Chapter 3 - Binding of T-rich oligonucleotides to bacterial cold shock proteins

3.1 Results

3.1.1 Crystallization and structure solution of a *Bs*-CspB·dT_6 complex

A complex of *Bs*-CspB, the major cold shock protein from *Bacillus subtilis*, and hexathymidine was crystallized in space group $C222_1$ (Figure 3.1). Diffraction data were collected up to a resolution of 1.78 Å at BESSY synchrotron beamline 14.1 [110] (Figure 3.2). Phases were determined by molecular replacement using a crystal structure of *Bs*-CspB without DNA (1CSP) as a template. An overview of the refinement is given in Figure 3.3. The structure was refined using REFMAC5 [121] to final R_{work} / R_{free} values of 18.9% and 22.3%, respectively (Table 3.1). 92.7% and 7.3% of all residues are in the allowed and additionally allowed regions of the Ramachandran plot. The asymmetric unit of the crystal contains one protein chain and one DNA molecule (Figure 3.5a). The resulting electron density of the complex is well defined, except for the protein sidechains of Lys5, Glu21 and Glu36, which are partially

Crystallization
reservoir solution 0.1 M sodium cacodylate (pH 6.5), 0.2 M calcium acetate, 18% (w/v) PEG 8000,
sample buffer 20 mM TRIS (pH 7.5) 50 mM NaCl 3 mM $MgCl_2$ 18 mg/ml *Bs*-CspB·dT_6 complex
crystallization setup 1 µl sample :1 µl reservoir solution hanging-drop crystallization, manual setup

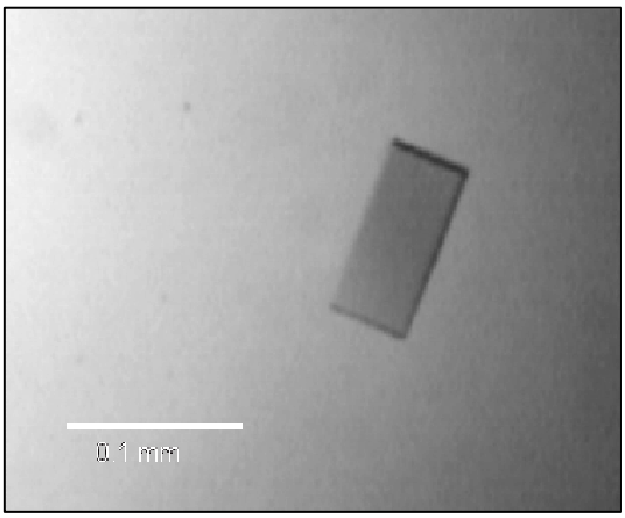


Figure 3.1 Protein crystallization of the *Bs*-CspB·dT_6 complex. The crystal shown on the right was crystallized according to the given conditions and used for data collection.

Data collection & processing		
wavelength (Å)		0.9184
resolution (Å)		19.25 - 1.78
resolution (Å)	(last shell)	2.00 - 1.78
space group		$C222_1$
temperature (K)		100
X-ray source		BESSY BL 14.1
detector		mar345IP
unit-cell parameters	*a* (Å)	49.0
	b (Å)	53.2
	c (Å)	77.0
unique reflections	(last shell)	9,861 (2,832)
$I/\sigma(I)$	(last shell)	12.6 (3.4)
% data completeness	(last shell)	99.2 (98.3)
R_{meas}[a] (%)	(last shell)	6.5 (42.8)

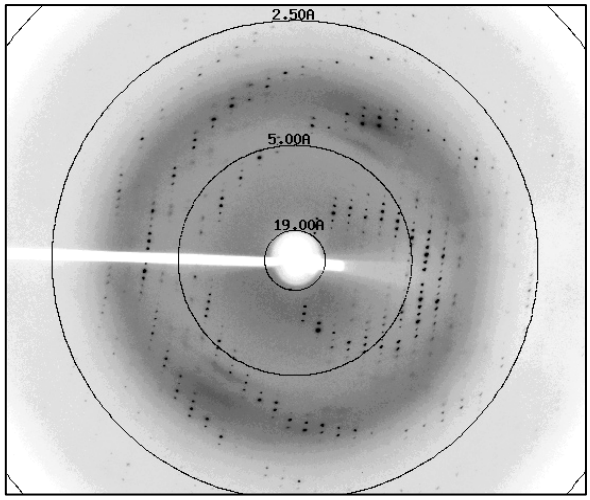


Figure 3.2: X-ray-diffraction-data statistics of a *Bs*-CspB·dT_6 complex crystal. One diffraction image is shown on the right. The angle of rotation is 1°. Resolution (Å) is indicated by rings. The outermost ring refers to 1.78 Å.
[a] R_{meas} is a redundancy-independent *R* factor which correlates intensities from symmetry-related reflections [3].

