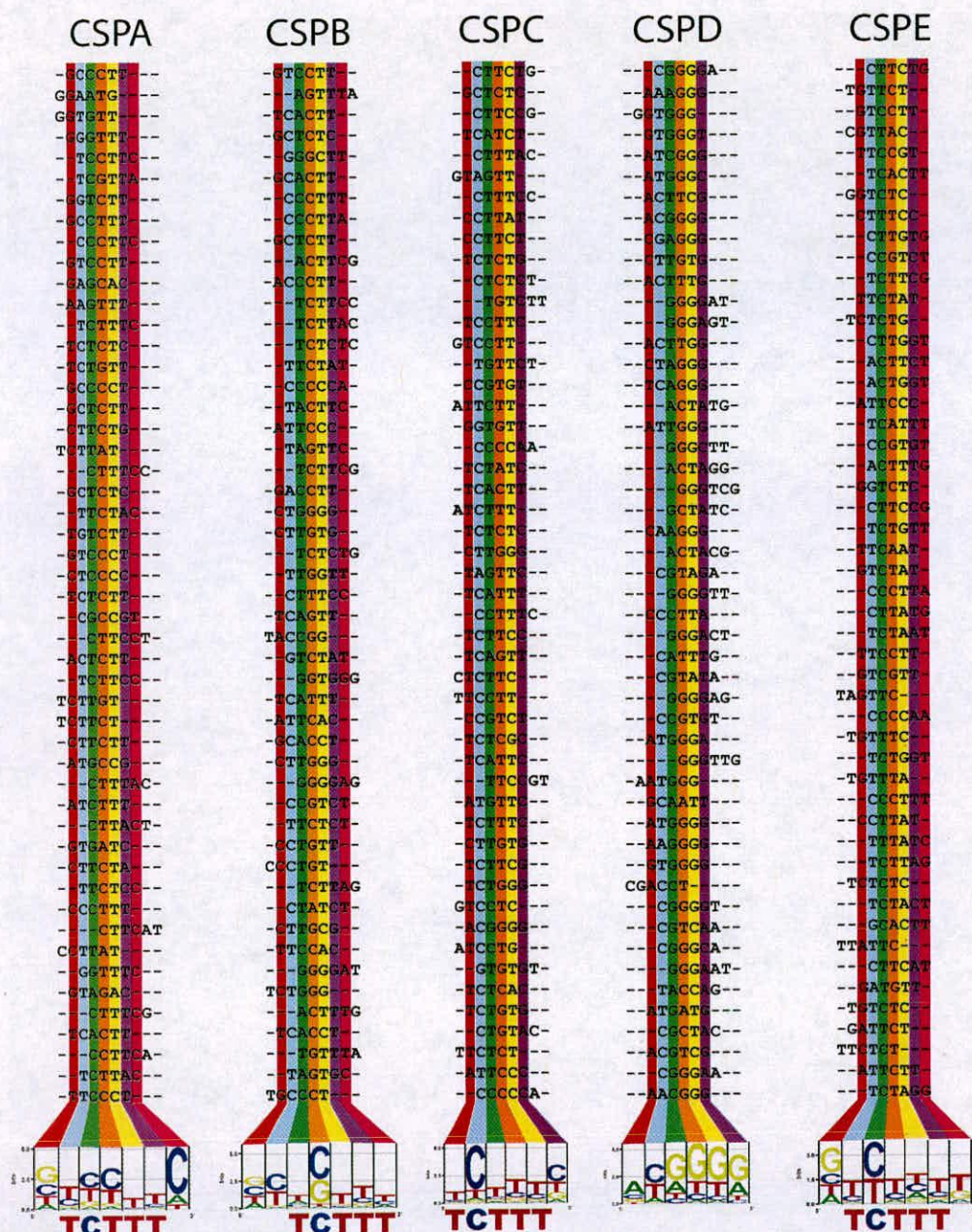
Figure 3-11. **The ss DNA sequence specificity of CSPs from *S. typhimurium*.**
Multiple sequence alignments of the top fifty 6-mer binding sequences for each CSP. The overall length of the sequence alignments have been reduced to highlight those bases which correspond to a meaningful number, i.e. >40% occupancy. The resulting sequence logos (generated using the EnoLOGOS program) correspond to the frequency of each base within the proposed binding sites. The core sequence, TCTTT, is shown below each column.

3.6.4 Two specific ss DNA motifs, 5-´ACGGGG-3´ and 5´-TCTTT-3´, emerge from the small family of S. typhimurium CSPs

Two clear protein groups have emerged from the CSPs examined, Poly T or Poly G binders, all of which show a strong affinity for cytosine bases at different positions. Although there are slight differences in the consensus binding sequences for *St*CspA, *St*CspB, *St*CspC and *St*CspE they all bind the same core sequence, TCTTT (Figure 3-11). This core binding sequence is identical to the core sequence of the high-affinity ligand, 5´-GTCTTTT-3´, determined for the *Bs*CspB (Figure 2-15 and [14]), suggesting similar functions for these CSPs. *St*CspE bound pyrimidine (thymine) rich sequences, with the consensus sequence, 5´-GTCTTTT-3´, which is identical to the high-affinity ligand determined for *Bs*CspB [174].

*St*CspD bound guanine rich sequences, resulting in the consensus sequence, 5´-ACGGGG-3´ (Figure 3-11). The difference in the DNA binding preference observed between CspD and the other CSPs examined is somewhat reflected in the dissimilarities in amino acid composition, with CspD possessing only one of two RNP motifs (Figure 3-12).
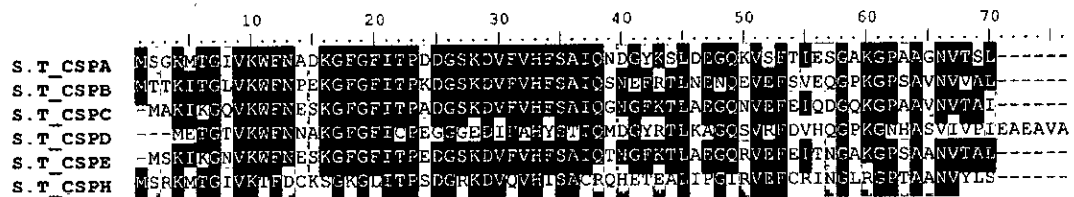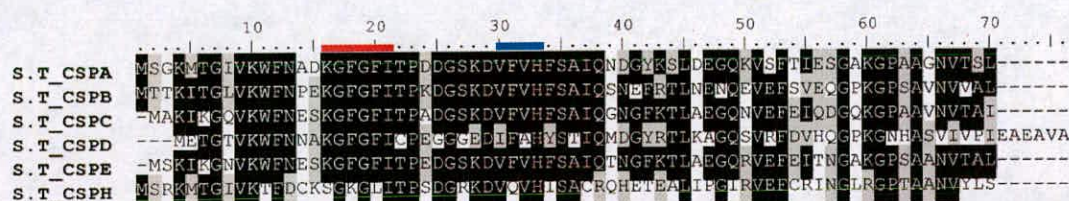

Figure 3-12. Alignment of the amino acid sequences of cold shock proteins from Salmonella typhimurium.

Identical residues are shown in black, similar residues in grey and non-identical in white. Ribonucleotide protein motifs 1 (red) and 2 (blue).

Based on DNA recognition, *St*CspD may be part of a minor sub-group of CSP homologues, of which *Ec*CspC is also part of as it binds preferentially to guanine rich sequences [33]. The mammalian homologue, Y-Box protein P50, also binds preferentially to guanine based sequences [189]. Both *Ec*CspD and *St*CspD may have similar functions to P50 and bind to mRNA not merely as an RNA chaperone but may inhibit or enhance translation [190, 191].

3.6.5 The differences in binding preferences may correspond to a difference in function.

CSPs have been characterised as RNA chaperones, preventing the formation of unfavourable RNA secondary structures during cold shock. The two different binding preferences seen here for *S. typhimurium* CSPs appear rational, with regards to their function as RNA chaperones, as both uracil (thymine) and guanine rich binding of RNA would be required to prevent all complementary base pairing. *St*CspA, *St*CspB, *St*CspC and *St*CspE could bind to uracil-rich stretches of mRNA preventing complementary base pairing with adenine (Figure 3-13b) and *St*CspD could bind to guanine-rich stretches preventing complementary base pairing with cytosine (Figure 3-13a).

Figure 3-13. **CspD (purine rich binder) and CspE (pyrimidine rich binder) bind ss RNA and prevent the formation of unfavourable secondary structures.**

(a) The secondary structure formed from a guanine-cytosine rich sequence at 15°C and the subsequent binding of CspD, preventing secondary structure formation. **(b)** The secondary structure formed from a thymine-adenine rich sequence at 15°C and the subsequent binding of CspE, preventing secondary structure formation.
RNA secondary structure was generated using MFOLD [192] with the following sequences:
GGGGCCCCCCGGGGGGGCCCCCGGGGGGGCCCCCCCGGGGGGCCCCCC
UUUUUUUAAAAAAUUUUUUUAAAAAAAUUUUUUUUUUUUAAAAUUUUUUUUU

3.7   Conclusions

Five (*St*CspA, *St*CspB, *St*CspC, *St*CspD, *St*CspE) out of six *S. typhimurium* CSPs were successfully cloned, expressed and purified using denaturing and refolding techniques, which removed any contaminating nucleic acids. The proteins were purified to an extremely high level and all bound ss DNA. The sequence specificity for each of the five CSPs were determined using the previously developed microarray assay, to reveal two different types of ss DNA binding preferences,

purine or pyrimidine rich sequences. *St*CspD bound purine (guanine) rich sequences, resulting in the consensus binding sequence, 5´-ACGGGG-3´. *St*CspA, *St*CspB, *St*CspC and *St*CspE bound pyrimidine (thymine) rich sequences, with an identical consensus core sequence, 5´-TCTTT-3´. Regarding RNA chaperone functions, these two types of binding proteins would be expected for the prevention of unfavourable RNA secondary structures, resulting from complementary base pairing.

CHAPTER 4. The biophysical characterisation of single-stranded DNA binding cold shock proteins from *Salmonella typhimurium*.

4.1 Aim

To investigate the interaction of *St*CspE and *St*CspD with various ss DNA sequences using surface plasmon resonance and isothermal titration calorimetry, in order to achieve a better understanding of the kinetics and thermodynamics of CSP/ss DNA interactions.

4.2 Summary

From the microarray assay results *St*CspE was found to bind preferable and with high affinity to the oligonucleotide sequence, 5′-GTCTTTT-3′ (Figure 3-11). SPR and ITC results for the His_6-CspE/5′-GTCTTTT-3′ interactions revealed temperature dependence of equilibrium dissociation constants (K_d), ranging from 8.71 ± 1.77 nM at 10°C to 258 ± 18 nM at 37°C, which was primarily due to a large increase in dissociation rate constant (k_d).

Problems arose when creating a stable Biacore chip surface to assay CspD/ss DNA interactions. A new and effective immobilisation strategy was achieved using a double-His-tagged version of the CspD protein. A stable NTA-Ni-double-His-CspD Biacore chip surface allowed for the accurate determination of kinetic and affinity data at low temperatures. *St*CspD was found to bind with highest affinity (19.16 nM) to the 5′-ACGGGG-3′ sequence and exhibited an extremely slow k_d (2.70 ± 0.22 x 10^{-4} s^{-1}) at 15°C.

4.3 Introduction

The interaction between various oligonucleotides and *Bs*CspB has been examined many times [49, 124, 130, 133, 170]. Recently high-affinity heptanucleotide sequences, 5´-GTCTTTT-3´ [174] and 5´-TTCTTTT-3´ [14] were examined, with K_d values of 30 ± 2 and 0.2 ± 0.007 nM, respectively. The kinetics for various *Bs*CspB/ss DNA interactions have been previously examined using tryptophan fluorescence titration experiments [14, 170], which provided affinity and stoichiometric data but did not fully characterise the kinetics and thermodynamics of this exemplary CSP/ss DNA interaction.

For this study *St*CspE and *St*CspD were used as representative members of the CSD containing protein family. Initial ss DNA binding results have shown that *St*CspE binds preferentially to the pyrimidine rich heptanucleotide, 5´-GTCTTTT-3´, while *St*CspD exhibits a binding preference for the purine rich hexanucleotide, 5´-ACGGGG-3´.

To provide a greater understanding of the function of CSPs and CSD-containing proteins an in-depth analysis of *St*CspE and *St*CspD interacting with their preferential binding sequences was performed using up-to-date biophysical techniques. The binding of various oligonucleotides to *St*CspE was analysed using Biacore surface plasmon resonance (SPR) method and by isothermal titration calorimetry (ITC) to generate kinetic and thermodynamic profiles for *St*CspE/ss DNA interactions. The binding of *St*CspD to various oligonucleotides was examined using the SPR method. The kinetics and thermodynamic profiles obtained for *St*CspE and *St*CspD were then used to compare and evaluate the binding characteristics of each CSP, which will provide a greater understanding of the function of CSPs.

4.4 Materials and Methods

4.4.1 Biacore chip preparation and standard procedures

SPR measurements were performed on a Biacore T100 instrument. All chemicals and reagents used were of the highest standard available. Prior to use, biosensor chips were docked into the instrument and both flow cell surfaces were washed with two sequential 60-sec injections at 10 μl.min^{-1} of 1 M NaCl; 50 mM NaOH and then HBS-P$^+$ (10 mM HEPES [pH 7.4], 150 mM NaCl and 0.05% v/v surfactant P20) at 50 μl.min^{-1} for at least 1800 sec.

All SPR binding experiments were performed in HBS-P$^+$ as the running buffer. The flow rate was 30 μl.min^{-1} in all experiments, as there was no evidence of mass transport issues at this flow rate. All procedures for binding studies were automated methods of repetitive cycles of calibration, sample injection and regeneration. The sensor surface was regenerated between binding experiments by dissociating any formed complex using the following regeneration buffers.

(1) For Ni-NTA-His$_6$CSP/ss DNA interactions, HBS-P$^+$ buffer containing 350 mM EDTA was used as a regeneration buffer.

(2) For SA-Biotinylated-ss DNA/His$_6$CSP interactions, 10 mM glycine-HCl, pH 2.5 was used as a regeneration buffer.

(3) Ni-NTA-double-His$_6$CSP/ss DNA interactions, 10 mM HEPES pH 7.4, 150 mM NaCl and 0.5 M imidazole was used as regeneration buffer.

All binding curves and thermodynamic parameters were calculated using a 1:1 interaction model (A + B = AB) and the evaluation software (version 1.1.1) provided

with the Biacore T-100 instrument. Van't Hoff plots were fitted to the non-linear integrated form of the equation (Equation 2).

Equation 2: $RT \ln K_d = \Delta H^{\circ}_T - T\Delta S^{\circ}_T + \Delta C_p^{\circ}.(T - T_0) - T C_p^{\circ}.\ln(T/T_0)$

4.4.2 SPR: Monitoring NTA-Ni-His₆CSP and double-His₆CSP/ss DNA interactions.

A NTA-Ni surface was generated by flow of a priming solution (HBS-P$^+$ buffer containing 500 μM NiSO$_4$) at a rate of between 30 μl.min^{-1} and 100 μl.min^{-1} over a single flow cell on the chip surface (Figure 4-1(1)). A NTA-Ni-His₆CSP surface was then generated by diluting His₆CSP (Stock = 0.4 mM) to 200 nM in HBS-P$^+$ buffer and injecting at 30 μl.min^{-1} over the NTA-Ni. One flow cell was left blank to serve as a control. Approximately 500-600 units of pure His₆CSP were captured on a NTA-Ni sensor surface using standard protocols (Figure 4-1(2)). To generate a binding response, two-fold serial dilutions of the test oligonucleotide were prepared in running buffer, typically spanning 2 – 1024 nM, and injected over both the control and the NTA-Ni-His₆CSP (Figure 4-1(3)). Temperature-dependency studies of His₆Csp/ss DNA interactions were conducted at 12, 16, 20 and 25°C. Surface regeneration and flow rates were as specified (section 4.4.1).
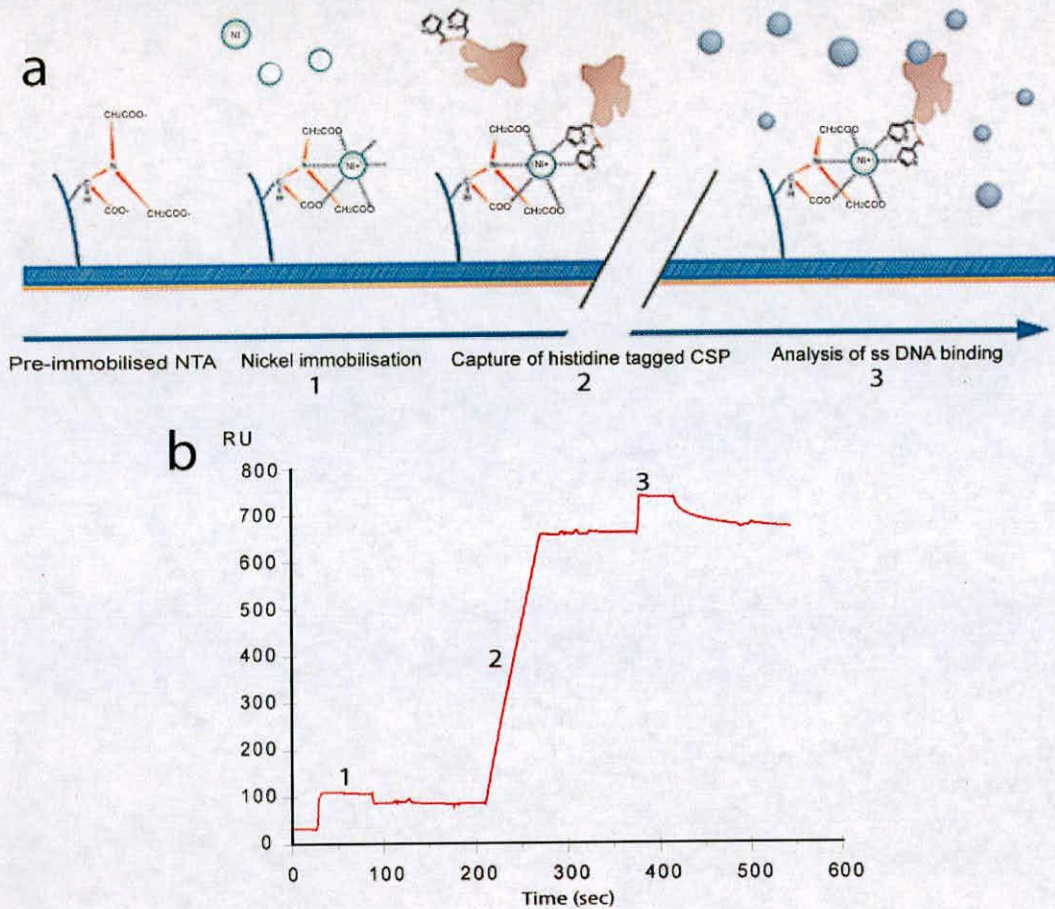
Figure 4-1. **A standard His$_6$CspE-ss DNA binding response generated using a Ni-NTA biosensor chip.**

(a) A schematic representation of a single experimental cycle; (1) Nickel (Ni^{2+}) immobilisation. (2) Immobilisation of His$_6$CspE. (3) ss DNA binding response. **(b)** The same process but showing the actual sensorgram readout from a Biacore machine.

4.4.3 SPR: Immobilisation of ss DNA on a streptavidin biosensor chip.

Three 5´-biotin-labeled ss DNA sequences (HPLC purified; Sigma-Aldrich) with a nine carbon spacer separating biotin from the DNA sequence (Figure 4-2) were used in surface plasmon resonance experiments: d(Biotin-C9-GTCTTTT-3´) (MW = 2778 Daltons) d(Biotin-C9-ACGGGG-3´) (MW = 2228) and d(Biotin-C9-AAAAAAA-3´) (MW = 2725 Daltons).

Biotinylated-DNA was diluted to 25 nM in HBS-P$^+$ buffer and injected manually at 25°C to a single flow cell in streptavidin-derivatised sensor chips (Biacore SA chips) at a flow-rate of 12 µl.min^{-1}, one flow cell was left blank to serve as a control. Typical non-covalent capture levels ranged from 37-43 RU of biotinylated-DNA, which would give a response of 130–152 RU$_{max}$, assuming His$_6$CspE (MW = 9614 Daltons) and His$_6$CspD (10048) were 100% active.
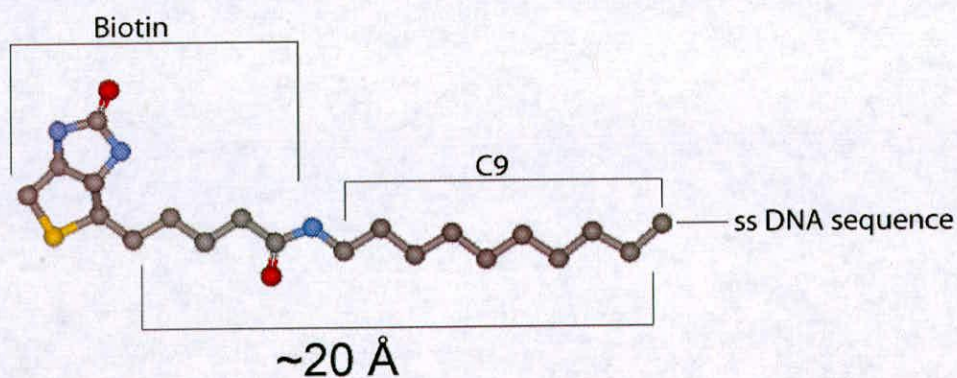
Figure 4-2. Chemical structure of the biotin and nine carbon spacer used to label the 5′ end of ss DNA sequences.

This figure was generated using DS Viewer pro from Accelry.

4.4.4 SPR: Monitoring SA-Biotin-C9-ss DNA/His$_6$CspE interactions.

For the His$_6$CspE study, a 0.4 mM (~4 mg/ml) stock solution was prepared directly in running buffer from which two-fold serial dilutions were made, typically spanning 2 nM to 4.096 µM. Temperature-dependent studies of His$_6$CspE/GTCTTTT-3′ interactions were conducted at 10, 15, 20, 25, 30, 35 and 40°C. His$_6$CspD/5′-ACGGGG-3′ interactions were conducted at 15°C only. Surface regeneration and flow rates were as specified (Standard procedures).

4.4.5   Isothermal Titration Calorimetry (ITC)

ITC experiments were carried out as described previously (section 2.4.11).

4.4.6   Construction of the C-termini double-Hexahistidine vector (pC_Term_2x vector)

The C-terminal double-hexahistidine sequence was created by overlap extension PCR using high-fidelity Taq polymerase (Roche). The resulting PCR product was digested with *XbaI* and *XhoI* (New England Biolabs). The digested insert was ligated into a *XbaI-XhoI*-digested pET28a vector (Novagen) and verified by DNA sequencing. The final vector is referred to as pC_Term_2xHis (see section A1.4 for details).

The following primers were used to create the C-Terminal sequence by overlap extension PCR.

Fwd_mid: (114 bases, % GC=68, Tm=87)
TCC GAA TTC GAG CTC CGT CGA CGC GGC CGC GGC GGA GGC AGC GGT GGT GGA AGC GGC
GGT GGA ACA GGC GGA GGA AGC GGC GGT GGA AAA CGC GCT GAT GCC GCT CAT CAT CAC

Rev mid: (114 bases, % GC=68, Tm=87)
GT GAT GAT GAG CGG CAT CAG CGC GTT TTC CAC CGC CGC TTC CTC CGC CTG TTC CAC
CGC CGC TTC CAC CAC CGC TGC CTC CGC CGC GGC CGC GTC GAC GGA GCT CGA ATT CGG
A

Fwd_C_5':   (81 bases, % GC=44, Tm=75)
CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT AGG CTA GCA AGC TTC GGA
TCC GAA TTC GAG CTC CGT CGA CGC

Rev_C_3':   (100 bases, % GC=59, Tm=82)
TA CTC GAG TTC AGT GGT GGT GGT GGT GGT GAC CAC CGA ACT GCG GGT GGC GCC AAG
CGC GAG AAT GAT GAT GGT GAT GAT GAG CGG CAT CAG CGC GTT TT

4.4.7   Overlap extension PCR experiments: Amplification of the C-Terminal insert

Overlap extension PCR reactions were performed using a thermocycler. PCR reactions were carried out as follows:

116

Reaction 1: 10 µl Elongase buffer A (Invitrogen), 2 µl FWD_mid primer [2 µM], 2 µl REV_mid primer [2 µM], 2 µl dNTPs [0.2 mM], 2 µl Elongase (Invitrogen), 32 µl H_2O. The reaction was incubated in a thermocycler using the following settings: (1) 95°C for 1 minute (2) 95°C for 30 seconds (3) 75°C for 1 minute (steps 2 and 3 were cycled 10 times).

Reaction 2 – Lane 1 (Figure 4-11) : 10 µl Elongase buffer A (Invitrogen), 2 µl FWD_C primer [350 nM], 2 µl REV_C primer [350 nM], 1 µl of reaction 1 as template DNA, 2 µl dNTPs [0.2 mM], 2 µl Elongase (Invitrogen), 32 µl H_2O. The reaction was incubated in a thermocycler using the following settings: (1) 95°C for 1 minute (2) 95°C for 30 seconds (3) 72°C for 1 minute (steps 2 and 3 were cycled 10 times). The C_Term product was gel purified and used as the template DNA for a third reaction.

Reaction 3 – Lane 2 (Figure 4-11): 5 µl PCR buffer (Roche), 1 µl FWD_mid primer [2 µM], 1 µl REV_mid primer [2 µM], 1.5 µl dNTPs [0.2 mM], 2 µl Taq polymerase (Roche), 1 µl purified template (reaction 2) 39.5 µl H_2O. The reaction was incubated in a thermocycler using the following settings: (1) 95°C for 3 minutes (2) 95°C for 30 seconds (3) 72°C for 1 minute (steps 2 and 3 were cycled 30 times) (4) 72°C for 10 minutes

Gel electrophoresis was used to confirm the size of all PCR products and when required PCR products were purified with the QIAquick PCR purification kit (Qiagen).

4.4.8 Cloning of double-His-tagged proteins

Primers (Table 4-1) were used to amplify the CspE and CspD genes by PCR using pET28a_CspE and pET28a_CspD DNA as the template (section 3.5.2). Sequences

were amplified using Elongase Taq polymerase (Invitrogen) and the corresponding

forward and reverse primers found in Table 3-1 in a standard PCR reaction.

PCR reaction – 10 µl of Elongase buffer A (Invitrogen), 1 µl Fwd primer [350 nM], 1 µl Rev primer [350 nM], 2 µl dNTPs [0.2 mM], 1 µl Template DNA, 2 µl Elongase (Invitrogen), 33 µl H$_2$O. The reaction was incubated in a thermocycler using the following settings: (1) 94°C for 1 minute (2) 94°C for 30 sec (3) 55°C for 1 minute (4) 68°C for 1 minute; steps 2-4 were cycled 30 times.

The resulting PCR products were cloned into a pCR2.1 TOPO vector using the

TOPO-TA Cloning® kit (Invitrogen) and digested with *NotI* and *HindIII* (New

England Biolabs). The digested inserts were ligated into the *NotI-HindIII*-digested

pC_Term_2x vector.

Table 4-1. **Primers used to created double-hexahistidine clones.**

Clone	Primer position	Primer sequence 5´ to 3´
pC_Term_2x His_CspE	Forward Reverse	AGG C*AA* GC*T T* ATG TCT AAG ATT AAA GGT AAC GTT TAT T*GC GGC CGC* G CAG AGC AGT TAC GTT TGC AGC GGA
pC_Term_2x His_CspD	Forward Reverse	AGG C*AA* GC*T T* ATG GAA ACG GGT ACT GTA AAG TGG TAT *TGC GGC CGC* G TGC AAC GGC CTC TGC TTC GAT GGG
pC_Term_2x His_CypA	Forward Reverse	ATTA *AAG CTT* ATG GTC AAC CCC ACC GTG TTC TTC TAT *GCG GCC GC* G TTC GAG TTG TCC ACA GTC AGC AAT

HindIII (shown in italics) and *NotI* (shown in bold italics) restriction sites were added to primers to allow for directional clonin
Note: The double-His$_6$-tagged cyclophilin A (CypA) was used as a control protein to examine the affects of adding many additional amino acids (~5 kDa) to the overall composition of a protein, while still retaining characteristic functionality.

4.4.9 Expression of double-His-tagged proteins

Chemically competent BL21 (DE3) cells were transformed with the verified

constructs and expressed as before (section 3.4.2).

4.4.10 Purification of double-His$_6$-tagged CspE

The purification was carried out as described previously for single-His$_6$-tagged CspE

(section 3.4.5) except the elution buffer contained 0.5 M imidazole).

4.4.11 Purification of double-His₆-tagged CspD

The purification was carried out as described previously for single-His₆-tagged CspD

(section 3.5.7) except the elution buffer contained 0.5 M imidazole)

4.5 Results and Discussion

4.5.1 CSPs have a common ss DNA binding preference

Two CSP (BsCspB and BcCsp)/T_6 (hexathymidine) complex X-ray structures have been published recently [14, 25], both of which show similar ss DNA-binding interfaces (Figure 1-11). The residues involved in ss DNA binding are highly conserved (Figure 4-3a and c) and the characteristic rigid DNA binding platform [14] is present in the StCspE structure, and in most other members of the highly homologous family of cold shock domain (CSD) containing proteins (Figure 4-3b). The high degree of conservation observed on the ligand binding surface of StCspE and the availability of extremely pure protein made StCspE an ideal candidate for an in-depth examination of ss DNA binding/recognition for this highly homologous family.

The sequence specificity of StHis$_6$-CspE was determined using our previously published microarray technique [174], to reveal a preferential binding sequence, 5´-GTCTTTT-3´. The results confirmed that StHis$_6$-CspE and BsHis$_6$-CspB (the major cold shock from *Bacillus subtilis*) had identical ss DNA binding preferences for 5´-GTCTTTT-3´. This was expected as the key DNA binding residues are identical (Figure 4-3a). The interaction between ss DNA and BsCspB has been examined many times [49, 124, 130, 133, 170]. Recently high-affinity heptanucleotide sequences, 5´-GTCTTTT-3´ [174] and 5´-TTCTTTT-3´ [14] were determined, with K_d values of 30 ± 2 and 0.2 ± 0.007 nM, respectively. The binding constants for various BsCspB/ss DNA interactions have been previously examined using tryptophan fluorescence titration experiments [14, 170], which provided affinity and

stoichiometric data but did not fully characterise the kinetics and thermodynamics of

the interaction.
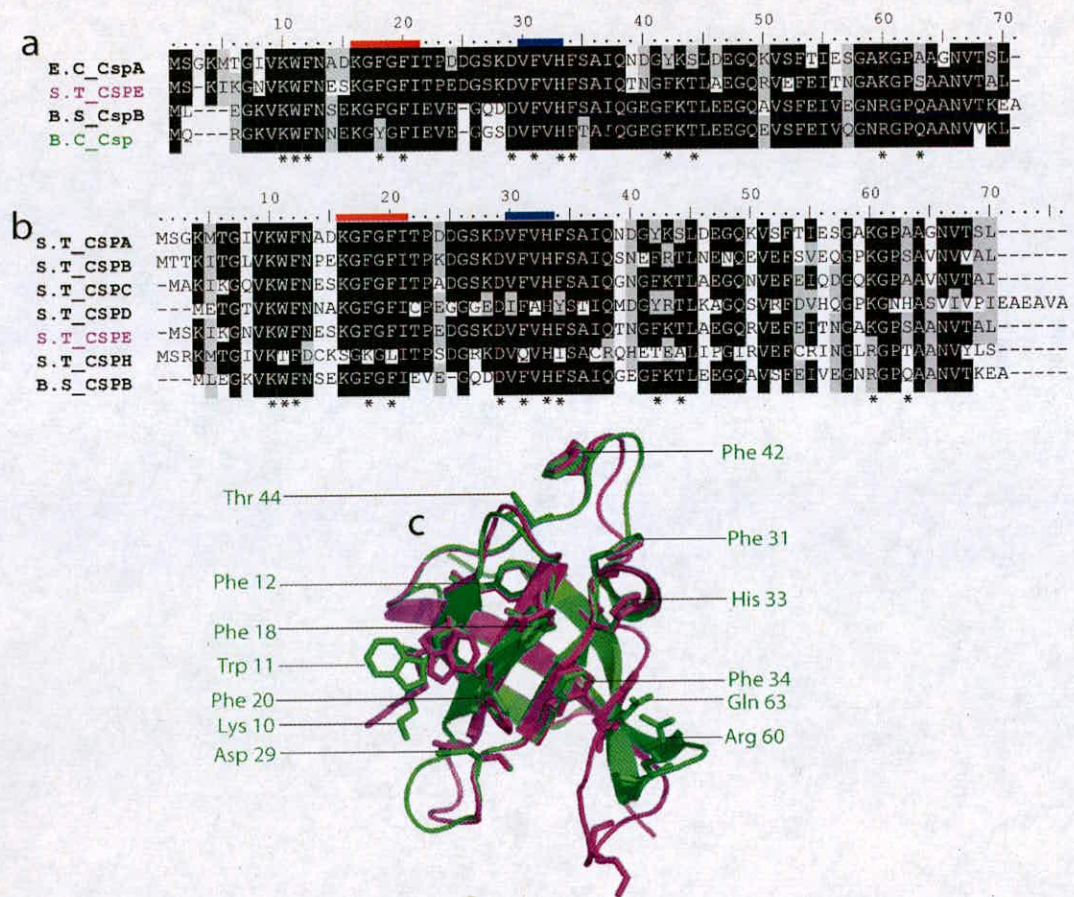
Figure 4-3. **Sequence analysis of CSPs.**
(a) Alignment of the amino acid sequences of cold shock proteins for which a x-ray crystal structure exists. (b) Alignment of the amino acid sequences of cold shock proteins from *Salmonella typhimurium* with the major cold shock protein from *Bacillus subtilis*. Identical residues are shown in black, similar residues in grey and non-identical in white. Residues involved in DNA binding are marked (*). Ribonucleotide protein motifs 1 (red) and 2 (blue). The abbreviations are as follow: *St*CSP, cold shock protein from *Salmonella typhimurium*; EcCspA, cold shock protein A from *Escherichia coli*; *Bs*CspB, cold shock protein B from *Bacillus subtilis*; *Bc*Csp, cold shock protein from *Bacillus caldolyticus*. (c) X-ray crystal structure of *Bs*CspB (green) superimposed on the structure of *St*CspE (purple - see chapter five for details of structure determination). The amino acid residues involved in ss DNA binding are shown as sticks and have been labelled according to the *Bs*CspB structure.

121

4.5.2 *SPR: Creating a sensor surface*

4.5.2.1 Ni-NTA surface

Measuring binding reactions using SPR biosensors requires that one of the binding partners be immobilised onto a surface. His_6CspE was immobilised onto the surface of a Ni-NTA biosensor chip and the previously identified consensus sequence; 5´-GTCTTTT-3´ was passed over the surface to generate a binding response. Binding data for the His_6CspE-5´/GTCTTTT-3´ interactions were collected at 12, 16, 20 and 25°C (Figure 4-4) on the biosensor. The response data for the binding of His_6CspE to the surface of immobilised 5´-GTCTTTT-3´ were fit to a 1:1 binding model. The local R_{max} values locally fitted (this helped compensate for the base-line drift), which provided a better fit as to the data, as shown by the overlay of the simulated binding responses (black lines in Figure 4-4a) with the experimental data (red lines in Figure 4-4a).

The data generated contained no useable kinetic data as a direct result of a steady base-line drift caused by the dissociation of the histidine tagged protein from the chip surface. The base-line drift became so rapid at temperatures above 20°C that results were not reproducible with any degree of accuracy (Figure 4-4a inset-25°C). The results recorded for His_6CspE-5´-GTCTTTT-3´ interactions contained steady-state data and apparent K_d values were obtained (Figure 4-4b). The affinities obtained for 12 and 16°C seemed reasonable as the problematic base line drift was not as visible at the binding plateau when compared with 20 and 25°C. This capture method is useable for the initial screening of various binding sequences but is by no means viable for the thermodynamic and kinetic characterisation of the His_6CspE-5´-GTCTTTT-3´ interactions. Clearly a more stable surface needed to be generated.
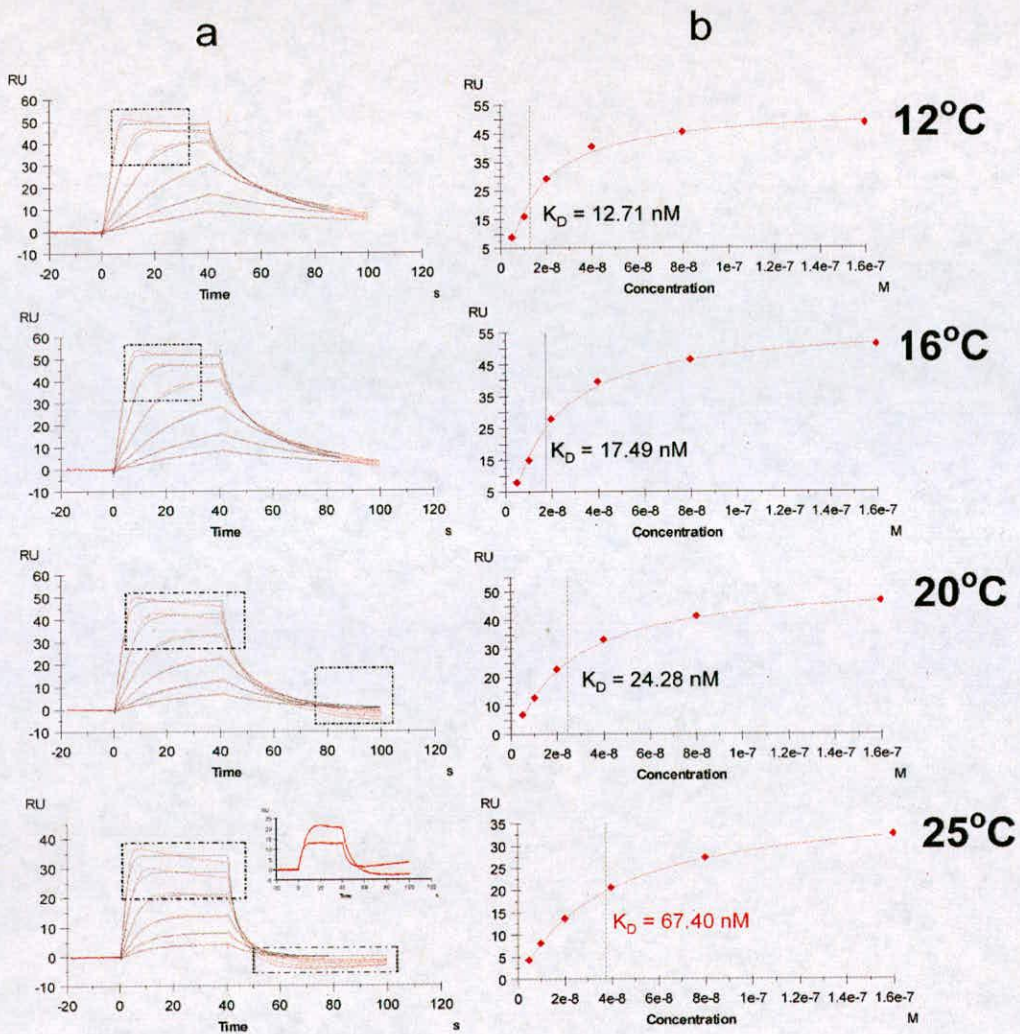
Figure 4-4. The characterization of the His$_6$CspE-5′-GTCTTTT-3′ interaction.

(a) Reference-corrected SPR binding curves (red), monitored on a surface with 550 RU of NTA-Ni stabilized His$_6$CspE, for various 5′-GTCTTTT-3′ concentrations (serial dilution from 160 nM to 5 nM) at the specified temperature in HBS-P$^+$. The apparent dissociation rate constant (k_d) and association rate (k_a) constants were determined by globally fitting (black) a simple 1:1 interaction model, A + B = AB, to the sensorgram using the software supplied with the instrument. Inset shows the normalized sensorgram overlay for 40 nM 5′-GTCTTTT-3′ injected over the same surface of His$_6$CspE, binding at 25°C. **(b)** Response *versus* concentration was plotted and the data was fitted using a 1:1 interaction model. The apparent disassociation constant is marked on each graph with a vertical line. Time is shown in seconds. Areas where the fit was poor have been high-lighted (Boxed areas).

4.5.2.2 CM5 surface

To stabilize the surface at higher temperatures His_6CspE was immobilised onto the dextran layer of a CM5 biosensor chip using amine coupling (Data not shown). Although adequate amounts of His_6CspE were obtained on the surface it was found to be inactive. This was perhaps a result of coupling a lysine within the DNA binding site to the chip surface, consequently eliminating accessibility to the active site.

4.5.2.3 SPR: Streptavidin-biotin immobilisation method produces a highly stable surface

Better results were achieved by immobilisation of ss DNA through streptavidin-biotin interaction. A single stranded DNA sequence, 5´-GTCTTTT-3´, was biotinylated at the 5´ end with a nine carbon spacer, separating biotin from the DNA. A nine carbon spacer was used to mitigate any possible effects from steric hindrance and maximize binding yield (based on microarray studies [193]). The Biotin-9C-GTCTTTT-3´ ligand was then immobilised onto the surface of a streptavidin biosensor chip. The His_6CspE (analyte) was passed over the biotin-C9-GTCTTTT-3´ surface to obtain a binding response. The reproducibility of binding responses from duplicate injections over the same surface before and after 100 cycles demonstrates the Biacore binding assay was very reproducible (Figure 4-5). No binding was recorded on the reference surface.
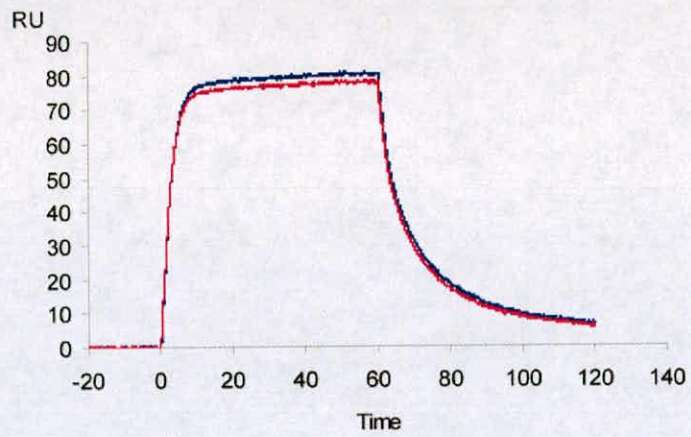
Figure 4-5. **Biacore sensorgram for the binding of His$_6$CspE to biotin-C9-GTCTTTT-3´ (40 RU on surface) at 25°C** immobilised on the sensor chip before (Pink) and after (Blue) 100 cycles.