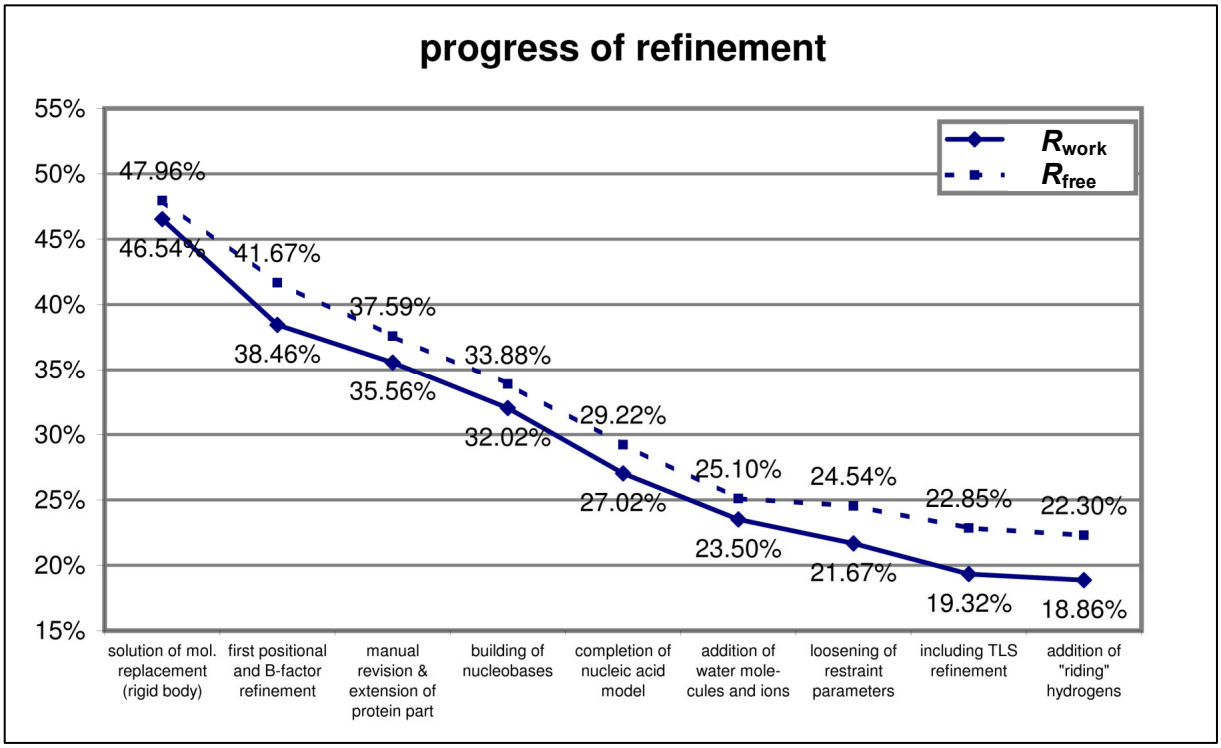


Figure 3.3: Refinement progress the *Bs*-CspB·dT_6 complex crystal structure.

Model building & refinement	
PDB model for mol. replacement	1CSP
PDB access. of the refined structure	2ES2
reflections - working set	9388
reflections - free set (= 4.8%)	473
R_{work}[a] (%)	18.9
R_{free}[a] (%)	22.3
Mean *B* factor (Å^2)	37.6
Number of non-hydrogen atoms	693
Number of protein molecules	1
Number of dT_6 molecules	1
Number of water molecules	63

Model evaluation	
root mean square deviations	
bond lengths (Å)	0.016
bond angles (°)	1.890
torsion angles (°)	5.560
planarity (Å)	0.008
Ramachandran statistics	
residues in allowed regions	92.7%
res. in add. allowed regions	7.3%
res. in generously all. regions	0%
res. in forbidden regions	0%
Real space correlation coefficient[b]	94.9%
ESD coordinate uncertainty (c)[b]	0.164
coord. uncert. from Luzzati P. (Å)[b]	0.249

Table 3.1: Building, refinement and evaluation of the atomic model based on diffraction data from *Bs*-CspB·dT_6.

[a] $R_{work,free} = \sum \frac{\left\| |F_{obs}| - |F_{calc}| \right\|}{|F_{obs}|}$ working and free *R* factors were calculated using the working and free reflection sets, respectively. The free reflections were held aside throughout refinement [2].

[b] Real-space correlation of structural model and experimental data was analyzed and coordinate uncertainties were determined using the SFCHECK software [15].

disordered. From the first nucleotide of the DNA only the phosphate group is ordered in the electron density. A double conformation is observed for the side chains of Glu42 and Asn62 as well as for parts of the sugar rings of the last two nucleotides and the adjacent sugarphosphate backbone of the ligand. The overall CSP structure is not significantly altered upon ligand binding: The superposition of structural models from the *Bs*-CspB·dT_6 complex and a free *Bs*-CspB (1CSP) structure shows a root mean square deviation (RMSD) of 0.88 Å for all C^{α} atoms. The strongest deviation between the two structures involves the C-terminal part of β4 and the subsequent loop L_{45}, which is reoriented by up to 4 Å (Figure 3.5a). This deviation is not observed between the *Bs*-CspB·dT_6 complex structure and a structure of free Csp from *Bacillus caldolyticus* (1C9O), however. In all four crystal structures the sidechain of Arg56, a central residue of loop L_{45}, forms different intermolecular contacts with symmetry-related molecules. Thus, the orientation of this loop is expected to be more strongly influenced by the molecular packing than by ligand binding. Minor deviations are observed for loops L_{23} and L_{34}. Most sidechains in dT_6-complexed *Bs*-CspB align quite well with their equivalents in the free protein (1CSP), with the exception of Trp8, whose orientation strongly deviates from

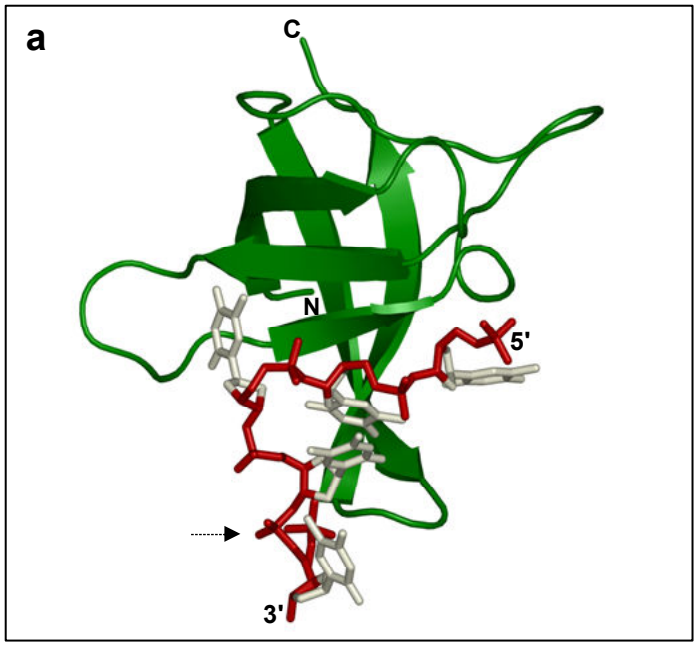
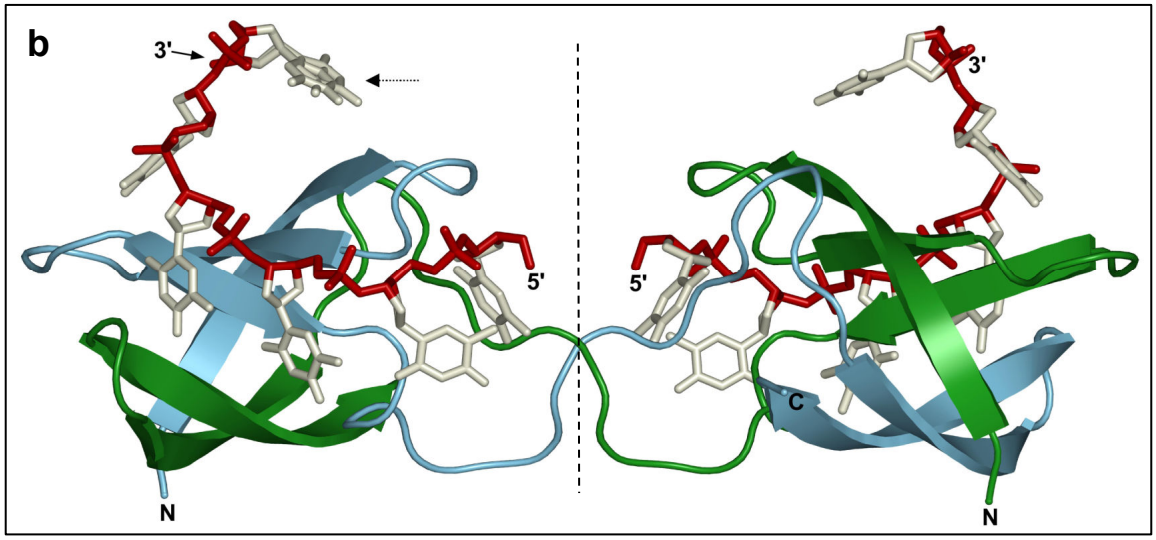


Figure 3.5: Unit cell content of the *Bs*-CspB·dT_6 and *Bc*-Csp·dT_6 crystal structures. The protein chains are shown as cartoons. DNA molecules are depicted as bonds connecting non-hydrogen atoms. The DNA backbone is colored red, the sugar-base moieties are colored gray.
a) *Bs*-CspB·dT_6; the phosphodiester group connecting the terminal nucleotides on the 3' end adopts a double conformation (dashed arrow).
b) *Bc*-Csp·dT_6 (protein chain A - green, protein chain B - blue); in ligand chain D a double conformation is observed for the 3'-terminal nucleobase (dashed arrow). A non-crystallographic twofold axis which refers the two swapped monomers to each other is depicted as a dashed line.