The His$_6$CspE concentration was 256 nM.
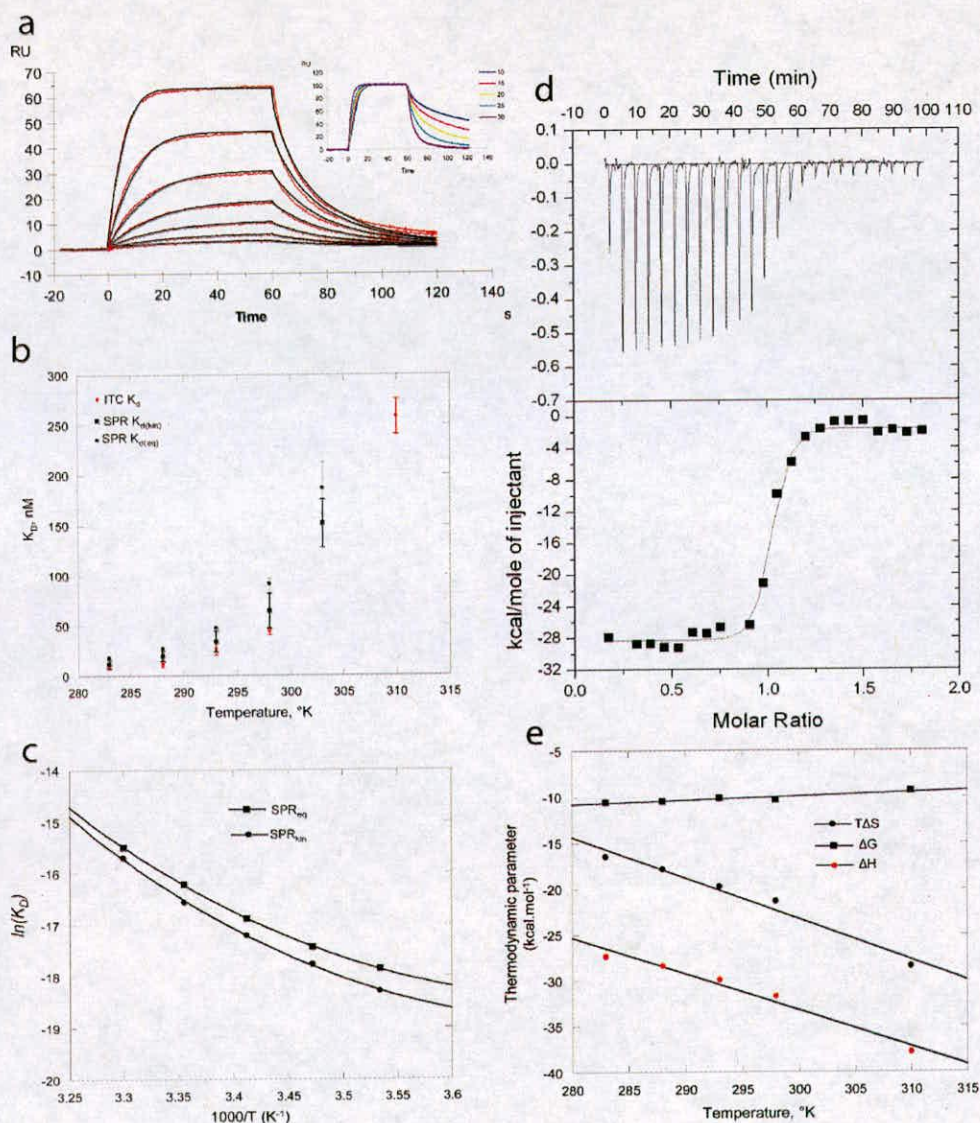
Figure 4-6. **Thermodynamic characterization of the His₆CspE-5′-GTCTTTT-3′ interaction.**
(a) Reference-corrected SPR binding curves (red), monitored on a surface with 40 RU of non-covalently stabilized 5′-GTCTTTT-3′, for various His₆CspE concentrations (serial dilution from 64 nM to 2 nM) at 25°C in HBS-P⁺ buffer. The apparent dissociation rate constant (K_d) and association rate (K_a) constants were determined by globally fitting (black) a simple 1:1 interaction model, A + B = AB, to the sensorgram using the software supplied with the instrument. Inset shows the normalized (100% at the end of injection) sensorgram overlay for 128 nM His₆CspE injected over the same surface of binding of 5′-GTCTTTT-3′ at the indicated temperatures. Temperatures are shown in degrees Celsius and time is in seconds. (b) Effect of temperature on the K_D for His₆CspE binding to 5′-GTCTTTT-3′ determined by SPR; SPR(kin) (filled squares) SPR(eq) (filled circles) and ITC (red circles). (c) Plot of lnK_DSPR(kin) and lnK_DSPR(eq) *versus* 1000/T for the SPR data (both lines are a fit to Eq.1 using the Kaleidagraph v4.03 software (Synergy Software, reading, PA). (d) Isothermal titration calorimetry for ss DNA binding to recombinant His₆CspE. Each exothermic heat pulse (upper panel) corresponds to an injection of 5 μl of oligonucleotide (100 μM) into the cell containing 5 μM His₆CspE at 15°C. Integrated heat data (lower panels) constitutes a differential binding curve, which was fitted to a single-site binding model to give, the stoichiometry of binding (N), binding affinity (K_d) and enthalpy of binding (ΔH) for each oligonucleotide (Tables III – IV). The heat of the reaction was obtained by integrating the peak after each injection of ss DNA oligonucleotide and correcting for the heats of dilution using the Origin® V.7 software. The resulting curve (lower panel) was fitted to a single class binding site model by non-linear least square analysis. (e) ΔH, ΔG and TΔS values from ITC experiments plotted *versus* temperature.

126

Table I. Binding kinetics determined for His$_6$-CspE/5'-GTCTTTT-3' interactions at various temperatures using SPR.

Temp (°C)	k_a (M^{-1}s^{-1})	k_d (s^{-1})	$K_{d(kin)}$ (nM)	$K_{d(eq)}$ (nM)
10	$(2.14 \pm 0.46) \times 10^6$	$(2.44 \pm 0.54) \times 10^{-2}$	11.4 ± 4.1	17.6 ± 0.6
15	$(2.12 \pm 0.42) \times 10^6$	$(4.06 \pm 0.52) \times 10^{-2}$	19.1 ± 5.2	26.7 ± 0.1
20	$(2.23 \pm 0.42) \times 10^6$	$(7.05 \pm 0.83) \times 10^{-2}$	33.7 ± 10.2	47.1 ± 0.3
25	$(3.21 \pm 0.63) \times 10^6$	$(2.06 \pm 0.27) \times 10^{-1}$	64.3 ± 17.7	91.0 ± 6.1
30	$(5.11 \pm 0.10) \times 10^6$	$(7.74 \pm 1.4) \times 10^{-1}$	151.4 ± 23.0	186.4 ± 26.4
35	N/D	N/D	N/D	N/D*
40	N/D	N/D	N/D	N/D*

Table II. Thermodynamic constants determined from the van't Hoff plots for the His$_6$-CspE/5'-GTCTTTT-3' interactions analysed by SPR.

Temp (°C)		ΔH° (kcal/mol^{-1})	ΔG° (kcal/mol^{-1})	$T\Delta S^\circ$ (kcal/mol^{-1})	ΔC_p° (kcal K^{-1}mol^{-1})
25	SPR$_{kin}$	-26.3 ± 0.4	-9.6 ± 0.01	-15.6 ± 0.4	-0.6 ± 0.2
25	SPR$_{eq}$	-28.7 ± 0.6	-9.6 ± 0.01	-18.7 ± 0.6	-0.7 ± 0.1
	Average	$-27.5 + 0.5$	-9.6 ± 0.01	$-17.1 + 0.5$	-0.6 ± 0.1

Table III. Thermodynamic parameters for His$_6$-CspE/5'-GTCTTTT-3' interactions using ITC.

Temp (°C)	K_d (nM)	Stoichiometry (n)	ΔH° (kcal/mol^{-1})	ΔG° (kcal/mol^{-1})	$T\Delta S^\circ$ (kcal/mol^{-1})	ΔC_p° (kcal K^{-1}mol^{-1})
10	8.7 ± 1.8	0.94 ± 0.01	-27.1 ± 0.3	-10.6 ± 0.5	-16.5 ± 0.2	-
15	11.3 ± 2.2	0.98 ± 0.01	-28.4 ± 0.4	-10.5 ± 0.7	-17.9 ± 0.3	-
20	27.2 ± 6.2	0.90 ± 0.01	-30.0 ± 0.4	-10.2 ± 0.7	-19.8 ± 0.3	-0.7 ± 0.1
25	44.4 ± 3.2	1.02 ± 0.01	-32.0 ± 0.2	-10.4 ± 0.4	-21.4 ± 0.2	-
37	258.4 ± 17.8	1.01 ± 0.01	-38.0 ± 0.4	-9.4 ± 0.7	-28.5 ± 0.3	-

Table IV. Thermodynamic parameters for His$_6$-CspE/5'-TTCTTTT-3' interactions using ITC.

Temp (°C)	K_d (nM)	Stoichiometry (n)	ΔH° (kcal/mol^{-1})	ΔG° (kcal/mol^{-1})	$T\Delta S^\circ$ (kcal/mol^{-1})
10	19.7 ± 2.1	0.97 ± 0.01	$-26.7 + 0.2$	-10.0 ± 0.5	$-16.7 + 0.3$
15	39.8 ± 4.5	0.98 ± 0.01	-28.9 ± 0.3	-9.9 ± 0.6	-19.2 ± 0.4
20	189.1 ± 29.8	1.05 ± 0.02	-29.8 ± 0.7	-9.0 ± 0.9	-20.7 ± 0.2
37	584.8 ± 76.7	0.95 ± 0.03	-31.6 ± 1.5	-8.9 ± 1.9	-22.7 ± 0.4

Table V. Thermodynamic parameters for His$_6$-CspE/ss DNA interactions at 15°C using ITC.

Sequence 5'-3'	K_d (nM)	Stoichiometry (n)	ΔH° (kcal/mol^{-1})	ΔG° (kcal/mol^{-1})	$T\Delta S^\circ$ (kcal/mol^{-1})
GTCTTTT	11.3 ± 2.2	0.98 ± 0.01	-28.4 ± 0.4	-10.5 ± 0.1	-17.9 ± 0.3
TTTTTTT	39.8 ± 4.5	0.98 ± 0.01	-28.9 ± 0.3	-9.8 ± 0.1	-19.1 ± 0.2
TTCTTTT	78.7 ± 10.5	0.92 ± 0.02	-37.6 ± 0.9	-9.4 ± 0.3	-28.2 ± 0.1
GTTTTTT	15.6 ± 4.3	0.45 ± 0.01	-31.6 ± 0.8	-10.5 ± 0.3	-21.9 ± 0.5
AATTGGA	1700.7 ± 46.2	0.40 ± 0.09	-12.0 ± 3.2	-7.4 ± 4.0	-4.6 ± 0.8
AAAAAAA	N/D	N/D	N/D	N/D	N/D

N/D not detectable, (SPR experiments, the K_a and K_d were too rapid for the SPR-T100 to obtain a measurement). N/D*, the $K_{D(eq)}$ could not be accurately measured as saturation point for the interaction was not reached. SPR K_D values (mean $\pm$ SE, $n = 3$) were obtained as follows; the $K_{D(kin)}$ value is calculated from kinetics binding data, where $K_D = K_d/K_a$, the $K_{D(eq)}$ value is calculated from steady state binding data.

Kinetic binding values were obtained from the global fitting of data from seven different concentrations of CspE (typically 2 nM, 4 nM, 6 nM, 8 nM, 16 nM, 32 nM, 64 nM) fit to a 1:1 interaction model, A + B = AB. Steady state affinity values were obtained from 12 different concentrations of CspE (2 nM – 4096 nM, concentration was doubled from 2 nM – 4096 nM). All fits had χ^2 value of less than 2, which is generally considered an excellent fit. Errors for ITC data are from the fitting program.

127

4.5.3 CspE interactions with ss DNA are temperature dependent

The ability to examine binding reactions effectively from 10°C to 40°C using Biacore makes it possible to record temperature-dependent binding events on the surface of the biosensor [194]. Binding data for the $His_6StCspE/5'-GTCTTTT-3'$ interactions were collected at 10°C to 40°C at 5°C (Table 1) increments on the biosensor. The response data for the binding of His_6CspE to the surface of immobilised 5'-GTCTTTT-3' were fit to a 1:1 binding model. The local R_{max} values were fitted locally, which provided a good fit to the data, as shown by the overlay of the simulated binding responses (black lines in Figure 4-6a) with the experimental data (red lines in Figure 4-6a). A total of three replicate temperature-dependent rate constants were collected for $His_6StCspE/5'-GTCTTTT-3'$ interactions and the average and standard error for all values are presented in Table I. Unfortunately adequate binding data could not be determined for 35°C and 40°C, due to machine limitations when measuring association and dissociation rates.

The SPR analysis for the $His_6StCspE-5'/GTCTTTT-3'$ interactions revealed that as the temperature was increased from 10°C to 30°C the interaction weakened. The mean equilibrium dissociation constant ($K_{d(eq)}$) for $His_6StCspE-5'/GTCTTTT-3'$ binding increased with temperature from 18 ± 1 nM at 10°C to 186 ± 26 nM at 30°C (Figure 4-6b and Table I), with the K_d approximately doubling for every 5°C increase in temperature up to 30°C, a similar trend was also seen for the solution based equilibrium constants (Figure 4-6b and Table III) determined from ITC experiments. The temperature dependence of the K_d was primarily due to a larger increase in the dissociation rate constant (k_d) than the association rate (k_a) for the $StCspE$-ss DNA interaction (Table I). The mean k_d increased 32-fold from $0.024 \pm$

0.0054 to 0.774 $\pm$ 0.14 s^{-1}, from 10 to 30°C respectively. The effect of temperature on the association rate constants was less profound. The mean k_a increased twofold from 2.14 $\pm$ 0.46 x 10^6 to 5.11 $\pm$ 0.10 x 10^6 M^{-1}s^{-1}, from 10 to 30°C respectively. Specific values for k_a and k_d could not be obtained at 35 and 40°C as the binding interaction was so rapid that it exceeded the limitations (Limits; k_a = 5.00 x 10^6 k_d = 1.00 s^{-1}) of the SPR machine. The difference in k_d and k_a can be seen visually in Figure 4-6a (inset). Minimal binding ($K_d \geq 1\mu M$) was observed for injecting 1 μM His$_6$CspE over a control [130] 5´-AAAAAAA-3' surface (data not shown).

4.5.4 *van't Hoff analysis of SPR data provides thermodynamic parameters very similar to ITC*

The equilibrium constants determined from temperature-dependent SPR analyses were used to determine van't Hoff enthalpies by plotting both ln($K_{d(eq)}$) and ln($K_{d(kin)}$) *versus* 1000/T. As shown in Figure 4-6c, the van't Hoff plots are clearly non-linear for His$_6$CspE-5´/GTCTTTT-3´ interactions, indicating a change in heat capacity upon binding. Fitting these data to Equation 2 (section 4.4.1) yielded thermodynamic parameters; $\Delta H°$ = -27.5 $\pm$ 0.5 kcal/mol^{-1}, T$\Delta S°$ = -17.1 $\pm$ 0.5 kcal/mol^{-1}, $\Delta G°$ = -9.6 $\pm$ 0.01 $\Delta C_p°$ = 0.64 + 0.10 kcal K^{-1}mol^{-1}. The average van't Hoff enthalpies and entropies determined from the SPR analysis were close to the thermodynamic constants measured by ITC at 25°C (Table III).

Thermodynamic constants obtained by ITC and SPR were similar but not identical. Two plausible explanations for the discrepancy between the two analysis of results for the same reaction is the difference in the buffers used for ITC (50 mM Tris, pH 7.5, 100 mM NaCl) and the standard HBS-P$^+$ buffer (10 mM HEPES, pH 7.4; 150

mM NaCl; 0.005% surfactant P20) used for SPR. ITC directly measured $\Delta H°$ values which included additional contributions from buffer interactions (temperature-dependent ionization of buffer ions etc) and as the standard HBS-P$^+$ buffer used for SPR experiments was different, we can assume that this will cause the small discrepancies in the resulting values. Another possible cause is that biosensors require immobilisation of one of the binding partners to onto the surface and the immobilisation moderately affected the binding constant.

4.5.5   ITC analysis of His₆CspE/ss DNA interactions confirm temperature dependence

The interactions of His$_6$$StCspE$ with the high-affinity ligand (5´-GTCTTTT-3´) and various ss DNA sequences (Table III-V) were assayed by ITC to determine solution-based equilibrium constants. ITC directly measures changes in heat during complex formation. A typical calorimetric analysis of His$_6$$StCspE$ binding to ss DNA is shown in Figure 4-6d. The interactions of His$_6$$StCspE$ with all oligonucleotides were exothermic, releasing heat during complex formation. The ITC assay for the His$_6$$StCspE$/5´-GTCTTTT-3´ interaction yielded equilibrium dissociation constants of 8.7 $\pm$ 1.8 nM at 10°C, 11.3 $\pm$ 2.2 nM at 15°C, 27.2 $\pm$ 6.2 nM at 20°C and 44.4 $\pm$ 3.2 nM at 25°C (Table III). Within experimental error, these affinity values are similar to the $K_{d(kin)}$ values determined by SPR (Table I). There are strong entropic and enthalpic contributions for the His$_6$$StCspE$/5´-GTCTTTT-3´ interaction which counterbalances the decrease in $\Delta G°$ (Figure 4-6e and Table III), indicating a change in heat capacity ($\Delta C_p°$). A plot of $\Delta H°$ $versus$ T is linear and from the slope of the line, $\Delta C_p°$ was calculated to be -0.7 $\pm$ 0.1 kcal K^{-1} mol^{-1} (Figure 4-6e), which is in good agreement with SPR data (Table II). The temperature dependence seen for the

His$_6$StCspE/5´-TTCTTTT-3´ interaction was more extreme with the K$_d$ increasing from 19.7 ± 2.1 nM at 10°C to 584.8 ± 76.7 nM at 37°C (Table IV), highlighting the importance of a guanine base at position one.

4.5.6 Analysis of His$_6$-CspE/heptanucleotide interactions identifies base specific sub-sites

A set of heptanucleotides (Table V) were used to analyse sequence specific binding of ss DNA to His$_6$StCspE and to highlight the importance of a key cytosine base at position 3 and a guanine at position 1 within the consensus binding sequence, 5´-GTCTTTT-3´ (Figure 4-6c). The guanine base preference at position 1 and the cytosine preference at position 3 were examined by systematically altering the base at position 1 or 3 for a thymine. The binding preference experiments showed that there was a preference for G at position 1 and C at position 3, with a maximum 7-fold increase seen in the K$_d$ for ON:TTTTTTT compared with the consensus sequence, ON:GTCTTTT. The affinities and stoichiometries of StCspE binding to the control Y-Box, 5´-A<u>ATTGG</u>A-3´ [86], and a polyadenine sequences were also determined. The apparent K$_d$ for the StCspE/5´-A<u>ATTGG</u>A-3 interaction was 1.7 ± 0.05 μM with a stoichiometry of 0.4 ± 0.09 (two StCspE to one ss DNA molecule) and minimal binding was detected for the StCspE/5´-AAAAAAA-3´ interaction (Table V).

4.5.7 Stoichiometries of His$_6$-CspE/ss DNA complex formation identifies the protein-oligonucleotide assemblies formed in solution

The binding stoichiometries were calculated to be between 0.94 and 1.02 to 1 (Table III) for the His$_6$StCspE/5´-GTCTTTT-3´ binding, indicating that each

heptanucleotide molecule populated a single site per His_6-CspE molecule, which was consistent with the 1:1 binding mechanism predicted by SPR analysis. The binding of His_6-CspE to various ss DNA sequences was also examined and the results are presented in Table V.

The X-ray structures of *Bacillus caldolyticus* Csp/T_6 (BcCSP/T_6) and BsCspB/T_6 complex have recently been published [14, 25] and a high degree of conservation is observed in the DNA binding site of His_6-CspE (Figure 4-3c). The BsCspB/T_6 crystal structure clearly shows that a single poly T oligonucleotide binds two protein molecules, resulting in a long continuous chain (Figure 4-7d). The authors assumed that this was a direct result of crystal packaging since a singe hexathymidine molecule bound to one globular functional unit in the BcCSP/T_6 crystal structure (e.g. Figure 4-7c), which has an identical DNA binding site. The His_6-CspE/5´-TTTTTTT-3´ complex must form a similar assembly in solution (Figure 4-7c) as the binding stoichiometry was 0.92 ± 0.02.

The binding stoichiometries for $His_6 St$CspE with 5´-TTCTTTTT-3´and 5´-GTCTTTT-3´were calculated to be 0.92 + 0.02 to 1 and 0.98 ± 0.01 to 1 respectively, consequently forming similar assemblies in solution (Figure 4-7a and 1b, respectively). From the ITC results the binding of 5´-GTTTTTT-3´ to His_6-CspE was clearly not a simple 1:1 binding interaction (Figure 4-7g), as binding occurred in two stages. There was an initial strong interaction (Figure 4-7g (1)) with a stoichiometry of ~0.5 (2 DNA molecules to 1 His_6-CspE), which may form the initial complex shown in Figure 4-7e. After the reaction reached baseline a weaker secondary binding event occurred (Figure 4-7g (2)), which may be caused by a

concentration dependent switch in the binding mode (Figure 4-7f), with a final stoichiometry of approximately 1:1. Due to the sequential binding observed for the His_6-CspE/5´-GTTTTTT-3´ interaction the resulting data could not be fitted accurately to a binding model.
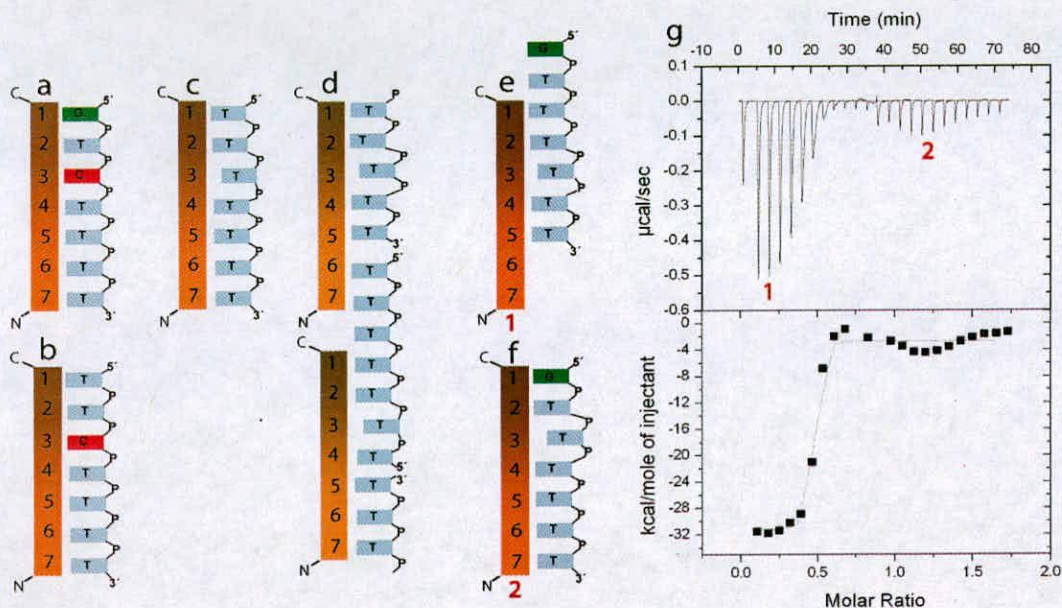
Figure 4-7. **Schematic overview of *St*CspE/oligonucleotide assemblies.**

Protein molecules (brown – orange objects) interact with nucleobases (boxed characters) from oligonucleotide molecules (-P-) at distinctive subsites (1-7). In solution, all seven subsites are occupied by either a single 5´-GTCTTTT-3´ **(a)**, 5´-TTCTTTT-3´ **(b)** or 5´-TTTTTTT-3´ **(c)** oligonucleotide molecule. **(d)** A continuous binding arrangement, as seen in the *Bs*CspB/T_6 crystal structure. **(g)** Isothermal titration calorimetry for 5´-GTTTTTT-3´ binding to recombinant His_6CspE. Each exothermic heat pulse (upper panel) corresponds to an injection of 5 µl of oligonucleotide (100 µM) into the cell containing 5 µM His_6CspE at 15°C. Integrated heat data (lower panels) constitutes a differential binding curve, which could not be adequately fitted to a theoretical binding model. The heat of the reaction was obtained by integrating the peak after each injection of ss DNA oligonucleotide and correcting for the heats of dilution using the Origin® V.7 software. The resulting curve (lower panel) clearly had 2 binding steps, one strong (1 and e), where n = ~0.5, and then shifting to a weaker interaction (2 and f), n = ~1.

4.6 Summary of CspE/ss DNA interactions

To summarise, both SPR and ITC show that StHis$_6$-CspE binds with highest affinity $(K_{d(ITC)} = 8.7 + 1.8$ nM and $K_{d(SPR-Kin)} = 11.4 \pm 4.1$ nM at 15°C) to the heptanucleotide, 5´-GTCTTTT-3´, with a stoichiometry of ~1:1, which fits with results obtained for the homologous protein, BsCspB [174]. ITC revealed temperature dependent changes in the relationship between ΔH, $T\Delta S$ and ΔCp. SPR experiments complemented values determined by ITC. SPR results show that the increase in K_d, as a result of temperature, was essentially due to the large increase in the k_d, which increased 32-fold from 0.024 ± 0.0054 to 0.774 ± 0.14 s^{-1}, from 10 to 30°C respectively.

The differences observed in solution-based equilibrium constants obtained for StCspE binding to various heptanucleotides by ITC (Table V) provides confirmation that there is a strong 1 to 1 binding preference for the 5´-GTCTTTT-3´ sequence. A cytosine at position 3 and guanine at position 1 are important for the high-affinity binding. Similar results were obtained elsewhere for BsCspB binding to heptanucleotides [14, 174]. The cytosine nucleobase at position three is a key recognition/landing point for strong 1 to 1 His$_6$CspE/ss DNA interaction as replacement with a thymine caused a 7-fold reduction in K_d (Table V).

The amino acids involved in CSP/ss DNA binding [14] are highly conserved in StCspE (Figure 4-3a). The nucleotide-binding interface for CSPs appears to be a highly conserved platform and shows no conformational change when DNA is bound [14] and unbound [119] (discussed further in chapter five). ITC and SPR experiments have revealed that binding mode and preference for StCspE is identical to BsCspB [14, 174]. We can assume that highly homologous CSP found in bacteria are similar

and could be attributed to similar functions, a finding that has been already demonstrated for the CSP paralogues in *B. subtilis* [74] and *E. coli* [69].

4.7   SPR: Analysis of CspD/5′-ACGGGG-3′ interactions

His$_6$StCspD was immobilised onto the surface of a Ni-NTA biosensor chip and the previously identified consensus sequence, 5′-ACGGGG-3′, was injected over the surface to generate a binding response. Binding data for the His$_6$StCspD/5′-ACGGGG-3′ (Figure 4-8c) and His$_6$StCspD/5′-GTCTTTT-3′ (Figure 4-8a) interactions were collected at 15°C on the biosensor. The response data for the binding of oligonucleotides to the surface of immobilised His$_6$StCspD were fit to a 1:1 binding model. The R$_{max}$ values were locally fitted, providing an excellent fit to the data, as shown by the overlay of the simulated binding responses (black lines in Figure 4-8a and c) with the experimental data (red lines in Figure 4-8a and c).

The SPR analysis of the His$_6$StCspD/5′-ACGGGG-3′ and His$_6$StCspD-5′-GTCTTTT-3′ interaction revealed apparent K$_{d(kin)}$ values of 109.4 nM and 75.5 nM, respectively at 15°C. The K$_d$ for both interactions are approximately the same but the kinetics of the interactions (only visible using SPR) are completely different. Visual inspection of the binding sensorgram in Figure 4-8d is sufficient to conclude that His$_6$StCspD exhibits slower dissociation and association rates for the 5′-ACGGGG-3′ interaction compared to the 5′-GTCTTTT-3′ interaction. The kinetic values obtained for the His$_6$StCspD/5′-ACGGGG-3′ interaction show a ten fold decrease in k$_a$ and k$_d$ compared to the His$_6$StCspD/5′-GTCTTTT-3′ interaction. Saturating amounts of 5′-ACGGGG-3′ could not be obtained on the His$_6$StCspD surface due to the slow k$_a$ of the His$_6$StCspD/5′-ACGGGG-3′ interaction and therefore K$_{d(eq)}$ values could not obtained.
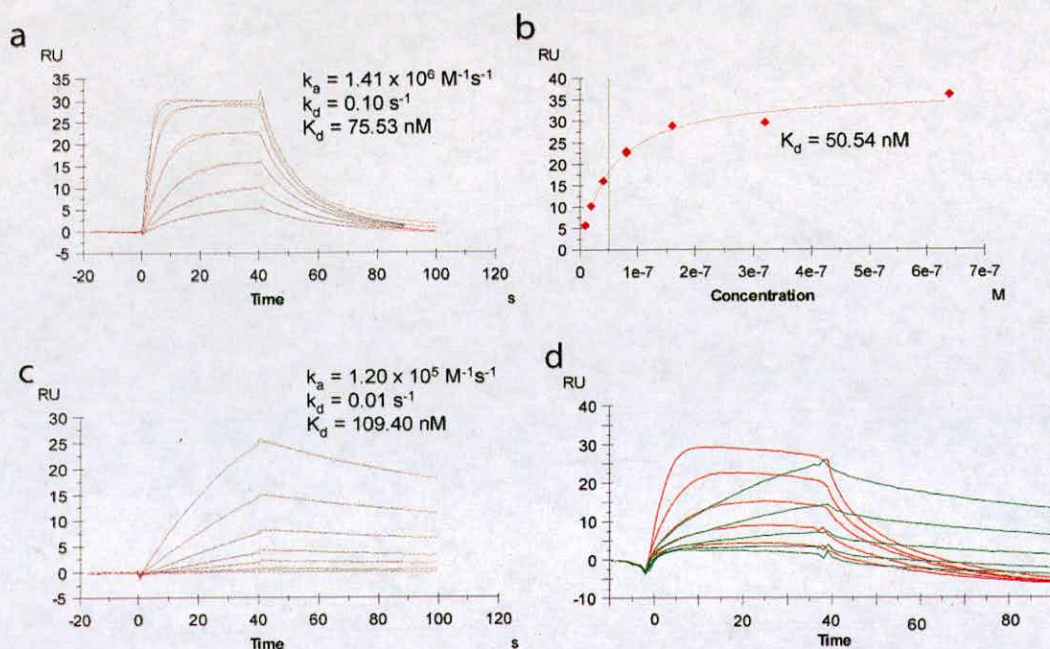
Figure 4-8. The characterization of the His$_6$*St*CspD-ss DNA interactions from a NTA-Ni-His$_6$CspD surface.

Reference-corrected SPR binding curves (red), monitored on a surface with 500 RU of NTA-Ni stabilized His$_6$*St*CspD, for various concentrations of 5′-GTCTTTT-3′ **(a)** (serial dilution from 160 nM to 10 nM) and 5′-ACGGGG-3′ **(c)** (serial dilution from 40 – 640 nM) at 15°C in HBS-P$^+$. The apparent dissociation rate constant (k$_d$) and association rate (k$_a$) constants were determined by globally fitting (black) a simple 1:1 interaction model, A + B = AB, to the sensorgram using the software supplied with the instrument. **(b)** Response *versus* concentration was plotted for the control reaction **(a)** and the data was fitted using a 1:1 interaction model. The apparent disassociation constant is marked on the graph with a vertical line. **(d)** Sensorgram overlay of **(a - red)** and **(c - green)**. Time is shown in seconds.

The Ni-NTA-His$_6$*St*CspD/ss DNA interactions were reproducible at 15°C, however the problems associated with Ni-NTA-His$_6$ immobilisation (see section 1.5.3.4), when obtaining viable kinetic and thermodynamic data, made the use of more covalent like immobilisation methods a compulsory step. Therefore, biotinylated oligonucleotides immobilised to a streptavidin sensor chip surface were decided upon, as this would provide a stable surface and more accurate results, especially at higher temperatures.

137

4.7.1 SPR: Streptavidin immobilisation of the 5′-ACGGGG-3′ sequence

To expand the analysis of DNA sequence recognition and further elucidate the mechanism of action of His₆StCspD, 5′-biotiylated oligomers (5′-ACGGGG-3′ and 5′-GTCTTTTT-3′) were immobilised to the surface of a streptavidin chip.

The SPR analysis of the His₆StCspD/5′-ACGGGG-3′ interaction clearly shows that it is not a 1:1 interaction, as the RUmax (400 RU) was ~3 times greater than the expected RUmax (as calculated by Equation 1: **RUmax = RU = (MW analyte / MW Ligand) x RU ligand immobilised**) of 130 RU (Figure 4-9a). The previously observed kinetics of the His₆StCspD/5′-ACGGGG-3′ interaction (Figure 4-8c) completely changed (Figure 4-9a). The resulting apparent K_d for the His₆StCspD/5′-ACGGGG-3′interaction was extremely weak with a K_d of 31.25 µM, previously results gave a value of 109 nM (Figure 4-8c). The K_d was even weaker for the His₆StCspE/5′-ACGGGG-3′ interaction, with values in the mM range (Figure 4-9c). The discrepancies seen in the results obtain here and when His₆StCspD was immobilised to the chip surface could be caused by; (1) the restricted movement of DNA immobilised to the chip surface is preventing the DNA from binding correctly to His₆StCspD; (2) the manufactured DNA sequence may be incorrect or partially degraded; (3) the hexahistidine tag may be partially binding/blocking the DNA binding site when His₆StCspD is free in solution. Unfortunately to trouble-shoot each of the possible causes would be extremely time consuming and expensive so a new method of protein immobilisation was devised.
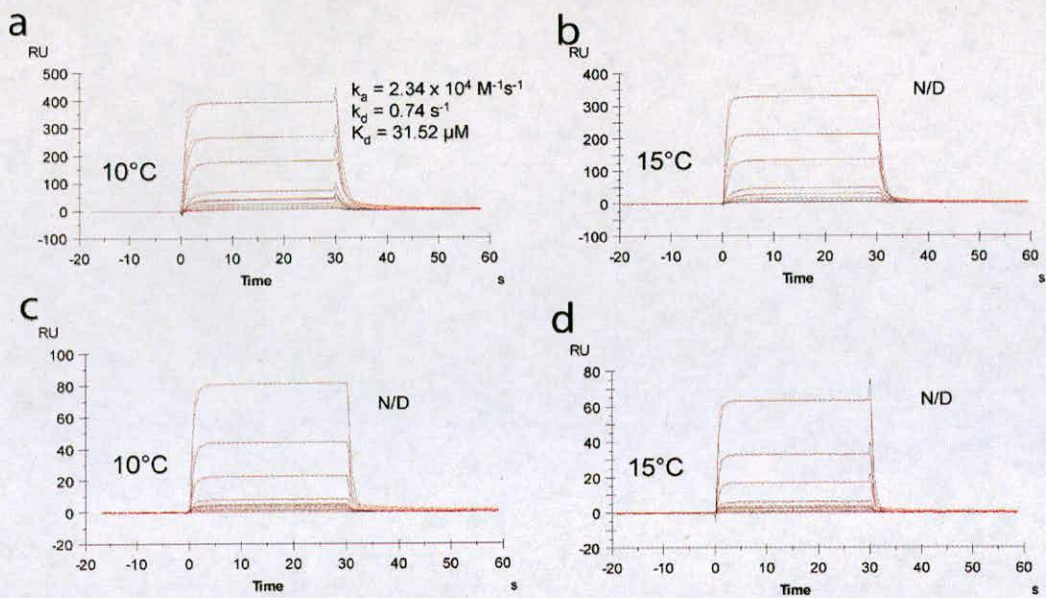
Figure 4-9. **The characterization of His₆*St*CspD/ss DNA interactions on a biotin-ss DNA surface.**

Reference-corrected SPR binding curves (red), monitored on a surface with 65 RU of non-covalently stabilized 5´-ACGGGG-3´, for various concentrations (20 µM, 10 µM 5 µM, 1µM, 500 nM, 250 nM, 125 nM, 60 nM, 30 nM) of His₆*St*CspD (a and b) and His₆CspE (c and d) at 10 and 15°C in HBS-P⁺ buffer. The apparent dissociation rate constant (K_d) and association rate (K_a) constants were determined by globally fitting (black) a simple 1:1 interaction model, A + B = AB, to the sensorgram using the software supplied with the instrument. Time in seconds.

4.8 Double-Hexahistidine tag increases Ni^{2+}-His_6 affinity

A novel strategy for the effective immobilisation of proteins onto the surface of an NTA chip was recently developed [195]. The method employs a double-hexahistidine tag sequence, which showed one order of magnitude stronger binding to Ni-NTA-modified Biacore surfaces than a conventional single-hexahistidine tag. This increase in affinity is not only enough to eliminate the baseline drift seen in SPR experiments (Figure 4-4) but makes protein purification a one-step procedure.

4.8.1 Details of pC_Term_2xHis vector construction

A C-terminal double-His_6-tagged vector sequence was cloned into a pET28a vector (Novagen) as described (section 4.4.6). The sequence (Figure 4-10) highlighted in grey (37 bases) was copied from the original pET28a vector; this contains an Xba restriction site (red) and a ribosome binding site (rbs) (bold). Two restriction sites, NdeI (blue) and HindIII (green), were added immediately after the rbs. The section of the multiple cloning site (MCS) highlighted in green corresponds to part of the MCS from pET28a, which ranges from BamHI to SalI (4 restriction sites in total = 25 bases). A NotI restriction site (red) was then added to the 3' end of the multiple cloning site. The double-His_6-tagged coding sequence (highlighted in purple) was manually constructed using codons from the StCspE sequence (A1.1). The amino acid sequence used here for the double-His_6-tagged region was determined originally in the Protein Technologies Laboratory [195]. A XhoI site (grey) was added to the end of the C-Terminal double-His_6-tagged vector sequence to allow for directional ligation of the sequence into *XbaI-XhoI*-digested pET28a vector. Except for the insertion of the C-Terminal double-His_6-tagged sequence between the *XbaI* and *XhoI* restrictions sites the pET28a vector was not modified in any other way.

```
1     CCC  TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT AGG CTA   45
1     Pro  Ser Arg Asn Asn Phe Val End Leu End Glu Gly Asp Arg Leu   15

46    GCA AGC TTC GGA TCC GAA TTC GAG CTC CGT CGA CGC GGC CGC CGC    90
16    Ala Ser Phe Gly Ser Glu Phe Glu Leu Arg Arg Arg Gly Arg Gly    30

91    CGA GGC AGC GGT GGT GGA AGC GGC GGT GGA ACA GGC GGA GGA AGC    135
31    Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Thr Gly Gly Gly Ser    45

136   GGC GGT GGA AAA CGC GCT GAT GCC GCT CAT CAT CAC CAT CAT CAT    180
46    Gly Gly Gly Lys Arg Ala Asp Ala Ala His His His His His His    60

181   TCT CGC GCT TGG CGC CAC CCG CAG TTC GGT GGT CAC CAC CAC CAC    225
61    Ser Arg Ala Trp Arg His Pro Gln Phe Gly Gly His His His His    75

226   CAC CAC TGA ACT CGA G AT   240
76    His His End Thr Arg
```

Figure 4-10. **The C-terminal double-hexahistidine tag sequence.**
Refer to text for a detailed description of the vector sequence.

4.8.2   *Amplification and cloning of the C-Terminal double his-tag sequence*

Specific oligonucleotide primers containing *XbaI* and *XhoI* restriction sites were used to create the C-terminal double-His$_6$ construct using overlap extension and PCR techniques (section 4.4.7). The resulting product was approximately 0.25 Kb (242bp) in size (Figure 4-11). The C-Term PCR product (Figure 4-11, lane 2) was digested using *XbaI* and *XhoI* restriction enzymes and purified from an agarose gel. The resulting fragment was ligated into *XbaI-XhoI*-digested pET28a vector using a rapid ligation kit (Roche). The resulting plasmids were then transformed into *E. coli* DH5-α. Transformants were isolated on L.B agar plates containing Kanamycin (50µg/ml). Plasmid DNA was isolated from several transformants and restriction analysis was carried out to confirm the presence of the C-Terminal DNA fragment (Figure 4-12a and b). Positive clones were sequenced and confirmed (Appendix 1.4) and the final vector is now referred to as pC-Term_2xHis.
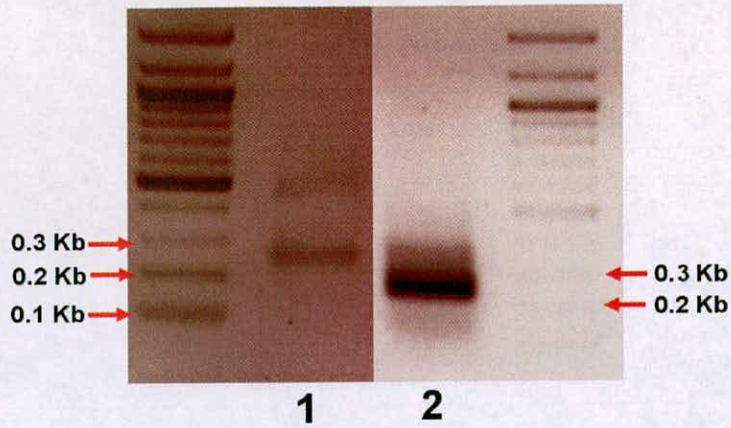
Figure 4-11. Overlap extension PCR of the C-Terminal double-His$_6$ sequence.

Lanes:
1) C-Terminal double-His$_6$ sequence
2) C-Terminal double-His$_6$ sequence after a further PCR step (section 4.4.7).
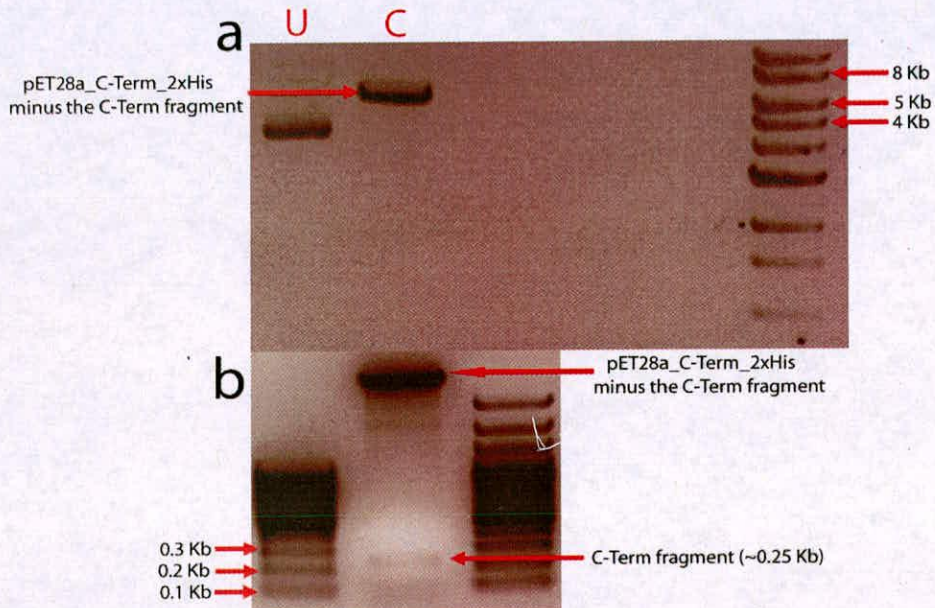
Figure 4-12. Confirmation of the pC-Term_2xHis vector sequence by restriction digest experiments.

(a) Agarose gel electrophoresis of *XhoI* and *XbaI* digested (C (linearised)) and undigested (U (coiled)) plasmid DNA isolated from *E. coli* DH5-α carrying the plasmid pET28a_C-Term_2xHis. (b) A 250 bp fragment (visualised by increasing the gain) resulting from *XhoI-XbaI* digested pET28a_C-Term_2xHis DNA. The resulting 5.2 kb and 0.25-kb fragments indicate that the C-Terminal insert is present.

142

4.8.3 Amplification and cloning of CspE and CspD double his-tag genes

Specific oligonucleotide primers containing *NotI* and *HindIII* restriction sites were used to amplify the open reading frames of the CspE and CspD genes from parental vectors, pET28a_CspE and pET28a_CspD (Figure 4-13).
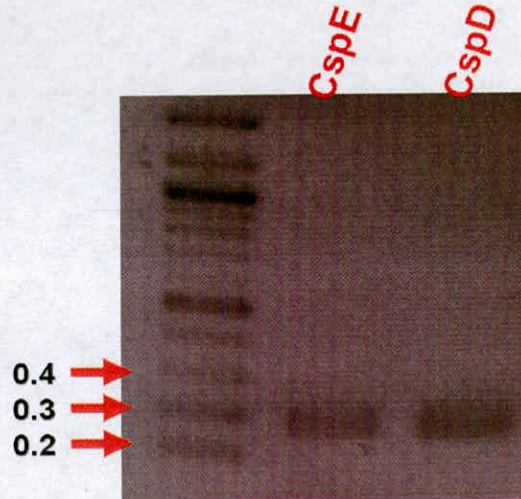
Figure 4-13. **PCR amplification of the *CspE and CspD* genes from pET28a_Csp DNA.**

100-bp DNA ladder used.

The amplification products were then ligated into the plasmid pCR®2.1-TOPO® and the resulting plasmid was transformed into chemically competent *E. coli* TOP10 cells. Transformants were isolated on LB agar plates containing X-gal (40 µg/ml) and kanamycin (50 µg/ml). Positive transformants were isolated using standard blue-white screening. Plasmid DNA was isolated from several transformants and digested with *NotI* and *HindIII* to confirm the presence of each *CSP* gene. The 0.25 Kb restriction fragments were purified from the agarose gel and ligated into *NotI-HindIII*-digested pC_Term_2xHis vector, upstream of the double-hexahistidine coding region (A1.4). Gel electrophoresis showing a restriction digest (*XbaI* and *XhoI)* of pC_Term_2xHis plasmid DNA isolated from transformants can be seen in

Figure 4-14. The correct sequences were verified by sequencing and are now referred to as; pC_Term_2xHis_CspE and pC_Term_2xHis_CspD.
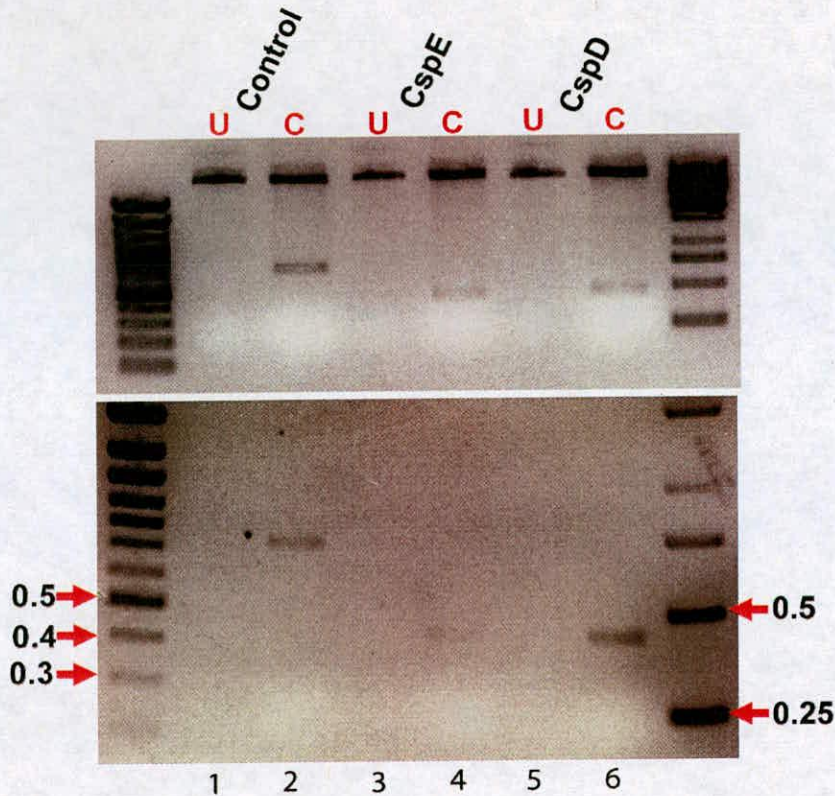
Figure 4-14. **Confirmation of the pC_Term_2xHis_CSP vector sequences by restriction digest experiments.**

Agarose gel electrophoresis of *XbaI* and *XhoI* digested (C) and undigested (U) plasmid DNA isolated from *E. coli* DH5-α carrying the plasmids pC_Term_2xHis_CSPE and D. Lanes: (3 and 5) plasmid DNA isolated from *E. coli* DH5-α containing the appropriate plasmid; (4 and 6) Plasmid DNA digested with *XbaI* and *XhoI*. The resulting 5-kb and 0.4-kb indicates that the *Csp* genes are present. 100 bp and 1 Kb MW ladder were used.