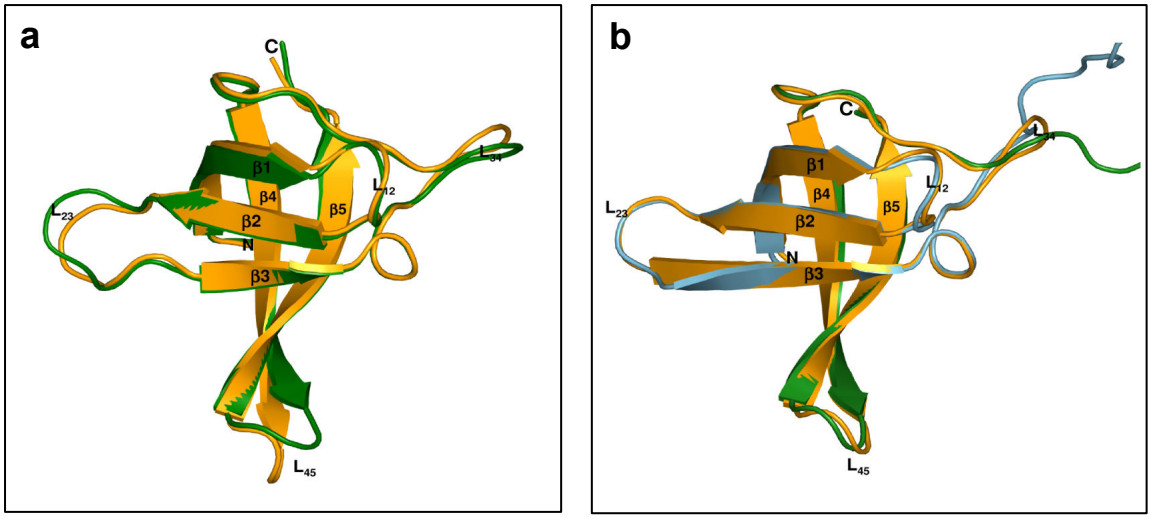


Figure 3.5: Superposition of ligand-free and dT_6-complexed CSP. The layout of protein structures follows that of Figure 3.5. Ligand-free CSP structures (1CSP & 1C9O) are depicted in orange. The superposition is based on α-carbon positions. a) *Bs*-CspB: 2ES2 (complex with dT_6) *vs.* 1CSP (no ligand), RMSD = 0.88 Å; b) *Bc*-Csp, 2HAX (complex with dT_6) chain A - green, chain B - blue, *vs.* 1C9O (no ligand), RMSD = 0.45Å. For calculation of RMSD values, equivalent α-carbon positions were used.

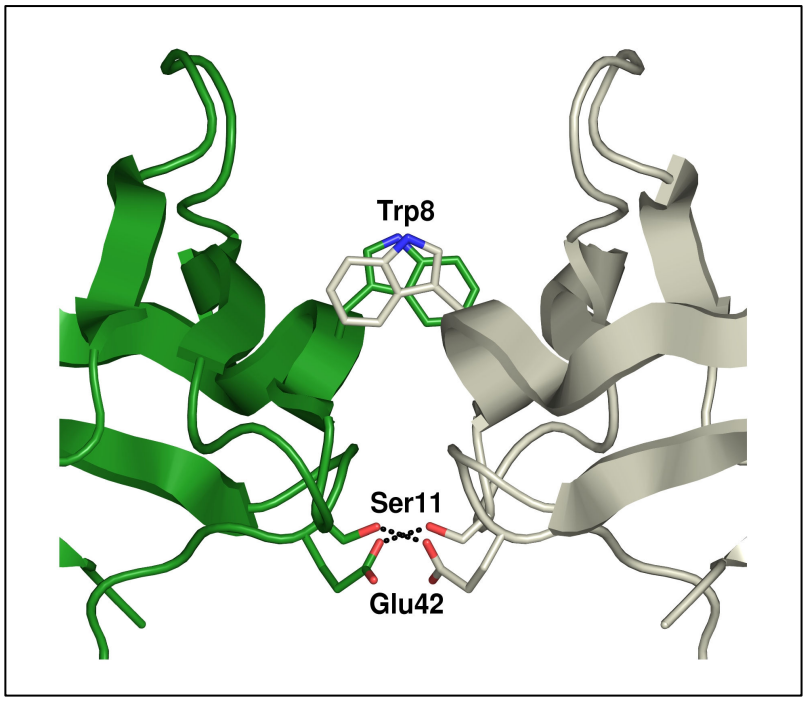


Figure 3.6: Trp8 forms a hydrophobic contact with its equivalent from a symmetry-related molecule in the *Bs*-CspB·dT_6 structure. This contact causes a reorientation of Trp8, which is not observed in any other CSP crystal structure (Figure 3.18). In addition, two hydrogen bonds (dashed lines) between Ser11 and Glu42 of the two symmetry-related molecules are formed.

that of all other *Bacillus* CSP structures reported so far. Its reorientation is associated with a hydrophobic interaction with its equivalent from a symmetry-related protein molecule (Figure 3.6). No significant structural rearrangement of the protein backbone or sidechain groups could be causally linked to ligand binding.

3.1.2 Crystallization and structure solution of a *Bc*-Csp·dT_6 complex (1)

Bc-Csp, the major cold shock protein from *Bacillus caldolyticus*, was crystallized in complex with hexathymidine (dT_6) in space group $P2_12_12$ (Figure 4.1), and diffraction data were collected up to 1.29 Å (Figure 4.2). Initial phases were obtained by molecular replacement using a crystal structure of *Bc*-Csp without a bound ligand (1C9O). The crystal's asymmetric unit contains a swapped dimer of *Bc*-Csp (Figure 3.5b, chains A and B) in contact with two DNA molecules (chains C and D). The structure was refined using REFMAC5 [121] (Table 4.1) to final R_{work} / R_{free} of 13.0% and 16.2%. 95.3% and 4.7% of all residues are in the allowed and additionally allowed regions of the Ramachandran plot. The electron density is well defined; all non-hydrogen atoms from protein and ligand molecules could be placed. Alternative sidechain conformations were built for Arg3, Glu36 and Gln46 in both protein chains, Gln19 and Gln39 in chain A, Glu12, Lys13, and Gln53 in chain B, and the nucleobase of T6 in chain D.