4.8.4 *Purification of double-His$_6$-tagged CSPs*

4.8.4.1 Purification of double-His$_6$-tagged CspE

Affinity purification from bacterial cell lysate of double-His$_6$-tagged *St*CspE on Ni-NTA agarose gave two elution peaks at 166 mM and 225 mM imidazole (Figure 4-15a). Samples from both peaks were analysed by SDS-PAGE and only peak 2

144

contained a protein corresponding to the size double-His$_6$-tagged *St*CspE (12.5 kDa), peak 2 was pooled and concentrated for further analysis. To remove bound nucleic acids, the protein sample was unfolded in 8 M urea overnight and refolded using size exclusion chromatography (Materials and Methods section 4.4.10). The purified protein was found to be monomeric, as determined by analytical gel filtration (Figure 4-15c) and had an experimentally determined molecular weight of 12.3 kDa (Figure 4-15c), which corresponds to the calculated molecular weight of double-His$_6$-tagged *St*CspE.
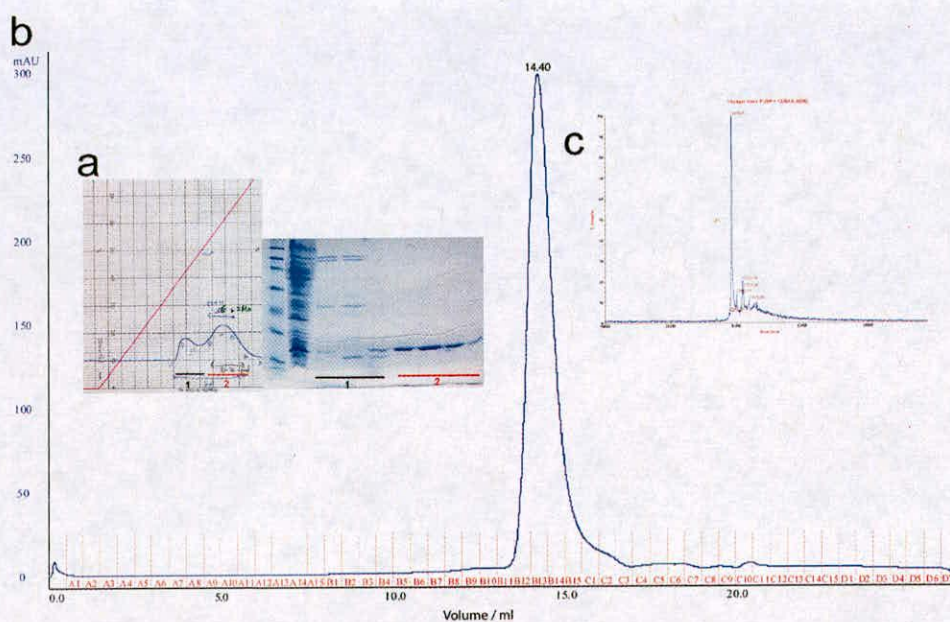
Figure 4-15. **Affinity purification of double-His$_6$-tagged *St*CspE on a Ni-NTA agarose column.**

Protein concentration was monitored by absorbance at A$_{280}$. **(a)** A gradient elution of 0-0.5 M imidazole (red line) gave two elution peaks, 1 and 2, which can be seen on the adjacent SDS-PAGE gel. **(b)** Elution profile of 87 µM double-His$_6$-tagged CspE as monitored by UV adsorption from a HR 10/30 Superdex-75 column. The peak retention volume was 14.4 ml which corresponds to an apparent molecular mass of 12 kDa, indicating double-His$_6$-tagged *St*CspE is monomeric (actual MW = 12.5 kDa). **(c)** MALDI-TOF spectrum for the double-His$_6$-tagged *St*CspE protein from 10 to 18 kDa, the Y axis is the relative intensity generated in the MALDI-TOF mass spectrometer. A clear mass peak resulted at 12305 Daltons (calculated mass for double-His$_6$-tagged *St*CspE is 12478 Daltons (A1.4).

4.8.4.2 Purification of double-His$_6$-tagged *St*CspD

Double-His$_6$-tagged *St*CspD from bacterial cell lysate was bound to a Ni-NTA agarose column. To remove bound nucleic acids, bound protein was slowly unfolded over a 10 column volume gradient (0-100%) with a buffer containing 6 M guanidine-hydrochloride, 20 mM HEPES, pH 8.0. The protein was refolded over a 10 column volume gradient with a buffer containing 20 mM HEPES pH8.0, 150 mM NaCl (100% 6 M guanidine-hydrochloride – 0% or 100 % 20 mM HEPES pH8.0, 150 mM NaCl buffer). The protein was purified from Ni-NTA agarose as described (section 4.4.11) and gave an elution peak at 250 mM imidazole. Samples containing double-His$_6$-tagged *St*CspD were concentrated and loaded onto a Superdex-75 column.

The purified protein was found to be monomeric, as determined by analytical gel filtration (Figure 4-16a), highly pure (Figure 4-16c) and was of a similar molecular weight to double-His$_6$-tagged CspD, as determined by MALDI-TOF mass spectrometry (Figure 4-16b).
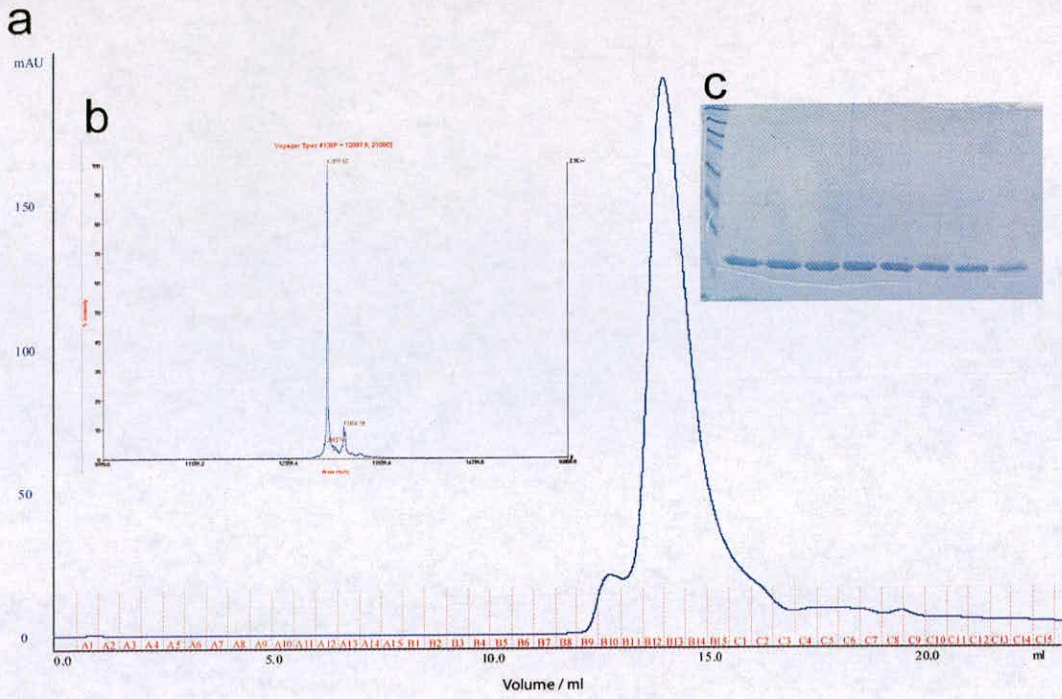
Figure 4-16. **Affinity purification of double-His$_6$-tagged *St*CspD.**

(a) Elution profile of double-His$_6$-tagged *St*CspD as monitored by UV adsorption from a HR 10/30 Superdex-75 column. Peak retention volume was 14.3 ml, which corresponds to an apparent molecular mass of 12.5 kDa, indicating double-His$_6$-tagged *St*CspD is monomeric (12.9 kDa). **(b)** SDS-PAGE analysis of elution peak. **(c)** MALDI-TOF spectrum for the double-His$_6$-tagged *St*CspD protein from 10 to 16 kDa, the Y axis is the relative intensity generated in the MALDI-TOF mass spectrometer. A clear mass peak resulted at 12907 Daltons (calculated mass for double-His$_6$-tagged *St*CspD is 12912 Daltons (A1.4).

4.8.5 Double-His₆-tagged CSPs show no base-line drift on a Biacore NTA chip

Binding curves were obtained with a Biacore NTA sensor chip. Figure 4-17a shows the binding and dissociation of single-His$_6$StCspD at various concentrations. The levels of protein immobilised were directly proportional to the concentration of protein injected. Protein was removed from the surface using 350 mM EDTA. Figure 4-17b shows the binding curves of double-His$_6$StCspD, where the dissociation rate was slower for most concentrations. Figure 4-17c compares dissociation curves for single-His$_6$-tagged StCspD (red) and double-His$_6$-tagged StCspD (green) at 50 to 400 nM, showing a nearly non-existent dissociation rate of double-His$_6$StCspD (7.1 x 10^{-7} M^{-1}s^{-1}) compared to the quicker dissociation rate of single-His$_6$StCspD (1.1 x 10^{-2} s^{-1}).
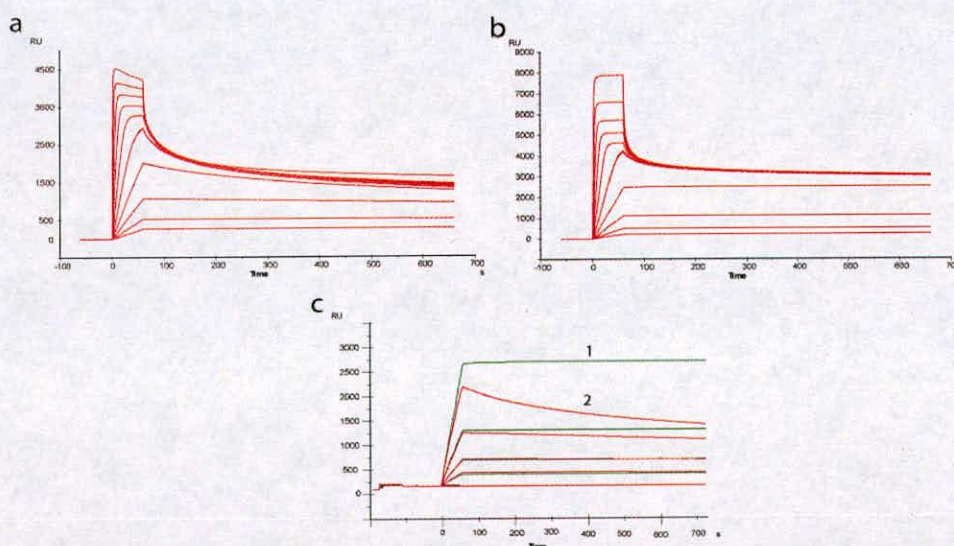
Figure 4-17. SPR analysis of binding of single and double-His₆-tagged CSPs.

Sensorgrams showing the binding to a Ni-NTA Biacore chip surface of **(a)** single-His$_6$-tagged StCspD and **(b)** double-His$_6$-tagged StCspD at different concentrations (25600, 12800, 6400, 3200, 1600, 800, 400, 200, 100, 50 nM) at 25°C. **(c)** Overlay of single-His$_6$-tagged CspD (red) and double-His$_6$-tagged StCspD (green) sensorgrams for 50, 100, 200 and 400 nM (1 and 2). Note that single-His$_6$-tagged StCspD dissociated rapidly over 10 minutes (2), whereas double-His$_6$-tagged StCspD shows no dissociation.

4.8.6 Regeneration of the double-His₆-tagged CSP chip surface

Both double-His$_6$-tagged *St*CspE and *St*CspD were immobilised onto the surface of a Ni-NTA biosensor chip and the previously identified consensus ss DNA sequences were injected over the surface to generate a binding response. Example binding data for a double-His$_6$-tagged CSP/ss DNA interaction at 15°C is shown in Figure 4-18. The sensorgrams could not be fitted to any binding model due to the elevated baseline drift observed after ss DNA binding. The observed upward drift was caused by insufficient regeneration of the chip surface between each cycle. A standard 350 mM EDTA injection over the surface was inadequate for the complete removal of nickel bound double-His$_6$-tagged CSP from the surface, resulting in a gradual residual increase of active CSP on the chip surface and worthless data.
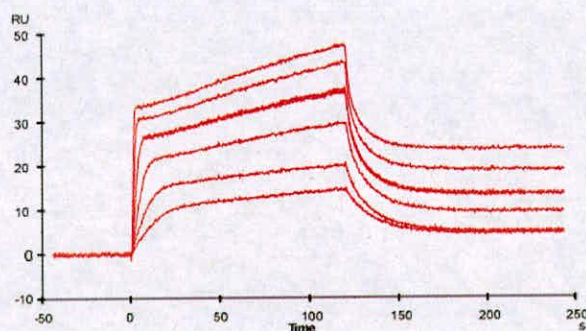
Figure 4-18. The initially characterization of double-His₆-tagged CSP/ss DNA interactions.

SPR binding curves (red), monitored on a surface with 400 RU of NTA-Ni stabilized double-His$_6$-tagged *St*CspE, for various 5′-GTCTTTT-3′ concentrations (serial dilution from 15.625 nM to 500 nM) at 15°C in HBS-P$^+$.

A buffer containing a high concentration of imidazole (0.5 M) at a constant pH was used to displace the histidine tag from the nickel particles, consequently regenerating the chip surface efficiently between assay cycles at 10 and 15°C (Figure 4-19a-b). Results obtained for 20°C and above were not reproducible (Figure 4-19c) due to an

increase in the rate of nickel dissociation from the NTA surface with increasing temperature. Without the replenishment of nickel on the chip surface between assay cycles (at temperatures above 20°C) the capture levels of protein on the surface were erratic and satisfactory results could not be obtained.
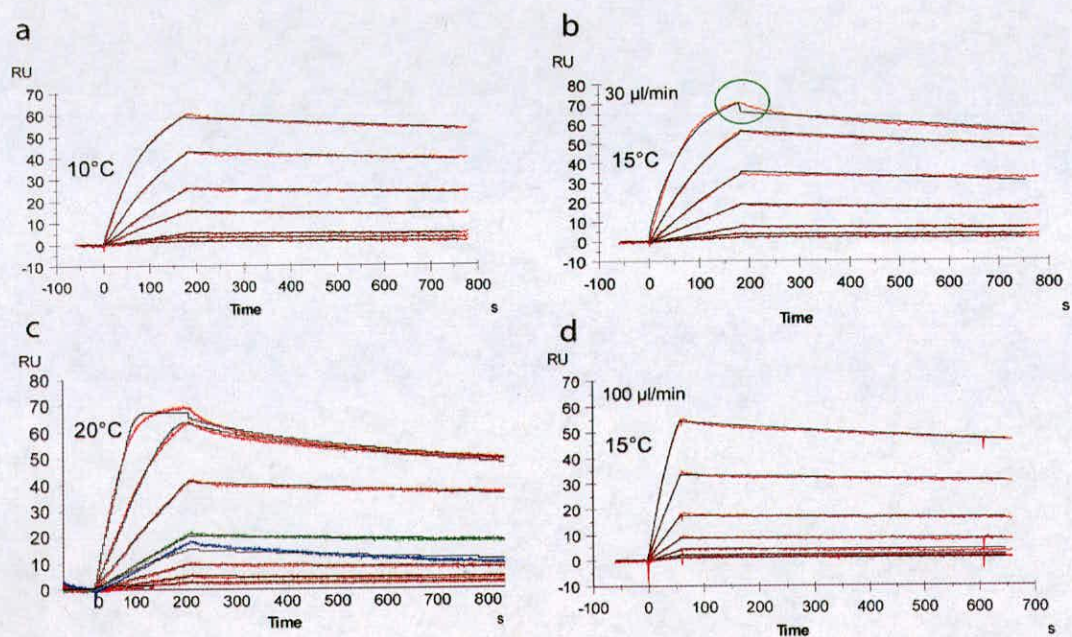
Figure 4-19. Characterization of the double-His$_6$-tagged *St*CspD/5´-ACGGGG-3´ interactions.

(a-d) Reference-corrected SPR binding curves (red), monitored on a surface with 350-400 RU of NTA-Ni stabilized double-His$_6$-tagged *St*CspD, for various 5´-ACGGGG-3´concentrations (serial dilution from 1000 nM to 15.625 nM) at the specified temperature in HBS-P$^+$. The apparent dissociation rate constant (K_d) and association rate (K_a) constants were determined by globally fitting (black) a simple 1:1 interaction model (A + B = AB) to the sensorgrams using the software supplied with the instrument. **(c)** A repeat concentration of 125 nM 5´-ACGGGG-3´ ran during (green) and after (blue) the concentration series. **(d)** The same assay as (b) but ran at 100 µl.min^{-1}, reducing the affects of mass transport (green circle, Fig. b).

The sensorgrams obtained for the double-His$_6$-tagged *St*CspD/5´-ACGGGG-3´ interactions at 15°C exhibited characteristics associated with mass transfer (Figure 4-19b – green circle) at the experimental flow rate used (30 µl.min^{-1}). Although the

150

software supplied with the Biacore instrument was able to compensate for the mass transfer issue, an increased flow rate of 100 μl.min^{-1} was used to eliminate the affects completely (Figure 4-19d).

To stabilize the surface at higher temperatures and obtain a complete thermodynamic profile for the StCspD/3´-ACGGGG-5´ interaction, double-His$_6$-tagged StCspD was immobilised onto the dextran layer of a NTA biosensor chip using amine coupling (Data not shown). The direct coupling of the primary amine of an engineered lysine between the double-His$_6$-tag, (Figure 4-10) to the dextran layer, using a slightly modified protocol [147] eliminated the coupling of lysine within the active site of the protein (see section 4.5.2.2). Although adequate amounts of active StCspD were obtained on the surface, suitable regeneration conditions could not be found. This was most likely owing to a strong hydrophobic nature of the StCspD/ss DNA interaction.

4.8.7   SPR analysis of double His$_6$-tagged CspD-ss DNA interactions

Binding data for the double-His$_6$-tagged StCspD/5´-ACGGGG-3´ interactions were collected at 10 and 15°C only. Binding data could not be obtained for temperatures above 15°C, for the reasons discussed previously (section 4.8.6). A total of three replicates were collected for double-His$_6$-tagged StCspD/5´-ACGGGG-3´ interactions and the values are presented in Table VI. The SPR analysis for the double-His$_6$-tagged StCspD/5´-ACGGGG-3´ interactions revealed little change in the $K_{d(kin)}$ around cold shock temperatures, with an increase of 16.54 ± 2.31 nM at 10°C to 19.16 ± 1.23 nM at 15°C.

Table VI. Binding kinetics determined for double-His$_6$-tagged CspD/5´-ACGGGG-3´ interactions at various temperatures using SPR.

Temp (°C)	k_a (M^{-1}s^{-1})	k_d (s^{-1})	$K_{d(kin)}$ (nM)
10	$(1.04 \pm 0.14) \times 10^4$	$(1.72 \pm 0.24) \times 10^{-4}$	16.54 ± 2.31
15	$(1.33 \pm 0.31) \times 10^4$	$(2.70 \pm 0.22) \times 10^{-4}$	19.16 ± 1.23

SPR K_D values (mean $\pm$ SE, $n = 3$) were obtained as follows; the $K_{D(kin)}$ value is calculated from kinetics binding data, where $K_D = K_d/K_a$. Kinetic binding values were obtained from the global fitting of data from seven different concentrations of 5'-ACGGGG-3´ (typically 15.625 nM, 31.25 nM, 62.5 nM, 125 nM, 250 nM, 500 nM, 1000 nM) fit to a 1:1 interaction model, A + B = AB. All fits had χ^2 values less than 2, which are generally considered an excellent fit.

The K_d for the single- and double-His$_6$-tagged StCspD/5´-ACGGGG-3´ interactions at 15°C were found to be 109.40 and 19.16 nM, respectively. Although these values vary slightly the characteristic kinetics observed (slow k_a and k_d) for the StCspD/5´-ACGGGG-3´ interactions are conserved for both the single- and double-His$_6$-tagged StCspD, indicating that addition of a double-His$_6$-tag has not changed the DNA binding function of StCspD but has improved the stability of the Biacore surface.

4.9 Conclusions: A comparison of StCspD and StCspE binding to ss DNA

Both the initial microarray assay results (Figure 3-11) and biophysical techniques show that StCspD binds preferentially, with a 1:1 stoichiometry and a high affinity (19.16 nM) to the 5´-ACGGGG-3´ sequence at 15°C. The StCspD/5´-ACGGGG-3´ interaction exhibits an extremely slow k_d ($2.70 \pm 0.22 \times 10^{-4}$ s^{-1}) under the experimental conditions used here, which may be a reflection of cellular nucleic acid binding kinetics. StCspD also binds the 5´-GTCTTTT-3´ sequence (StCspE consensus binding sequence) with high affinity (75.53 nM) but with a completely different set of kinetics, exhibiting very fast k_a (1.41×10^6 M^{-1} s^{-1}) and k_d (0.1 s^{-1}) rates at 15°C.

Both the initial microarray assay results (Figure 3-11) and biophysical techniques

show that StCspE binds preferentially, with a 1:1 stoichiometry and a high affinity

(11.27 nM) to the 5´-GTCTTTT-3´ sequence at 15°C. StCspE also bound the 5´-

ACGGGG-3´ sequence (StCspD consensus binding sequence) at 15°C and 37°C

(Figure 4-20) with K_ds of approximately 1 µM and 1 mM, respectively. The

StCspE/5´-ACGGGG-3´ interaction was extremely weak at 37°C, which can be seen

visually in Figure 4-20b. The StCspE/5´-ACGGGG-3´ interaction was noticeable

stronger at 15°C (Figure 4-20a) but was more than just a simple 1:1 interaction. The

interaction could only be fitted to a complex four sequential binding site model,

which is logical as StCspE binds preferential to poly-thymidine sequences (5´-

GTCTTTT-3´ (Figure 3-11)) and not the guanine rich sequence, 5´-ACGGGG-3´,

bound by StCspD.
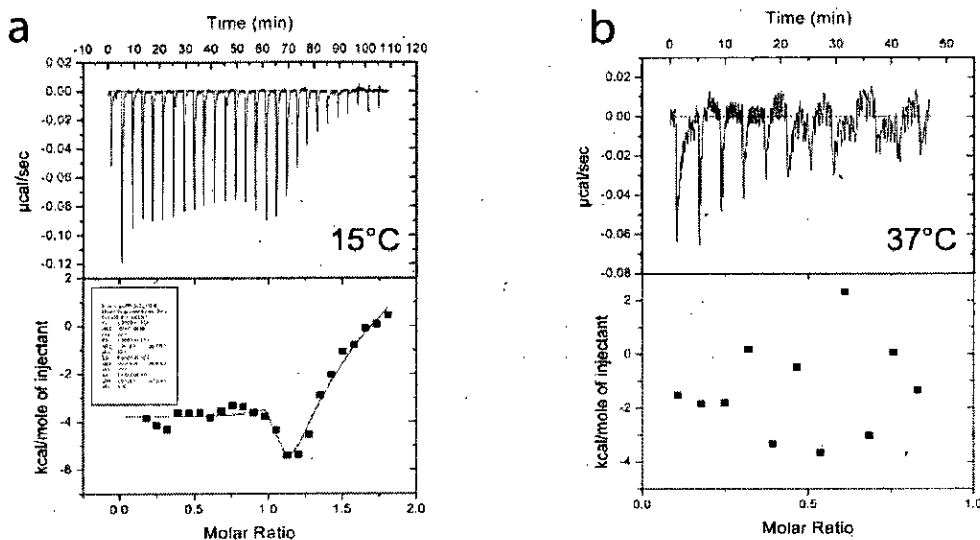
Figure 4-20. **Isothermal titration calorimetry for 5´-ACGGGG-3´ binding to recombinant His₆StCspE.**

Each exothermic heat pulse (upper panel) corresponds to an injection of 5 µl of oligonucleotide (100 µM) into the cell containing 5 µM His₆StCspE at 15°C (a) and 37°C (a). Integrated heat data (lower panels) constitutes a differential binding curve, which was fitted to a sequential binding site model with four sites.

CHAPTER 5. Structural studies of CSPs from *S. typhimurium*.

5.1 Aim

To obtain X-ray crystal structures of *St*CspE and *St*CspD, which were identified as the two definitive CSPs from the *S. typhimurium* paralogues examined, regarding ss DNA sequence specificity (Figure 3-11). To obtain complex structures of *St*CspE/5´-GTCTTTT-3´ and *St*CspD/5´-ACGGGG-3´, which will provide further insight in CSP/ss DNA recognition at a molecular level and explain the binding preferences of these two CSPs, as observed during ITC and SPR experiments.

5.2 Summary

The crystal structure of cold shock protein E form *Salmonella typhimurium* (*St*CspE) has been determined at 1.1 Å resolution and refined to R = 0.203. The three-dimensional structure is similar to previously determined CSPs and is composed of five anti-parallel β-strands forming a classic OB-fold/five-stranded β-barrel. CspE formed a crystallographic dimer via the electropositive DNA binding face, which could suggest a possible self-regulation mechanism, preventing non-specific binding of single stranded nucleic acids within the cell, something that has not been seen before in the structural studies of CSPs. This first structure of a CSP from *S. typhimurium* provides a new insight into the cold shock response in *S. typhimurium* and may lead to the development of potential CSP inhibitors, vital to the food industry (see section 1.4). *St*CspE/5´-GTCTTTT-3´ proved to be difficult to crystallise and as a result the 5´-GTCTTTT-3´ sequence was modelled onto the structure of *St*CspE. A cytosine specific binding pocket was identified at DNA subsite 3, explaining the preference for a cytosine base. The replacement of a

thymine with a guanine at subsite 1 resulted in the gain of a hydrogen bond, explaining the preference for a guanine base at position this position.

Many attempts were made to obtain crystals of *St*CspD but were unsuccessful. A computer generated model was created and used to generate the *St*CspD/5´-ACGGGG-3´ complex model. This model provided explanations for base preference at each of the DNA binding subsites, as observed in previous biochemical experiments.

5.3   Introduction

As discussed in the previous chapters (3 and 4), two different types of recognition sequences have been determined for CSPs. SPR and ITC results have shown that CSPs recognise either purine (e.g. *St*CspD) or pyrimidine (e.g. *St*CspE and *Bs*CspB) rich sequences in a sequence specific binding manner. Examination of the structure of CSPs in a DNA bound and unbound state, at an atomic level, would provide a detailed insight of the CSP/ss DNA interactions. One method that allows for such an analysis is X-ray crystallography.

To date only three cold shock proteins and two ss DNA complexes have been crystallised and their structures determined (Table 5-1), all of which have been produced by Professor Udo Heinemann's laboratory. The next challenge in the current work was to obtain apo crystal structures of *St*CspE and *St*CspD, and their complexes with ss DNA. Both ITC and Biacore studies have shown that the preferential ss DNA binding sequences, 5´GTCTTTT-3´ and 5´-ACGGGG-3´ (as determined by microarray analysis; section 3.6.3, Figure 3-11), bind with low nM affinity to *St*CspE and *St*CspD, respectively. This knowledge provides an advantageous start to crystallising the CSP/ss DNA complexes as the sequence that is

most likely to keep the complex intact throughout the crystallisation process is

known.

Table 5-1. **Crystallisation conditions of cold shock proteins and their DNA complexes**

Protein	Reservoir	Reference	PDB ID
CspA (*E. coli*) 10mg/ml in 10 mM Tris pH 7.5/50 mM NaCl	26% Peg 1500	[118]	1MJC
CspB (*B. subtilis*) 10 mg/ml in 50 mM Tris/HCl pH 8.5 and 100 mM Na-acetate	1.3-1.5 M (NH4)2SO4 or 22-24% Peg 4000 in 50 mM Tris pH 8.5 & 100mM Na-acetate	[119]	1CSP
B. caldolyticus-Csp 10mg/ml in 10 mM Tris-HCl pH 7.5	100 mM HEPES (pH 7.5) and 65 % (v/v) 2-methyl 2,4-pentanediol	[78]	1HZA
CspB (*B. subtilis*) dT6 complex 6-11 mg/ml in 20 mM Tris pH7.5, 50 mM KCl and 3 mM MgCl2	0.1 M Sodium cacodylate pH 6.5, 0.2 M calcium acetate and 18% (w/v) peg 8,000 2 Protein:1 reservoir	[14]	2ES2
Csp (*B. caldolyticus*) dT6 complex in (as above)	35% (w/v) 2-methly-2, 4-pentandiol, 0.1 M sodium acetate and 0.02 M CaCl2	[25]	2HAX

5.4 Materials and Methods

5.4.1 Purification of StCspE

His$_6$StCspE was purified as described previously (section 3.4.5). To obtain adequate amounts of protein for crystallization the refolding procedure was repeated 10-15 times. Protein samples were concentrated and buffer exchanged using standard protocols (PD-10 column; Amersham Biosciences) into the required buffer. For cleavage of the histidine tag the protein was buffer exchanged (PD-10 column) into buffer E (20 mM Tris, pH 7.5). 1.5 units of thrombin protease was added per mg of His$_6$StCspE and incubated at RT for 4 hrs with gentle agitation. Thrombin was removed by loading the protein sample onto a self-packed 0.5 ml benzamidine gravity flow column, pre-equilibrated in a buffer E. The protein was eluted by adding 2 ml of buffer E to the column. To remove the free histidine-tag the protein was concentrated to 20 mg/ml and 200 µl of protein sample was loaded onto a HiLoad 26/60 Superdex 75 column, pre-equilibrated with a buffer E. StCspE was quantified using both the Bradford method and the theoretical extinction coefficient, ε280nm = 5500 M^{-1}.cm^{-1}.

5.4.2 Purification of StCspE-ss DNA complex

Concentrated recombinant CspE [~20 mg/ml] was added to HPLC purified oligonucleotide sequence 5′-GTCTTTT-3′ (resuspended in TEN buffer) in a 1:1.3 molar ratio and incubated on ice for 30 minutes. To remove the unbound ss DNA, the complex was concentrated (between 20 and 30 mg/ml) and 200 µl of sample was loaded onto a HiLoad 26/60 Superdex 75 column, pre-equilibrated with TEN buffer.

*St*CspE-ss DNA complex was concentrated and quantified using the Bradford method.

5.4.3   Crystallisation screening procedures

5.4.3.1   Protein Screening:

Protein solution was mixed with equal volumes (1:1 ratio) of reservoir solution, which was used for all screening conditions. Structure screens 1 and 2 (Molecular Dimensions), crystallization screens 1 and 2 (Hampton Research), various precipitate/pH screens including PEG (1.5 K to 20 K) and ammonium sulphate were initially used in the screening of crystallisation conditions at 4°C and 17°C.

5.4.3.2   CSP-ss DNA complex screening:

Protein solution (10 and 20 mg/ml) was mixed with equal volumes (1:1 ratio) of reservoir solution, which was used for all screening conditions. In addition to standard protein screening a Natrix screen (a specialised screen for Protein-DNA complexes from Hampton Research) was also used to screen for crystallisation conditions at 4°C and 17°C.

5.4.4   Data collection and processing

All data were collected at either the European synchrotron radiation facility (ERSF) in Grenoble, France on beamline BM14 or 10.1, SRS, Daresbury, UK. The data collection process is illustrated in Figure 5-1. All data were collected as 1.5° oscillation images with the crystal maintained at 100 K in a nitrogen gas stream and processed with MOSFLM [196]. The data was scaled using SCALA [197].

DATA COLLECTION PROCESS
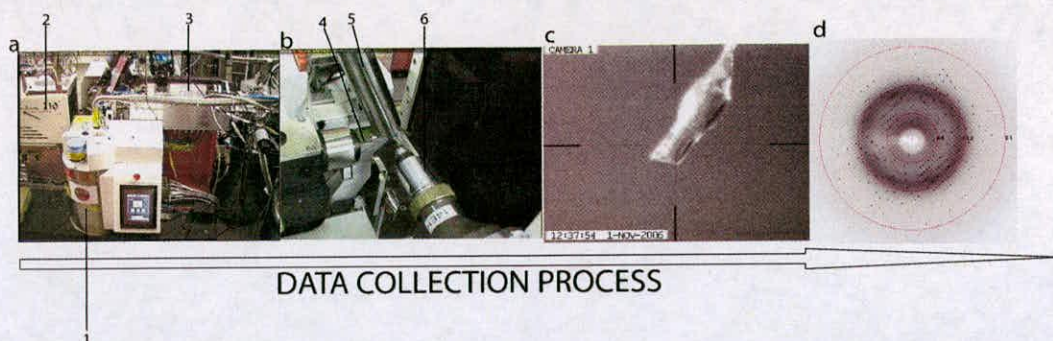
Figure 5-1. **Data collection process at a synchrotron work station.**

Panels are as follows; (a) Hutch view; (b) Crystal mounting area; (c) Crystal in a nylon loop; (d) Diffraction image. Labels; (1) Sample changing robot; (2) CCD X-ray detector; (3) Synchrotron beam source; (4) Pin-hole camera; (5) Nitrogen source (cryostream); (6) Goniometer head.

5.4.5 *Phasing*

Phasing was performed with ACORN [198], using the positions of 10 amino acids (PHE12 to ILE21) from the homologous structure of the *E. coli* CspA previously solved at 2.0 Å resolution [118]. No reflections were omitted for the purpose of calculating the normalized structure factors input to ACORN.

5.4.6 *Model building and refinement*

Using the ACORN phases, an initial structure was built automatically using ARPwarp [199] to give 112 residues out of a total of 144 for the asymmetric unit cell. The structure was edited using the program COOT [200] to add missing residues that were clearly defined in the electron density maps. The model was then subjected to 10 cycles in REFMAC [201] yielding R of 21.10% and R_{free} of 24.02%. After several cycles of REFMAC refinement with isotropic B-factors, including automatic water molecule placement using ARPwarp, manual rebuilding with COOT and construction of alternative side chain conformations, R and R_{free} values decreased to 20.80% and 23.89%, respectively. Anisotropic B-factors were then

refined and additional water molecules were added on the basis of Fo-Fc difference maps. Water molecules were added if the electron density was greater than 2.5 σ and the surrounding hydrogen bonding network made sense (greater than 2.5 Å apart). This resulted in a final model of 138 residues, composed to 2 CspE molecules (Molecule 1/Chain A = SER-1 to LEU 69 and Molecule 2/Chain B = Lys3 to LEU 69) in the asymmetric unit cell, 100 water molecules and a final R and R_{free} value 20.3% and 23.42%, respectively at 1.1 Å resolution. Coordinates for the 1.1 Å structure have been deposited in the Protein Data Bank with the code 2p6q.

5.4.7 Cytosine soaking experiments

Either 2′-deoxycytidine 5′-monophosphate (dCMP) (Biochemika) or 2′-deoxycytidine 5′-triphosphate (dCTP) (Invitrogen) was used for soaking experiments with CspE. Lyophilised dCMP (32 mM) and dCTP (100 mM) were dissolved in 28% PEG 20,000, 0.05 M AMPSO, pH 9 and used to prepare a series of dilutions for soaking experiments, which were carried out at 17°C. Native crystals of CspE were either soaked in 5 μl drops of well solution containing increasing concentrations of either dCMP or dCTP or small amounts of lyophilised dCMP or dCTP were added directly to a drop containing a single CspE crystal. Crystals were soaked from one minute to 24 hours.

5.4.8 Analytical Gel Filtration

Gel filtration studies were carried out on an AKTA explorer FPLC using a Superdex 75 HR 30/10 column at 4°C. The column was equilibrated with a buffer containing, 100 mM NaCl and 20 mM Tris, pH 7.5 and calibrated using the following molecular weight standards –thyroglobulin (690 kDa), catalase (232 kDa), aldolase (158 kDa),

albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa), ribonuclease A (13.7 kDa), vitamin-B (1.35 kDa), ATP (0.55 kDa) from Amersham Biosciences. 200 µl protein concentrations ranging from 20-800 µM was applied to the column and run at 0.5 ml.min^{-1}. Calibration curves for molecular weight standards were generated using Kaleidagraph v4.03 software (Synergy Software, reading, PA). For molecular weight a plot K_{av} against log molecular weight was used where K_{av} = (Ve-Vo)/(Vt-Vo); Ve = retention volume, Vo = void volume, Vt = column bed volume.

5.4.9 Glutaraldehyde cross-linking

For glutaraldehyde crosslinking experiments, 70 µg of CspE protein in 10 mM HEPES buffer (pH 7.5) in a total volume of 100 µl was treated with 5 µl of 2.4% freshly prepared solution of glutaraldehyde at room temperature. 10 µl aliquots were removed at various time intervals and added to 1 µl of 1 M Tris-HCl, pH 8.0. An equal volume of 2X SDS-sample buffer was added to the cross-linked proteins and electrophoresis was conducted in 18% SDS-polyacrylamide gels at 200 V.

5.4.10 Trypsin digest

Protein bands were cut tightly from a SDS-PAGE gel using a sterile scalpel. Each of the gel pieces were then incubate in 300 µl of 200 mM NH_4HCO_3 (ABC) in 50% acetonitrile (ACN) for 30 minutes at 30°C (x2) to remove any SDS. The samples were centrifuged at 13,000 rpm for 2 minutes and the covered with ACN (must turn white). ACN was decanted and gel pieces were allowed to dry. A stock solution of trypsin was prepared by adding 50 µl of 50 mM ABC to a new vial of trypsin (Promega) at 4°C and 1 µl aliquots were then stored at -20 °C. 29 µl of cold 50 mM ABC was added to each trypsin aliquot (as required). The final 30 µl solution was

added to the dry gel piece and incubated for 5 minutes at 4°C. The tops of the tubes were sealed with Nescofilm and incubate at 32°C for 16-24 hrs.

5.5 Results and discussion: Section one

5.5.1 Crystallisation of His₆StCspE

Many attempts were made to crystallise His$_6$StCspE but were all unsuccessful. This was considered to be a direct result of the flexible histidine tag region of the recombinant protein.

5.5.2 Purification of StCspE

Large amounts of His$_6$StCspE were purified and the His-tag was removed by specific cleavage using the protease thrombin (Figure 5-2b). On completion of the cleavage reaction thrombin was removed by passing the solution through a benzamidine column. The protein sample was concentrated and gel filtration with a HiLoad 26/60 Superdex 75 column was performed. Recombinant StCspE eluted as two peaks (Figure 5-2a), only StCspE from the major peak (16.07 ml) was used for crystallisation trails. StCspE was more than 95% pure, as judged by SDS-PAGE (Figure 5-2c), with an experimentally determined molecular weight of 7.7 kDa (Figure 5-2d).
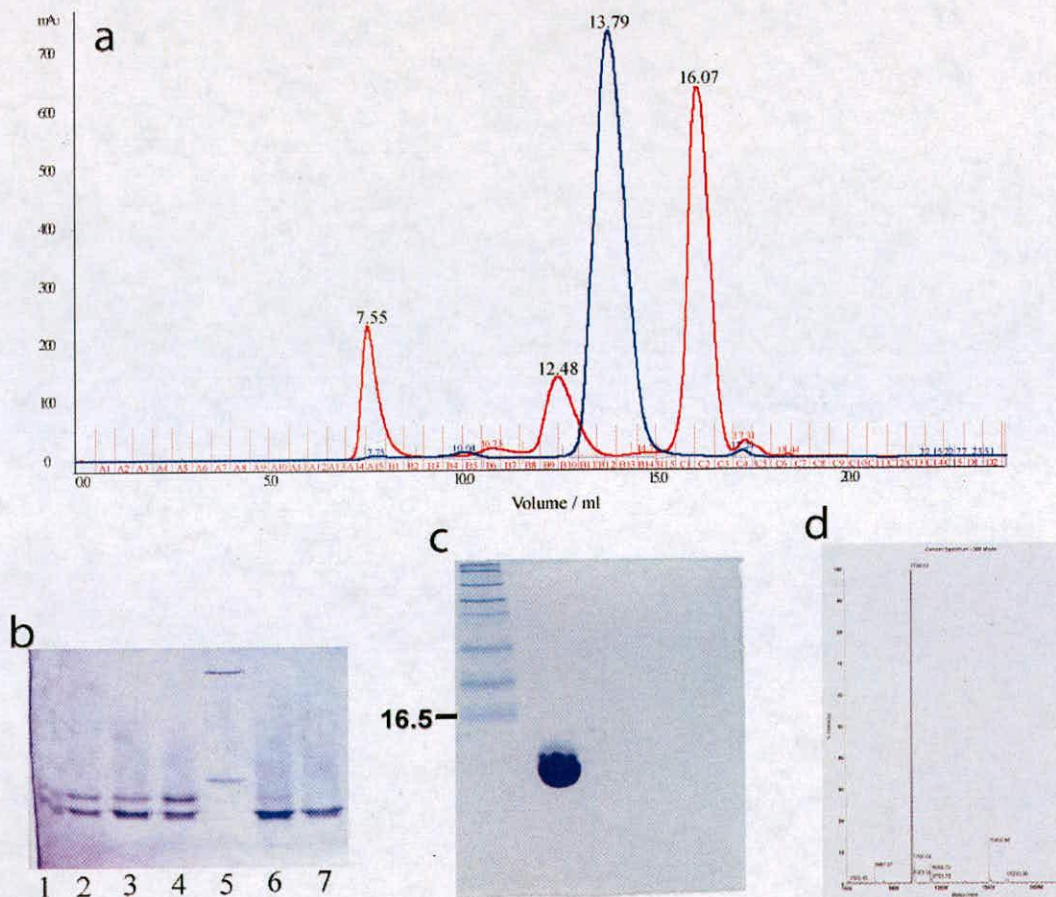
Figure 5-2. **Purification of *St*CspE.**

(a) Gel filtration with Superdex 200 column. Chromatographs for refolded His$_6$-CspE (blue, retention volume = 13.79 ml/12 kDa) and CspE (red, retention volume 16.07 ml/7kDa). MW of His$_6$-CspE = 9614 Da; MW of CspE = 7732 Da. **(b)** Thrombin cleavage of 6xHistidine tag. Lanes- cleavage after; (1) 0.5 hr, (2) 1hr, (3) 1.5 hr, (4) 2hr, (5) control protein, (6) 3 hrs, (7) 4 hrs. **(c)** SDS-PAGE analysis of purified CspE. **(d)** The resulting MALDI-Tof mass spectrum for *St*CspE; observed = 7728 Daltons; expected 7732 Daltons.

5.5.3 *Crystallisation of CspE from S. typhimurium*

*St*CspE was crystallised using the hanging drop vapour diffusion method [202].