In the *Bc*-Csp·dT_6 crystal structure, two protein chains form a swapped dimer with two glo-

bular functional units, which are composed of residues 1 to 35 from one protein chain, and residues 38 to 65 from another protein chain (Figure 3.5b). The architecture these units closely resembles that of all other structural models of CSP, which feature five highly curved antiparallel β-strands, connected by four loops and a short 3_{10} helix at the C-terminus of β3. In the non-swapped (closed monomeric) structures the respective β-strands 1-3 and 4+5 are arranged as two β-sheets which form a closed β-barrel. In the domain-swapped structure the first β-sheet of one chain assembles with a second β-sheet from a different chain and *vice versa*. The swap occurs in the middle of loop L_{34} and is promoted by a unique combination of torsion angles Glu36 ψ and Gly37 φ as compared to closed monomers. The functional units align well with two structural models of the closed protein (1C9O), giving RMSD values of less than 0.50 Å for equivalent α-carbon atoms (Figure 3.5b). Both functional units superimpose with a RMSD of 0.11 Å, and the phosphorus atoms of the sugarphosphate backbone from both DNA chains superimpose with a RMSD of 0.23 Å. While this chapter focuses on the ability of the CSP to bind single-stranded nucleic acids, the domain-swapped architecture and its consequences will be analyzed and discussed in Chapter 4.

3.1.3 The ligand binding sites of *Bs*-CspB and *Bc*-Csp

The globular domains of *Bs*-CspB and the functional units of *Bc*-Csp have a strongly dipolar surface: One side has a prominent negative surface potential, which is derived from acidic sidechains. On the opposite side several aromatic and hydrophobic solvent-exposed sidechains form a hydrophobic platform, which is surrounded by basic and polar groups (Figure 3.7). The platform starts at Trp8 of strand β1, proceeds along Phe15, Phe17, Val26, Phe27, Val28, and His29, all anchored in β2 and β3, and finally ends inside loop L_{34}, close to Phe30 and Phe38 (Figure 3.7, Figure 3.15), which are separated from each other by a shallow polar cleft. Many of these residues are located within the RNP1 (ribonucleoprotein motif 1) (Lys13 - Val20) and RNP2 (ribonucleoprotein motif 2) (Val26 - Phe30) motifs, which are conserved in various RNA binding proteins [81-83]. Following the molecular orientation of Figure 3.7 and Figure 3.15, the upper side of the platform is defined by a small protrusion, formed by residues from the termini of strands β1 and β2 with their connecting loop L_{12}. On the opposite side the platform is bordered by a larger protrusion of amino acids from β4 and β5 and their connecting turn (L_{45}). The platform is surrounded by polar groups from the protein backbone and the sidechains of Lys7, Lys13, Asp25, His29, and Arg56, resulting in a predominantly positive surface potential around the platform (Figure 3.7, Figure 3.15).

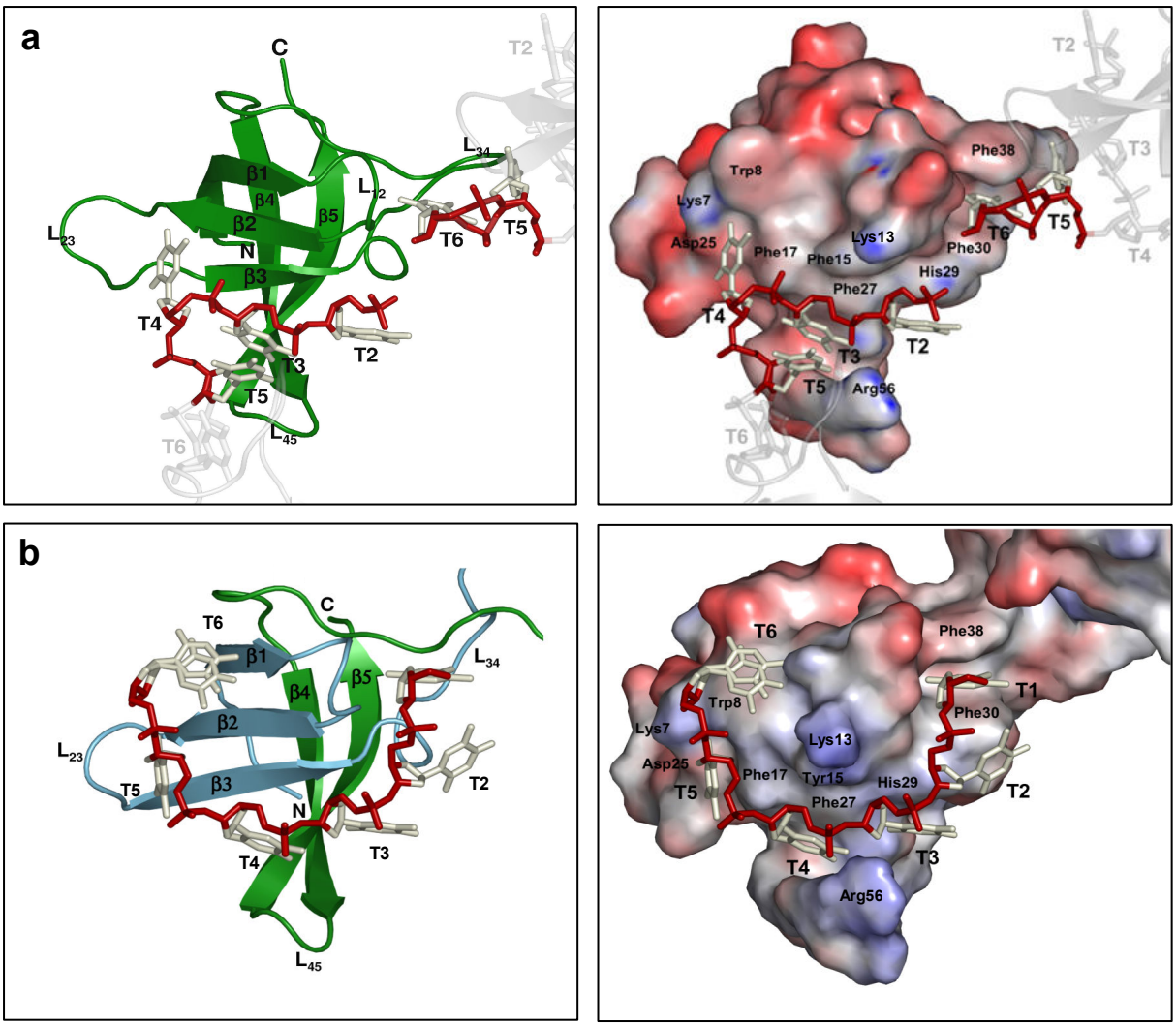


Figure 3.7: The ligand nucleobases interact with an extended hydrophobic platform and surrounding polar groups on the CSP surface. Left: Schematic overview of protein-ligand assemblies according to Figure 3.5. Right: Electrostatic surface potential ranging from -10 *kT* (red) to +10 *kT* (blue) mapped onto the protein contact surface. The potential was calculated for pH 7.0 using APBS [10]. a) *Bs*-CspB·dT_6, b) *Bc*-Csp·dT_6.

3.1.4 Protein-ligand interactions in CSP·dT_6 structures

Although the structures of the dT_6 complex share large parts of the CSP binding interface, their arrangements of DNA and protein molecules differ considerably. As will be discussed later, these differences are expected to be caused by different crystal packing environments favored by the low complexity of the dT_6 oligonucleotide ligand. In order to allow a common evaluation of the ligand binding for both structures, references to individual nucleobases are parenthesized and specified like this: (reference to the *Bs*-CspB·dT_6 structure / reference to *Bc*-Csp·dT_6 structure).

DNA and protein molecules bind to each other through hydrophobic interactions and hydrogen bonds (Figure 3.7, Figure 3.15, Figure 3.18). Hydrophobic interactions mainly involve

residues Phe17, Phe27, His29, and Phe30 / Phe38, which stack with nucleobases. A stack between Trp8 and a nucleobase is only observed in the *Bc*-Csp·dT_6 complex. In both structures individual stacks add up to form tetrastacks, which involve successive sidechain and nucleobase groups from Phe38, a thymine (T6 / T1), Phe30, His29, and another thymine (T2 / T3). In the *Bs*-CspB·dT_6 structure an edge-on contact is formed between Phe38 and the nucleobase of T5. A different edge-on contact is formed between Phe30 and T2 in the *Bc*-Csp·dT_6 structure. Apart from these directed hydrophobic stacking interactions the shielding of further hydrophobic groups (Phe15/Tyr15 and Val26, Val28) of the ligand binding interface from the solvent upon ligand binding is expected to stabilize the complexes, because their solvent-exposed location is expected to be energetically unfavorable.

In both crystal structures a common set of 7 hydrogen bonds between polar protein and nucleobase groups were identified, which involve sidechains from Lys7, Trp8, Asp25 Gln59 and the backbone of Lys39 (Figure 3.15). Few additional hydrogen bonds addressing nucleobase groups, single hydrogen bonds addressing the ligand's sugarphosphate backbone, and some water-mediated interactions between protein and ligand groups differ somewhat between the two complex structures. All protein residues involved in hydrophobic and polar ligand interactions in the two crystal structures also show up in NMR-based analyses of *Bs*-CspB and dT_7, either by showing extreme line-broadening in the presence of low ligand concentrations, and / or by standing out from the mean weighted change in chemical shifts induced by ligand binding [132]. This suggests that in solution the same interactions contribute to ligand binding of thymine bases as observed in the crystal structures. Interestingly, in both complex structures interactions between protein and thymine methyl-groups, which would discriminate this nucleobase from its RNA equivalent uracil, are not observed (Figure 3.15). Likewise, the charged and polar groups from the sugarphosphate backbone of the ssDNA, which enclose the hydrophobic nucleobases and the hydrophobic binding platform of the protein below them, are exposed to the solvent and not involved in extensive protein contacts. 2'-hydroxyls, attached with standard stereochemistry to the deoxyribose $C^{2'}$ atoms of the bound dT_6, would be solvent-exposed, potentially allowing ssRNA oligonucleotides to bind to *Bs*-CspB in the same way as ssDNA.

The size of the solvent accessible surface of *Bs*-CspB occupied by the ligand is 663 $Å^2$, or 15.0% of the total solvent accessible surface. In the *Bc*-Csp·dT_6 structure the solvent-accessible surface is reduced by 696 A^2 or 16.2% of the total solvent-accessible surface per functional unit upon ligand binding. In both structures about two thirds of this area is formed by hydrophobic groups.

In the *Bs*-CspB·dT_6 crystal, ligand and protein molecules form a continuum

In the *Bs*-CspB·dT_6 structure nucleotides 2 to 6 were completely built into the electron density; of the first nucleotide only the 3'-phosphate group could be fitted (Figure 3.8a). The DNA ligand binds to two protein molecules, which are related by crystal symmetry (Figure 3.7a, Figure 3.9, Figure 3.15a, c). The bases of T2, T3, and T4 make contact with one protein molecule, T5 bridges between two protein molecules, and T6 is bound to the next protein molecule, which makes contact with another DNA molecule by interacting with its nucleobases T2, T3, T4, and T5. In this way, protein and DNA form a continuous arrangement inside the crystal. The 5'-to-3'-polarity of the ligand is the same as described for most other OB-fold proteins in complex with nucleic acids [133]: The 5'-end of the oligonucleotide is in the vicinity of L_{12}, the DNA strand proceeds *via* β1