*St*CspE crystals were grown in 28% PEG 20,000, 0.05 M AMPSO (pH 9) and 1%

glycerol at 17°C. Crystals appeared after 24 hrs (Figure 5-3a) and grew to maximum

dimensions of 1.3 x 0.2 x 0.05 mm after 72 hrs (Figure 5-3b). Different *St*CspE

crystal forms were obtained during crystallisation (Figure 5-3c-f). All crystals were flash cooled in liquid nitrogen directly from the mother liquor, prior to data collection.
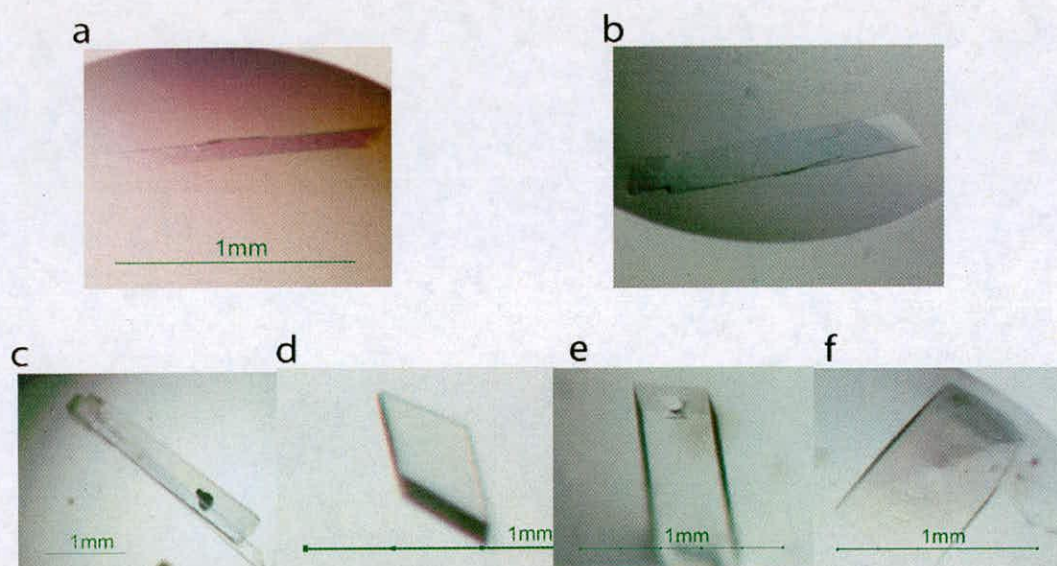
Figure 5-3. **Examples of CspE crystals grown in 28% PEG 20,000, 1% glycerol and buffered with 0.05 M AMPSO, pH 9.**

CspE crystal after 24 hrs growth (a) and 72 hrs growth (b). (c-f) Examples of various crystal forms obtained.

5.5.4 *X-ray data analysis*

*St*CspE crystals flash cooled directly from the mother liquor (28% PEG 20,000, 0.05 M AMPSO, pH 9 and 1% glycerol) were mounted and diffracted X-rays were collected to a maximum resolution of 1.1 Å (Figure 5-4). Data were indexed in the monoclinic space group $P2_1$, with unit cell dimensions $a = 29.40$, $b = 46.80$, $c = 46.72$ Å and angles of $\alpha = 90.0°$, $\beta = 103.4°$ and $\gamma = 90.0°$.
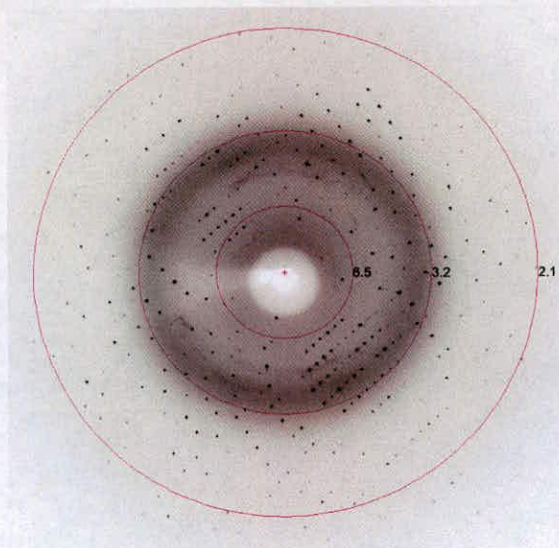
Figure 5-4. **Diffraction pattern for a CspE crystal (crystallised at 28% PEG 20,000, 0.05 M AMPSO, pH 9 and 1% glycerol).**

Rings have been added to provide an indicator of resolution (values = Å). Diffraction extended out to 1.1 Å.

5.5.5   *Content of the asymmetric unit*

For a protein with a molecular weight of 7728 Da, the Matthew's equation [203] indicates there to be two (V_m = 2.05, solvent 51%) monomers per asymmetric unit for space-group $P2_1$. The homologous CspE from *Escherichia coli* has been shown to form dimers in solution [124] and differs in only 4 amino acids.

A self-rotation Patterson map calculated using the data processed in space-group $P2_1$ revealed no obvious non-crystallographic symmetry (NCS) (Figure 5-5) between two molecules in an asymmetric unit. The two molecules are related by a simple rotation and translation movement (as illustrated in Figure 5-6).
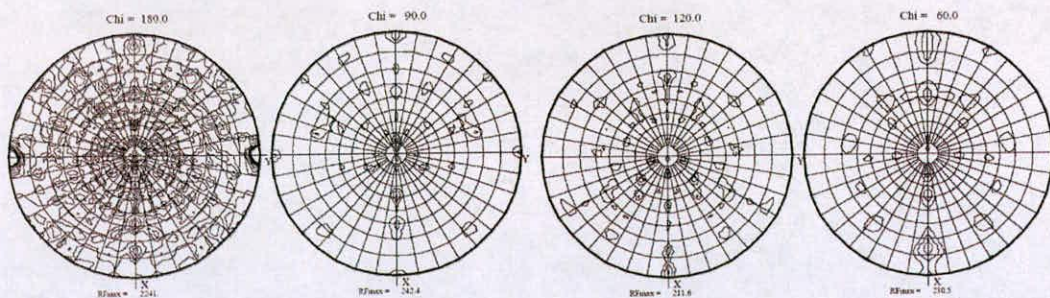
Figure 5-5. Self-rotation analysis data processed in space-group P2₁.

Self rotation map showing peaks at Chi = 180°, 90°, 120° and 60°.
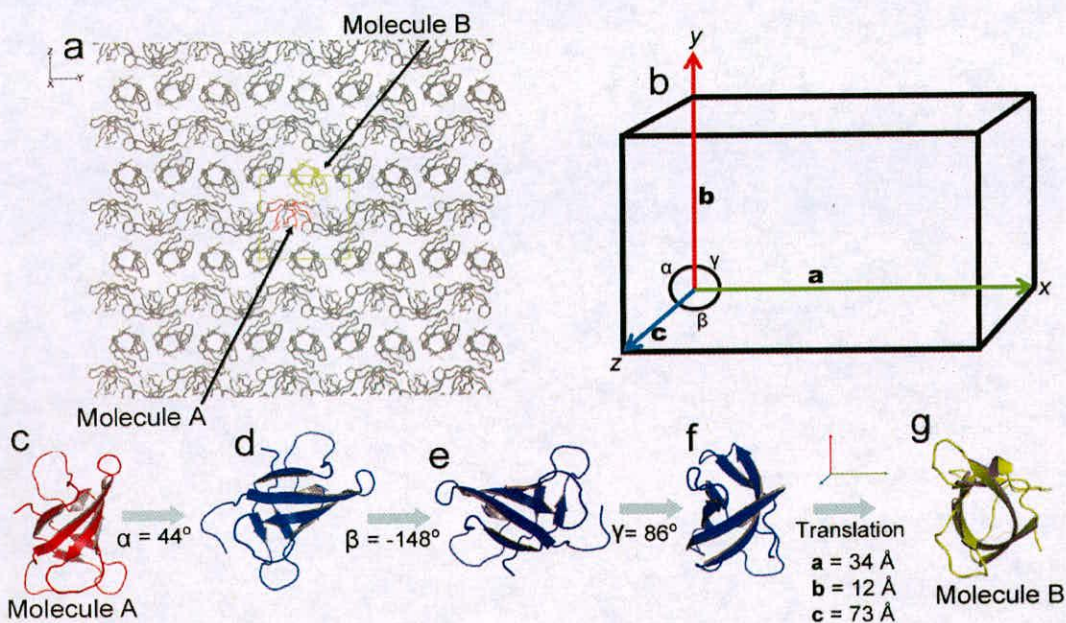
Figure 5-6. The relationship between the two *St*CspE molecules in the asymmetric unit.

(a) The crystal packing arrangement of *St*CspE, shown as backbone trace. The unit cell is highlighted with a yellow box and the two *St*CspE molecules present in the asymmetric unit have been coloured red (molecule A) and yellow (molecule b). **(b)** A representation of a unit cell, with edges **a, b, c** and angles α, β and γ, which illustrates the coordinate system used to describe the movement of molecule A to the position of molecule B in the asymmetric unit, as illustrated **(steps c to g).**

167

5.5.6 Obtaining Phases

Phases were determined with ACORN [198], using the positions of a decapeptide (PHE12 to ILE21) from the homologous structure of the *E. coli* CspA [118]. Electron density maps resulting from ACORN phasing, using a minimum of 10 amino acids were of sufficient quality (Figure 5-7a) to begin model building. All of the electron density maps derived from ACORN were examined in COOT and distinctive shapes of key amino acids were clearly visible (Figure 5-7).
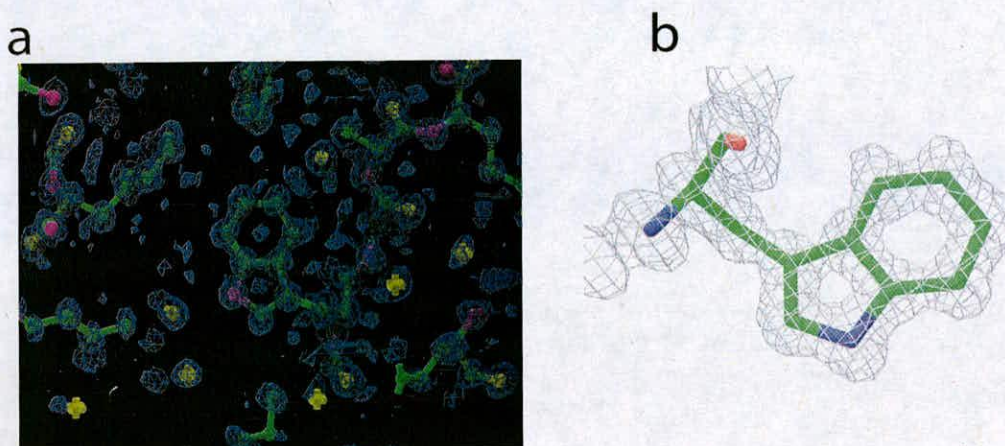
Figure 5-7. **Phases obtained using ACORN.**

(a) A portion of the electron density map generated using experimentally derived phases with a decapeptide from *E. coli* CspA. The single tryptophan present in the CspE structure is clearly visible. **(b)** The same tryptophan after refinement. Images were generated using Povray (a) and Pymol (b).

5.5.7 Structure solution of StCspE

Structure refinement for *St*CspE converged at an R value of 0.20 with good model stereochemistry (Table 5-3). The final model was made up of two molecules of *St*CspE (Figure 5-8a; Chain A and B) comprising 1056 protein atoms and 100 water molecules. The electron density was poorly defined for residues 38 to 40 on chain B, where the some of the highest thermal (B) factors are observed, with an average of

168

30.10 Å^2, in comparison the average B factor for the entire protein is 18.86 Å^2. In chain B the electron density is poorly defined for the N-terminal, as a result the model could only be built from lysine six onwards (Figure 5-8c).

Table 5-3. StCspE – Data collection and refinement statistics

Data statistics		
Space group	P2$_1$	
Unit-cell parameters (Å)		
a	29.40	
b	46.80	
c	46.72	
γ, β, α	90.0°, 103.4°, 90.0°	
Resolution range (Å)	45.36-1.10	(1.16-1.10)
Molecules per asymmetric unit	2	
Observed reflections	260904	(27528)
Unique reflections	47339	(7261)
$<I>/\sigma(I)$	15.7	(2.5)
Completeness (%)	99.82	(100)
Multiplicity	5.2	(3.8)
R_{sym} (%)	5	(41)
Refinement statistics		
R factor[‡] (%)	20.33	(34.90)
R free[§] (%)	23.42	(35.3)
Correlation coefficient	0.95	
Number of protein atoms	1073	
Number of water molecules	100	
Mean B factor for protein	18.86	
R.m.s deviations from ideal geometry		
Bond distance	0.02	
Bond angle	1.94	
Ramachandran statistics		
Most favoured regions (%)	93.90	
Additional allowed regions (%)	6.10	
Generously allowed regions (%)	0	
Residues in disallowed regions (%)	0	

The outer shell statistics are shown in brackets.
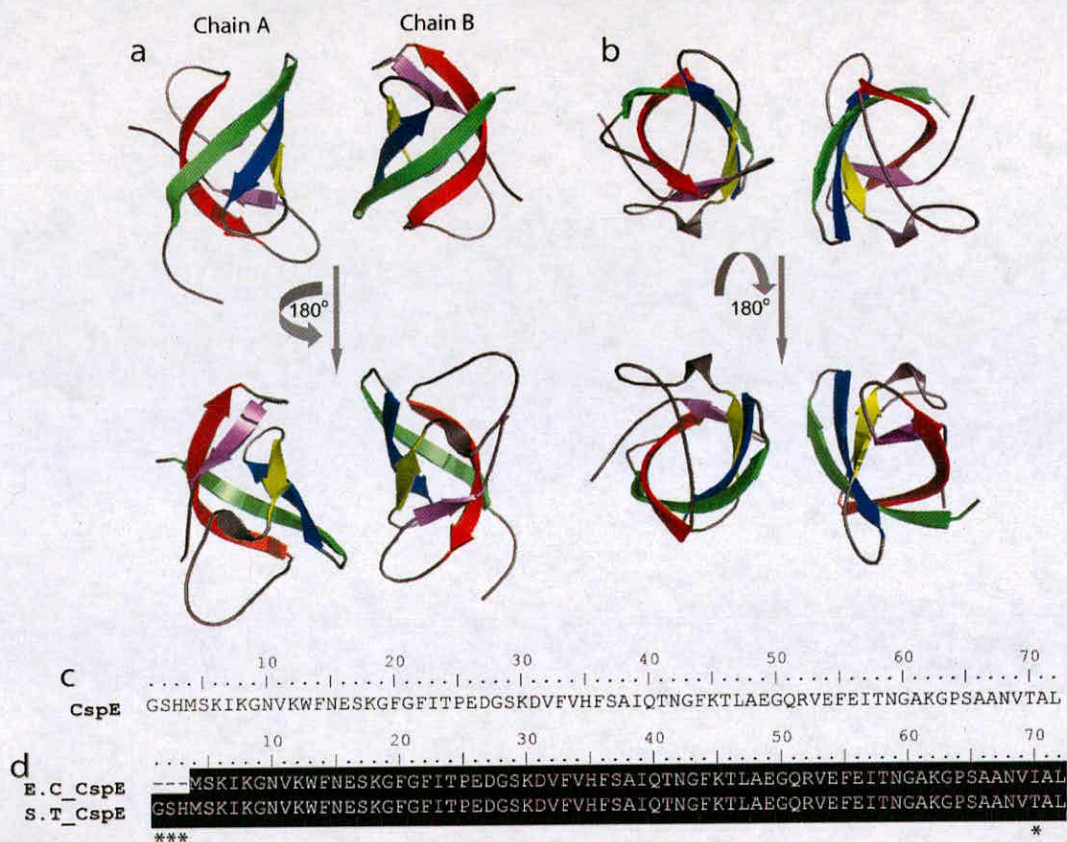
Figure 5-8. **Three-dimensional ribbon diagram of *St*CspE, showing the classic five stranded β barrel structure/OB fold.**

The five β-strands have been coloured to highlight their positions; β-1 (green), β-2 (blue), β-3 (yellow), β-4 (red) and β-5 (magenta). **(a)** Face-on view of and top view **(b)** of *St*CspE. **(c)** The amino acid sequence of the crystallized recombinant CspE. **(d)** Alignment of the amino acid sequence of *St*CspE with EcCspE. Identical residues are shown in black and non-identical are marked by an asterix. The abbreviations are as follow: *St*CspE, cold shock protein E from *Salmonella typhimurium*; *Ec*CspA, cold shock protein A from *Escherichia coli*;

For Chain A of the model, the electron density for the N-terminal is well defined and all residues except the first glycine residue (Figure 5-8c) were modelled. The results of the PROCHECK analysis indicated that no residues were found outside the allowed region of Ramachandran diagram (Figure 5-9).

171
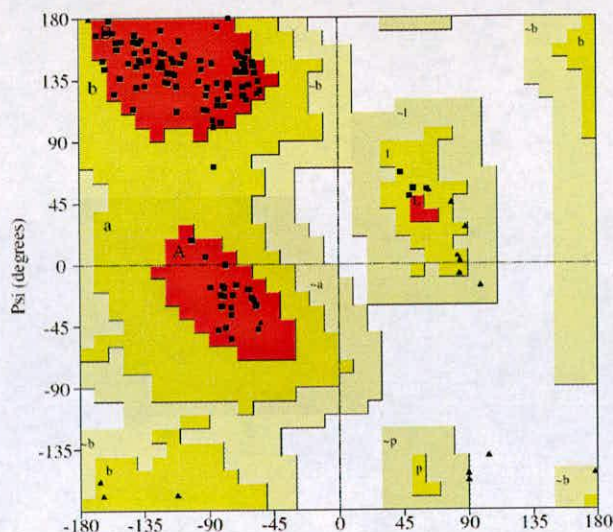
Figure 5-9. **Analysis of sterochemical properties of CspE model with PROCHECK.**

93.9% of residues in the most favoured region of the Ramachandran plot.

5.5.8 *The overall structure of CspE*

The three crystal structures of CSPs [78, 118, 119] and in complex with dT6 [14, 25] superpose well with *St*CspE with average r.m.s deviation values of 0.63 and 0.67 respectively, for 64 C$^\alpha$ atoms out of a possible 67 C$^\alpha$ atoms. Loop four of *St*CspE was excluded as it adopted an unusual conformation compared to all other CSPs (Figure 5-10). Two *St*CspE molecules in the asymmetric unit pack with their DNA binding faces [133] held together by seven potential hydrogen bonds (Figure 5-11). The crystal structure of *St*CspE presented here suggests that it may exist as a dimer.
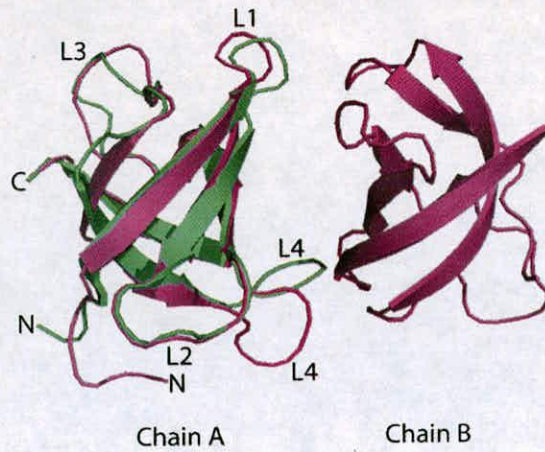
Figure 5-10. **Three-dimensional ribbon diagram of *St*CspE (Purple) superposed with *Ec*CspA (Green).**

Loops between β strands are numbered L1-L4. Loop 4 of *St*CspE follows a different direction to all other CSP structures.
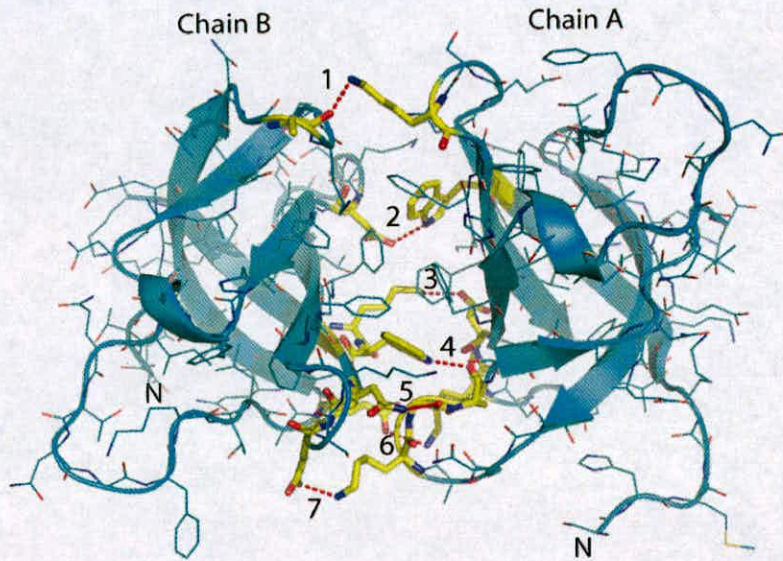
Figure 5-11. **Hydrogen bonds formed between the *St*CspE dimer.**

The amino acids (shown in yellow) which form potential hydrogen bonds (depicted as red lines) between the two *St*CspE monomers, calculated using the PISA programme [204]. Residues which potentially hydrogen bond with one another are; (1) Ala58-Lys15; (2) Asp28-Trp10; (3) Lys9-Asp28; (4) Trp10-Pro61; (5) Asn12-Gly60; (6) Glu46-Lys27; (7) Glu13-Lys59.

173

5.5.9 StCspE, dimer or monomer?

5.5.9.1 Dimerisation of CSPs

The highly homologous CspE from *Escherichia coli*, differing in only four amino acid, forms dimers in solution [124]. Therefore one can assume that the crystallographic *St*CspE dimer may be equivalent to the dimer formed in solution.

The highly homologous CSP from *B. subtilis*, *Bs*CspB (Figure 5-12a), and *B. cereus* CspE, have been reported to form dimers in solution [123, 125]. Only *Bs*CspB has been shown to form dimers in both the crystal structure and in solution [119, 123]. Figure 5-12c shows the structure of the *Bs*CspB dimer and the six hydrogen bonds (Figure 1-2b and d) that hold the dimer intact at β-strand-4 of each CspB molecule (Figure 5-12c). The four amino acids involved in *Bs*CspB dimer formation (Figure 5-12a, red arrows) are not all present in the amino acid sequences of *E. coli* CspA (monomer), *B. cereus* CspE (dimer) or *S. typhimurium* CspE (crystallographic dimer). The fact that other CSP form dimers in solution without the presence of these four amino acids, indicates that dimer formation is not merely the results of conserved amino-acids forming a linking hydrogen-bond network. Dimer formation between CSPs may either form through the linking of β-strands four (Figure 5-12c) or similar to that of the *St*CspE structure presented here (Figure 5-11).
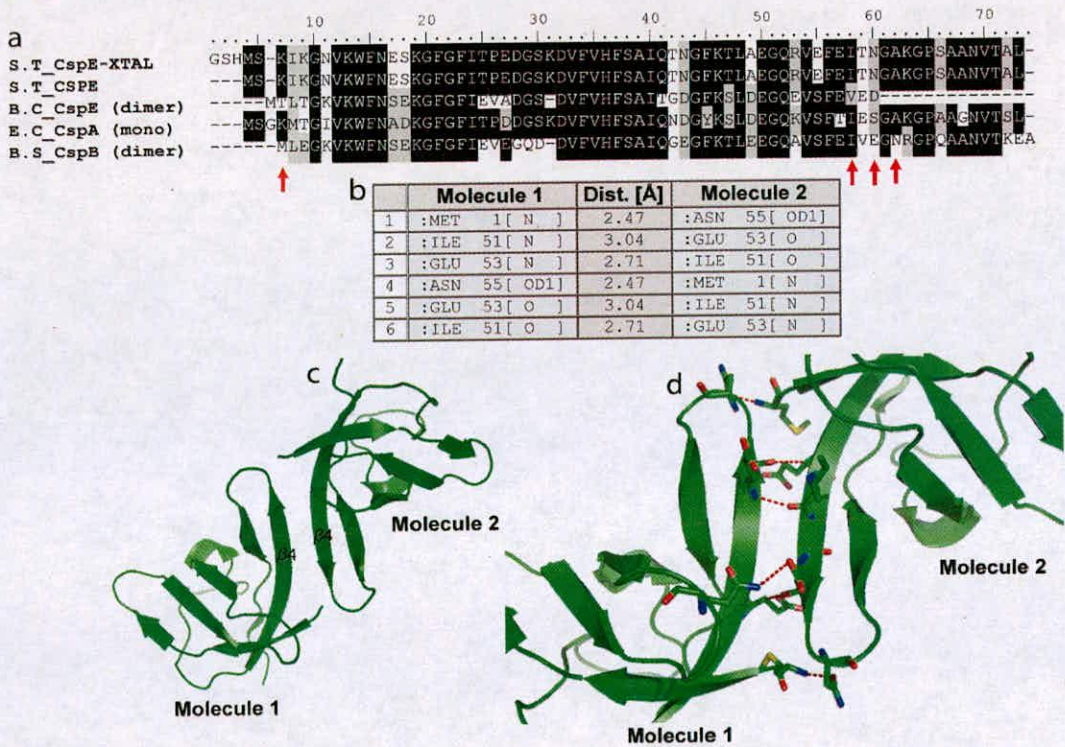
Figure 5-12. **CSP dimer formation.**

(a) Multiple sequence alignment of, *S. typhimurium* CspE, both the recombinant (XTAL) and wildtype, *E. coli* CspA, *B. subtilis* CspB and *B. cereus* Csp. The recorded forms of each protein, either as a monomer (mono) or a dimer is indicated. Identical residues are shaded black and residues shown to be involved in dimer formation for *Bs*CspB are highlighted with red arrows. (b) The residues which form hydrogen bonds between two *Bs*CspB molecules to form the dimer structure (c), which is made stable by the bonds connecting β-strands 4 on both subunits. (d) The six hydrogen bonds (dashed red lines) at the interface of the CspB dimer.

5.5.9.2 Recombinant *St*CspE forms oligomeric species in solution

The oligomerisation state of CspE was investigated using size exclusion chromatography. His$_6$*St*CspE elutes as a single species with a retention volume of 13.79 ml (Figure 5-13), which corresponds to molecular weight of 12.5 kDa, demonstrating that His$_6$*St*CspE exists as a monomer. After the removal of the histidine tag, *St*CspE eluted as two distinct species with retention volumes of 12.48 and 16.07 ml (Figure 5-13). On comparison with molecular weight standards, the

16.07 ml peak corresponds to a molecular weight of 7 kDa or a single *St*CspE molecule (actual CspE MW = 7.7 kDa). The 12.48 ml peak corresponds to a molecular weight of 25 kDa or three *St*CspE molecules, which may also correspond to a possible *St*CspE dimer or even a higher-order oligomeric species.
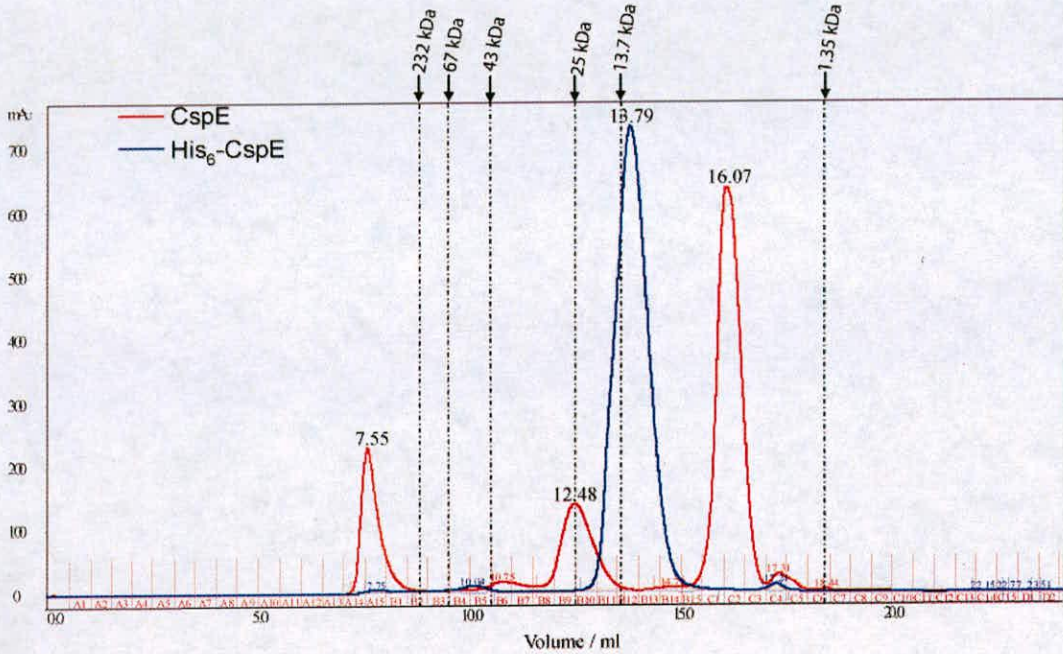
Figure 5-13. Gel filtration analysis of His₆*St*CspE and *St*CspE.

Recombinant His₆*St*CspE and *St*CspE resolved on a HiLoad 26/60 Superdex-75 column. The column was calibrated with the molecular weight standards, catalase (232 kDa), albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa), ribonuclease A (13.7 kDa) and vitamin-B (1.35 kDa) from Amersham Biosciences. The dashed lines on the trace correspond to the elution maxima of the standards. His₆*St*CspE eluted as a single peak with a retention volume of 13.79 ml, corresponding to a molecular weight standard of 12.5 kDa. *St*CspE (after cleavage) eluted as one major and one minor peak with retention volumes of 16.07 and 12.48 ml, corresponding to molecular weight standards of 7.0 kDa and 25 kDa, respectively.

176

5.5.9.3 Chemical cross-linking of *St*CspE indicates potential dimer formation

To further examine potential *St*CspE dimer formation in solution, a series of chemical cross-linking experiments were performed. Chemical cross-linking offers a direct method for identifying stable interactions. The technique involves the formation of a covalent bond between two proteins by using bifunctional reagents containing reactive end groups that react with functional groups, e.g. primary amine and sulfhydryls of amino acids. If two proteins physically interact with each other then they can be covalently cross-linked, providing convincing evidence of their close proximity.

The cross-linking agent glutaraldehyde ($C_5H_8O_2$) reacts with primary amine groups (ε-amino groups of lysine residues) to form stable covalent links. If *St*CspE exists as a dimer in solution then the majority primary amines, within the dimer interface (Figure 5-14a), should be susceptible to cross-linking. Cross-linking results in the formation of various high-order oligomeric species including a dimeric form (Figure 5-14b). The presence of higher-order oligomeric species suggests dimer formation and further oligomerisation perhaps highlights the promiscuous nature of glutaraldehyde as a cross-linker.
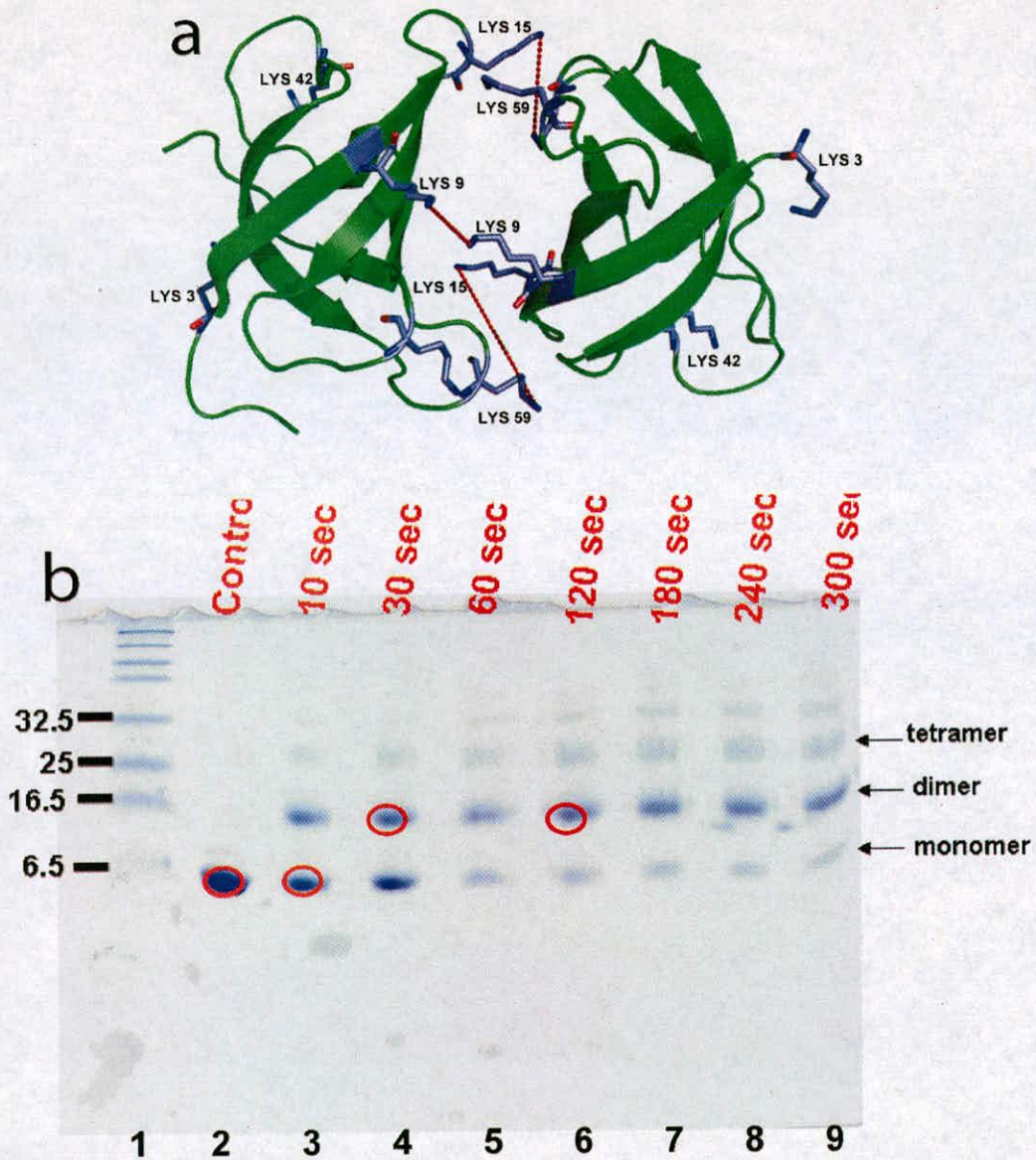
Figure 5-14. **Chemical cross-linking of *St*CspE.**

(a) The lysine residues (blue) present in the CspE structure. The amine groups which should react with glutaraldehyde to from stable covalent linkers (red) between the dimer interface. (b) Crosslinking *St*CspE using glutaraldehyde. Lane 1, Protein molecular weight standards (shown in kDa). Lane 2, 7 µg of CspE untreated; Lane 3-9, 7 µg of *St*CspE crosslinked for various time periods. Bands circled red were excised from the gel for trypsin digest experiments.

5.5.9.4 Trypsin digestion of the cross-linked CspE dimer.

A series of trypsin digest experiments were performed to determine if the dimer seen in cross-linking experiments (Figure 5-14b) was the result of cross-linking lysine residues at the dimer interface (Figure 5-14a). Trypsin is a protease that recognises lysine and arginine residues and cleaves the chain at the C-terminal end, resulting in a peptide fingerprint specific to individual proteins. Protein bands were excised from the SDS-PAGE gel (Figure 5-14b, red circles) and digested with trypsin. Peptide fragments were analysed using MALDI-Tof mass spectrometry. The calculated peptide masses (Figure 5-15e) correspond well with the observed masses for the control (Figure 5-15f – shaded grey). A mass of 1997.9 Da was observed in the MALDI-spectrum for the cross-linked dimer band (Figure 5-14b; lanes 4 and 6 – red circle), which corresponds to an expected cross-linked peptide fragment (Figure 5-15c and d (2), Lys15 on monomer 1 is cross-linked with Lys59 on monomer 2) with a calculated molecular weight of 1997.9 Da. Irritatingly, an identical peptide peak was also observed in the MALDI-Tof spectrum corresponding to the monomeric sample from lane 3 (Figure 5-14b). The 1997.9 Da mass peak was not observed in the MALDI-Tof spectrum for the untreated control sample, corresponding to lane 2 (Figure 5-14b). The presence of this cross-linked peptide in the control sample may be a direct result of the presence of glutaraldehyde in the sample, small traces of cross-linked dimer present or intramolecular cross-linking. The experiment was repeated three times but no significant differences were observed in the results obtained. Conclusive evidence that StCspE is forming dimers in solution still remains to be determined.

a

```
         -2        10        20        30        40        50        60
          |....|...|....|....|....|....|....|....|....|....|....|....|....|....|....
CspE     GSHMSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
```

b

```
CspE     GSHMSKIK        WFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
                 GNVK
                 GNVK    1
CspE     GSHMSKIK        WFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
```

c

```
CspE     GSHMSKIKGNVK        GFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
              WFNESK                                      2
                                                          VEFEITNGAK
CspE     GSHMSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQR        GPSAANVTAL
```

d

e

Expected peptide molecular weights

Start	End	Peptides	Calculated Mass (Da)
-2	3	GSHMSK	764.46
4	5	IK	259.35
6	9	GNVK	416.48
10	15	WFNESK	809.88
16	27	GFGFITPEDGSK	1254.36
28	42	DVFVHFSAIQTNGFK	1709.92
43	49	TLAEGQR	773.84
50	59	VEFEITNGAK	1107.23
60	69	GPSAANVTAL	900.00
2 (10) 50	(15) 59	(WFNESK) VEFEITNGAK	1899.09 + 98 = 1997.09
1 (6) 6	(9) 9	GNVK GNVK	814.94 + 98 = 912.94

f

Resulting peptide molecular weights

Observed: Mass Peak (Da)	Expected/matched Mass (Da)	Peptide sequence
774.40	774.41	(K)TLAEGQR(V)
810.37	810.3786	(K)WFNESK(G)
1107.5742	1107.57	(R)VEFEITNGAK(G)
1254.60	1254.60	(K)GFGFITPEDGSK(D)
1709.86	1709.87	(K)DVFVHFSAIQTNGFK(T)
1994.85	1997.09	WFNESK VEFEITNGAK

Figure 5-15. **Trypsin digestion of cross-linked *St*CspE.**

(a) Amino acid sequence of recombinant *St*CspE. Trypsin cleavage sites (Lys or Arg) are shown by a red arrow. (b,c) Potential peptide fragments resulting from digestion of cross-linked CspE (d). (e) Expected peptide masses after trypsin digest of CspE and cross-linked *St*CspE (Shaded grey). (f) Resulting peptide masses determined from MALDI-Toff analysis. Masses that were observed in both the cross-linked and control samples are shaded grey.

5.5.9.5 A truncated version of CspE forms dimers in solution

*St*CspE was resolved on a Superdex-75 column (Figure 5-16a). *St*CspE eluted as two peaks, suggesting a population of both monomeric and dimeric forms (discussed previously). SDS-PAGE analysis (Figure 5-16a) of protein samples taken from each peak clearly shows that the protein sample corresponding to the dimer peak (green) runs at a lower molecular weight than the monomer (black). The molecular weights

of these proteins were further analysed by mass spectrometry and the resulting

molecular weights were, 7728.30 (monomer peak) and 6808.07 (dimer peak), which

correspond to two forms of *St*CspE (Figure 5-16b), a cleaved (7732.67 Da) and a

truncated (6806.57 Da) form. It is the truncated form that has the ability to form

dimers in solution (Figure 5-16a); whether the dimer structure is identical to the

crystallographic *St*CspE dimer (Figure 5-11) or resembles the *Bs*CspB dimer (Figure

5-12) still remains to be determined.

b

```
1   MGSSHHHHHHSSGLVPRGSHMSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
2   ------------------GSHMSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
3   ------------------------NVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
```
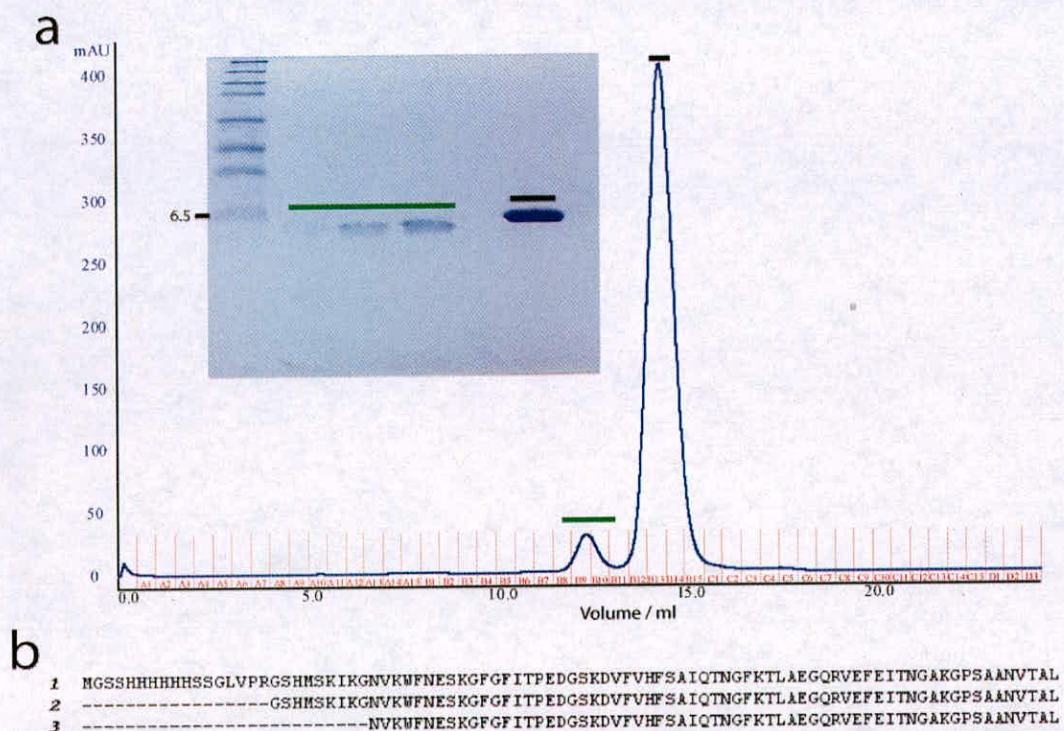
Figure 5-16. **Gel-filtration analysis of truncated *St*CspE.**

(a) *St*CspE resolved on a HiLoad 26/60 Superdex-75 column equilibrated in PBS buffer (pH 7.5). *St*CspE eluted as two distinct peaks and a SDS-PAGE analysis of both peaks (colour coded, green and black) are shown (inset). **(b)** Amino acid composition of *St*CspE forms; (1) His$_6$-CspE - MW = 9614.7 Da; (2) Recombinant *St*CspE – MW = 7732.67 Da; (3) Truncated CspE – MW = 6806.57 Da.

5.5.10 Purification and crystallisation of CspE-ss DNA complex

Both ITC and Biacore studies (Chapter 4) have shown that the preferential ss DNA binding sequence, 5′-GTCTTTT-3′ (as determined by microarray analysis – Chapter 3), binds with low nM affinity to *St*CspE. This knowledge provided an advantageous start to crystallising the *St*CspE-ss DNA complex as this sequence is most likely to keep the complex intact throughout the crystallisation process.

A complex was formed between *St*CspE and oligonucleotide sequence 5′-GTCTTTT-3′. The *St*CspE-ss DNA complex was resolved on a Superdex-75 gel-filtration column, which removed any unbound DNA from the complex. The recombinant *St*CspE-ss DNA complex eluted as two peaks (Figure 5-17a), the latter corresponding to unbound ss DNA. The *St*CspE-ss DNA complex was more than 95% pure, as judged by SDS-PAGE (Fig. 5-16b). Crystallisation trails of purified *St*CspE-5′-GTCTTTT-3′ complex trails failed to produce crystals or promising hits.
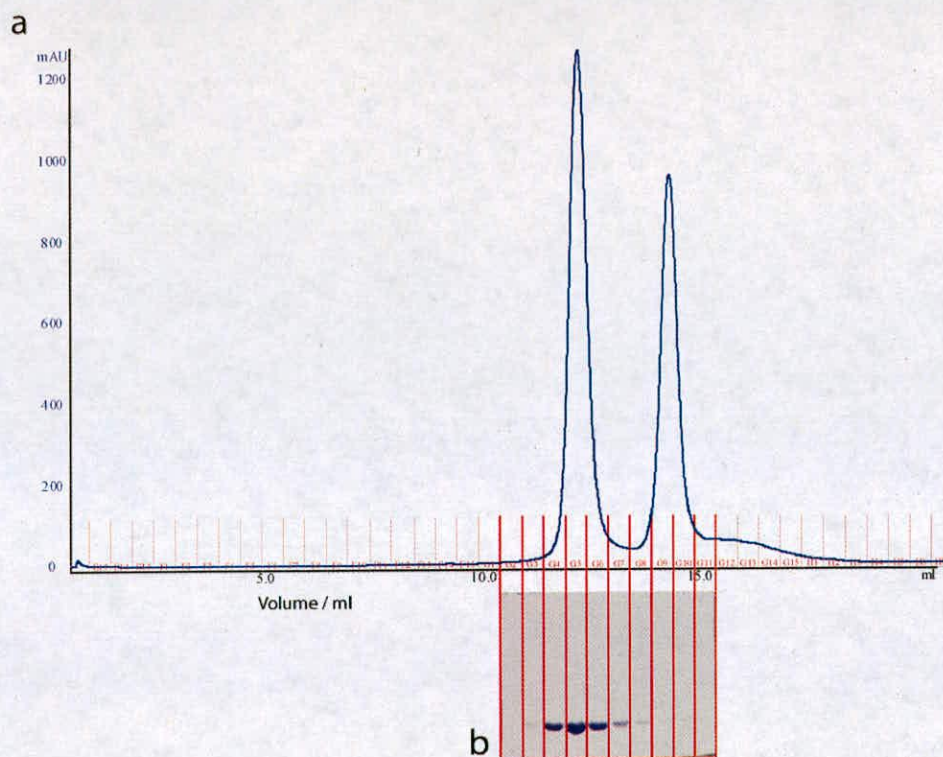
Figure 5-17. **Purification of *St*CspE-5′-GTCTTTT-3′ complex resolved on a HiLoad 26/60 Superdex-75 column (a) equilibrated in TEN buffer (pH 7.5).** (b) SDS-PAGE analysis of eluted peaks.