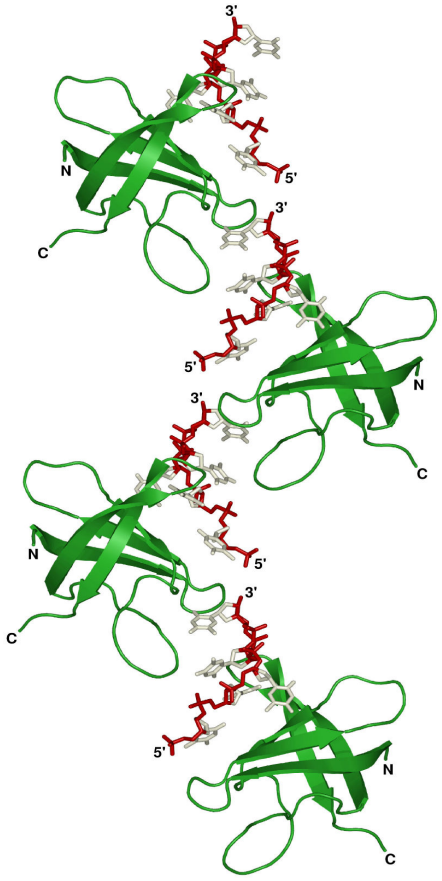


Figure 3.9: A continuous arrangement is formed by *Bs*-CspB (green) and dT_6 molecules (backbone red, bases beige).

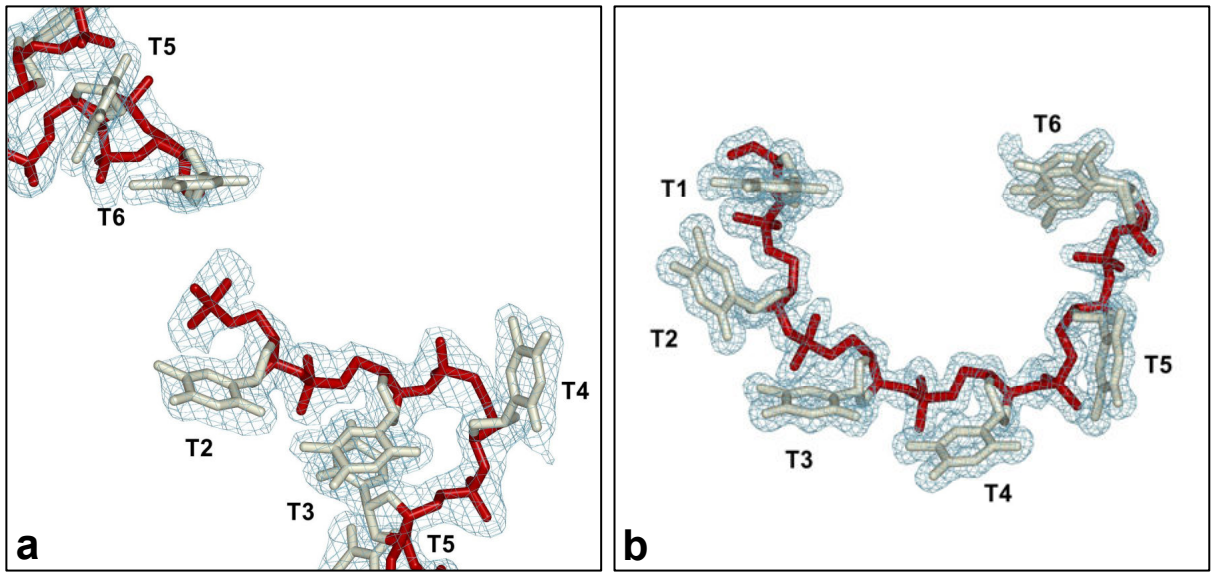


Figure 3.8: DNA single strands adopt irregular conformations upon binding to CSP. Ligand molecules are surrounded by their $2F_o$-F_c difference electron density (mesh). All nucleobases are unstacked with respect to each other. The nucleotides are in *anti*-conformation. a) *Bs*-CspB·dT_6 complex; the difference density is contoured at 1.0 σ. b) *Bc*-CspBdT_6 complex; the difference density is contoured at 1.2 σ.

and β2, follows the N-C polarity of β2, and points towards the kink within β1. The DNA is bound to the protein in an irregular extended conformation. With respect to each other all nucleobases are unstacked and oriented towards the hydrophobic protein surface, the nucleosides are in *anti* conformation, and the sugarphosphate backbone is exposed to the solvent. At nucleotide T4, prior to the transition of the DNA from one protein molecule to the other, the sugar phosphate backbone is bent almost at a right angle and protrudes towards the consecutive protein molecule. The sugar rings in nucleotides T2, T3, T4 and T5 pucker in the energetically favorable $C^{2'}$-*endo* conformation (typical for double-stranded B-DNA) while T5 displays $C^{4'}$-*exo* pucker of the pentose pseudo-rotation cycle [134]. $C^{2'}$-*endo* puckers are not possible in double-helical RNA, due to steric clashes resulting from the 2'OH groups. The exocyclic angles of the sugarphosphate backbone in the *Bs*-CspB·dT_6 crystal, however, differ significantly from those of double-stranded DNA and agree with the limitations of backbone torsion angles of tRNA [135]. The sugar puckers of the two conformers of T6 occupy energetically disfavored $O^{1'}$-*endo* and $C^{1'}$-*endo* states. Their high energy states are expected to be a consequence of molecular packing in the crystal, as these two nucleotides are located in the area of transition between two protein molecules.