During this work the *Bs*CspB in complex with T_6 (a six stretch of thymine nucleobases) was published [14]. The authors also observed difficulties in trying to crystallise *Bs*CspB with the high-affinity ligand 5′-TTCTTTT-3′, which is similar to the ligand used here. Interestingly, the T_6 ligand bridges between two *Bs*CspB molecules forming long continuous chains of complex molecules within the crystal structure (Figure 5-18a). The long chains formed in the *Bs*CspB/T_6 structure may be a requirement for crystallisation. Taking this into account a similar bridging ligand was designed (Figure 5-18b – 5′-CTTTTTGT-3′), which should retain the high-affinity interaction and promote the formation of a crystal lattice. Unfortunately

183

economic restraints have not allowed for crystallisation trails of the CspE/5′-CTTTTTGT-3′ complex to proceed at this time.
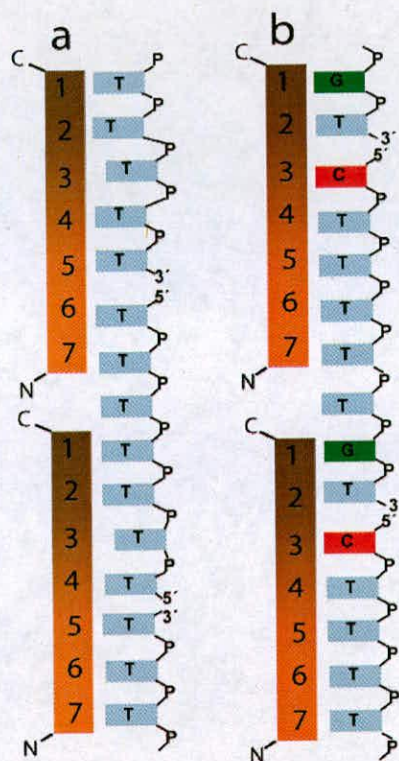
Figure 5-18. **Schematic overview of CSP/oligonucleotide assemblies.**

Protein molecules (brown – orange objects) interact with nucleobases (boxed characters) from oligonucleotide molecules (-P-) at distinctive subsites (1-7). **(a)** A continuous binding arrangement, as seen in the BsCspB/T$_6$ crystal structure. **(b)** Proposed ligand, 5′-CTTTTTGT-3′, used to induce a continuous binding arrangement but retain the high-affinity recognition sequence.

5.5.11 Cytosine soaking experiments

There is a strong binding preference seen for a cytosine nucleobase at position three for both the BsCspB protein and the StCspE protein (Figure 2-15 and Figure 4-3). A series of soaking experiments were performed using crystals of StCspE and various concentrations of 2′-deoxycytidine 5′-monophosphate (dCMP) or 2′-deoxycytidine 5′-triphosphate (dCTP) to obtain a StCspE-cytosine complex. Cytosine soaking

experiments had detrimental effects on *St*CspE crystals (cracked and dissolved), which was thought a good indicator that the cytosine base was binding to *St*CspE. Although processable data were obtained there was no evidence of electron density for a bound cytosine. This was assumed to be a result of the inaccessibility of the DNA binding interface caused by crystal packing (the highly aromatic DNA binding interfaces of both *St*CspE were locked together, Figure 5-11).

5.5.12 Crystallisation of His$_6$CspD

Many attempts were made to crystallise His$_6$*St*CspD but were all unsuccessful, this was thought to be a direct result of the flexible histidine tag.

5.5.13 Purification of CspD

Large amounts of His$_6$*St*CspD were purified and the His-tag was removed by the specific cleavage using the protease thrombin. The protein sample was concentrated and gel filtration with a HiLoad 26/60 Superdex 75 column was performed. Recombinant *St*CspD eluted as a single peak (Figure 5-19a), with a retention volume of 11.54 ml. *St*CspD was more than 95% pure, as judged by SDS-PAGE (Figure 5-19b), with an apparent molecular weight of 8.0 kDa.
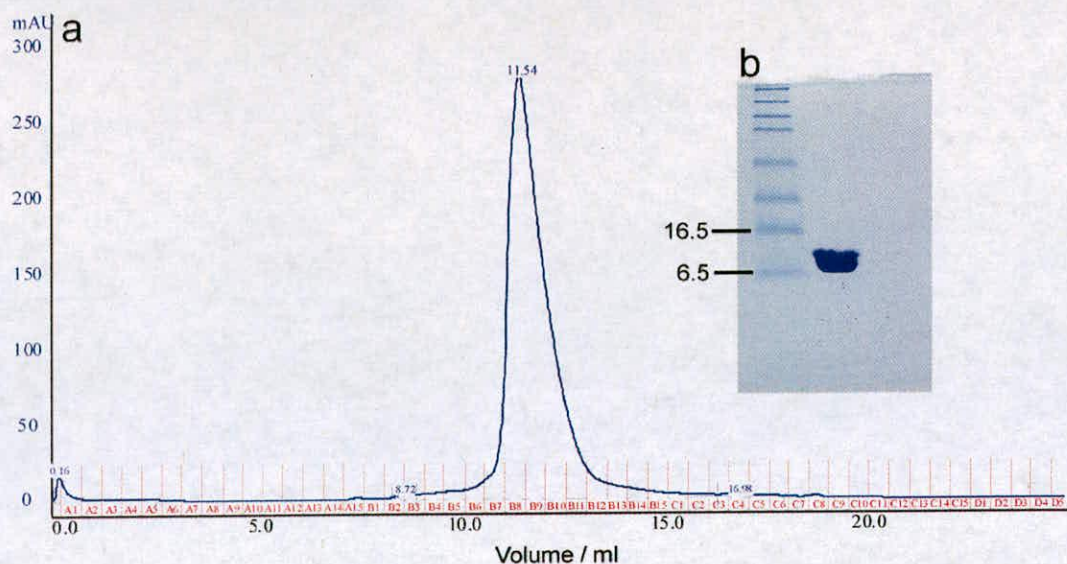
Figure 5-19. **The purification of *St*CspD.**

(a) Purification of *St*CspD, resolved on a HiLoad 26/60 Superdex-75 column equilibrated in 10 mM Tris buffer (pH 7.5). **(b)** SDS-PAGE analysis of eluted *St*CspD protein.

5.5.14 Crystallisation trails of StCspD

Crystallisation screens of pure *St*CspD failed to produce crystals or promising hits, as the protein precipitated out of solution with most precipitants used. Either the protein and/or the precipitant concentrations were reduced to try and reduce protein precipitation but nearly all conditions produced a gelatinous precipitate, similar to that seen in Figure 5-20.

5.5.15 Purification and crystallisation trails of CspD-ss DNA complex

Biacore studies (Chapter 4) have shown that the preferential ss DNA binding sequence, 5′-ACGGGG-3′ (as determined by microarray analysis – Chapter 3), binds with to *St*CspD with low nM affinity. For financial reasons the crystallisation of the *St*CspD-5′-ACGGGG-3′ complex was not attempted.
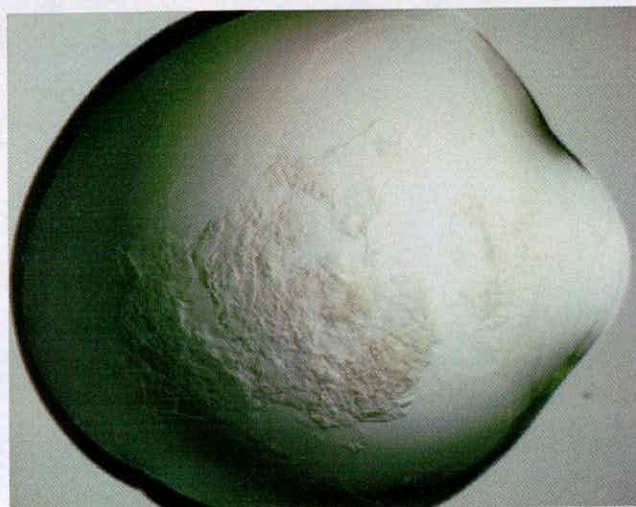
Figure 5-20. **Crystallisation of CspD.**

An example of the gelatinous precipitate observed for most screening conditions.

5.5.16 Conclusions

The structure of the *S. typhimurium* cold shock protein E (*St*CspE) has been solved. The final model was refined to R/R$_{free}$ values of 20.3% and 23.42%. The five-stranded β-barrel structure of *St*CspE is highly similar to other CSP structures. The X-ray crystal structure of *St*CspE revealed a dimeric structure, with two *St*CspE molecules in the asymmetric unit, packed with their DNA binding faces hydrogen bonded together, which until now has never before observed in a CSP structure. Although the highly homologous *Ec*CspE forms dimers in solution, no structure of such a dimer has been solved. The formation of a dimer between the ss DNA binding face of two *St*CspE molecules could suggest a possible self control or feedback mechanism, which would explain why many CSPs are transiently expressed at all times [60] but do not randomly bind single stranded nucleic acids, interrupting translational/transcriptional machinery. Studies of recombinant *St*CspE dimer formation in solution have shown that an N-terminal truncated version of *St*CspE

187

was forming dimers, whether native *St*CspE forms a dimer structure similar to that seen in the crystal still remains to be determined.

5.6 Results and discussion: Structural analysis of CSP/ss DNA complexes

5.6.1 Overview

During this work two CSP/hexathymidine (dT_6) complex structures have been published [14, 25], both of which show identical ss DNA-binding interfaces. The ligand positions from both structures were combined to create a model of a BsCspB/T_7 complex structure. The StCspE crystal structure was superimposed onto the BsCspB/T_7 structure to obtain the StCspE/5´-TTTTTTT-3´ model, with identical interactions observed at each DNA subsite. The heptathymidine ligand was mutated to obtain the StCspE/5´-GTCTTTT-3´ complex structure. The derived StCspE/5´-GTCTTTT-3´ model was used to explain the base preference seen at each of the seven DNA binding sub-sites, as shown from ITC and SPR experiments. The StCspE/5´-GTCTTTT-3´ structure was superimposed onto the homology modelled structure of StCspD to obtain potential DNA subsite positions and a final StCspD/5´-GTCTTTT-3´ model. The ss DNA ligand was mutated to obtain a model of StCspD in complex with 5´-ACGGGG-3´. The StCspD/5´-ACGGGG-3´ model was used to explain the base preference seen at each of the six DNA binding subsites, as observed in previous biochemical experiments.

5.6.2 Analysis of evolutionary conservation and electrostatic properties of the S.T CspE ss DNA binding site.

Consurf, a web-server for analysis of evolutionary conservation, was used to predict functionally and structurally important residues. The analysis was carried out using all experimentally validated prokaryotic CSPs (for alignments refer to A1.5). Amino acids with above average conservation scores could be grouped into two categories; those involved in the defining overall structure and those predicted to participate in

189

DNA binding (Figure 5-21b and c, highlights the electronegative DNA binding face). All residues involved in DNA binding were found to be highly conserved (Figure 5-21b), with above average conservation levels for the residues that make up the RNP-1 and 2 motifs. The least conserved residues could be seen on opposite face of the protein (180°) (Figure 5-21b), which have no designated function and we could expect these residues to vary a lot in organisms with evolution as long as they function to conserve to overall fold of the CSP. The high-level of conservation observed in the DNA binding site of *St*CspE with many other CSP homologues and the use of predefined ligand positions (section 5.6.3) provides an excellent basis for the creation of a valid *St*CspE/ss DNA model.
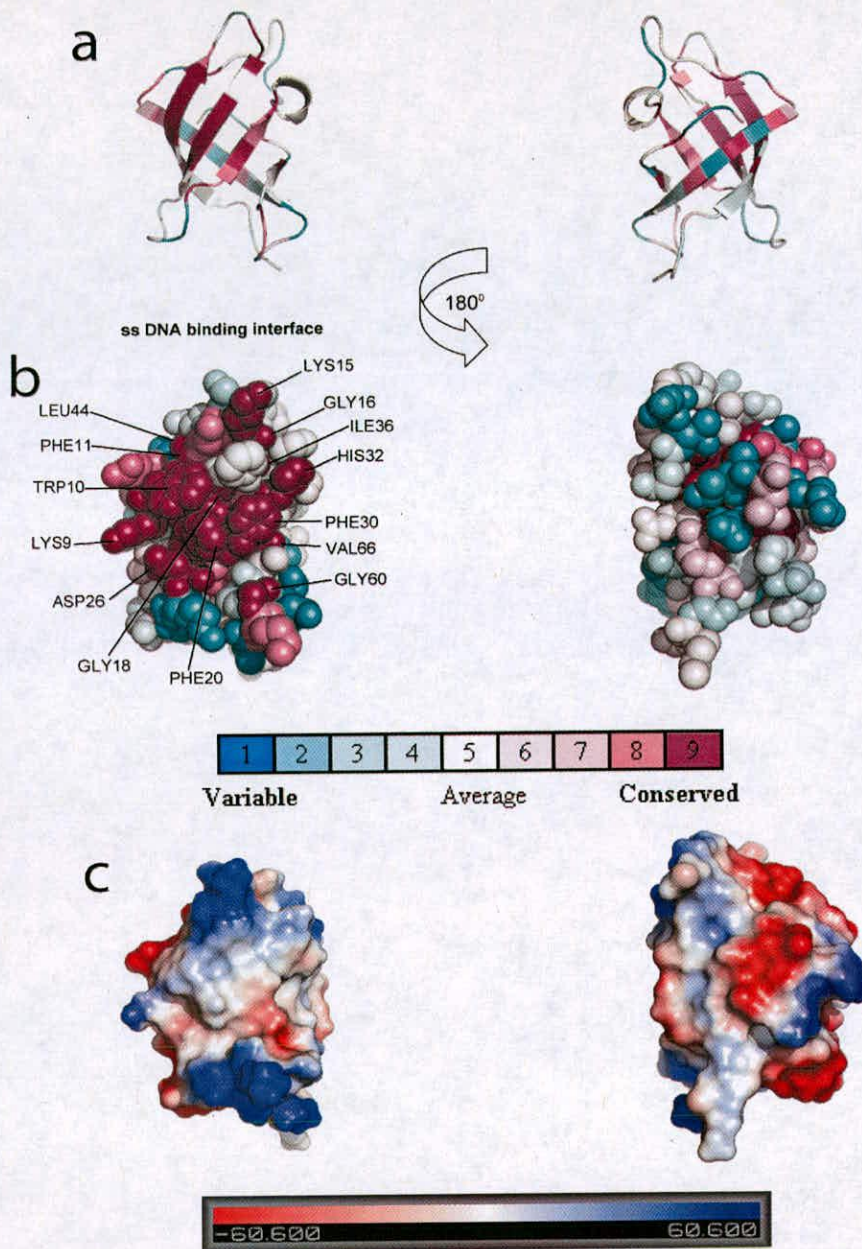
Figure 5-21. **Evolutionary conservation and electrostatic properties of *St*CspE.**

(a) Cartoon representation of the *St*CspE X-ray structure (present earlier in this chapter) coloured according to ConSurf analysis at 0 and 180°. **(b)** Sphere representation in same orientation as (a). **(c)** Electrostatic surface calculated using Pymol in the same orientation as **(a)**. Highly electropositive DNA binding surface shows a high level of conservation.

5.6.3   Generation of a CSP/dT₇ model

Two recent publications of *B. caldolyticus* and *B. subtilis* in complex with hexathymidine (dT_6) provide the position of five and six DNA-binding subsites [14, 25], respectively, which can be used in conjunction to obtain a valid model of seven complete DNA-binding subsites. The *B. caldolyticus*/dT_6 structure provides the position of DNA-binding subsites 2-7 and the *B. subtilis*/dT_6 structure provides the position of DNA-binding subsite 1, which is occupied by T5 of the dT_6 ligand. Figure 5-21a-b shows an overlay of the dT_6 ligands from each of the structures, bound to *Bs*CspB. Combination of all seven DNA-binding subsites resulted in a convincing *Bs*CspB/dT_7 model (Fig. 5-21c-e).

Starting at subsite 1 of the dT_7 bound ligand and following the 5′-3′ polarity of the DNA, the following interactions are made between the protein and DNA. At subsite 1 (Figure 5-22f), T1 makes an edge-on stack with Phe38. At subsite 2 (Figure 5-22f), T2 forms a double stack with Phe38 and Phe30 and the nucleobase head groups make specific hydrogen bonds with the Lys39 backbone and a water, which coordinates two more hydrogen bonds with Ser11 and Thr40. At subsite 3 (Figure 5-22g), T3 makes an edge-on contact with His29 and a hydrogen bond with the deoxyribose ring oxygen of the nucleoside. At subsite 4 (Figure 5-22g), T4 forms a stack with His29 and Gln59 makes a hydrogen bond via the nucleobase head group. At subsite 5 (Figure 5-22h), T5 stacks with Phe27 and Arg56 hydrogen bonds via the nucleobase head group and a phosphate oxygen. At subsite 6 (Figure 5-22h), T6 stacks against Phe17 and its head groups interact with the side chains of Lys7 and Trp8 via hydrogen bonds. At subsite 7 (Figure 5-22i), T7 forms a stacking interaction Trp8.
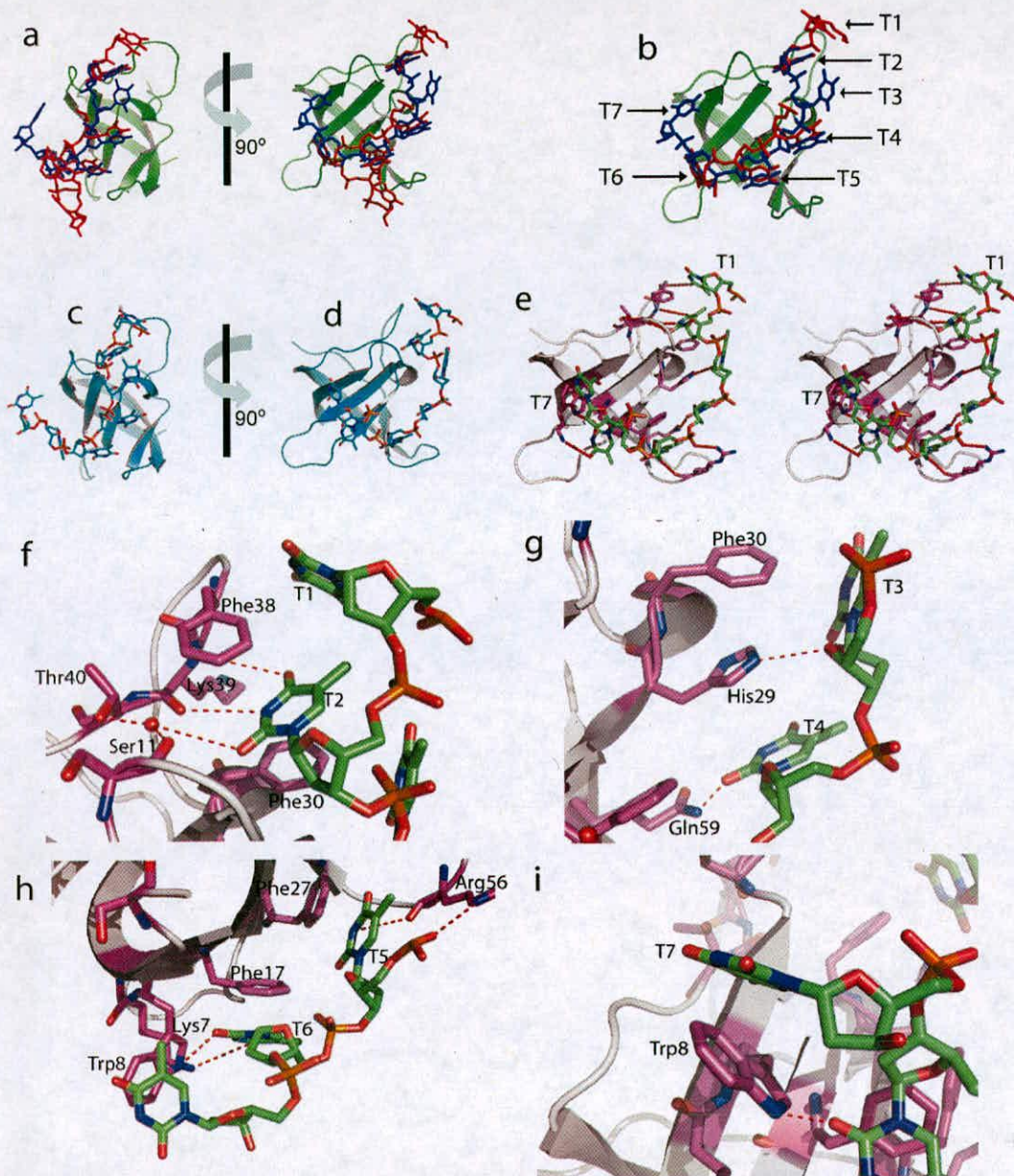
Figure 5-22. **Creation of the *Bs*CspB/heptathymine complex model.**

(a) *Bs*CspB in complex with dT$_6$ (red) and the positions of the dT$_6$ ligand from the *B. caldolyticus*/dT$_6$ complex, overlaid (blue). **(b)** The position of each nucleobase subsite from the combined T$_6$ ligands. **(c)** The *Bs*CspB/dT$_7$ model generated by combining the interacting subsites of both dT$_6$ ligands. **(e)** Stereo image of the *Bs*CspB/dT$_7$ model. **(f-i)** The interactions made between the protein and dT$_7$ ligand at each of the DNA-binding subsites 1-7. All interactions are described in the text above.

5.6.4   Generation of the StCspE/dT7 model

The BsCspB/dT7 model was superimposed onto the experimental derived crystal structure of StCspE (Figure 5-23a). All of the BsCspB/dT7 DNA binding subsites (Figure 5-22f-i) were highly conserved in the StCspE structure, with two minor discrepancies. In the BsCspB/dT7 structure, T5 hydrogen bonds with Arg56 via the nucleobase head group and phosphate oxygen (Figure 5-22h) but in the StCspE structure this residue is replaced by a lysine (Lys59) (Figure 5-23a).
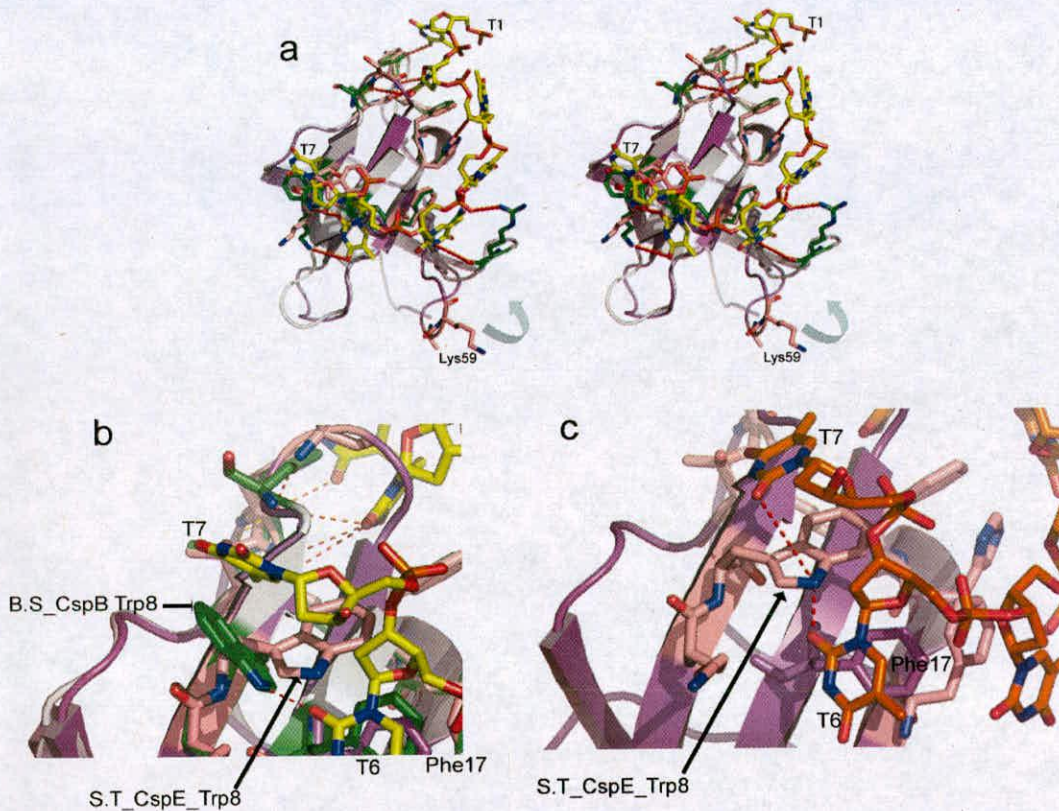
Figure 5-23. Superimposition of the StCspE structure on to the BsCspB/dT7 model.

(a) Stereo image of the BsCspB/dT7 model (grey) superimposed on the crystal structure of StCspE (Purple). StCspE amino acid side chains are coloured pink and BsCspB are coloured green. A blue arrow shows the movement in loop required for Lys59 of StCspE to substitute the contacts made between Arg56 of BsCspB and the dT7 ligand. **(b)** The stacking interaction made between T7 and Trp8 of BsCspB. **(c)** The new position of T7 in the StCspE/dT7 model, creating a similar stacking interaction.

An upward movement of the same loop region in the *St*CspE structure, containing Lys59, to the position of Arg56 would recreate the same hydrogen bond made at this subsite. In the *Bs*CspB/dT₇ structure at subsite 7 (Figure 5-22i), T7 forms a stacking interaction with Trp8. The position of the tryptophan is flipped out in the *St*CspE structure (Figure 5-23b). In order to mimic this stacking interaction in the *St*CspE/dT₇ model, the position of T7 was adjusted (Figure 5-23c). This adjustment resulted in the formation of an additional hydrogen bond, between the NH group of the tryptophan and the nucleobase head group oxygen (Figure 5-23c).

5.6.5 A StCspE/5′-GTCTTTT-3′ model explains sequence specificity

Previous microarray binding studies (Chapter 3) for *St*CspE identified the consensus ss DNA binding sequence, 5′-GTCTTTT-3′, with a binding preference for guanine (G) at position 1 and strong preference for cytosine (C) at position 3. The G base preference at position 1 and the C preference at position 3 were examined using ITC by systematically altering the base at position 1 or 3 for a thymine. The results clearly show that there is a preference for G at position 1 and C at position 3, with a 7-fold increase seen in K_d for the consensus sequence, 5′-GTCTTTT-3′, when compared to 5′-TTTTTTT-3′. To explain the base preferences of G and C, the heptathymidine ligand from the *St*CspE/dT₇ model was mutated to 5′-GTCTTTT-3′ (Figure 5-24a).

On replacement of the thymine with a cytosine at subsite 3 of the dT₇ ligand, the apparent K_d decreased from 78.74 ± 10.47 to 39.84 ± 4.47 nM. In the *Bs*CspB/dT₇ and *St*CspE/dT₇ model the thymine base at position 3 makes an edge-on contact with

His29 and a hydrogen bond with the deoxyribose ring oxygen of the nucleotide (Figure 5-22g).
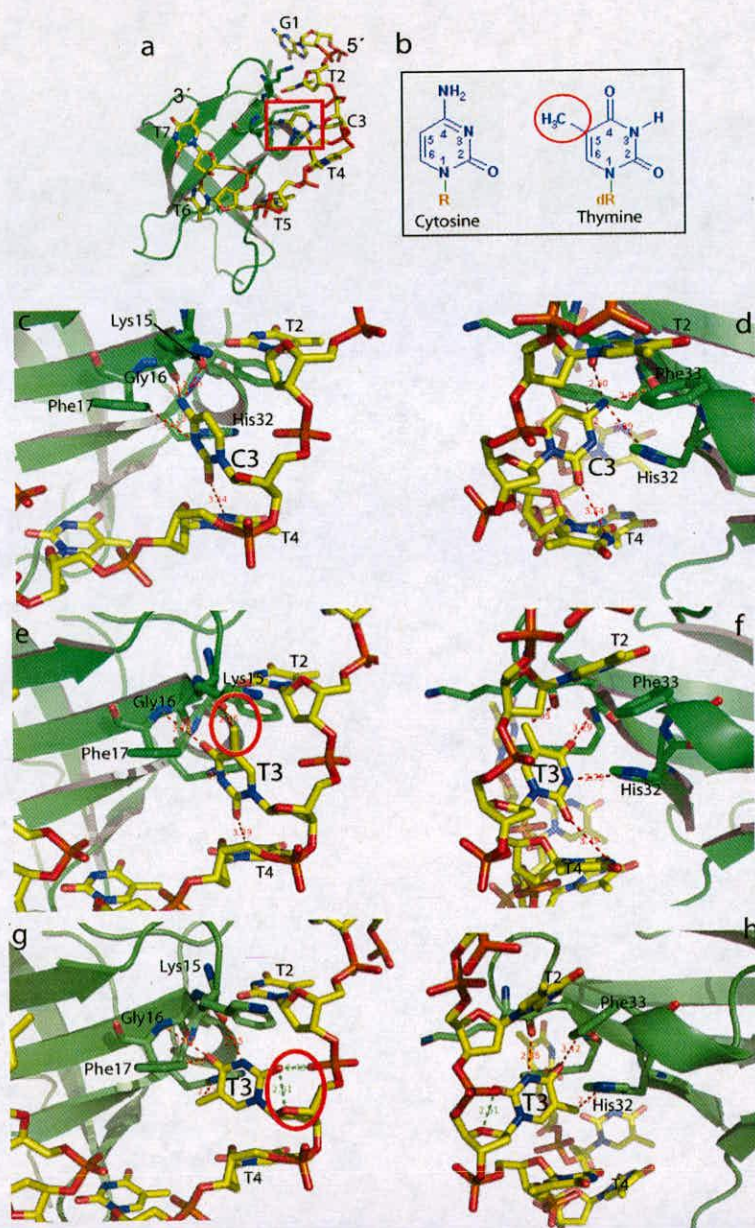
Figure 5-24. **Cytosine preference in the *St*CspE/5′-GTCTTTT-3′ structure.**

(a) *St*CspE in complex with 5′-GTCTTTT-3′ (yellow). Subsite 3 is highlighted by a red box. (b) Chemical structure of cytosine and thymine nucleobases. (c) Subsite 3, showing a bound cytosine nucleobase, and rotated 120° (d). (e) Subsite 3, showing a bound thymine nucleobase, and rotated 120° (f). (g) Subsite 3, showing a bound thymine nucleobase bound in the opposite conformation to (c) and rotated 120° (h). Hydrogen bonds are shown as dashed red lines and interactions < 2.5 Å have been highlighted with dashed green lines and circled in red. All interactions are described in the text below.

Changing the T3 to C3 at this position could not explain the specificity or base preference observed at this subsite. A new adjacent subsite was discovered which provides an ideal shape and hydrogen-bonding complementarity for the specific binding of cytosine over thymine (Figure 5-24c). At subsite 3, C3 head groups make three specific hydrogen bonds, two of which are via the backbone oxygens of Gly 16 and Lys15. A third hydrogen bond is also made between the cytosine head group oxygen and the NH group of T4. The binding of thymine nucleobase, in both possible conformations, to this cytosine specific pocket was examined (Figure 5-24e-h). Upon substitution of a thymine for the cytosine in subsite 3, the methyl group sterically inhibits the binding of a thymine base (Figure 5-24e – circled red). Rotation of the thymine base 180° in the pocket (Figure 5-24g) eliminated the initially steric hindrance observed but resulted in opposing charges from the oxygen of the thymine base head group and a backbone phosphate oxygen and deoxyribose oxygen (Figure 5-24f – red circle). These opposing charges would prevent the thymine nucleobase from binding to the cytosine specific binding pocket. The interactions observed for the docked cytosine and thymine provide evidence for specificity of cytosine nucleobase at subsite 3, which is in good agreement with ITC result.

On replacement of the thymine with a guanine at subsite 1 of the 5'-GTCTTTT-3' ligand, the apparent K_d decreases from 39.84 ± 4.47 to 11.27 ± 2.22 nM. In the BsCspB/dT$_7$ and StCspE/dT$_7$ model the thymine base at position 1 makes a side on stacking interaction with (Figure 5-25c). Changing T1 to G1 at this position retains the stacking interaction and results in the formation of an additional hydrogen bond between the backbone NH group of Gly37 and the headgroup oxygen of the guanine

nucleobase (Figure 5-25b). The single hydrogen bond difference between the docked guanine and thymine provides evidence for the differences in binding affinity observed during ITC binding experiments between 5′-TTCTTTT-3′ and 5′-GTCTTTT-3′.
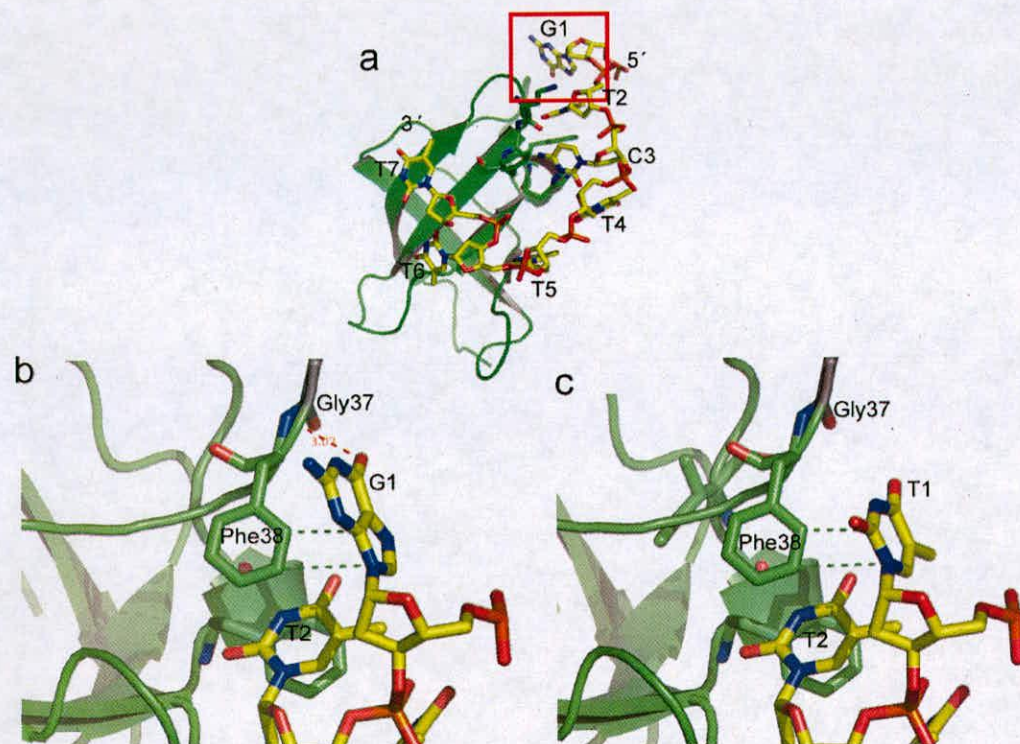
Figure 5-25. **Guanine preference in the *St*CspE/5′-GTCTTTT-3′ structure.**

(a) *St*CspD in complex with 5′-GTCTTTT-3′ (yellow). Subsite 1 is highlighted by a red box. (b) Subsite 1, showing a bound guanine nucleobase. (c) Subsite 1, showing a bound thymine nucleobase. Hydrogen bonds are shown as dashed red lines and stacking interactions have been highlighted with dashed green lines. All interactions are described in the text above.

5.6.6 Generation of the StCspD/5′-ACGGGG-3′ model

A computer generated model [205] of *St*CspD was superimposed onto the *St*CspE/5′-GTCTTTT-3′ complex model to obtain docking positions for the experimentally determined high-affinity ligand, 5′-ACGGGG-3′, resulting in the *St*CspD/5′-ACGGGG-3′ model (Figure 5-26a).

Starting at subsite 1 of the bound ligand and following the 5′-3′ polarity of the DNA, the following interactions are made between StCspD and DNA. At subsite 1 (Figure 5-26c), A1 forms a double stack with Tyr39 and Try31 and the nucleobase head groups make specific hydrogen bonds with the Arg40 backbone oxygen and the OH groups of both tyrosines. At subsite 2 (Figure 5-26d), C2 head groups make three hydrogen bonds via the backbone oxygens of Lys13 and Gly14 and the nitrogen of His30. At subsite 3 (Figure 5-26e), G3 stacks with His30 and the nucleobase forms a hydrogen bond via the nucleobase head group of G4. The head groups of G3 have the potential to form additional hydrogen bonds with nearby His60 via water molecules (Figure 5-26e - dashed green lines). At subsite 4 (Figure 5-26f), G4 forms a stacking interaction with Phe28 and forms several hydrogen bonds via the backbone oxygen and nitrogen of Lys57 and the backbone oxygen of Asn59. At subsite 5 (Figure 5-26g), G5 stacks with Phe17 and head groups hydrogen bond with the amino acid side chains, Trp8, Lys7 and Asp26. At subsite 6 (Figure 5-26h), G6 forms a stacking interaction with Trp8 and two base specific hydrogen bonds with Asn10.

Previous microarray binding studies (Chapter 3) for StCspD identified the consensus ss DNA binding sequence, 5′-ACGGGG-3′, which was later identified by SPR as an extremely tight binding ligand (K_d = 19.16 nM). The StCspD/5′-ACGGGG-3′ model clearly shows how the DNA ligand could specifically bind to CspD but without data this is merely speculation.
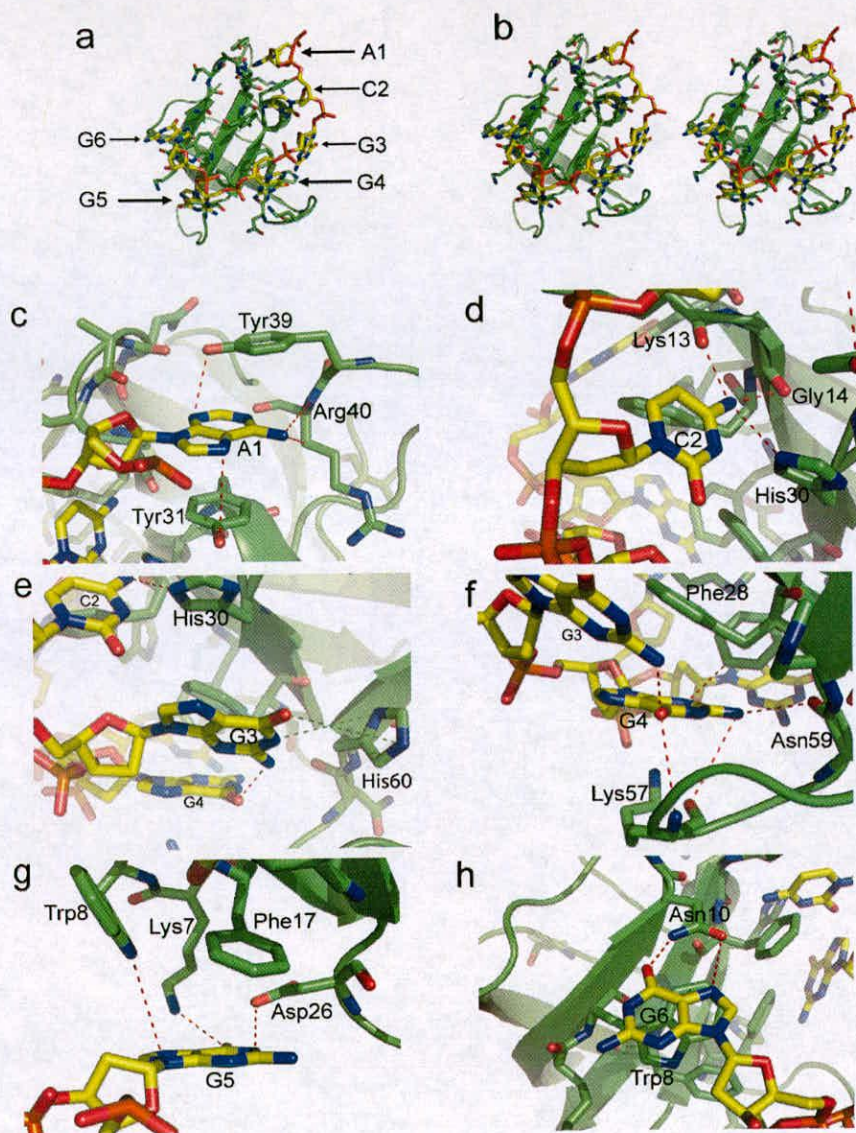
Figure 5-26. **A model of the *St*CspD in complex with 5′-ACGGGG-3′.**

(a) *St*CspE in complex with 5′-ACGGGG-3′ (yellow), showing the position of each nucleobase subsite. **(b)** Stereo image of the *St*CspE/5′-ACGGGG-3′ complex. **(c-h)** The interactions made between the protein and 5′-ACGGGG-3′ ligand at each of the DNA-binding subsites 1-6. Hydrogen bonds are shown as dashed red lines and potential hydrogen via water molecules have been highlighted with dashed green lines. All interactions are described in the text above.

5.7 Conclusions

The 5′-GTCTTTT-3′ ligand was docked onto the *St*CspE crystal structure. ITC

results clearly show that there is a preference for G at position 1 and C at position 3

over a thymine base. Comparison of the *St*CspE/5´-GTCTTTT-3´ with the *St*CspE/5´-TTTTTTT-3´ structure has clearly revealed key differences in the interactions at subsite one and three, which explains the difference in binding preference and affinity for the 5´-GTCTTTT-3´ over the 5´-TTTTTTT-3´ ligand.

Although the computer generated *St*CspD/5´-ACGGGG-3´ model provides a hypothetical explanation of how the experimentally determined DNA binding preference, 5´-ACGGGG-3´, could bind to *St*CspD, it cannot be used with any degree of confidence to clarify differences in the binding preference between *St*CspE and *St*CspD. This is primarily due to the uncertainty associated with modelling the protein structure of *St*CspD, as gel filtration results have suggested *St*CspD formed dimeric structures at high protein concentrations in solution (Figure 3-8a). The dimeric structure may be similar to that of the domain swapped dimer of the *Bc*Csp/T6 structure [25], making it extremely difficult to accurately predict a ss DNA binding mode for *St*CspD.

CHAPTER 6. Summary and future work

This thesis has described the structural and biochemical studies of cold shock domain containing proteins from *Bacillus subtilis* and *Salmonella typhimurium*. The work was broadly divided into four main projects:

1. A novel DNA microarray approach for the identification of sequence specificity of single stranded DNA binding proteins (Chapter 2).
2. The purification and characterisation of a family of CSPs from *salmonella typhimurium* (Chapter 3).
3. The biochemical characterisation of two key CSPs from the family of *Salmonella typhimurium* CSPs (Chapter 4).
4. The structural studies of CspE and CspD from *salmonella typhimurium* (Chapter 5).

This chapter summarises the major findings and conclusions of the projects and also outlines the areas which require further work.

6.1 Sequence specificity of single stranded DNA binding proteins: A novel DNA microarray approach.

As described in chapter one, SNABPs have been shown to bind with high affinity, non-specifically and specifically, to ss DNA which has been shown to regulate gene expression both positively and negatively. To understand how binding of SNABPs to ss nucleic acids regulates transcription and translation, the regions of sequence specificity must be identified. The identification of preferred binding sequences of single stranded nucleic acid binding proteins can be a long laborious process using standard techniques. A high-throughput method is required which allows the user to screen multiple binding sequences for the candidate SNABP. Many SNABPs contain a small structural motif known as the <u>O</u>ligonucleotide/oligosaccharide <u>b</u>inding fold (OB-fold). Cold shock proteins (CSPs) are the simplest examples of an OB-fold

containing proteins, containing only a single OB-fold domain, which provide an exemplary test protein for the development of a high-through screening assay. CSPs have been shown to be directly or indirectly involved in protein transcription and translation but a clear mechanism has yet to be elucidated. Initially CSPs were thought to bind preferentially to the Y-Box sequence, ATTGG, as well as the reverse complement, CCAAT [86]. The interactions between CSPs have been further examined over the years using various experimental techniques and have clearly demonstrated that highly homologous CSPs bind more than just the Y-Box sequence.

6.1.1 Project aims

To examine the use of a generic oligonucleotide microchip as the basis for the development of a high-throughput assay for the identification of sequence specific ss DNA protein interactions, using the major CspB from *B. subtilis*.

6.1.2 Major findings and conclusions presented in this thesis

- Proteins could be labelled appropriately (either directly labelled with Cy5 or detected using Cy3-penta-His antibodies) and the subsequent binding to the chip could then be visualised using a standard microarray scanner.

- Two recombinant forms of the major cold shock protein from *B. subtilis* (an N-terminal 6xHis tag and a non-tagged construct) were expressed and purified. Both forms of the CspB bound oligonucleotides that were structurally consistent with the oligonucleotides found on the array surface with similar affinity.

- CspB binds with high-affinity to pyrimidine-rich sequences on the surface of the chip, with a binding preference for the consensus sequence, 5´-GTCTTTG/T-3´.

- The sequence was modelled onto the known structure of CspB and a cytosine-binding pocket was identified, which explains the strong preference for a cytosine base at position 3.

- This novel unbiased microarray assay provides a means of rapidly identify optimal binding sequences for SNABPs *in vitro*, which may provide insight into their role in regulating cellular functions.

6.1.3 Future work

The next step in this project is to screen a whole repertoire of SNABPs from many different organisms. The results may identify SNABP promoter regions, as yet previously unseen, which may highlight new gene regulation pathways and provide a clear insight into their role in regulating cellular functions.

The human Y-Box protein, P50 (a eukaryotic CSP homologue), has been initially screened for ss DNA sequence specificity using the novel DNA microarray assay (refer to section A1.7 for a summary of the completed work). The results were promising and have been therefore included as an appendix (Appendix 1.7). These preliminary results indicate that P50 has a binding preference for the sequence, 5´-ACGGGGT-3´, which is very similar to the consensus binding sequence identified for *St*CspD (5´-ACGGGG-3´) from *S. typhimurium*, again highlighting a new polyguanine binding subgroup with CSD containing proteins. *St*CspD may therefore have similar functions to P50 and bind to mRNA not merely as an RNA chaperone but may inhibit or enhance translation [190, 191].

6.2 The purification and initial characterisation of a family of CSPs from *salmonella typhimurium*.

In *Salmonella typhimurium*, six CSP (*St*CspA, *St*CspB, *St*CspC, *St*CspD, *St*CspE, *St*CspH) have been identified and the cold inducibility of *St*CspA , *St*CspB and *St*CspH has been reported [157-160], although their functions are yet to be clearly elucidated. The complete ss DNA binding characterisation of this family may highlight differences/similarities in DNA binding, corresponding to the potential functions of each CSP.

6.2.1 Project aims

To clone, express and purify each of the six CSPs from *S. typhimurium* and to determine the specific ss DNA sequence recognition using the previously developed microarray assay.

6.2.2 Major findings and conclusions

- Recombinant N-Terminal His-tagged constructs of all six CSPs from *S. typhimurium* were successfully generated.

- All *St*CSPs, with the exception of *St*CspH, were successfully expressed, purified and the sequence specificity for ss DNA binding was determined using the DNA microarray assay.

- *St*CspD bound purine (guanine) rich sequences, resulting in the consensus binding sequence, 5′-ACGGGG-3′. *St*CspA, *St*CspB, *St*CspC and CspE bound pyrimidine (thymine) rich sequences, with an identical core consensus sequence, 5′-TCTTT-3′.

6.2.3 Future work

The two types of ss DNA binding specificities (5′-ACGGGG-3′ and 5′-TCTTT-3′) were observed in the small group of CSPs analysed in this work, which was suggestive that there may be differences in function. These differences in function have been investigated in a series of preliminary complementation experiments.

Table 6-1. Growth of CSP mutants after 72 hours at 15°C.

Strain	Growth*
ΔCspA	YES
ΔCspB	YES
ΔCspC	YES
ΔCspD	NO
ΔCspE	YES
ΔCspH	NO
ΔNull	NO
Wild-type	YES

*Growth comparable to wild-type after 72 hrs
Δ = all other CSP have been mutated with the exception of CspX
Null = all CSP had been mutated.

As discussed previously, (Chapter 1) CSPs are essential for cellular growth at low temperatures in prokaryotes [74] and recent experiments have shown that the presence of certain individual *St*CSP have the ability to restore wild-type cellular growth of *S. typhimurium* at 15°C (Table 6-1).

The only CSPs that did not restore wild-type growth of *S. typhimurium* after 72 hours at 15°C were *St*CspD and *St*CspH. Interestingly, *St*CspD also showed a completely different ss DNA binding specificity (5′-ACGGGG-3′) when compared with *St*CspA, *St*CspB, *St*CspC and *St*CspE, which bound similar ss DNA sequences (5′-TCTTT-3′) and also restored wild-type growth at 15°C (Table 6-1). Clearly *St*CspD has a different function in *S. typhimurium*, which has been clearly highlighted throughout this thesis. CSP paralogues in *B. subtilis* [74] and *E. coli* [69] have been shown to have similar functions but this work clearly demonstrates that *St*CspD (and

most likely *St*CspH) have different functions, which are yet unclear but may be similar to that of their eukaryotic relatives, the Y-Box binding proteins [190, 191]. Further studies of *St*CspD and *St*CspH will hopefully determine the function of this new class of CSPs.

6.3   The biochemical characterisation of two key CSPs from the family of *Salmonella typhimurium* CSPs.

For this study, *St*CspE and *St*CspD were used as representative members of the CSD containing protein family. Initial ss DNA binding results have shown that *St*CspE binds preferentially to the pyrimidine rich heptanucleotide, 5´-GTCTTTT-3´, while *St*CspD exhibits a binding preference for the purine rich hexanucleotide, 5´-ACGGGG-3´.

6.3.1   Project aims

To investigate the interaction of *St*CspE and *St*CspD with various ss DNA sequences using surface plasmon resonance method and isothermal titration calorimetry and consequently gain a better understanding of the kinetics and thermodynamics of CSP/ss DNA interactions.

6.3.2   Major findings and conclusions

- *St*CspE was found to bind with highest affinity to the oligonucleotide sequence, 5´-GTCTTTT-3´.

- SPR and ITC results for the His_6*St*CspE/5´-GTCTTTT-3´ interactions revealed temperature dependence of equilibrium dissociation constants (K_d), ranging from 8.71 $\pm$ 1.77 nM at 10°C to 258 $\pm$ 17.73 nM at 37°C, which was primarily due to a large increase in dissociation rate constant (k_d).

- Problems arose when creating a stable Biacore chip surface to assay StCspD/ss DNA interactions. A covalent like immobilisation strategy was achieved using a double-His-tagged version of the StCspD protein.

- Double-his-tagged StCspD was found to bind with highest affinity (19.16 nM) to the 5´-ACGGGG-3´ sequence and with an extremely slow k_d (2.70 $\pm$ 0.22 x 10^{-4} s^{-1}) at 15°C.

- StCspD and StCspE bind to different oligonucleotide sequences with very high affinity and with a completely different set of kinetics.

6.3.3   Future work

The double histidine tag vector is to be developed further and used to create stable covalent-like immobilisation of proteins on the surface of Biacore NTA chips. The stable surface would allow users to screen various ligands but would also allow for the removal of the double-his-tagged protein from the surface. The replacement of one protein for another protein of interest would reduce the overall cost of Biacore screening experiments as many proteins could be screened on a single NTA chip.

6.4   The structural studies of CspE and CspD from *salmonella typhimurium*.

Two different types of recognition sequences were determined for StCspE and StCspD. Previous results have shown that they recognise either purine (e.g. StCspD) or pyrimidine (e.g. StCspE) rich sequences in a sequence specific binding manner. Examination of the structure of CSPs in a DNA bound and unbound state, at an atomic level, would provide a detailed insight of the CSP-ss DNA interaction, one method that allows for such an analysis is X-ray crystallography.

6.4.1 Project Aims

- To obtain X-ray crystal structures of *St*CspE and *St*CspD.

- To obtain X-ray crystal structures of *St*CspE/5´-GTCTTTT-3´ and *St*CspD/5´-ACGGGG-3´.

6.4.2 Major findings and conclusions

- *St*CspE was crystallised in the monoclinic space group *P2₁* and the final model, refined to R/R$_{free}$ values of 20.3%/23.42%, consists of two *St*CspE molecules in the asymmetric unit packed DNA binding face to face, held together by six potential hydrogen bonds.

- The recombinant *St*CspE formed a crystallographic dimers but it was an N-terminal truncated version of the protein that formed dimers in solution. Whether the additional *St*CspE N-terminal amino acids (which were a result of cloning) are preventing *St*CspE forming dimers in solution, similar to that of the crystallographic structure, still remains to be determined.

- *St*CspE/5´-GTCTTT-3´ proved to be difficult to crystallise and as a result the 5´-GTCTTTT-3´ sequence was modelled onto the structure of *St*CspE. A cytosine specific binding pocket was identified at DNA subsite 3, explaining the preference for a cytosine base. The replacement of a thymine with a guanine at subsite 1 resulted in the gain of a hydrogen bond, explaining the preference for a guanine base at position this position. This structural analysis of *St*CspE specificity for the 5´-GTCTTTT-3´ sequence was clearly confirmed in earlier ITC experiments (see section 4.5.5).

- Attempts to obtain crystals of *St*CspD were unsuccessful but a computer generated model was created which was used to generate the *St*CspD/5´-

ACGGGG-3´ complex model, this provided reasoning for the base preference seen at each of the DNA binding subsites, as observed in previous biochemical experiments.

6.4.3 Future Work

The crystallisation of StCspD is ongoing using different crystallisation techniques (the batch method and sitting drop) and screening procedures to obtain diffraction quality crystals. Structural differences in StCspD may highlight why StCspD binds specifically to polyguanine rich sequences and why functionally it is unable to restore the cold shock phenotype. To purify full length (untagged) StCspE; the purification of full length StCspE would allow for the determination of the oligomeric state of StCspE in solution.

1   Appendix

1.1   Sequence details of the CSPs cloned from *S. typhimurium.*

The nucleotide sequences of the cold shock proteins from *Salmonella typhimurium* strain1344, which were cloned during this work.
(Sequenced by the Sanger centre: http://www.sanger.ac.uk/Projects/Salmonella/)
The nucleotide sequences have been colour coordinated with the primers used for cloning, which shows the sequence incorporated into the primers.

>CSPA
ATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGCTGATAAAGGCTTCGGCTTTA
TTACTCCTGATGACGGTTCTAAAGACGTGTTCGTACACTTCTCCGCTATTCAGAACGATGG
TTACAAATCTCTGGACGAAGGTCAGAAAGTTTCCTTCACCATCGAAAGCGGCGCTAAGGG
CCCGGCAGCTGGCAACGTAACCAGCCTGTAA

>CSPB
ATGACGACGAAAATCACTGGTTTAGTAAAATGGTTTAACCCTGAAAAGGGCTTTGGTTTC
ATTACGCCTAAAGATGGCAGCAAAGATGTGTTTGTGCATTTTTCAGCCATTCAAAGTAATG
AATTCCGCACTCTGAATGAAAATCAGGAAGTGGAGTTTTCAGTAGAGCAGGGACCAAAAG
GTCCATCAGCGGTCAACGTTGTGGCGCTTTAA

>CSPC
ATGGCAAAGATTAAAGGTCAAGTTAAGTGGTTCAACGAATCCAAAGGTTTTGGCTTCATT
ACTCCGGCTGACGGCAGCAAAGATGTGTTCGTACACTTCTCCGCTATCCAGGGCAATGGTT
TCAAAACTCTGGCTGAAGGCCAGAACGTTGAGTTCGAAATTCAGGACGGCCAGAAAGGTC
CGGCCGCTGTTAACGTAACAGCTATCTGA

>CSPD
ATGGAAACGGGTACTGTAAAGTGGTTCAACAATGCCAAAGGGTTTGGTTTCATCTGCCCT
GAAGGCGGCGGCGAGGATATTTTCGCCCATTATTCCACCATTCAAATGGATGGTTACAGA
ACGCTTAAAGCCGGACAGTCTGTCCGGTTTGATGTCCACCAGGGGCCAAAAGGCAATCAC
GCCAGCGTCATCGTGCCCATCGAAGCAGAGGCCGTTGCATAG

>CSPE
ATGTCTAAGATTAAAGGTAACGTTAAGTGGTTTAATGAATCCAAAGGATTCGGTTTCATTA
CTCCGGAAGATGGCAGCAAAGACGTGTTTGTACACTTCTCTGCAATCCAGACCAATGGTTT
TAAAACTCTGGCTGAAGGTCAGCGCGTAGAGTTCGAAATCACTAACGGTGCCAAAGGCCC
TTCCGCTGCAAACGTAACTGCTCTGTAA

>CSPH
ATGTCTCGTAAAATGACAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGGTCTC
ATCACCCCCTCCGATGGACGCAAAGATGTTCAGGTCCACATTTCAGCATGTCGCCAACAC
GAAACAGAAGCGCTTATCCCCGGTATACGCGTTGAGTTTTGTCGTATTAATGGCCTCCGCG
GACCTACCGCCGCCAACGTTTATCTTTCATAA

The following primers were used to clone the *S. typhimurium csp* genes. HindIII and NdeI sites have been added to all primers (shown in bold), primers are colour coordinate with the previous CSP nucleotide sequences.

FWD_CSPA 5'- ACACA **CAT** ATG TCC GGT AAA ATG ACT GGT ATC
32 bases long, GC content = 41%, Tm=61
REV_CSPA 5'- CA **AAGCTT** TTA CAG GCT GGT TAC GTT GCC AGC
32 bases long, GC content = 50%, Tm=64

FWD_CSPB 5'- ACGCA **CAT** ATG ACG ACG AAA ATC ACT GGT TTA
32 bases long, GC content = 41%, Tm=61
REV_CSPB 5'- CGA **AAGCTT** TTA AAG CGC CAC AAC GTT GAC CGC
32 bases long, GC content = 50%, Tm=64

FWD_CSPC 5'- ACGCA **CAT** ATG GCA AAG ATT AAA GGT CAA GTT
32 bases long, GC content = 38%, Tm=59
REV_CSPC 5'- CGA **AAGCTT** TCA GAT AGC TGT TAC GTT AAC AGC
33 bases long, GC content = 42%, Tm=62

FWD_CSPD 5'- ACACA **CAT** ATG GAA ACG GGT ACT GTA AAG TGG
32bases long, GC content = 44%, Tm=62
REV_CSPD 5'- CA **AAGCTT** CTA TGC AAC GGC CTC TGC TTC GAT
32bases long, GC content = 50%, Tm=64

FWD_CSPE 5' - AGGCA **CAT** ATG TCT AAG ATT AAA GGT AAC GTT
32bases long, GC content = 34%, Tm=58
REV_CSPE 5' – CGA **AAGCTT** TTA CAG AGC AGT TAC GTT TGC AGC
33bases long, GC content = 45%, Tm=63

FWD_CSPH 5' - AGGCA **CAT** ATG TCT CGT AAA ATG ACA GGA ATT
32bases long, GC content = 38%, Tm=59

REV_CSPH 5'- CGA **AAGCTT** TTA TGA AAG ATA AAC GTT GGC GGC
33bases long, GC content = 42%, Tm=62

1.2 Cloning of the six *S. typhimurium* CSPs, sequencing results.

The following sequence alignments show each of the *S. typhimurium* CSP nucleotide sequences aligned with the cloned sequences. Initially the *csp* genes were cloned into the pCR2.1-TOPO (TOPO 2.1) shuttle vector (Invitrogen) and were then subsequently cloned into the final pET28a expression vector.

CSPA in TOPO 2.1

```
            10        20        30        40        50        60
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd ATHHTGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATAT
CSPA         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

            70        80        90       100       110       120
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd CTGCATAAHTCGCCCTTACACACATATGTCCGGTAAAATGACTGGTATCGTAAAATGGTT
CSPA         ~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGTCCGGTAAAATGACTGGTATCGTAAAATGGTT

            130       140       150       160       170       180
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd CAACGCTGATAAAGGCTTCGGCTTTATTACTCCTGATGACGGTTCTAAAGACGTGTTCGT
CSPA         CAACGCTGATAAAGGCTTCGGCTTTATTACTCCTGATGACGGTTCTAAAGACGTGTTCGT

            190       200       210       220       230       240
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd ACACTTCTCCGCTATTCAGAACGATGGTTACAAATCTCTGGACGAAGGTCAGAAAGTTTC
CSPA         ACACTTCTCCGCTATTCAGAACGATGGTTACAAATCTCTGGACGAAGGTCAGAAAGTTTC

            250       260       270       280       290       300
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd CTTCACCATCGAAAGCGGCGCTAAGGGCCCGGCAGCTGGCAACGTAACCAGCCTGTAAAA
CSPA         CTTCACCATCGAAAGCGGCGCTAAGGGCCCGGCAGCTGGCAACGTAACCAGCCTGTAA

            310       320       330       340       350       360
     ....|....|....|....|....|....|....|....|....|....|....|....|
HughM_A2_fwd GCTTTGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTAC
CSPA
```

CSPB in TOPO 2.1

```
            10        20        30        40        50        60
     ....|....|....|....|....|....|....|....|....|....|....|....|
HM_B1_fwd  HHCHTGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATAT
CSPB         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

            70        80        90       100       110       120
     ....|....|....|....|....|....|....|....|....|....|....|....|
HM_B1_fwd  CTHCAHAATTCGCCCTTACGCACATATGACGACGAAAATCACTGGTTTAGTAAAATGGTT
CSPB         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGACGACGAAAATCACTGGTTTAGTAAAATGGTT

            130       140       150       160       170       180
     ....|....|....|....|....|....|....|....|....|....|....|....|
HM_B1_fwd  TAACCCTGAAAAGGGCTTTGGTTTCATTACGCCTAAAGATGGCAGCAAAGATGTGTTTGT
CSPB         TAACCCTGAAAAGGGCTTTGGTTTCATTACGCCTAAAGATGGCAGCAAAGATGTGTTTGT

            190       200       210       220       230       240
     ....|....|....|....|....|....|....|....|....|....|....|....|
HM_B1_fwd  GCATTCTTCAGCCATTCAAAGTAATGAATTCCGCACTCTGAATGAAAATCAGGAAGTGGA
CSPB         GCATTTTTCAGCCATTCAAAGTAATGAATTCCGCACTCTGAATGAAAATCAGGAAGTGGA

            250       260       270       280       290       300
     ....|....|....|....|....|....|....|....|....|....|....|....|
HM_B1_fwd  GTTTTCAGTAGAGCAGGGACCAAAAGGTCCATCAGCGGTCAACGTTGTGGCGCTTTAAAA
CSPB         GTTTTCAGTAGAGCAGGGACCAAAAGGTCCATCAGCGGTCAACGTTGTGGCGCTTTAA
```

CspC in TOPO 2.1

```
                  10        20        30        40        50        60
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     GNATNTNTAGGGCGNATTGGGACCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGG
CSPC          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                  70        80        90        100       110       120
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     ATATCNCCCNTAATNNGCCCTTACGCACATATGGCAAAGATTAAAGGTCNAGTTAAGTGG
CSPC          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGGCAAAGATTAAAGGTCAAGTTAAGTGG

                  130       140       150       160       170       180
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     TTCAACGAATCCAAAGGTTTTGGCTTCATTACTCCGGCTGACGGCAGCAAAGATGTGTTC
CSPC          TTCAACGAATCCAAAGGTTTTGGCTTCATTACTCCGGCTGACGGCAGCAAAGATGTGTTC

                  190       200       210       220       230       240
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     GTACACTTCTCCGCTATCCAGGGCAATGGTTTCAAAACTCTGGCTGAAGGCCAGAACGTT
CSPC          GTACACTTCTCCGCTATCCAGGGCAATGGTTTCAAAACTCTGGCTGAAGGCCAGAACGTT

                  250       260       270       280       290       300
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     GAGTTCGAAATTCAGGACGGCCAGAAAGGTCCGGCTGCTGTTAACGTAACAGCTATCTGA
CSPC          GAGTTCGAAATTCAGGACGGCCAGAAAGGTCCGGCCGCTGTTAACGTAACAGCTATCTGA

                  310       320       330       340       350       360
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_C1_fwd     AAGCTTTCGAAGGGCGAATTCCAGCACACTGGCNGCCGTTACTAGTGNATCCGAGCTCGG
CSPC
```

CspD in TOPO 2.1

```
                  10        20        30        40        50        60
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     ATNCTTGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGAIA
CSPD          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                  70        80        90        100       110       120
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     TCTNCANAATTCGCCCTTACCACATATGCAAACGGGTACTGTAAAGTGGTTCAACAATGC
CSPD          ~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGGAAACGGGTACTGTAAAGTGGTTCAACAATGC

                  130       140       150       160       170       180
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     CAAAGGGTTTGGTTTCATCTGCCCTGAAGGCGGCGGCGAGGATATTTTCGCCCATTATTC
CSPD          CAAAGGGTTTGGTTTCATCTGCCCTGAAGGCGGCGGCGAGGATATTTTCGCCCATTATTC

                  190       200       210       220       230       240
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     CACCATTCAAATGGATGGTTACAGAACGCTTAAAGCCGGACAGTCTGTCCGGTTTGATGT
CSPD          CACCATTCAAATGGATGGTTACAGAACGCTTAAAGCCGGACAGTCTGTCCGGTTTGATGT

                  250       260       270       280       290       300
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     CCACCAGGGGCCAAAAGGCAATCACGCCAGCGTCATCGTGCCCATCGAAGCAGAGGCCGT
CSPD          CCACCAGGGGCCAAAAGGCAATCACGCCAGCGTCATCGTGCCCATCGAAGCAGAGGCCGT

                  310       320       330       340       350       360
              ....|....|....|....|....|....|....|....|....|....|....|....|
HM_D1_fwd     TGCATAGAAGCTTTGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGA
CSPD          TGCATAG
```

214

CSPE in TOPO 2.1

```
              10        20        30        40        50        60
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd ATCHNTGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATA
CSPE      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
              70        80        90       100       110       120
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd TCTHCAHAATTCGCCCTTAGGCACATATGTCTAAGATTAAAGGTAACGTTAAGTGGTTTA
CSPE      ~~~~~~~~~~~~~~~~~~~~~~~~~ATGTCTAAGATTAAAGGTAACGTTAAGTGGTTTA
```

```
             130       140       150       160       170       180
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd ATGAATCCAAAGGATTCGGTTTCATTACTCCGGAAGATGGCAGCAAAGACGTGTTTGTAC
CSPE      ATGAATCCAAAGGATTCGGTTTCATTACTCCGGAAGATGGCAGCAAAGACGTGTTTGTAC
```

```
             190       200       210       220       230       240
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd ACTTCTCTGCAATCCAGACCAATGGTTTTAAAACTCTGGCTGAAGGTCAGCGCGTAGAGT
CSPE      ACTTCTCTGCAATCCAGACCAATGGTTTTAAAACTCTGGCTGAAGGTCAGCGCGTAGAGT
```

```
             250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd TCGAAATCACTAACGGTGCCAAAGGCCCTTCCGCTGCAAACGTAACTGCTCTGTAAAAGC
CSPE      TCGAAATCACTAACGGTGCCAAAGGCCCTTCCGCTGCAAACGTAACTGCTCTGTAA
```

```
             310       320       330       340       350       360
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_E2_fwd TTTCGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACC
CSPE
```

CSPH in TOPO 2.1(5′-3′ sequence)

```
              10        20        30        40        50        60
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev NGCTGGTCGAGCTGGATCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTAGGCA
CSPH      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
              70        80        90       100       110       120
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev CAHCTHTTAHHGHAAAATGACAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGG
CSPH      ~~~~ATGTCTCGTAAAATGACAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGG
```

```
             130       140       150       160       170       180
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev TCTCATCACCCCCTCCGATGGACGCAAAGATGTTCAGGTCCACATTTCAGCATGTCGCCA
CSPH      TCTCATCACCCCCTCCGATGGACGCAAAGATGTTCAGGTCCACATTTCAGCATGTCGCCA
```

```
             190       200       210       220       230       240
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev ACACGAAACAGAAGCGCTTATCCCCGGTATACGCGTTGAGTTTTGTCGTATTAATGGCCT
CSPH      ACACGAAACAGAAGCGCTTATCCCCGGTATACGCGTTGAGTTTTGTCGTATTAATGGCCT
```

```
             250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev CCGCGGACCTACCGCCGCCAACGTTTATCTTTCATAAAAGCTTTCGAAGGGCGAATTCTG
CSPH      CCGCGGACCTACCGCCGCCAACGTTTATCTTTCATAA
```

```
             310       320       330       340       350       360
         ....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_rev CAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTA
CSPH
```

CSPH in TOPO 2.1 (3′-5′ sequence)

```
            10        20        30        40        50        60
         ...|....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_fwd ATNNTGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATAT
CSPH      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


            70        80        90       100       110       120
         ...|....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_fwd CTCCNTAANNCGCCCTTCGAAAGCTTTTATGAAAGATAAACGTTGGCGGCGGTAGGTCCG
CSPH      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TTATGAAAGATAAACGTTGGCGGCGGTAGGTCCG


           130       140       150       160       170       180
         ...|....|....|....|....|....|....:|....|....|....|....|....|....|.
HM_H1_fwd CGGAGGCCATTAATACGACAAAACTCAACGCGTATACCGGGGATAAGCGCTTCTGTTTCG
CSPH      CGGAGGCCATTAATACGACAAAACTCAACGCGTATACCGGGGATAAGCGCTTCTGTTTCG


           190       200       210       220       230       240
         ...|....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_fwd TGTTGGCGACATGCTGAAATGTGGACCTGAACATCTTTGCGTCCATCGGAGGGGGTGATG
CSPH      TGTTGGCGACATGCTGAAATGTGGACCTGAACATCTTTGCGTCCATCGGAGGGGGTGATG


           250       260       270       280       290       300
         ...|....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_fwd AGACCTTTACCGCTCTTACAATCAAAGGTTTTGACAATTCCTGTCATTTTACGAGACATA
CSPH      AGACCTTTACCGCTCTTACAATCAAAGGTTTTGACAATTCCTGTCATTTTACGAGACAT


           310       320       330       340       350       360
         ...|....|....|....|....|....|....|....|....|....|....|....|....|
HM_H1_fwd TGTGCCTAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTA
CSPH
```

CSPS in pET28a

Sequencing results for CspA in pET28, using the T7fwd primer for sequencing.

```
HughM_CSP_A_fwd  GCCCCTTAAAGNHCATCATCACAGCAGCGGCCTGGTGCCGCGCCGCHACC
CSPA             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


HughM_CSP_A_fwd  ATATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGCTGATAAA
CSPA             ~~ATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGCTGATAAA


HughM_CSP_A_fwd  GGCTTCGGCTTTATTACTCCTGATGACGGTTCTAAAGACGTGTTCGTACA
CSPA             GGCTTCGGCTTTATTACTCCTGATGACGGTTCTAAAGACGTGTTCGTACA


HughM_CSP_A_fwd  CTTCTCCGCTATTCAGAACGATGGTTACAAATCTCTGGACGAAGGTCAGA
CSPA             CTTCTCCGCTATTCAGAACGATGGTTACAAATCTCTGGACGAAGGTCAGA


HughM_CSP_A_fwd  AAGTTTCCTTCACCATCGAAAGCGGCGCTAAGGGCCCGGCAGCTGGCAAC
CSPA             AAGTTTCCTTCACCATCGAAAGCGGCGCTAAGGGCCCGGCAGCTGGCAAC


HughM_CSP_A_fwd  GTAACCAGCCTGTAAAAGCTTGCGGCCGCACTCGAGCACCACCACCACCA
CSPA             GTAACCAGCCTGTAA


HughM_CSP_A_fwd  CCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG
CSPA


HughM_CSP_A_fwd  CTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGG
CSPA
```

Sequencing results for CspB in pET28, using the T7fwd primer for sequencing.

```
              10        20        30        40        50        60        70
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_B_fwd  GCCCCTTNAGGNNCATCATCACAGCAGCGGCCTGGTGCCGCGCGCCNGCCATATGACGACGAAAATCACT
CSPB          ------------------------------------------------ATGACGACGAAAATCACT


              80        90        100       110       120       130       140
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_B_fwd  GGTTTAGTAAAATGGTTTAACCCTGAAAAGGGCTTTGGTTTCATTACGCCTAAAGATGGCAGCAAAGATG
CSPB          GGTTTAGTAAAATGGTTTAACCCTGAAAAGGGCTTTGGTTTCATTACGCCTAAAGATGGCAGCAAAGATG


              150       160       170       180       190       200       210
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_B_fwd  TGTTTGTGCATTCTTCAGCCATTCAAAGTAATGAATTCCGCACTCTGAATGAAAATCAGGAAGTGGAGTT
CSPB          TGTTTGTGCATTTTTCAGCCATTCAAAGTAATGAATTCCGCACTCTGAATGAAAATCAGGAAGTGGAGTT


              220       230       240       250       260       270       280
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_B_fwd  TTCAGTAGAGCAGGGACCAAAAGGTCCATCAGCGGTCAACGTTGTGGCGCTTTAAAAGCTTGCGGCCGCA
CSPB          TTCAGTAGAGCAGGGACCAAAAGGTCCATCAGCGGTCAACGTTGTGGCGCTTTAA--------------


              290       300       310       320       330       340       350
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_B_fwd  CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGANTTGGCTG
CSPB          --------------------------------------------------------------------
```

Sequencing results for CspC (5′-3′) in pET28, using the T7fwd primer for sequencing.

```
              10        20        30        40        50        60        70
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_C_fwd  GCCCCNTTAAAGNNNCATCATCACAGCAGCGGCCTGGTGCCGCGCGNCCNGNNATATGGCAAAGATTAAAGG
CSPC          -------------------------------------------------ATGGCAAAGATTAAAGG


              80        90        100       110       120       130       140
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_C_fwd  TCAAGTTAAGTGGTTCAACCGAATCCAAAGGTTTTGGCTTCATTACTCCGGCTGACGGCAGCAAAGATGTG
CSPC          TCAAGTTAAGTGGTTCAACCGAATCCAAAGGTTTTGGCTTCATTACTCCGGCTGACGGCAGCAAAGATGTG


              150       160       170       180       190       200       210
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_C_fwd  TTCGTACACTTCTCCGCTATCCAGGGCAATGGTTTCAAAACTCTCGGCTGAAGGCCAGAACGTTGAGTTCG
CSPC          TTCGTACACTTCTCCGCTATCCAGGGCAATGGTTTCAAAACTCTCGGCTGAAGGCCAGAACGTTGAGTTCG


              220       230       240       250       260       270       280
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_C_fwd  AAATTCAGGACGGCCAGAAAGGTCCGGCTGCTGTTAACGTAACAGCTATCTGAAAGCTTGCGGCCGCACT
CSPC          AAATTCAGGACGGCCAGAAAGGTCCGGCCGCTGCTGTTAACGTAACAGCTATCTGA-------------


              290       300       310       320       330       340       350
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_C_fwd  CGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT
CSPC          --------------------------------------------------------------------
```

217

Sequencing results for CspD (5′-3′) in pET28, using the T7fwd primer for sequencing.

```
                10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_D_fwd GCCCCCHTAAAGHHCATCATCACAGCAGCGGCCTGGTGCCGCGCGCCNGHCATATGGAAACGGGTACTGT
CSPD         -----------------------------------------------ATGGAAACGGGTACTGT

                80        90        100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_D_fwd AAAGTGGTTCAACAATGCCAAAGGGTTTGGTTTCATCTGCCCTGAAGGCGGCGGCCAGGATATTTTCGCC
CSPD         AAAGTGGTTCAACAATGCCAAAGGGTTTGGTTTCATCTGCCCTGAAGGCGGCGGCGAGGATATTTTCGCC

                150       160       170       180       190       200       210
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_D_fwd CATTATTCCACCATTCAAATGGATGGTTACAGAACGCTTAAAGCCGGACAGTCTGTCCGGTTTGATGTCC
CSPD         CATTATTCCACCATTCAAATGGATGGTTACAGAACGCTTAAAGCCGGACAGTCTGTCCGGTTTGATGTCC

                220       230       240       250       260       270       280
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_D_fwd ACCAGGGGCCAAAAGGCAATCACGCCAGCGTCATCGTGCCCATCGAAGCAGAGGCCGTTGCATAGAAGCT
CSPD         ACCAGGGGCCAAAAGGCAATCACGCCAGCGTCATCGTGCCCATCGAAGCAGAGGCCGTTGCATAG-----

                290       300       310       320       330       340       350
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_D_fwd TGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCT
CSPD         ---------------------------------------------------------------------
```

Sequencing results for CspE (5′-3′) in pET28, using the T7fwd primer for sequencing.

```
                10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_E_fwd AGCCCCHHHAAAGGGGCATCATCACAGCAGCGGCCTGGTGCCGCGCNCCCTHAHATATGTCTAAGATTAA
CSPE         ------------------------------------------------------ATGTCTAAGATTAA

                80        90        100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_E_fwd AGGTAACGTTAAGTGGTTTAATGAATCCAAAGGATTCGGTTTCATTACTCCGGAAGATGGCAGCAAAGAC
CSPE         AGGTAACGTTAAGTGGTTTAATGAATCCAAAGGATTCGGTTTCATTACTCCGGAAGATGGCAGCAAAGAC

                150       160       170       180       190       200       210
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_E_fwd GTGTTTGTACACTTCTCTGCAATCCAGACCAATGGTTTTAAAACTCTGGCTGAAGGTCAGCGCGTAGAGT
CSPE         GTGTTTGTACACTTCTCTGCAATCCAGACCAATGGTTTTAAAACTCTGGCTGAAGGTCAGCGCGTAGAGT

                220       230       240       250       260       270       280
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_E_fwd TCGAAATCACTAACGGTGCCAAAGGCCCTTCCGCTGCAAACGTAACTGCTCTGTAAAAGCTTGCGGCCGC
CSPE         TCGAAATCACTAACGGTGCCAAAGGCCCTTCCGCTGCAAACGTAACTGCTCTGTAA-------------

                290       300       310       320       330       340       350
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM_CSP_E_fwd ACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCT
CSPE         ---------------------------------------------------------------------
```

218

Sequencing results for CspH (5′-3′) in pET28, using the T7fwd primer for sequencing.

```
              10        20        30        40        50        60        70
         ...I....I....I....I....I....I....I....I....I....I....I....I....I....I
HM_CSP_H_fwd AGCCCCHNTAANGNNCATCATCACAGCAGCGGCCTGGTGCCGCGCCCTNNCCATATGTCTCGTAAAATGA
CSPH         --------------------------------------------------ATGTCTCGTAAAATGA

              80        90       100       110       120       130       140
         ...I....I....I....I....I....I....I....I....I....I....I....I....I....I
HM_CSP_H_fwd CAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGGTCTCATCACCCCCTCCGATGGACGCAAAGA
CSPH         CAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGGTCTCATCACCCCCTCCGATGGACGCAAAGA

             150       160       170       180       190       200       210
         ...I....I....I....I....I....I....I....I....I....I....I....I....I....I
HM_CSP_H_fwd TGTTCAGGTCCACATTTCAGCATGTCGCCAACACGAAACAGAAGCGCTTATCCCCGGTATACGCGTTGAG
CSPH         TGTTCAGGTCCACATTTCAGCATGTCGCCAACACGAAACAGAAGCGCTTATCCCCGGTATACGCGTTGAG

             220       230       240       250       260       270       280
         ...I....I....I....I....I....I....I....I....I....I....I....I....I....I
HM_CSP_H_fwd TTTTGTCGTATTAATGGCCTCCGCGGACCTACCGCCGCCAACGTTTATCTTTCATAAAAGCTTGCGGCCG
CSPH         TTTTGTCGTATTAATGGCCTCCGCGGACCTACCGCCGCCAACGTTTATCTTTCATAA-----------
```

1.3 *S. typhimurium* CSP amino acid sequences

>CSPA (Theoretical pI/Mw: 5.58 / 7403.28)
MSGKMTGIVKWFNADKGFGFITPDDGSKDVFVHFSAIQNDGYKSLDEGQKV
SFTIESGAKGPAAGNVTSL

>CSPB (Theoretical pI/Mw: 5.68 / 7744.76)
MTTKITGLVKWFNPEKGFGFITPKDGSKDVFVHFSAIQSNEFRTLNENQEVEF
SVEQGPKGPSAVNVVAL

>CSPC (Theoretical pI/Mw: 6.54 / 7402.37) (lots)
MAKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVE
FEIQDGQKGPAAVNVTAI

>CSPD (Theoretical pI/Mw: 5.81 / 7884.86)
METGTVKWFNNAKGFGFICPEGGGEDIFAHYSTIQMDGYRTLKAGQSVRFD
VHQGPKGNHASVIVPIEAEAVA

>CSPE (Theoretical pI/Mw: 8.09 / 7451.40) (Lots)
MSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEF
EITNGAKGPSAANVTAL

>CSPH (Theoretical pI/Mw: 9.57 / 7659.91)
MSRKMTGIVKTFDCKSGKGLITPSDGRKDVQVHISACRQHETEALIPGIRVEF
CRINGLRGPTAANVYLS

S. typhimurium His-Tagged CSP amino acid sequences

His-CspA (Theoretical pI/Mw: 7.10/9566.6)
Ext. coefficient = 6990 M^{-1} cm^{-1}
MGSSHHHHHHSSGLVPRGSHMSGKMTGIVKWFNADKGFGFITPDDGSKDVF
VHFSAIQNDGYKSLDEGQKVSFTIESGAKGPAAGNVTSL

His-CspB (Theoretical pI/Mw: 7.11 / 9908.0)
Ext. coefficient = 5500 M^{-1} cm^{-1}
MGSSHHHHHHSSGLVPRGSHMTTKITGLVKWFNPEKGFGFITPKDGSKDVF
VHFSAIQSNEFRTLNENQEVEFSVEQGPKGPSAVNVVAL

His-CspC (Theoretical pI/Mw: 8.20 / 9565.6)
Ext. coefficient = 5500 M^{-1} cm^{-1}
MGSSHHHHHHSSGLVPRGSHMAKIKGQVKWFNESKGFGFITPADGSKDVFV
HFSAIQGNGFKTLAEGQNVEFEIQDGQKGPAAVNVTAI

His-CspD (Theoretical pI/Mw: 6.70 / 10048.1) **(1 cysteine)**
Ext. coefficient = 8480 M^{-1} cm^{-1}
Cleaved = Theoretical pI/Mw: 6.03 / 8166.13
MGSSHHHHHHSSGLVPRGSHMETGTVKWFNNAKGFGFICPEGGGEDIFAHY
STIQMDGYRTLKAGQSVRFDVHQGPKGNHASVIVPIEAEAVA

His-CspE (Theorectical pI/Mw: 9.23/9614.7)
Ext. coefficient = 5500 M^{-1} cm^{-1}
MGSSHHHHHHSSGLVPRGSHMSKIKGNVKWFNESKGFGFITPEDGSKDVFV
HFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL

His-CSPH (Theoretical pI/Mw: 9.81 / 9823.2) 3 x cysteines
Ext. coefficient = 1615 M^{-1} cm^{-1}
MGSSHHHHHHSSGLVPRGSHMSRKMTGIVKTFDCKSGKGLITPSDGRKDVQ
VHISACRQHETEALIPGIRVEFCRINGLRGPTAANVYLS

1.4 C-Terminal Double-Hexahistidine vector

The pC_Term_2xHis vector carries a C-terminal double-hexahistidine (His$_6$) tag sequence, comprising two hexahistidines separated by an 11-amino acid spacer. Unique sites are shown on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. Sequencing of the cloning region should be performed using standard T7 promoter (Novagen, Cat. No. 69348-3) and terminator (Novagen, Cat. No. 69337-3) primers.

pC_Term_2xHis cloning/expression region
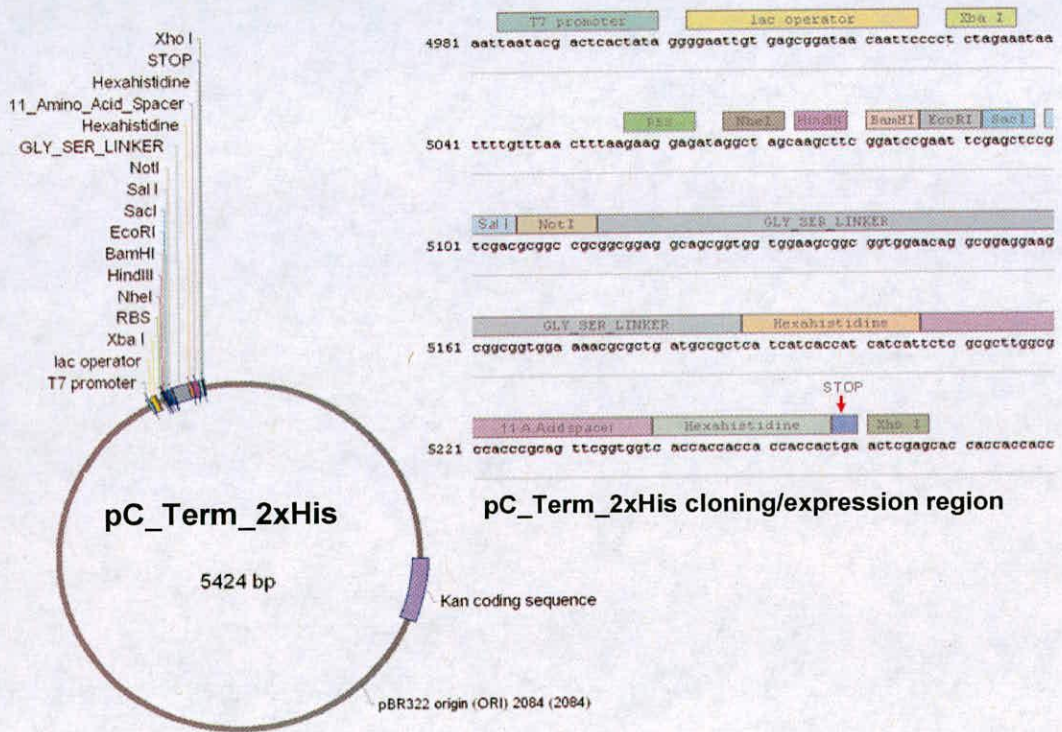
pC_Term_2xHis sequence landmarks	
T7 promoter	4984-5000
T7 transcription start	4983
His-Tag coding sequences	5192-5206 (1) 5240-5257 (2)
Multiple cloning sites	
Nhe I-Not I	5068-5111
T7 terminator	5352-5399
C-Terminal stop codon	5258-5260
rbs	5057-5063
Kan coding sequence	562-1374
F1 origin	11-466

The pC_Term_2xHis vector sequence (5424 bp)

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGG
GCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCC
ATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAA
ACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGC
AGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATCAACCTATTAATTTCCCCTCGT
CAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCC
AGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCT
GAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCG
CATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACC
ATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCAT
CTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTG
TCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAG
AGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT
CTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTT
TTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACC
GGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG
TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGT
ATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAG
CGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTG
ATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGC
TGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAG
GTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGC
CTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGT
TTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGAT
GCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCG
GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCC
TGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCAT
TCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACC
AGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGAT
AATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGC
AAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTAC
GAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAA
GGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTT
CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGG
TGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCA
CGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGGCTGTCTTCGGTATCGTCGT
ATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGG
CAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTT
CCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTA
ATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGG
CATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTGCG
ACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGG
GAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGG
AAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCT
CTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGAC
TCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCG
CCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCT
TCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAG
AGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCC**TCTAGA** AATA
ATTTTGTTTAACTTTAAGAAGGAGATAGGCTAGCAAGCTTCGGATCCGAATTCGAGCTCCGTCGACGCGGCCGC GGC GGA G
GC AGC GGT GGT GGA AGC GGC GGT GGA ACA GGC GGA GGA AGC GGC GGT GGA AAA CGC GCT GAT G
CC GCT CAT CAT CAC CAT CAT CAT TCT CGC GCT TGG CGC CAC CCG CAG TTC GGT GGT CAC CAC C
AC CAC CAC CAC TGA **ACTCGAG** C ACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGA
GTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGG
AGGAACTATATCCGGAT
```

The start and end of the inserted double his-tagged coding sequence are shown in
bold.

Single cut restriction sites for pC_Term_2xHis

AcuI
(CTGAAGNNNNNNNNNNNNNNNN_NN^)
1599/1597

AlwNI (CAG_NNN^CTG)
1732/1729

ApaI (G_GGCC^C)
*4039/4035

AsiSI (GCG_AT^CGC)
*945/943

BamHI (G^GATC_C)
5081/5085

BclI (T^GATC_A)
#4228/4232

BglI (GCCN_NNN^NGGC)
*3185/3182

BglII (A^GATC_T)
4964/4968

BlpI (GC^TNA_GC)
5341/5344

BmtI (G_CTAG^C)
5072/5068

BspDI (AT^CG_AT)
*1250/1252

BspQI (GCTCTTCN^NNN_)
2258/2261

BssHII (G^CGCG_C)
*3831/3835

BssSI (C^ACGA_G)
1968/1972

BstAPI (GCAN_NNN^NTGC)
4566/4563

BstZ17I (GTA|TAC)
2374/2374

ClaI (AT^CG_AT)
*1250/1252

DraIII
(CAC_NNN^GTG)
*245/242

EagI (C^GGCC_G)
*5107/5111

EcoRI (G^AATT_C)
*5087/5091

EcoRV (GAT|ATC)
3796/3796

FspI (TGC|GCA)
*3164/3164

HindIII (A^AGCT_T)
5074/5078

HpaI (GTT|AAC)
3740/3740

MluI (A^CGCG_T)
*4242/4246

NheI (G^CTAG_C)
5068/5072

NotI (GC^GGCC_GC)
*5107/5111

NruI (TCG|CGA)
*1286/1286

PaeR7I (C^TCGA_G)
*5262/5266

PciI (A^CATG_T)
2141/2145

PflFI (GACN^N_NGTC)
2399/2400

PpuMI (RG^GWC_CY)
3136/3139

PshAI
(GACNN|NNGTC)
*3401/3401

PsiI (TTA|TAA)
370/370

PspOMI (G^GGCC_C)
*4035/4039

PspXI (VC^TCGA_GB)
*5262/5266

PvuI (CG_AT^CG)
*945/943

SacI (G_AGCT^C)
5097/5093

SacII (CC_GC^GG)
*5113/5111

SalI (G^TCGA_C)
*5100/5104

SapI
(GCTCTTCN^NNN_)
2258/2261

SgrAI (CR^CCGG_YG)
*4923/4927

SmaI (CCC|GGG)
*1069/1069

SphI (G_CATG^C)
4775/4771

StyI (C^CWWG_G)
5363/5367

TliI (C^TCGA_G)
*5262/5266

TspMI (C^CCGG_G)
*1067/1071

Tth111I
(GACN^N_NGTC)
2399/2400

XbaI (T^CTAG_A)
5030/5034

XhoI (C^TCGA_G)
*5262/5266

XmaI (C^CCGG_G)
*1067/1071

The amino acid sequences for double-His$_6$-tagged CspE (a) and CspD (b).
(a) CspE_2xHis; MW = 12478 Daltons, pI = 9.99 (Expasy)
(b) CspD_2xHis; MW = 12912 Daltons, pI = 8.94 (Expasy)

a

```
            10        20        30        40        50        60        70
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CspE_2His  MSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTALG


            80        90        100       110       120
   ....|....|....|....|....|....|....|....|....|....|
CspE_2His  GRGGGSGGGSGGGTGGGSGGGKRADAAHHHHHHSRAWRHPQFGGHHHHHH
```

b

```
            10        20        30        40        50        60        70
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CspD_2His  METGTVKWFNNAKGFGFICPEGGGEDIFAHYSTIQMDGYRTLKAGQSVRFDVHQGPKGNHASVIVPIEAE


            80        90        100       110       120
   ....|....|....|....|....|....|....|....|....|....|
CspD_2His  AVAGGRGGGSGGGSGGGTGGGSGGGKRADAAHHHHHHSRAWRHPQFGGHHHHHH
```

1.5 CspE sequence alignment results from Consurf.

```
                    10        20        30        40        50        60        70
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
CSPE_ST     SHMSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVTAL
GRP2_NICSY  ----RAKGTVKWFSDKGFGFITPDDGGEDLFVRQSGIRSEGFRSLAEGETVEFEVESGGDGRTKA-----
CSPA_BORBR  ------GVVKWFNAEKGYGFITPEAGGKDLFAHFSEIQANGFKSLEENQRVSFVTAMGPKGPQATKIQIL
CSPD_ECO57  ------KGTVKWFNNAKGFGFICPEGGGEDIFAHYSTIQMDGYRTLKAGQSVQFDVHQGPKGNHAS-----
CSP_AQUAE   ------RGTVKWFSKDKGYGFITREDTNADVFVHFTDIQMEGFKTLQKGQKVEFDVVEDTKGPRAKNVRVL
CSP_ARTGO   ------QGTVKWFNAEKGFGFITPDDSDGDVFVHYSELQTGGFKTLDENARVQFEIGQGAKGPQATGVT--
CSPA_MYCBO  ------QGTVKWFNAEKGFGFIAPEDGSADVFVHYTEIQGTGFRTLEENQKVEFEIGHSPKGPQATGVRSL
CSPA_STIAU  -------GTVKWFNDAKGFGFITQDGGEDVFCHFSAINMDGFRTIQEGQKVEFEVTRGPKGLQAQNVRSL
CSPA_PSEAE  -------GTVKWFNDAKGFGFITPESGNDLFVHFRSIQGTGFKSLQEGQKVSFVVVNGQKGLQADEV---
CSPA_MICLU  -------GTVKWFNAEKGYGFIAPEDNSADVFVHFSAIQGNGFKELQENDRVEFSTQDGPKGLQAANVTKL
CSPE_BUCAI  --MSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQSNGFKTLAEGQSVEFEITEGAKGPSAANVISL
CSPE_BUCBP  --MSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQSNGFKTLSEGQSVEFEITEGAKGPSAANVISL
CSPE_ECO57  --MSKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAANVIAL
CSPC_BUCAI  --MAKIKGQVKWFNESKGFGFITPSDGSKDVFVHFSSIQGNGFKTLTEGQNVEFEIQDGQKGPAAVNVFSL
CSPC_SHIFL  --MAKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGLKTLAEGQNVEFEIQDGQKGPAAVNVTAI
CSPC_ECO57  --MAKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVEFEIQDGQKGPAAVNVTAI
CSPB_YEREN  ---NKMTGLVKWFDAGKGFGFISPADGSKDVFVHFSAIQGNDYKTLDEGQNVEFSIEQGQKGPSAVNVVAL
CSPA_ECO57  ----KMTGIVKWFNADKGFGFITPDDGSKDVFVHFSAIQNDGYKSLDEGQKVSFTIESGAKGPAAGNVTSL
CSPB_ECOLI  ---NKMTGLVKWFNADKGFGFISPVDGSKDVFVHFSAIQNDNYRTIPEGQKVEFSIESGAKGPAAANV---
CSPG_ECO57  ---NKMTGLVKWFNADKGFGFITPDDGSSKDVFVFTAIQSNEFRTLNENQKVEFSIEQGQRGPAAANVVTL
CSPI_ECOL6  ---NKMTGLVKWFNPEKGFGFITPKDGSKDVFVHFSAIQSNDFKTLTENQEVEFGIENGPKGPAAVHVVAL
CSPJ_SALTY  ---TKITGLVKWFNPEKGFGFITPKDGSKDVFVHFSAIQSNEFRTLNENQEVEFSVEQGPKGPSVNVVAL
CSPJ_SALTI  ---TKITGLVKWFNPEKGFGFITPKDGSKDVFVHFSAIQSNEFRTLNENQEVEFSAEQGPKGPSVNVVAL
CSP_THEMA   -----MRGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHV---
CSPA_SHEVI  -------GTVKWFNEDKGFGFLTQDNGGADVFVHFRAIASEGFKTLDEGQKVLFEVVEQGPKGLQASNVIAL
CSPG_SHEVI  -------GLVKWFNEEKGFGFITQDNGGDDVFVHFRGITSDGFKTLAEGQKVGFEVVEQGQKGLQAANVVAL
CSPV_VIBCH  ---TKMTGSVKWFNETKGFGFLTQDNGGNDVFVHFNSIQSEGFKTLAEGQRVSFIVEQGKKGFQASNVVAL
CSPA_VIBCH  ----KMTGSVKWFNETKGFGFISQDNGGQDVFVHFKSIVSEGFKTLAEGQRVSFTVEQGKKGPQAAQVTAL
CSP2_LACPL  -------GTVKWFNADKGFGFITGEDGT-DVFVHFSAIQTDGFKTLDEGQKVIYDEEQGDRGPQATNV---
CSP1_LACPL  -------GTVKWFNADKGYGFITGEDGN-DVFVHFSAIQTDGFKTLEEGQKVEFDEESSDRGPQAANV---
CSPA_BACCE  -----VTGQVKWFNNEKGFGFIEVPGEN-DVFVHFSAIETDGFKSLEEGQKVSFEIEDGNRGPQAKNVIKL
CSPA_BACAN  -----VTGQVKWFNNEKGFGFIEVPGEN-DVFVHFSAIETDGFKSLEEGQKVSFEIEEGNRGPQAKNVIKL
CSPB_BACCE  -----MNGKVKWFNNEKGFGFIEMGSE-DVFVHFSAIQSDGYKALEEGQEVSFDITEGNRGPQAANVAKL
CSPE_BACCR  -----LTGKVKWFNSEKGFGFIEVADGN-DVFVHFSAITGDGFKSLDEGQEVSFEVEEGNRGPQAKNVVKL
CSPE_BACAN  -----LTGKVKWFNSEKGFGFIEVEGGN-DVFVHFSAITGDGFKSLDEGQEVSFEVEDGNRGPQAKNVVKL
CSPB_LISIN  -------GTVKWFNSEKGFGFIEVEGGD-DVFVHFSAIEGEGFKTLDEGQSVEFEIVEGQRGPQAEKVTKL
CSPD_BACCE  -------GKVKWFNGEKGFGFIEVEGGE-DVFVHFSAIQGDGFKTLEEGQEVSFEIVDGNRGPQAANVT--
CSPB_BACSU  -----LEGKVKWFNSEKGFGFIEVEGQD-DVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVT--
CSPD_BACAN  -------GKVKWFNSEKGFGFIEVEGGD-DVFVHFSAIQGDGFKTLEEGQEVSFEIVEGNRGPQAANVT--
CSPD_BACSU  -------GKVKWFNNEKGFGFIEVEGGD-DVFVHFTAIEGDGYKSLEEGQSVEFEIVEGNRGPQASNVVKL
CSPB_BACCL  ------REKVKWFNNEKGYGFIEVEGGS-DVFVHFTAIQGEGFKTLEEGQEVSFEIVQSNRPQAANVVKL
CSPB_BACST  ------RGKVKWFNNEKGYGFIEVEGGS-DVFVHFTAIQGEGFKSLEEGQEVSFEIVVGNRGPQAANVVKL
CSPA_STRP1  ------QGTVKWFNAEKGFGFISTENGQ-DVFAHFSAIQTNGFKTLEEGQKVEFDVEEGQRGPQAVNITKL
CSPA_STRP6  ------QGTVKWFNAEKGFGFISTENGQ-DVFAHFSAIQTSGFKTLEEGQKVAFDVEEGQRGPQAVNITKL
CSPC_BACSU  ------QGTVKWFNAEKGFGFIERENQD-DVFVHFSAIQSDGFKSLDEGQKVSFDVEQGARGAQAANV---
CSPA_LISIN  ------QGTVKWFNAEKGFGFIERSNGD-DVFVHFSAIQGDGFKSLDEGQAVEFDVEEGQRGPQAANV---
CSPC_BACAN  -----MQGRVKWFNAEKGFGFIEREDED-DVFVHFSAIQQDGYKSLEEGQQVEFDIVDGARGPQAANVVKL
CSPA_STAEQ  ------QGTVKWFNAEKGFGFIEVEGEN-DVFVHFSAINQEGYKSLEEGQSVEFEVVEGDRGPQAANVVKL
CSPA_STAA3  ------QGTVKWFNAEKGFGFIEVEGEN-DVFVHFSAINQDGYKSLEEGQAVEFEVVEGDRGPQAANVVKL
CSPA_STAHJ  ------QGTVKWFNAEKGFGFIEVEGEN-DVFVHFSAINQDGYKSLEEGQSVEFEVVEGDRGPQAANVVKL
```

1.6 Masspectrometry results for Cy5-labeled and unlabelled *Bs*CspB.

Figure A1.6. **Cy5 covalent labelling of *Bs*CspB**

(a) MALDI-Tof mass spectrum of BsCspB pre labelling with Cy5, observed 7377 Daltons, calculated 7365 Daltons. (b) Post labelling with Cy5, observed 8015, calculated 8016 Daltons (Protein + Cy5 = 7365 + 650 = 8015).

1.7 The sequence specificity of binding of the mammalian Y-Box binding protein, analysed by competitive binding to the oligonucleotide chip.

1.7.1 Overview

The binding of the P50 protein (*Oryctolagus cuniculus*) to hexadeoxynucleotides was previously analysed [189] using an oligonucleotide chip similar in structure to the chip used here. The purpose of this section was to analyse the binding of the P50 protein using the competitive microarray assay. A comparison of both sets of results will serve as a control and help validate previous experiments.

1.7.2 Introduction

The rabbit P50 protein contains an Oligonucleotide/oligosaccharide binding fold (OB-fold), which is a known characteristic of many DNA/RNA binding proteins found in many species[35], including the family of cold shock proteins found in bacteria. Both bacterial cold shock proteins (CSPs) and eukaryotic cold shock domain (CSD) containing proteins have an affinity for both RNA and DNA and contain the RNP-1 (KGFGFI) and RNP-2 (VFVH) signatures of RNA binding proteins (Figure A5-1). The rabbit P50 protein and its counterparts from various mammalian sources share about 96% sequence identity (Figure A5-1 & Table AI). They contain a CSD (~amino acids 58-129) and two functional regions [52, 206-208]. The calculated (ExPASY) pI/Mw for P50 is 9.82 / 35824.07 daltons.

P50 and related Y-Box proteins are structural mRNP proteins associated with almost all mRNAs [50, 51]. Once bound to mRNA, these proteins facilitate secondary structural changes which contribute mRNA transport, translation, processing and

stabilisation [52-54]. P50 is therefore believed to bind relatively non-specifically [50] to RNA/DNA.

Y-box binding proteins have been show to interact with DNA and affect transcription of many genes. They also bind specifically to many sequences including the Y-box. Examination of P50 binding using the microarray assay will provide clues into its participation in gene regulatory networks and validate the microarray assay used.

Table AI. Similarities in the amino acid sequence of the P50 protein and other mammalian proteins.

Gene	% Identity	% Similarity	Organism	Accession	Reference
P50	-	-	*Oryctolagus cuniculus*	AAA66069	[50]
dbpB	98	99	*Homo sapiens*	AAA35750	[209]
YB-1	98	99	*Homo sapiens*	P67809	[210]
YB-1	96	97	*Rattus norvegicus*	BAC99313	[211]
EF1A	96	97	*Rattus norvegicus*	AAA41108	[212]
TSEP-1	96	97	*Rattus norvegicus*	AAB46889	[213]
rcNSEP1	94	95	*Homo sapiens*	ABB92460	[214]
mYB-1a	93	95	*Mus musculus*	AAA75475	DS

```
                    10        20        30        40        50        60        70        80
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P50      --------------------------------------MSSEAETQQPPAAPPAAAL SAADTKPGTTGSGAGSGGPG
dbpB     EFGGPQRALSSPTAAAGLVTITPREEPQLPQPAPVTITATMSSEAETQQPPAAPPAAAL SAADTKPGTTGSGAGSGGPG
YB-1 h.sapien  ----------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGTTGSGAGSGGPG
YB-1 r.norvegious ------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGSTGSGAGSGGPG
EF1A     ----------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGSTGSGAGSGGPG
TSEP-1   ----------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGTTGSGAGSGGPG
rcNSEP1  ----------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGTTGSGAGSGGPG
mYB-1a   ----------------------------------------MSSEAETQQPPAAP--AAAL SAADTKPGSTGSGAGSGGPG

                    90       100       110       120       130       140       150       160
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P50      GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
dbpB     GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFIHRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
YB-1 h.sapien  GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
YB-1 r.norvegious GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
EF1A     GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
TSEP-1   GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
rcNSEP1  GLTSAXPAGGIKKVIATKVLGTVKWFNVRNGYGFIHRNDTKEDVFVHXTAXKKHNPRKYLRSVGDGEXVEFDVVEGEKGA
mYB-1a   GLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGA
                                                   RNP-1              RNP-2

                   170       180       190       200       210       220       230       240
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P50      EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQAQQRRPYRRRRFPPYYMRR
dbpB     EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKHEGSESAPEGQAQQRRPYRRRRFPPYYMRR
YB-1 h.sapien  EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQAQQRRPYRRRRFPPYYMRR
YB-1 r.norvegious EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQAQQRRPYRRRRFPPYYMRR
EF1A     EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQAQQRRPYRRRRFPPYYMRR
TSEP-1   EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQAQQRRPGPYRSRRFPPYYMRR
rcNSEP1  EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKHEGSESAPEGQAQQRRPYRRRRFPPYYMRR
mYB-1a   EAANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNESSAPEGQAQQRRPYRRRRFPPYYMRR

                   250       260       270       280       290       300       310       320
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P50      PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDGWEEDKENQGDETQGQQPPPSP
dbpB     PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDGWEEDKENQGDETQGQQPPQRR
YB-1 h.sapien  PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDGWEEDKENQGDETQGQQPPQRR
YB-1 r.norvegious PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDGWEEDKENQGDETQGQQPPQRR
EF1A     PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDGWEEDKENQGDETQGQQPPQRR
TSEP-1   PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPLQRQPREDGWEEDKENQGDETQGQQPPQRR
rcNSEP1  PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDDWEEDKENQGDETQGQQPPQRR
mYB-1a   PYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVRQNHYRGYRPRFRRGPPRQRQPREDDWEEDKENQGDETQGQQPPQRR

                   330       340       350       360
         ....|....|....|....|....|....|....|....|
P50      YRRNFNYRRRRPDNPKPQDGKETKAADPPAEHSSAPEAEQGGAE
dbpB     YRRNFNYRRRRPENPKPQDGKETKAADPPAEHSSAPEAEQGGAE
YB-1 h.sapien  YRRNFNYRRRRPENPKPQDGKETKAADPPAEHSSAPEAEQGGAE
YB-1 r.norvegious YRRNFNYRRRRPENPKPQDGKETKAADPPAEHSSAPEAEQGGAE
EF1A     YRRNFNYRRRRPENPKPQDGKETKAADSPAEHSSAPEAEQGGAE
TSEP-1   YRRNFNYRRRRPENPKPQDGKETKAADPPAEHSSAPEAEQGGAE
rcNSEP1  YRRNFNYRRRRPENPKPQDGKETKAAPPAEHSSAPEAEQGGAE
mYB-1a   YRRNFNYRRRRPENPKSKDGKETKATHPPTEHSSAPEAEQGGAE
```

Figure A5-1. **Multiple sequence alignment** of the amino acid sequences of P50 with related mammalian proteins. Residues that are identical in all sequences are within grey boxes.

1.7.3 Cloning of His-p50

The full coding cDNA (1.5-kb) of p50 was previously isolated from rabbit reticulocytes and inserted into *NdeI-XhoI* sites of pET-3-1 to create pET-3-1-p50[50]. pET-3-1-p50 was digested with *NdeI* and *XhoI* and yielded two fragments approximately 4-kb and 1.5-kb. The 1500 base pair (bp) restriction fragment was

purified from the agarose gel and was ligated into a *NdeI-XhoI*-digested pET28a vector, resulting in the plasmid pET28a_p50. pET28a_p50 was then transformed into *E. coli* DH5-α. Positive transformants were isolated on L.B agar plates containing kanamycin (50μg/ml). Plasmid DNA was isolated from several transformants and restriction analysis was carried out to confirm the presence of the *p50* gene. Gel electrophoresis showing a restriction digest (*NdeI* and *XhoI*) of pET28a_p50 plasmid DNA isolated from transformants can be seen in Figure A5-2. Positive clones were carried forward for expression of the p50 protein.

Figure A5-2. **Agarose gel electrophoresis** of *NdeI* and *XhoI* digested and undigested plasmid DNA isolated from *E. coli* DH5-α carrying the plasmid pET28a_p50. Lanes: (1, 3, 5 & 7) plasmid DNA isolated from *E. coli* DH5-α; (2, 4, 6 & 8) Plasmid DNA digested with *NdeI* and *XhoI*. The resulting 5-kb and 1.5-kb fragments indicates that the *p50* gene is present. Scale is in <u>K</u>ilo<u>B</u>ases.

1.7.4   Small Scale expression trails of His₆-p50

BL21 (DE3)pLysS *E. coli* was transformed with pET28a_p50 and expression of the P50 gene was induced by adding IPTG (1 mM). Cell where grown at 37°C for 4 hrs or 20°C overnight and expression of P50 was analysed by SDS-PAGE (Figure A5-3).
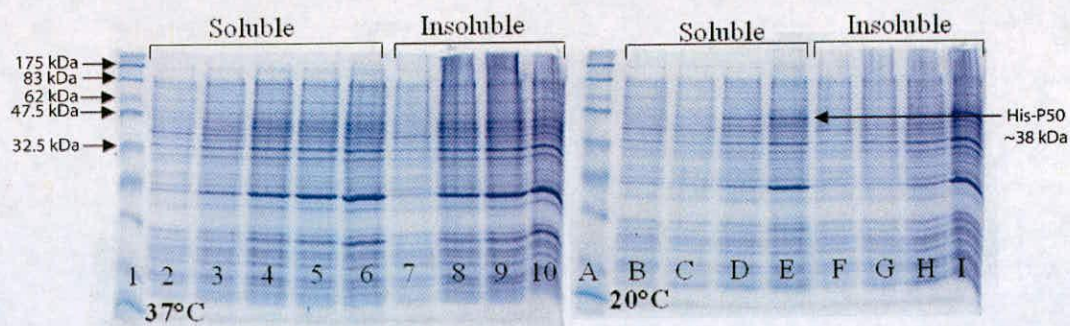
Figure A5-3. **Expression of P50 at 37°C and 20°C.**
Samples were analysed by 15% SDS-PAGE and visualised by Coomassie staining. Cell extracts from BL21 (DE3)pLysS harbouring pET28_P50. Lanes 2, 7, B and F have not been induced. Lanes 3, 4, 5 & 6 show soluble cell extract after induction at 37°C for 1, 2,3 and 4 hrs, respectively and lanes 8, 9 and 10 show insoluble cell extracts after induction, for 2,3 and 4 hrs. Lanes C, D, and E show soluble cell extract after induction at 20°C for 1, 4 and 16 hrs. Lanes G, H and I show insoluble extracts after induction, for 1, 4 and 16 hrs. Lanes 2, 7, B and F have not been induced.

The band corresponding to the P50 protein can be seen in the soluble cell extract when the protein was expressed at 20°C for 4-16 hrs. By SDS-PAGE analysis P50 has an apparent molecular mass of 50 KDa (calculated 37987.39 Da), this anomalous electrophoretic mobility is due to the proteins high isoelectric point of 9.5. The resulting expression conditions for the production of P50 were kept constant throughout.

1.7.5 Purification of His-P50

BL21 (DE3)pLysS *E. coli* was transformed with the verified constructs and expressed by inoculating Luria-Bertani medium, containing 50 µg/ml Kanamycin, at 37°C with agitation. One-litre cultures were grown to an OD_{600} of 0.5-0.7, growth was the stopped by incubating on ice for 30mins. IPTG (isopropyl-β-D-thiogalactosidase) was then added to a final concentration of 1 mM and incubation was then continued overnight at 20°C. The cells were harvested at 8,000 rpm in a

231

JLA-9.1000 rotor for 12 minutes at 10°C. Pellets were then frozen in liquid nitrogen and stored at -80°C.

Cell pellets were resuspended in buffer A (20 mM Tris HCl pH 7.6, 0.025% Thesit, 0.1 mM PMSF, a spatula tip of lysozyme, DNaseI [40 µg/ml] and fresh protease inhibitors) to a final volume of 30 ml per tube. Pellets were incubated at 4°C for 1hr with gentle agitation. The fusion protein was purified using a talon resin affinity column (Figure A5-4), as per manufacturer's instructions (Qiagen). Fractions containing the recombinant protein were combined and concentrated until they reached a final volume of 2.5 ml. The protein sample was exchanged into buffer B (20 mM Tris-HCl pH 7.6, 0.025% Thesit and 0.1 mM PMSF) using a PD-10 column. In order to remove any contaminating DNA, the sample was applied onto a heparin-Sepharose column (Amersham), equilibrated with buffer B. The column was eluted with 5 volumes of buffer B containing 500 mM NaCl (Figure A5-4); 2 ml fractions were collected during subsequent elution with buffer B containing 1.5 M NaCl (Figure A5-4). The protein fractions were concentrated, buffer exchanged (PD-10) into buffer B and applied onto a superdex 200 column equilibrated with the buffer C (10 mM HEPES-KOH (pH 7.6), 250 mM KCl, 0.5 mM DTT and 0.025% Thesit). Fractions containing the largest amounts of protein in question were combined. The protein was concentrated using a Viva Spin column (MWCO = 10KDa), aliquots were then stored at -80°C. Purity was determined by SDS gel electrophoresis (Figure A5-4). The molecular mass of the protein was confirmed by MALDI-Tof analysis (Figure A5-4). Protein concentration was determined using the Bradford method.
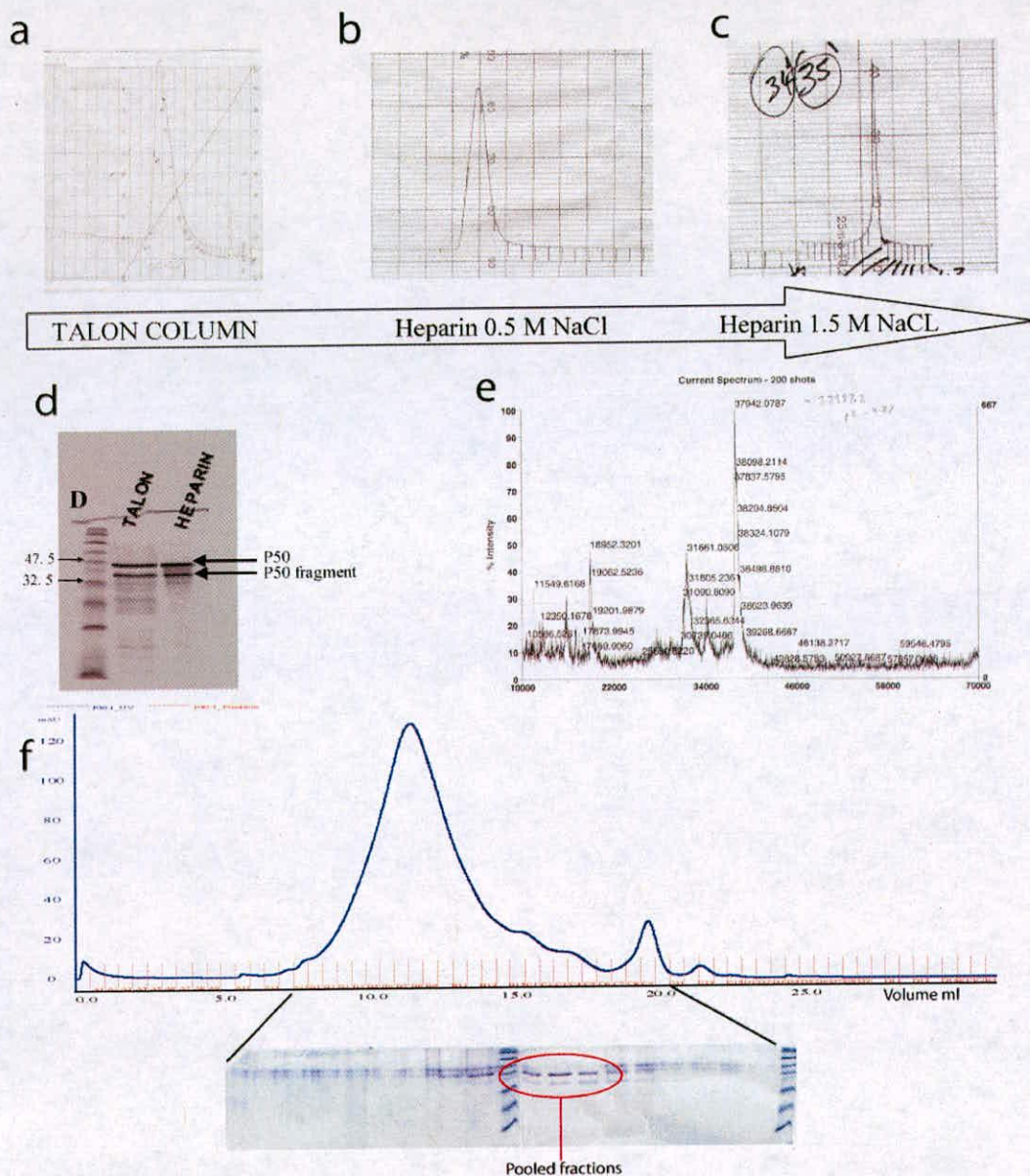
Figure A5-4. **Purification of the P50 protein.**

(A) Elution profile of P50 from Talon column.

(B) Fractions containing His-P50 from talon purification were concentrated and bound to a heparin column to remove DNA. Elution profile from a heparin column when washed with 0.5 M NaCl, DNA corresponds to the absorbance peak seen.

(C) Elution from heparin column when washed with 1.5 M NaCl. Protein was eluted, desalted and buffer exchanged into 20 mM Tris-HCl pH 7.6.

(D) SDS-PAGE analysis of His-P50 purification after each column.

(E) Mass spectrum of His-P50 at final stage of purification (37942 Da)

(F) Gel filtration of analysis of P50 on a Superdex 200 column, fractions were pooled, dialysed into a buffer containing 10 mM HEPES-KOH (pH 7.6), 250 mM KCl, 0.5 mM DTT, concentrated to 2 mg/ml and stored at -80°C.[190]

1.7.6   *His-P50 binds ss DNA.*

An electrophoretic mobility shift assay was conducted to ascertain if the P50 protein
had retained its DNA binding activity subsequent to the addition of a His-tag. A ss
25-mer (P50-ONBSeq-5'–AAAAAAAAAA**TGGGGG**AAAAAAAAA-'3), labelled
at the 5'-end with γ-^{32}P was used to assess if His$_6$-P50 was still capable of binding to
its previously identified ss DNA binding site [189], 5'-TGGGGG-3'. Figure A5-5
shows a gel-shift experiment performed for P50-ONBSeq in the presence of
decreasing amounts of His-P50.

The gel-shift results for the His-P50 protein suggest that the addition of a His-tag
does not significantly affect the binding affinity or specificity of the protein as the
relative amounts of His-P50 used to shift the DNA correspond to the amounts of
untagged-P50 used for the same experiment [189].
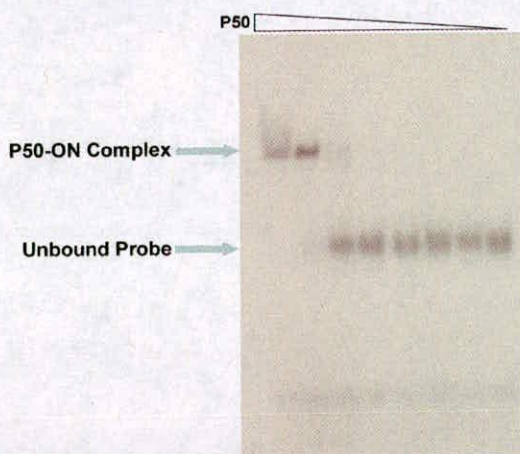
**Figure A5-5. Electrophoretic mobility shift assay of P50 and its complexes with
P50-ONBSeq.**
A fixed concentration (80 pI) of 5'-end γ-^{32}P-labelled oligonucleotide was incubated
with decreasing amounts of P50 (Concentrations of P50: 650 pM, 130 pM, 90 pM,
50 pM, 10 pM, 5 pM, 1 pM, 0 pM).
All 20 μL reactions mixtures, containing 10 mM HEPES-KOH (pH 7.6), 100 mM
KCl, 0.5 mM DTT, 1 mg/ml BSA, were incubated for 20 min at 4°C. Protein-DNA
complexes were separated on a 6% native polyacrylamide gel at 4°C.
(P50-ONBSeq-5'–AAAAAAAAAA**TGGGGG**AAAAAAAAA-'3)

1.7.7 Competitive binding of the P50 protein, monitored by EMSA

A competitive electrophoretic mobility shift assay was used to examine if the His-P50 can successfully out compete the *Sso*SSB protein for its consensus binding motif (TGGGG). The 25-mer, P50-ONBSeq, which contains the consensus binding motif, was labelled at the 5'-end with γ-^{32}P. This probe was used to assess if the His-P50 protein can compete with the *Sso*SSB protein. Figure A5-6 shows a gel-shift experiment performed with P50-ONBSeq. The *Sso*SSB protein was gradually added to a fixed amount of the His-P50 protein. The *Sso*SSB protein binds non-specifically to ss DNA with a binding density of 5 nt per monomer and therefore the shift in lane one (Figure A5-6) corresponds to five monomers of *Sso*SSB.
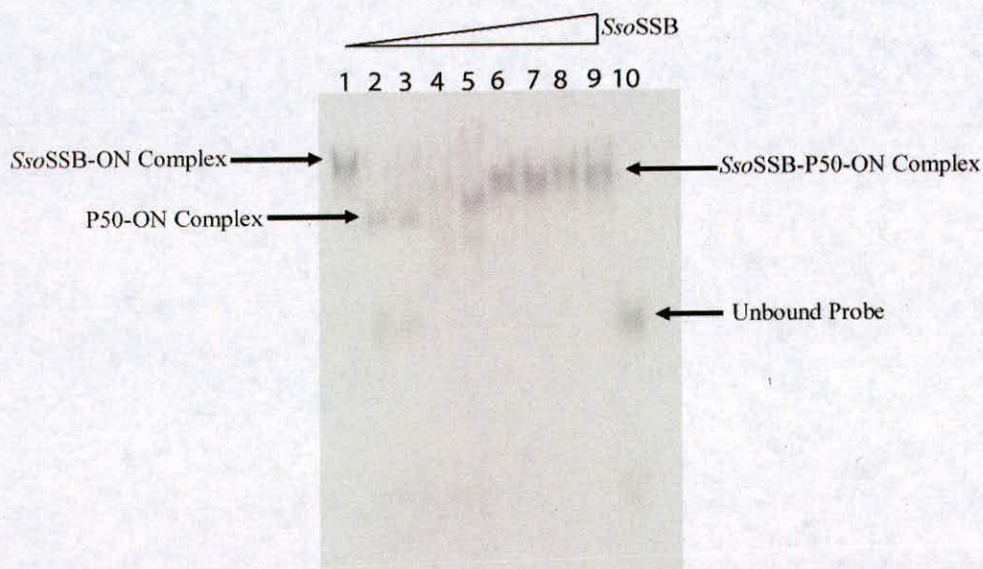
Figure A5-6. Electrophoretic mobility shift assay of *Sso*SSB, P50 and the complexes formed with ON-P50BSeq.
A fixed concentration (80 pM) of 5'-end γ-^{32}P-labelled ON-P50BSeq was used at all times.
Lanes; (1) 12 μM *Sso*SSB; (2) 1.3 μM P50; (3) 31 pM *Sso*SSB & 1.3 μM P50; (4) 310 pM *Sso*SSB & 1.3 μM P50; (5) 620 pM *Sso*SSB & 1.3 μM P50; (6) 2.48 μM *Sso*SSB & 1.3 μM P50; (7) 3 μM *Sso*SSB & 1.3 μM P50; (8) 4.5 μM *Sso*SSB & 1.3 μM P50; (9) 6 μM *Sso*SSB & 1.3 μM P50; (10) Free ONc probe.
All 20 μL reaction mixtures were incubated for 20 min at 4°C. Protein-DNA complexes were separated on a 10% native polyacrylamide gel at 4°C.

The P50 protein is known to bind specifically to ss DNA containing the core region, TGGGGG. The shift in lane 2 (Figure A5-6) most likely corresponds to a total of one monomer of His-P50. An intermediate band can be seen in lanes 6-9, which is shown as the SsoSSB-P50- ON complex. Successful formation of the P50-SsoSSB-ONBSeq complex occurs when the proteins are in a molar ratio of roughly 2 SsoSSB:1 P50 (Lane 6) and continues to be seen until lane 9. The appearance of this intermediate band is indicative that the P50 protein can compete with the SsoSSB protein for its high affinity binding site. These defined conditions (pH, NaCl and molar ratio of proteins) under which the formation of this third complex can be seen will allow for the transfer to a microarray format assay.

1.7.8   Competitive microarray assay

A 2:1 molar excess mixture (as determined by gel-shift, Figure A5-6) of SsoSSB and His-P50 proteins were bound to the oligonucleotide chip at 4°C for one hour. The chip was then washed and protein-bound microarrays were labelled with Alexa 532-conjugated antibody to His_5 (Molecular Probes) and scanned with an Axon 4200 AL microarray scanner. The microarray TIF images were quantified using GenPix Pro version 6 software. All experiments were performed in duplicate in order to validate results (Figure A5-7). High intensity fluorescence spots on the array are generally indicative of higher DNA-protein binding affinity.[169]

5′-CGGGGA-3′
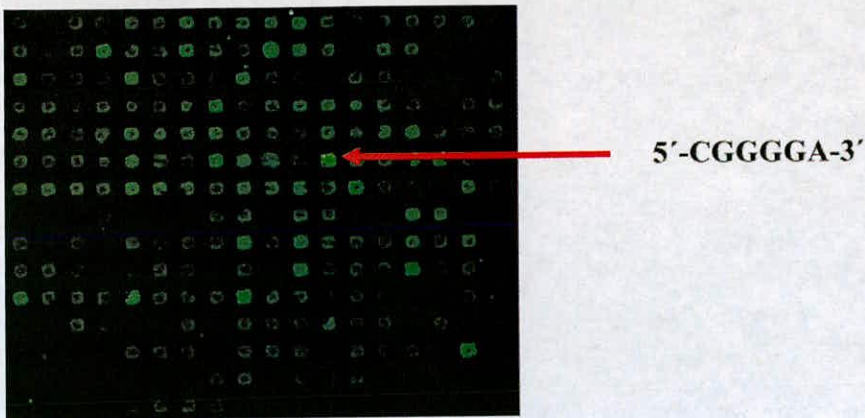
Figure A5-7. **Competitive microarray assay.**
Array results for the competitive protein binding assay between His-P50 (specific binding protein) and *Sso*SSB (non-specific binding protein).
An enlargement of sub-array four from each of the two competitive protein binding experiments. The oligonucleotide, CGGGGA, appears as a bright spot.
100 μL of protein binding reaction (0 mM HEPES, 100 mM KCl, pH 7.6, 2% non-fat dried milk and proteins were added to a final concentration of *Sso*SSB [12 μM], P50 [6 μM].

1.7.9   The specificity of P50 binding to the oligonucleotide chip

High intensity spots which occurred on both arrays were recorded and the top fifty high-affinity binding sequences were aligned using ClustalW (Figure A5-8a). The overall length of the sequence alignment is nine nucleotides in length but only seven nucleotides (positions 2 to 8) are occupied by a meaningful number of bases, which we have used as an indicator of the maximum size of the nucleotide binding site for P50. The frequency of each base for positions one to seven, in the proposed nucleotide binding site, was calculated (Figure A5-8b) and used to calculate the relative base distribution (Figure A5-8c and d). The results indicate that there is a preference for the sequence, **5′-ACGGGGT-3′.**

Absolute Base Distribution

Position	1	2	3	4	5	6	7
Base							
A	17	3	10	5	2	7	6
C	4	29	5	6	11	11	2
G	2	7	24	25	22	25	8
T	2	2	11	14	15	2	9
Total	25	41	50	50	50	45	25

Relative Base Distribution

Position	1	2	3	4	5	6	7
Base							
A	68%	7%	20%	10%	4%	16%	24%
C	16%	71%	10%	12%	22%	24%	8%
G	8%	17%	48%	50%	44%	56%	32%
T	8%	5%	22%	28%	30%	4%	36%

Figure A5-8. **(a)** Multiple sequence alignment of the top fifty 6-mer binding sequences for His-P50. The overall length of the sequence alignment is numbered 1-9 and positions 2-8 have been coloured and number 1-7 in italics, as they are occupied by a meaningful number of bases (>20). **(b)** The absolute base distribution for P50 that was obtained from the alignment of all 50 sequences shown in figure 9B. **(c)** The relative base distribution for P50 was obtained from absolute base distribution. **(d)** A 3-D bar chat showing the relative base distribution for each base within the proposed heptanucleotide binding site.

238

1.7.10 Conclusions

The progress of this work was limited by the extremely adhesive nature of the P50 protein, which prevented any biophysical characterisation of the P50/ss DNA interaction using SPR and ITC.

1.8 The P50 amino acid sequence (His tag region shown in red)

Theoretical pI/Mw for His-P50: 9.87 / **37987.39** (ExPASY)
Theoretical pI/Mw for P50 (untagged): 9.82 / **35824.07**
MGSSHHHHHHSSGLVPRGSHMSSEAETQQPPAAPPAAPALSAAETKPGTTGS
GAGSGGPGGLTSAAPAGGDKKVIATKVLGTVKWFNVRNGYGFINRNDTKE
DVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKGAEAANVTGPGGVP
VQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSESGEKNEGSESAPEGQA
QQRRPYRRRRFPPYYMRRPYGRRPQYSNPPVQGEVMEGADNQGAGEQGRP
VRQNMYRGYRPRFRRGPPRQRQPREDGNEEDKENQGDETQGQQPPPSRYRR
NFNYRRRRPDNPKPQDGKETKAADPPAENSSAPEAEQGGAE

Reference List

1. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.

2. Chase, J.W. and K.R. Williams, *Single-stranded DNA binding proteins required for DNA replication.* Annu Rev Biochem, 1986. **55**: p. 103-36.

3. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annu Rev Biochem, 1997. **66**: p. 61-92.

4. Umezu, K., N.W. Chi, and R.D. Kolodner, *Biochemical interaction of the Escherichia coli RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein.* Proc Natl Acad Sci U S A, 1993. **90**(9): p. 3875-9.

5. Kornberg, A., *DNA replication.* J Biol Chem, 1988. **263**(1): p. 1-4.

6. Meyer, R.R. and P.S. Laine, *The single-stranded DNA-binding protein of Escherichia coli.* Microbiol Rev, 1990. **54**(4): p. 342-80.

7. Matsumoto, K. and A.P. Wolffe, *Gene regulation by Y-box proteins: coupling control of transcription and translation.* Trends Cell Biol, 1998. **8**(8): p. 318-23.

8. Kowalczykowski, S.C., et al., *Biochemistry of homologous recombination in Escherichia coli.* Microbiology and Molecular Biology Reviews, 1994. **58**(3): p. 401-465.

9. Coverley, D., et al., *A role for the human single-stranded DNA binding protein HSSB/RPA in an early stage of nucleotide excision repair.* Nucleic Acids Res, 1992. **20**(15): p. 3873-80.

10. Konkel, L.M., et al., *A class of single-stranded telomeric DNA-binding proteins required for Rap1p localization in yeast nuclei.* Proc Natl Acad Sci U S A, 1995. **92**(12): p. 5558-62.

11. Thieringer, H.A., P.G. Jones, and M. Inouye, *Cold shock and adaptation.* BioEssays, 1998. **20**(1): p. 49-57.

12. Antson, A.A., et al., *Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA.* Nature, 1999. **401**(6750): p. 235-42.

13. Messias, A.C. and M. Sattler, *Structural basis of single-stranded RNA recognition.* Acc. Chem. Res, 2004. **37**(5): p. 279-287.

14. Max, K.E., et al., *T-rich DNA Single Strands Bind to a Preformed Site on the Bacterial Cold Shock Protein Bs-CspB.* J Mol Biol, 2006. **360**((3)): p. 702-14.

15. Changela, A., R.J. DiGate, and A. Mondragon, *Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule.* Nature, 2001. **411**(6841): p. 1077-81.

16. Tucker, P.A., et al., *Crystal structure of the adenovirus DNA binding protein reveals a hook-on model for cooperative DNA binding*. Embo J, 1994. **13**(13): p. 2994-3002.

17. Morellet, N., et al., *Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC)*. J Mol Biol, 1998. **283**(2): p. 419-34.

18. Myers, J.C. and Y. Shamoo, *Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM)*. J Mol Biol, 2004. **342**(3): p. 743-56.

19. Braddock, D.T., et al., *Structure and dynamics of KH domains from FBP bound to single-stranded DNA*. Nature, 2002. **415**(6875): p. 1051-6.

20. Bustamante, C., Z. Bryant, and S.B. Smith, *Ten years of tension: single-molecule DNA mechanics*. Nature, 2003. **421**(6921): p. 423-7.

21. Schnuchel, A., et al., *Structure in solution of the major cold-shock protein from Bacillus subtilis*. Nature, 1993. **364**(6433): p. 169-71.

22. Raghunathan, S., et al., *Structure of the DNA binding domain of E. coli SSB bound to ssDNA*. Nature Structural Biology, 2000. **7**: p. 648-652.

23. Matsumoto, T., et al., *Roles of functional loops and the C-terminal segment of a single-stranded DNA binding protein elucidated by X-ray structure analysis*. J Biochem, 2000. **127**: p. 329-335.

24. Bochkarev, A., et al., *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA*. Nature, 1997. **385**(6612): p. 176-181.

25. Max, K.E., et al., *Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from Bacillus caldolyticus*. Febs J, 2007. **274**(5): p. 1265-79.

26. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, DeLano Scientific, San Carlos, CA.

27. Deo, R.C., et al., *Recognition of polyadenylate RNA by the poly(A)-binding protein*. Cell, 1999. **98**(6): p. 835-45.

28. Lopez, M.M. and G.I. Makhatadze, *Major cold shock proteins, CspA from Escherichia coli and CspB from Bacillus subtilis, interact differently with single-stranded DNA templates*. Biochim Biophys Acta, 2000. **1479**(1-2): p. 196-202.

29. Jiang, W., Y. Hou, and M. Inouye, *CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone*. J Biol Chem, 1997. **272**(1): p. 196-202.

30. Ghosh, A. and M. Bansal, *A glossary of DNA structures from A to Z*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 4): p. 620-6.

31. Bustamante, C., et al., *Single-molecule studies of DNA mechanics*. Curr Opin Struct Biol, 2000. **10**(3): p. 279-85.

32.   Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. **95**(7): p. 963-74.

33.   Phadtare, S. and M. Inouye, *Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of Escherichia coli.* Mol Microbiol, 1999. **33**(5): p. 1004-14.

34.   Brandi, A., et al., *Post-transcriptional regulation of CspA expression in Escherichia coli.* Mol Microbiol, 1996. **19**(2): p. 231-40.

35.   Theobald, D.L., R.M. Mitton-Fry, and D.S. Wuttke, *Nucleic acid recognition by OB-fold proteins.* Annu Rev Biophys Biomol Struct, 2003. **32**: p. 115-33.

36.   Murzin, A.G., et al., *SCOP: a structural classification of proteins database for the investigation of sequences and structures.* J Mol Biol, 1995. **247**(4): p. 536-40.

37.   Su, S., et al., *Analyses of the stability and function of three surface mutants (R82C, K69H, and L32R) of the gene V protein from Ff phage by X-ray crystallography.* Protein Sci, 1997. **6**(4): p. 771-80.

38.   Mitton-Fry, R.M., et al., *Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13.* J Mol Biol, 2004. **338**(2): p. 241-55.

39.   Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.

40.   Shamoo, Y., et al., *Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA.* Nature, 1995. **376**(6538): p. 362-6.

41.   Hollis, T., et al., *Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9557-62.

42.   Bochkareva, E., et al., *Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA.* Embo J, 2002. **21**(7): p. 1855-63.

43.   Kerr, I.D., et al., *Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB protein.* Embo J, 2003. **22**(11): p. 2561-70.

44.   Venclovas, C., K. Ginalski, and C. Kang, *Sequence-structure mapping errors in the PDB: OB-fold domains.* Protein Sci, 2004. **13**(6): p. 1594-602.

45.   Bernstein, D.A., et al., *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, 2004. **101**(23): p. 8575-80.

46.   Saikrishnan, K., et al., *Structure of Mycobacterium tuberculosis single-stranded DNA-binding protein. Variability in quaternary structure and its implications.* J Mol Biol, 2003. **331**(2): p. 385-93.

47. Huang, C.Y., et al., *Complexed crystal structure of replication restart primosome protein PriB reveals a novel single-stranded DNA-binding mode.* Nucleic Acids Res, 2006. **34**(14): p. 3878-86.

48. Folmer, R.H., et al., *Solution structure of the single-stranded DNA binding protein of the filamentous Pseudomonas phage Pf3: similarity to other proteins binding to single-stranded nucleic acids.* Embo J, 1995. **14**(17): p. 4132-42.

49. Lopez, M.M., K. Yutani, and G.I. Makhatadze, *Interactions of the cold shock protein CspB from Bacillus subtilis with single-stranded DNA. Importance of the T base content and position within the template.* J Biol Chem, 2001. **276**(18): p. 15511-8.

50. Evdokimova, V.M., et al., *The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family.* J Biol Chem, 1995. **270**(7): p. 3186-92.

51. Sommerville, J., *Activities of cold-shock domain proteins in translation control.* Bioessays, 1999. **21**(4): p. 319-25.

52. Ruzanov, P.V., et al., *Interaction of the universal mRNA-binding protein, p50, with actin: a possible link between mRNA and microfilaments.* J Cell Sci, 1999. **112** (**Pt 20**): p. 3487-96.

53. Evdokimova, V.M. and L.P. Ovchinnikov, *Translational regulation by Y-box transcription factor: involvement of the major mRNA-associated protein, p50.* Int J Biochem Cell Biol, 1999. **31**(1): p. 139-49.

54. Evdokimova, V., et al., *The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer.* Embo J, 2001. **20**(19): p. 5491-502.

55. Tezcan, E.F., *Cold shock proteins.* Turk J Med Sci, 2001. **31**: p. 283-290.

56. Jones, P.G. and M. Inouye, *The cold-shock response- A hot topic.* Molecular Microbiology, 1994. **11**(5): p. 811-818.

57. Jones, P.G. and M. Inouye, *RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response.* Mol Microbiol, 1996. **21**(6): p. 1207-18.

58. Graumann, P.L. and M.A. Marahiel, *A superfamily of proteins that contain the cold-shock domain.* Trends Biochem Sci, 1998. **23**(8): p. 286-90.

59. Perl, D., et al., *Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins.* Nat Struct Biol, 1998. **5**(3): p. 229-35.

60. Yamanaka, K., et al., *Cloning, sequencing, and characterization of multicopy suppressors of a mukB mutation in Escherichia coli.* Mol Microbiol, 1994. **13**(2): p. 301-12.

61. Brandi, A., et al., *Massive presence of the Escherichia coli 'major cold-shock protein' CspA under non-stress conditions.* Embo J, 1999. **18**(6): p. 1653-9.

62.   Yamanaka, K. and M. Inouye, *Growth-phase-dependent expression of cspD, encoding a member of the CspA family in Escherichia coli.* J Bacteriol, 1997. **179**(16): p. 5126-30.

63.   Lee, S.J., et al., *Family of the major cold-shock protein, CspA (CS7.4), of Escherichia coli, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins.* Mol Microbiol, 1994. **11**(5): p. 833-9.

64.   Nakashima, K., et al., *A novel member of the cspA family of genes that is induced by cold shock in Escherichia coli.* J Bacteriol, 1996. **178**(10): p. 2994-7.

65.   Wang, N., K. Yamanaka, and M. Inouye, *CspI, the ninth member of the CspA family of Escherichia coli, is induced upon cold shock.* J Bacteriol, 1999. **181**(5): p. 1603-9.

66.   Gualerzi, C.O., A.M. Giuliodori, and C.L. Pon, *Transcriptional and post-transcriptional control of cold-shock genes.* J Mol Biol, 2003. **331**(3): p. 527-39.

67.   Mitta, M., L. Fang, and M. Inouye, *Deletion analysis of cspA of Escherichia coli: requirement of the AT-rich UP element for cspA transcription and the downstream box in the coding region for its cold shock induction.* Mol Microbiol, 1997. **26**(2): p. 321-35.

68.   Bae, W., P.G. Jones, and M. Inouye, *CspA, the major cold shock protein of Escherichia coli, negatively regulates its own gene expression.* J Bacteriol, 1997. **179**(22): p. 7081-8.

69.   Xia, B., H. Ke, and M. Inouye, *Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in Escherichia coli.* Mol Microbiol, 2001. **40**(1): p. 179-88.

70.   Feng, Y., et al., *Escherichia coli poly(A)-binding proteins that interact with components of degradosomes or impede RNA decay mediated by polynucleotide phosphorylase and RNase E.* J Biol Chem, 2001. **276**(34): p. 31651-6.

71.   Hanna, M.M. and K. Liu, *Nascent RNA in transcription complexes interacts with CspE, a small protein in E. coli implicated in chromatin condensation.* J Mol Biol, 1998. **282**(2): p. 227-39.

72.   Bae, W., et al., *Escherichia coli CspA-family RNA chaperones are transcription antiterminators.* Proc Natl Acad Sci U S A, 2000. **97**(14): p. 7784-9.

73.   Kunst, F., et al., *The complete genome sequence of the gram-positive bacterium Bacillus subtilis.* Nature, 1997. **390**(6657): p. 249-56.

74.   Graumann, P., et al., *A family of cold shock proteins in Bacillus subtilis is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures.* Mol Microbiol, 1997. **25**(4): p. 741-56.

75.   Yamanaka, K., et al., *CspD, a novel DNA replication inhibitor induced during the stationary phase in Escherichia coli.* Mol Microbiol, 2001. **39**(6): p. 1572-84.

76.   Willimsky, G., et al., *Characterization of cspB, a Bacillus subtilis inducible cold shock gene affecting cell viability at low temperatures.* J Bacteriol, 1992. **174**(20): p. 6326-35.

77.   Yamanaka, K., L. Fang, and M. Inouye, *The CspA family in Escherichia coli: multiple gene duplication for stress adaptation.* Mol Microbiol, 1998. **27**(2): p. 247-55.

78.   Delbruck, H., et al., *Crystal structures of mutant forms of the Bacillus caldolyticus cold shock protein differing in thermal stability.* J Mol Biol, 2001. **313**(2): p. 359-69.

79.   Bayles, D.O., B.A. Annous, and B.J. Wilkinson, *Cold stress proteins induced in Listeria monocytogenes in response to temperature downshock and growth at low temperatures.* Applied and Environmental Microbiology, 1996. **62**(3): p. 1116-1119.

80.   Kremer, W., et al., *Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima.* 2001, FEBS.

81.   Giaquinto, L., et al., *Structure and function of cold shock proteins in archaea.* J Bacteriol, 2007. **189**(15): p. 5738-48.

82.   Rothman-Denes, L.B., et al., *Transcriptional regulation by DNA structural transitions and single-stranded DNA-binding proteins.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 63-73.

83.   Didier, D.K., et al., *Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box.* Proc Natl Acad Sci U S A, 1988. **85**(19): p. 7322-6.

84.   Ting, J.P., et al., *YB-1 DNA-binding protein represses interferon gamma activation of class II major histocompatibility complex genes.* J Exp Med, 1994. **179**(5): p. 1605-11.

85.   Sapru, M.K., et al., *Cloning and characterization of a novel transcriptional repressor of the nicotinic acetylcholine receptor delta-subunit gene.* J Biol Chem, 1996. **271**(12): p. 7203-11.

86.   Graumann, P. and M.A. Marahiel, *The major cold shock protein of Bacillus subtilis CspB binds with high affinity to the ATTGG- and CCAAT sequences in single stranded oligonucleotides.* FEBS Lett, 1994. **338**(2): p. 157-60.

87.   Pon, C.L., R.A. Calogero, and C.O. Gualerzi, *Identification, cloning, nucleotide sequence and chromosomal map location of hns, the structural gene for Escherichia coli DNA-binding protein H-NS.* Mol Gen Genet, 1988. **212**(2): p. 199-202.

88.   Friedrich, K., et al., *Proteins from the prokaryotic nucleoid. Interaction of nucleic acids with the 15 kDa Escherichia coli histone-like protein H-NS.* FEBS Lett, 1988. **229**(1): p. 197-202.

89.   La Teana, A., et al., *Identification of a cold shock transcriptional enhancer of the Escherichia coli gene encoding nucleoid protein H-NS.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10907-11.

90.   Espinoza, C.A., et al., *B2 RNA binds directly to RNA polymerase II to repress transcript synthesis.* Nat Struct Mol Biol, 2004. **11**(9): p. 822-9.

91.   Hofweber, R., et al., *The influence of cold shock proteins on transcription and translation studied in cell-free model systems.* Febs J, 2005. **272**(18): p. 4691-702.

92.   La Teana, A., et al., *Translation during cold adaptation does not involve mRNA-rRNA base pairing through the downstream box.* Rna, 2000. **6**(10): p. 1393-402.

93.   Etchegaray, J.P., et al., *Downstream box: a hidden translational enhancer.* Mol Microbiol, 1998. **27**(4): p. 873-4.

94.   Etchegaray, J.P. and M. Inouye, *DB or not DB in translation?* Mol Microbiol, 1999. **33**(2): p. 438-9.

95.   Ermolenko, D.N. and G.I. Makhatadze, *Bacterial cold-shock proteins.* Cell Mol Life Sci, 2002. **59**(11): p. 1902-13.

96.   Broeze, R.J., C.J. Solomon, and D.H. Pope, *Effects of low temperature on in vivo and in vitro protein synthesis in Escherichia coli and Pseudomonas fluorescens.* J Bacteriol, 1978. **134**(3): p. 861-74.

97.   Jones, P.G., et al., *Function of a relaxed-like state following temperature downshifts in Escherichia coli.* J Bacteriol, 1992. **174**(12): p. 3903-14.

98.   Spirin, A., *Ribosomes.* 1999, New York.

99.   Szer, W. and S. Ochoa, *Complexing Ability and Coding Properties of Synthetic Polynucleotides.* J Mol Biol, 1964. **8**: p. 823-34.

100.   Giuliodori, A.M., et al., *Preferential translation of cold-shock mRNAs during cold adaptation.* Rna, 2004. **10**(2): p. 265-76.

101.   Agafonov, D.E., V.A. Kolb, and A.S. Spirin, *A novel stress-response protein that binds at the ribosomal subunit interface and arrests translation.* Cold Spring Harb Symp Quant Biol, 2001. **66**: p. 509-14.

102.   Agafonov, D.E., V.A. Kolb, and A.S. Spirin, *Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage.* EMBO Rep, 2001. **2**(5): p. 399-402.

103.   Cordin, O., et al., *The DEAD-box protein family of RNA helicases.* Gene, 2006. **367**: p. 17-37.

104.   Charollais, J., M. Dreyfus, and I. Iost, *CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit.* Nucleic Acids Res, 2004. **32**(9): p. 2751-9.

105.   Fairman, M.E., et al., *Protein displacement by DExH/D "RNA helicases" without duplex unwinding.* Science, 2004. **304**(5671): p. 730-4.

106. de la Cruz, J., D. Kressler, and P. Linder, *Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families.* Trends Biochem Sci, 1999. **24**(5): p. 192-8.

107. Tanner, N.K. and P. Linder, *DExD/H box RNA helicases: from generic motors to specific dissociation functions.* Mol Cell, 2001. **8**(2): p. 251-62.

108. Schmid, S.R. and P. Linder, *D-E-A-D protein family of putative RNA helicases.* Mol Microbiol, 1992. **6**(3): p. 283-91.

109. Iost, I. and M. Dreyfus, *mRNAs can be stabilized by DEAD-box proteins.* Nature, 1994. **372**(6502): p. 193-6.

110. Jones, P.G., et al., *Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 76-80.

111. Miczak, A., et al., *Proteins associated with RNase E in a multicomponent ribonucleolytic complex.* Proc Natl Acad Sci U S A, 1996. **93**(9): p. 3865-9.

112. Py, B., et al., *A DEAD-box RNA helicase in the Escherichia coli RNA degradosome.* Nature, 1996. **381**(6578): p. 169-72.

113. Grunberg-Manago, M., *Messenger RNA stability and its role in control of gene expression in bacteria and phages.* Annu Rev Genet, 1999. **33**: p. 193-227.

114. Prud'homme-Genereux, A., et al., *Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a'cold shock degradosome'.* Molecular Microbiology, 2004. **54**(5): p. 1409-1421.

115. Awano, N., et al., *Complementation analysis of the cold-sensitive phenotype of the Escherichia coli csdA deletion strain.* J Bacteriol, 2007.

116. Hunger, K., et al., *Cold-induced putative DEAD box RNA helicases CshA and CshB are essential for cold adaptation and interact with cold shock protein B in Bacillus subtilis.* J Bacteriol, 2006. **188**(1): p. 240-8.

117. Kremer, W., et al., *Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima.* Eur J Biochem, 2001. **268**(9): p. 2527-39.

118. Schindelin, H., et al., *Crystal structure of CspA, the major cold shock protein of Escherichia coli.* Proc Natl Acad Sci U S A, 1994. **91**(11): p. 5119-23.

119. Schindelin, H., M.A. Marahiel, and U. Heinemann, *Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein.* Nature, 1993. **364**(6433): p. 164-8.

120. Newkirk, K., et al., *Solution NMR structure of the major cold shock protein (CspA) from Escherichia coli: identification of a binding epitope for DNA.* Proc Natl Acad Sci U S A, 1994. **91**(11): p. 5114-8.

121. Jung, A., et al., *High-temperature solution NMR structure of TmCsp.* Protein Sci, 2004. **13**(2): p. 342-50.

122. Mueller, U., et al., *Thermal stability and atomic-resolution crystal structure of the Bacillus caldolyticus cold shock protein.* J Mol Biol, 2000. **297**(4): p. 975-88.

123. Makhatadze, G.I. and M.A. Marahiel, *Effect of pH and phosphate ions on self-association properties of the major cold-shock protein from Bacillus subtilis.* Protein Sci, 1994. **3**(11): p. 2144-7.

124. Johnston, D., et al., *Specificity of DNA binding and dimerization by CspE from Escherichia coli.* J Biol Chem, 2006. **281**(52): p. 40208-15.

125. Mayr, B., et al., *Identification and purification of a family of dimeric major cold shock protein homologs from the psychrotrophic Bacillus cereus WSBC 10201.* J Bacteriol, 1996. **178**(10): p. 2916-25.

126. Wolffe, A.P., et al., *The Y-box factors: a family of nucleic acid binding proteins conserved from Escherichia coli to man.* New Biol, 1992. **4**(4): p. 290-8.

127. Wilkinson, M.F. and A.B. Shyu, *Multifunctional regulatory proteins that control gene expression in both the nucleus and the cytoplasm.* Bioessays, 2001. **23**(9): p. 775-87.

128. Horn, G., et al., *Structure and function of bacterial cold shock proteins.* Cell Mol Life Sci, 2007. **64**(12): p. 1457-70.

129. Landsman, D., *RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain.* Nucleic Acids Res, 1992. **20**(11): p. 2861-4.

130. Lopez, M.M., K. Yutani, and G.I. Makhatadze, *Interactions of the major cold shock protein of Bacillus subtilis CspB with single-stranded DNA templates of different base composition.* J Biol Chem, 1999. **274**(47): p. 33601-8.

131. Falsone, S.F., et al., *Unfolding and double-stranded DNA binding of the cold shock protein homologue Cla h 8 from Cladosporium herbarum.* J Biol Chem, 2002. **277**(19): p. 16512-6.

132. Phadtare, S., et al., *CspB and CspL, thermostable cold-shock proteins from Thermotoga maritima.* Genes Cells, 2003. **8**(10): p. 801-10.

133. Zeeb, M. and J. Balbach, *Single-stranded DNA binding of the cold-shock protein CspB from Bacillus subtilis: NMR mapping and mutational characterization.* Protein Sci, 2003. **12**(1): p. 112-23.

134. Karpel, R.L., L.E. Henderson, and S. Oroszlan, *Interactions of retroviral structural proteins with single-stranded nucleic acids.* J Biol Chem, 1987. **262**(11): p. 4961-7.

135. Namsaraev, E.A. and P. Berg, *Branch migration during Rad51-promoted strand exchange proceeds in either direction.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10477-81.

136. Kelm, R.J., Jr., et al., *Structure/function analysis of mouse Purbeta, a single-stranded DNA-binding repressor of vascular smooth muscle alpha-actin gene transcription.* J Biol Chem, 2003. **278**(40): p. 38749-57.

137.   Simon, M.D., et al., *A phage display selection of engrailed homeodomain mutants and the importance of residue Q50.* Nucleic Acids Res, 2004. **32**(12): p. 3623-31.

138.   Papatsenko, D.A., et al., *Mapping of DNA-binding proteins along the yeast genome by UV-induced DNA-protein crosslinking.* FEBS Lett, 1996. **381**(1-2): p. 103-5.

139.   Ding, J., et al., *Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA.* Genes Dev, 1999. **13**(9): p. 1102-15.

140.   Wang, J., et al., *Exonuclease III protection assay with FRET probe for detecting DNA-binding proteins.* Nucleic Acids Res, 2005. **33**(2): p. e23.

141.   Walker, G.T., C.P. Linn, and J.G. Nadeau, *DNA detection by strand displacement amplification and fluorescence polarization with signal enhancement using a DNA binding protein.* Nucleic Acids Res, 1996. **24**(2): p. 348-53.

142.   Kozlov, A.G. and T.M. Lohman, *Large contributions of coupled protonation equilibria to the observed enthalpy and heat capacity changes for ssDNA binding to Escherichia coli SSB protein.* Proteins, 2000. **Suppl 4**: p. 8-22.

143.   Witte, G., C. Urbanke, and U. Curth, *DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery.* Nucleic Acids Res, 2003. **31**(15): p. 4434-40.

144.   Klug, S.J. and M. Famulok, *All you wanted to know about SELEX.* Mol Biol Rep, 1994. **20**(2): p. 97-107.

145.   van der Merwe, P.A., *Surface Plasmon Resonance in Protein-Ligand interactions: hydrodynamics and calorimetry.*, in *Practical Approach series.* 2001, Oxford University Press. p. 137-170.

146.   Wear, M.A. and M.D. Walkinshaw, *Thermodynamics of the cyclophilin-A/cyclosporin-A interaction: a direct comparison of parameters determined by surface plasmon resonance using Biacore T100 and isothermal titration calorimetry.* Anal Biochem, 2006. **359**(2): p. 285-7.

147.   Wear, M.A., et al., *A surface plasmon resonance-based assay for small molecule inhibitors of human cyclophilin A.* Anal Biochem, 2005. **345**(2): p. 214-26.

148.   Kinoshita, T., *[Application and development of structure-based drug design using X-ray analysis].* Nippon Yakurigaku Zasshi, 2007. **129**(3): p. 186-90.

149.   Remenyi, A., et al., *Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers.* Genes Dev, 2003. **17**(16): p. 2048-59.

150.   van Pouderoyen, G., et al., *Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA.* Embo J, 1997. **16**(19): p. 6044-54.

151. Rhodes, G., *Crystallography made crystal clear.* Third ed. 2006: Elsevier Inc. 306.

152. Chayen, N.E., *Methods for separating nucleation and growth in protein crystallisation.* Prog Biophys Mol Biol, 2005. **88**(3): p. 329-37.

153. McNae, I., et al., *Study of protein-ligand inetractions using protein crystallography.* Crystallography Reviews., 2005. **11**: p. 61-71.

154. Skarzynski, T. and J. Thorpe, *Industrial perspective on X-ray data collection and analysis.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 102-7.

155. Sorensen, T.L., et al., *New light for science: synchrotron radiation in structural medicine.* Trends Biotechnol, 2006. **24**(11): p. 500-8.

156. Minich, W.B. and L.P. Ovchinnikov, *Role of cytoplasmic mRNP proteins in translation.* Biochimie, 1992. **74**(5): p. 477-83.

157. Craig, J.E., et al., *Expression of the cold-shock gene cspB in Salmonella typhimurium occurs below a threshold temperature.* Microbiology, 1998. **144** (**Pt 3**): p. 697-704.

158. Jeffreys, A.G., et al., *Growth, survival and characterization of cspA in Salmonella enteritidis following cold shock.* Curr Microbiol, 1998. **36**(1): p. 29-35.

159. Horton, A.J., et al., *Adaptive response to cold temperatures and characterization of cspA in Salmonella typhimurium LT2.* Antonie Van Leeuwenhoek, 2000. **77**(1): p. 13-20.

160. Kim, B.H., et al., *Expression of cspH, encoding the cold shock protein in Salmonella enterica serovar Typhimurium UK-1.* J Bacteriol, 2001. **183**(19): p. 5580-8.

161. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin.* Nature, 1992. **355**(6360): p. 564-6.

162. Wang, C., et al., *SELEX Screening and Characterization of Small RNA Molecules That Specifically Bind the Reactive Blue Dye.* Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai), 1999. **31**(5): p. 504-508.

163. Warren, C.L., et al., *Defining the sequence-recognition profile of DNA-binding molecules.* Proc Natl Acad Sci U S A, 2006. **103**(4): p. 867-72.

164. Berger, M.F., et al., *Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities.* Nat Biotechnol, 2006. **24**(11): p. 1429-1435.

165. Estibeiro, P., *ACCESSarray suplier.* 2005, ETTC Biospace.

166. Cubeddu, L. and M.F. White, *An E145C mutation in the SsoSSB protein (St Andrews University).* Unpublished, 2007.

167. Wadsworth, R.I. and M.F. White, *Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus.* Nucleic Acids Res, 2001. **29**(4): p. 914-20.

168. Schindelin, H., et al., *Overproduction, crystallization, and preliminary X-ray diffraction studies of the major cold shock protein from Bacillus subtilis, CspB.* Proteins, 1992. **14**(1): p. 120-4.

169. Bulyk, M.L., et al., *Exploring the DNA-binding specificities of zinc fingers with DNA microarrays.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7158-63.

170. Zeeb, M., et al., *Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution.* Nucleic Acids Res, 2006. **34**(16): p. 4561-71.

171. Santoro, I.M., T.M. Yi, and K. Walsh, *Identification of single-stranded-DNA-binding proteins that interact with muscle gene elements.* Mol Cell Biol, 1991. **11**(4): p. 1944-53.

172. Gerrero, M.R., et al., *Brn-3.0: a POU-domain protein expressed in the sensory, immune, and endocrine systems that functions on elements distinct from known octamer motifs.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10841-5.

173. Kaan, T., et al., *Genome-wide transcriptional profiling of the Bacillus subtilis cold-shock response.* Microbiology, 2002. **148**(Pt 11): p. 3441-55.

174. Morgan, H.P., et al., *Sequence specificity of single-stranded DNA-binding proteins: a novel DNA microarray approach.* Nucleic Acids Res, 2007. **35**(10): p. e75.

175. Higgs, P.G., *RNA secondary structure: physical and computational aspects.* Quarterly Reviews of Biophysics, 2001. **33**(03): p. 199-253.

176. Kloks, C.P., et al., *The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1.* J Mol Biol, 2002. **316**(2): p. 317-26.

177. Yip, T.T., Y. Nakagawa, and J. Porath, *Evaluation of the interaction of peptides with Cu(II), Ni(II), and Zn(II) by high-performance immobilized metal ion affinity chromatography.* Anal Biochem, 1989. **183**(1): p. 159-71.

178. Alberts, B.M. and L. Frey, *T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA.* Nature, 1970. **227**(5265): p. 1313-8.

179. Reuben, R.C. and M.L. Gefter, *A deoxyribonucleic acid-binding protein induced by bacteriophage T7. Purification and properties of the protein.* J Biol Chem, 1974. **249**(12): p. 3843-50.

180. Schaller, H., et al., *Affinity Chromatography of DNA-Binding Enzymes on Single-Stranded DNA-Agarose Columns.* European Journal of Biochemistry, 1972. **26**(4): p. 474-481.

181. Cadman, C.J. and P. McGlynn, *PriA helicase and SSB interact physically and functionally.* Nucleic Acids Research, 2004. **32**(21): p. 6378.

182. Clark, E.D.B., *Refolding of recombinant proteins.* Curr. Opin. Biotechnol, 1998. **9**(2): p. 157-163.

183. Goldberg, M.E., R. Rudolph, and R. Jaenicke, *A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme.* Biochemistry, 1991. **30**(11): p. 2790-7.

184. Batas, B. and J.B. Chaudhuri, *Protein refolding at high concentration using size-exclusion chromatography.* Biotechnology and Bioengineering, 1996. **50**(1): p. 16-23.

185. Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E. coli.* Curr Opin Biotechnol, 1998. **9**(5): p. 497-501.

186. Xie, B., et al., *The functional domains of human ventricular myosin light chain 1.* Biophysical Chemistry, 2003. **106**(1): p. 57-66.

187. Wallqvist, A., D.G. Covell, and D. Thirumalai, *Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation.* J. Am. Chem. Soc, 1998. **120**: p. 427-428.

188. Bhuyan, A.K., *Protein stabilization by urea and guanidine hydrochloride.* Biochemistry, 2002. **41**(45): p. 13386-94.

189. Zasedateleva, O.A., et al., *Specificity of mammalian Y-box binding protein p50 in interaction with ss and ds DNA analyzed with generic oligonucleotide microchip.* J Mol Biol, 2002. **324**(1): p. 73-87.

190. Evdokimova, V.M., et al., *The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro.* J Biol Chem, 1998. **273**(6): p. 3574-81.

191. Davydova, E.K., et al., *Overexpression in COS cells of p50, the major core protein associated with mRNA, results in translation inhibition.* Nucleic Acids Res, 1997. **25**(14): p. 2911-6.

192. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research, 2003. **31**(13): p. 3406-3415.

193. Shchepinov, M.S., et al., *Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays.* Nucleic Acids Research. **25**(6): p. 1155-1161.

194. Roos, H., et al., *Thermodynamic analysis of protein interactions with biosensor technology.* J Mol Recognit, 1998. **11**(1-6): p. 204-10.

195. Khan, F., M. He, and M.J. Taussig, *Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces.* Anal Chem, 2006. **78**(9): p. 3072-9.

196. Leslie, A.G.W., *Recent changes to the MOSFLM package for processing film and image plate data.* Joint CCP4+ ESF-EAMCB Newsletter on Protein Crystallography, 1992. **26**: p. 11-20.

197. Project, T.S.C.C., *CCP4.* Science and Engineering Research Council, Warrington, U.K., 1979: p. Project #4.

198.   Cowtan, K.D. and K.Y. Zhang, *Density modification for macromolecular phase improvement.* Prog Biophys Mol Biol, 1999. **72**(3): p. 245-70.

199.   Perrakis, A., R. Morris, and V.S. Lamzin, *Automated protein model building combined with iterative structure refinement.* Nat Struct Biol, 1999. **6**(5): p. 458-63.

200.   Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

201.   Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

202.   McPherson, A., Methods in enzymology, 1985. **11**: p. 112-128.

203.   Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.

204.   Krissinel, E. and K. Henrick, *Detection of Protein Assemblies in Crystals.*, in *CompLife* 2005, Springer-Verlag Berlin Heidelberg. p. 163-174.

205.   Bates, P.A., et al., *Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM.* Proteins, 2001. **Suppl 5**: p. 39-46.

206.   Okamoto, T., et al., *Direct interaction of p53 with the Y-box binding protein, YB-1: a mechanism for regulation of human gene expression.* Oncogene, 2000. **19**(54): p. 6194-202.

207.   Shnyreva, M., et al., *Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein.* J Biol Chem, 2000. **275**(20): p. 15498-503.

208.   Izumi, H., et al., *Y box-binding protein-1 binds preferentially to single-stranded nucleic acids and exhibits 3'-->5' exonuclease activity.* Nucleic Acids Res, 2001. **29**(5): p. 1200-7.

209.   Sakura, H., et al., *Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology.* Gene, 1988. **73**(2): p. 499-507.

210.   Horwitz, E.M., K.A. Maloney, and T.J. Ley, *A human protein containing a "cold shock" domain binds specifically to H-DNA upstream from the human gamma-globin genes.* J Biol Chem, 1994. **269**(19): p. 14130-9.

211.   Funakoshi, T., et al., *Isolation and characterization of brain Y-box protein: developmentally regulated expression, polyribosomal association and dendritic localization.* Brain Res Mol Brain Res, 2003. **118**(1-2): p. 1-9.

212.   Ozer, J., et al., *Isolation and characterization of a cDNA clone for the CCAAT transcription factor EFIA reveals a novel structural motif.* J Biol Chem, 1990. **265**(36): p. 22143-52.

213. Ohmori, M., et al., *A Y-box protein is a suppressor factor that decreases thyrotropin receptor gene expression.* Mol Endocrinol, 1996. **10**(1): p. 76-89.

214. Marques, A.C., et al., *Emergence of young human genes after a burst of retroposition in primates.* PLoS Biol, 2005. **3**(11): p. e357.

122. Mueller, U., et al., *Thermal stability and atomic-resolution crystal structure of the Bacillus caldolyticus cold shock protein.* J Mol Biol, 2000. **297**(4): p. 975-88.

123. Makhatadze, G.I. and M.A. Marahiel, *Effect of pH and phosphate ions on self-association properties of the major cold-shock protein from Bacillus subtilis.* Protein Sci, 1994. **3**(11): p. 2144-7.

124. Johnston, D., et al., *Specificity of DNA binding and dimerization by CspE from Escherichia coli.* J Biol Chem, 2006. **281**(52): p. 40208-15.

125. Mayr, B., et al., *Identification and purification of a family of dimeric major cold shock protein homologs from the psychrotrophic Bacillus cereus WSBC 10201.* J Bacteriol, 1996. **178**(10): p. 2916-25.

126. Wolffe, A.P., et al., *The Y-box factors: a family of nucleic acid binding proteins conserved from Escherichia coli to man.* New Biol, 1992. **4**(4): p. 290-8.

127. Wilkinson, M.F. and A.B. Shyu, *Multifunctional regulatory proteins that control gene expression in both the nucleus and the cytoplasm.* Bioessays, 2001. **23**(9): p. 775-87.

128. Horn, G., et al., *Structure and function of bacterial cold shock proteins.* Cell Mol Life Sci, 2007. **64**(12): p. 1457-70.

129. Landsman, D., *RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain.* Nucleic Acids Res, 1992. **20**(11): p. 2861-4.

130. Lopez, M.M., K. Yutani, and G.I. Makhatadze, *Interactions of the major cold shock protein of Bacillus subtilis CspB with single-stranded DNA templates of different base composition.* J Biol Chem, 1999. **274**(47): p. 33601-8.

131. Falsone, S.F., et al., *Unfolding and double-stranded DNA binding of the cold shock protein homologue Cla h 8 from Cladosporium herbarum.* J Biol Chem, 2002. **277**(19): p. 16512-6.

132. Phadtare, S., et al., *CspB and CspL, thermostable cold-shock proteins from Thermotoga maritima.* Genes Cells, 2003. **8**(10): p. 801-10.

133. Zeeb, M. and J. Balbach, *Single-stranded DNA binding of the cold-shock protein CspB from Bacillus subtilis: NMR mapping and mutational characterization.* Protein Sci, 2003. **12**(1): p. 112-23.

134. Karpel, R.L., L.E. Henderson, and S. Oroszlan, *Interactions of retroviral structural proteins with single-stranded nucleic acids.* J Biol Chem, 1987. **262**(11): p. 4961-7.

135. Namsaraev, E.A. and P. Berg, *Branch migration during Rad51-promoted strand exchange proceeds in either direction.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10477-81.

136. Kelm, R.J., Jr., et al., *Structure/function analysis of mouse Purbeta, a single-stranded DNA-binding repressor of vascular smooth muscle alpha-actin gene transcription.* J Biol Chem, 2003. **278**(40): p. 38749-57.

137.   Simon, M.D., et al., *A phage display selection of engrailed homeodomain mutants and the importance of residue Q50*. Nucleic Acids Res, 2004. **32**(12): p. 3623-31.

138.   Papatsenko, D.A., et al., *Mapping of DNA-binding proteins along the yeast genome by UV-induced DNA-protein crosslinking*. FEBS Lett, 1996. **381**(1-2): p. 103-5.

139.   Ding, J., et al., *Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA*. Genes Dev, 1999. **13**(9): p. 1102-15.

140.   Wang, J., et al., *Exonuclease III protection assay with FRET probe for detecting DNA-binding proteins*. Nucleic Acids Res, 2005. **33**(2): p. e23.

141.   Walker, G.T., C.P. Linn, and J.G. Nadeau, *DNA detection by strand displacement amplification and fluorescence polarization with signal enhancement using a DNA binding protein*. Nucleic Acids Res, 1996. **24**(2): p. 348-53.

142.   Kozlov, A.G. and T.M. Lohman, *Large contributions of coupled protonation equilibria to the observed enthalpy and heat capacity changes for ssDNA binding to Escherichia coli SSB protein*. Proteins, 2000. **Suppl 4**: p. 8-22.

143.   Witte, G., C. Urbanke, and U. Curth, *DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery*. Nucleic Acids Res, 2003. **31**(15): p. 4434-40.

144.   Klug, S.J. and M. Famulok, *All you wanted to know about SELEX*. Mol Biol Rep, 1994. **20**(2): p. 97-107.

145.   van der Merwe, P.A., *Surface Plasmon Resonance in Protein-Ligand interactions: hydrodynamics and calorimetry.*, in *Practical Approach series*. 2001, Oxford University Press. p. 137-170.

146.   Wear, M.A. and M.D. Walkinshaw, *Thermodynamics of the cyclophilin-A/cyclosporin-A interaction: a direct comparison of parameters determined by surface plasmon resonance using Biacore T100 and isothermal titration calorimetry*. Anal Biochem, 2006. **359**(2): p. 285-7.

147.   Wear, M.A., et al., *A surface plasmon resonance-based assay for small molecule inhibitors of human cyclophilin A*. Anal Biochem, 2005. **345**(2): p. 214-26.

148.   Kinoshita, T., *[Application and development of structure-based drug design using X-ray analysis]*. Nippon Yakurigaku Zasshi, 2007. **129**(3): p. 186-90.

149.   Remenyi, A., et al., *Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers*. Genes Dev, 2003. **17**(16): p. 2048-59.

150.   van Pouderoyen, G., et al., *Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA*. Embo J, 1997. **16**(19): p. 6044-54.

151. Rhodes, G., *Crystallography made crystal clear.* Third ed. 2006: Elsevier Inc. 306.

152. Chayen, N.E., *Methods for separating nucleation and growth in protein crystallisation.* Prog Biophys Mol Biol, 2005. **88**(3): p. 329-37.

153. McNae, I., et al., *Study of protein-ligand inetractions using protein crystallography.* Crystallography Reviews., 2005. **11**: p. 61-71.

154. Skarzynski, T. and J. Thorpe, *Industrial perspective on X-ray data collection and analysis.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 102-7.

155. Sorensen, T.L., et al., *New light for science: synchrotron radiation in structural medicine.* Trends Biotechnol, 2006. **24**(11): p. 500-8.

156. Minich, W.B. and L.P. Ovchinnikov, *Role of cytoplasmic mRNP proteins in translation.* Biochimie, 1992. **74**(5): p. 477-83.

157. Craig, J.E., et al., *Expression of the cold-shock gene cspB in Salmonella typhimurium occurs below a threshold temperature.* Microbiology, 1998. **144** (**Pt 3**): p. 697-704.

158. Jeffreys, A.G., et al., *Growth, survival and characterization of cspA in Salmonella enteritidis following cold shock.* Curr Microbiol, 1998. **36**(1): p. 29-35.

159. Horton, A.J., et al., *Adaptive response to cold temperatures and characterization of cspA in Salmonella typhimurium LT2.* Antonie Van Leeuwenhoek, 2000. **77**(1): p. 13-20.

160. Kim, B.H.,.et al., *Expression of cspH, encoding the cold shock protein in Salmonella enterica serovar Typhimurium UK-1.* J Bacteriol, 2001. **183**(19): p. 5580-8.

161. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin.* Nature, 1992. **355**(6360): p. 564-6.

162. Wang, C., et al., *SELEX Screening and Characterization of Small RNA Molecules That Specifically Bind the Reactive Blue Dye.* Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai), 1999. **31**(5): p. 504-508.

163. Warren, C.L., et al., *Defining the sequence-recognition profile of DNA-binding molecules.* Proc Natl Acad Sci U S A, 2006. **103**(4): p. 867-72.

164. Berger, M.F., et al., *Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities.* Nat Biotechnol, 2006. **24**(11): p. 1429-1435.

165. Estibeiro, P., *ACCESSarray suplier.* 2005, ETTC Biospace.

166. Cubeddu, L. and M.F. White, *An E145C mutation in the SsoSSB protein (St Andrews University).* Unpublished, 2007.

167. Wadsworth, R.I. and M.F. White, *Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus.* Nucleic Acids Res, 2001. **29**(4): p. 914-20.

168. Schindelin, H., et al., *Overproduction, crystallization, and preliminary X-ray diffraction studies of the major cold shock protein from Bacillus subtilis, CspB.* Proteins, 1992. **14**(1): p. 120-4.

169. Bulyk, M.L., et al., *Exploring the DNA-binding specificities of zinc fingers with DNA microarrays.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7158-63.

170. Zeeb, M., et al., *Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution.* Nucleic Acids Res, 2006. **34**(16): p. 4561-71.

171. Santoro, I.M., T.M. Yi, and K. Walsh, *Identification of single-stranded-DNA-binding proteins that interact with muscle gene elements.* Mol Cell Biol, 1991. **11**(4): p. 1944-53.

172. Gerrero, M.R., et al., *Brn-3.0: a POU-domain protein expressed in the sensory, immune, and endocrine systems that functions on elements distinct from known octamer motifs.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10841-5.

173. Kaan, T., et al., *Genome-wide transcriptional profiling of the Bacillus subtilis cold-shock response.* Microbiology, 2002. **148**(Pt 11): p. 3441-55.

174. Morgan, H.P., et al., *Sequence specificity of single-stranded DNA-binding proteins: a novel DNA microarray approach.* Nucleic Acids Res, 2007. **35**(10): p. e75.

175. Higgs, P.G., *RNA secondary structure: physical and computational aspects.* Quarterly Reviews of Biophysics, 2001. **33**(03): p. 199-253.

176. Kloks, C.P., et al., *The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1.* J Mol Biol, 2002. **316**(2): p. 317-26.

177. Yip, T.T., Y. Nakagawa, and J. Porath, *Evaluation of the interaction of peptides with Cu(II), Ni(II), and Zn(II) by high-performance immobilized metal ion affinity chromatography.* Anal Biochem, 1989. **183**(1): p. 159-71.

178. Alberts, B.M. and L. Frey, *T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA.* Nature, 1970. **227**(5265): p. 1313-8.

179. Reuben, R.C. and M.L. Gefter, *A deoxyribonucleic acid-binding protein induced by bacteriophage T7. Purification and properties of the protein.* J Biol Chem, 1974. **249**(12): p. 3843-50.

180. Schaller, H., et al., *Affinity Chromatography of DNA-Binding Enzymes on Single-Stranded DNA-Agarose Columns.* European Journal of Biochemistry, 1972. **26**(4): p. 474-481.

181. Cadman, C.J. and P. McGlynn, *PriA helicase and SSB interact physically and functionally.* Nucleic Acids Research, 2004. **32**(21): p. 6378.

182. Clark, E.D.B., *Refolding of recombinant proteins.* Curr. Opin. Biotechnol, 1998. **9**(2): p. 157-163.

183. Goldberg, M.E., R. Rudolph, and R. Jaenicke, *A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme.* Biochemistry, 1991. **30**(11): p. 2790-7.

184. Batas, B. and J.B. Chaudhuri, *Protein refolding at high concentration using size-exclusion chromatography.* Biotechnology and Bioengineering, 1996. **50**(1): p. 16-23.

185. Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E. coli.* Curr Opin Biotechnol, 1998. **9**(5): p. 497-501.

186. Xie, B., et al., *The functional domains of human ventricular myosin light chain 1.* Biophysical Chemistry, 2003. **106**(1): p. 57-66.

187. Wallqvist, A., D.G. Covell, and D. Thirumalai, *Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation.* J. Am. Chem. Soc, 1998. **120**: p. 427-428.

188. Bhuyan, A.K., *Protein stabilization by urea and guanidine hydrochloride.* Biochemistry, 2002. **41**(45): p. 13386-94.

189. Zasedateleva, O.A., et al., *Specificity of mammalian Y-box binding protein p50 in interaction with ss and ds DNA analyzed with generic oligonucleotide microchip.* J Mol Biol, 2002. **324**(1): p. 73-87.

190. Evdokimova, V.M., et al., *The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro.* J Biol Chem, 1998. **273**(6): p. 3574-81.

191. Davydova, E.K., et al., *Overexpression in COS cells of p50, the major core protein associated with mRNA, results in translation inhibition.* Nucleic Acids Res, 1997. **25**(14): p. 2911-6.

192. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research, 2003. **31**(13): p. 3406-3415.

193. Shchepinov, M.S., et al., *Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays.* Nucleic Acids Research. **25**(6): p. 1155-1161.

194. Roos, H., et al., *Thermodynamic analysis of protein interactions with biosensor technology.* J Mol Recognit, 1998. **11**(1-6): p. 204-10.

195. Khan, F., M. He, and M.J. Taussig, *Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces.* Anal Chem, 2006. **78**(9): p. 3072-9.

196. Leslie, A.G.W., *Recent changes to the MOSFLM package for processing film and image plate data.* Joint CCP4+ ESF-EAMCB Newsletter on Protein Crystallography, 1992. **26**: p. 11-20.

197. Project, T.S.C.C., *CCP4.* Science and Engineering Research Council, Warrington, U.K., 1979: p. Project #4.

198. Cowtan, K.D. and K.Y. Zhang, *Density modification for macromolecular phase improvement.* Prog Biophys Mol Biol, 1999. **72**(3): p. 245-70.

199. Perrakis, A., R. Morris, and V.S. Lamzin, *Automated protein model building combined with iterative structure refinement.* Nat Struct Biol, 1999. **6**(5): p. 458-63.

200. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

201. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

202. McPherson, A., Methods in enzymology, 1985. **11**: p. 112-128.

203. Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.

204. Krissinel, E. and K. Henrick, *Detection of Protein Assemblies in Crystals.*, in *CompLife* 2005, Springer-Verlag Berlin Heidelberg. p. 163-174.

205. Bates, P.A., et al., *Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM.* Proteins, 2001. **Suppl 5**: p. 39-46.

206. Okamoto, T., et al., *Direct interaction of p53 with the Y-box binding protein, YB-1: a mechanism for regulation of human gene expression.* Oncogene, 2000. **19**(54): p. 6194-202.

207. Shnyreva, M., et al., *Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein.* J Biol Chem, 2000. **275**(20): p. 15498-503.

208. Izumi, H., et al., *Y box-binding protein-1 binds preferentially to single-stranded nucleic acids and exhibits 3'-->5' exonuclease activity.* Nucleic Acids Res, 2001. **29**(5): p. 1200-7.

209. Sakura, H., et al., *Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology.* Gene, 1988. **73**(2): p. 499-507.

210. Horwitz, E.M., K.A. Maloney, and T.J. Ley, *A human protein containing a "cold shock" domain binds specifically to H-DNA upstream from the human gamma-globin genes.* J Biol Chem, 1994. **269**(19): p. 14130-9.

211. Funakoshi, T., et al., *Isolation and characterization of brain Y-box protein: developmentally regulated expression, polyribosomal association and dendritic localization.* Brain Res Mol Brain Res, 2003. **118**(1-2): p. 1-9.

212. Ozer, J., et al., *Isolation and characterization of a cDNA clone for the CCAAT transcription factor EFIA reveals a novel structural motif.* J Biol Chem, 1990. **265**(36): p. 22143-52.

213.   Ohmori, M., et al., *A Y-box protein is a suppressor factor that decreases thyrotropin receptor gene expression.* Mol Endocrinol, 1996. **10**(1): p. 76-89.

214.   Marques, A.C., et al., *Emergence of young human genes after a burst of retroposition in primates.* PLoS Biol, 2005. **3**(11): p. e357.