In the *Bc*-Csp·dT_6 crystal, hexathymidines adopt curved conformations

In the *Bc*-Csp·dT_6 structure the two ligand molecules adopt very similar conformations. Looking from the protein surface towards the ligand, their sugarphosphate backbone appears curved like a "C" character (Figure 3.7b, Figure 3.8b, Figure 3.15b), its nucleobases point towards the protein surface. As observed in the *Bs*-CspB·dT_6 structure, no stacking occurs between nucleobases and all nucleosides are in *anti* conformation. The sugar of T1 from chain C maintains a $C^{4'}$-*exo* pucker while the remaining sugars adopt $C^{2'}$-*endo* puckers, typical for double-stranded B-DNA. In ligand chain D the terminal nucleotides adopt $C^{3'}$-*endo* conformation, which is typical for the double-stranded A-DNA and RNA, whereas all other nucleotides display $C^{2'}$-*endo* puckers. All sugar puckers observed in the *Bc*-Csp·dT_6 structure are inside energetically favorable regions of the pentose pseudo-rotation cycle [134], and the exocyclic angles of the sugarphosphate backbone are within limits observed in tRNA structures [135].

3.1.5 Contributions of individual sidechains in *Bs*-CspB to ligand binding

In order to analyze the contributions of individual sidechains to the binding of thymine-rich oligonucleotides, mutant *Bs*-CspB protein variants were generated (K7A, D25A, R56A, Q59A). Because dT_7 is bound with a ~200-fold higher K_D value than dT_6 [132], and heptanucleotides were suspected to completely occupy the CSP binding site [63], it was used for the determination of affinity data of the *Bs*-CspB mutants). K_D values were determined by fluorescence titration using Trp8, which is part of the binding site in both CSP complexes, as a fluorophore. Its fluorescence signal is quenched by more than 80% upon ligand binding (see Figure 3.11). In order to exclude the possibility that the mutant variants are significantly destabilized under conditions used for studying their interactions, which would result in erroneous K_D data, melting curves of all mutants produced in this work were recorded and melting temperatures were determined by CD spectroscopy (Figure 3.10).

Affinity and stability data for all other mutants (F17A, F27A, F30A, F38A, H29Q) and for *Bs*-CspB were generously provided by Markus Zeeb [84, 132] who used the same experimental conditions. Although the stability of some mutants is significantly reduced, their melting temperatures are at least 21 °C (Table 3.2) above the temperature used for interaction studies (15 °C). Their melting curves suggest that the mutants are well folded at this temperature (Figure 3.10). Substitutions of Phe17, Phe27, Phe30 and Phe38 by alanines and the substitution of His29 by glutamine resulted in 55- to 190-fold increases in K_D (Table 3.2). All these sidechains are involved in nucleobase stacking interactions (Figure 3.15). The replacement of

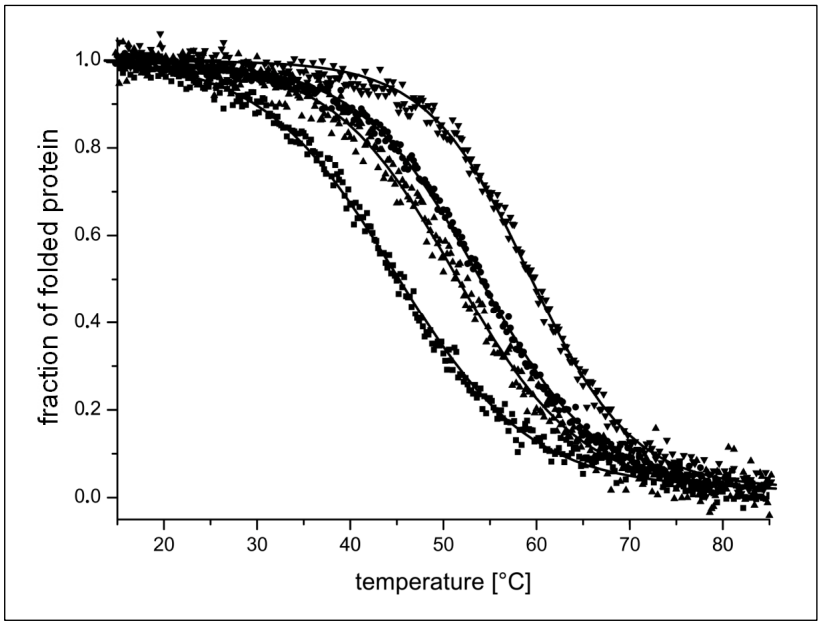


Figure 3.10: Melting curves of *Bs*-CspB mutants determined by CD spectroscopy (Chapter 2.3.3). θ_λ was determined at 222.6 nm in 50 mM cacodylate, 100 mM KCl, pH 7.0. The protein concentration was 12 µM.
Melting temperatures were determined according to Schindler & Perl [16, 17] (Chapter 2.3.2).

■ - K7A, T_M = 44.8 °C ± 0.1 °C,
▲ - D25A, T_M = 51.3 °C ± 0.1 °C,
● - Q59A, T_M = 53.9 °C ± 0.1 °C,
▼ - R56A, T_M = 59.2°C ± 0.1 °C.

Phe15, which does not stack with a nucleotide but is shielded by the sugarphosphate backbone connecting T2 and T3 in the complex structure, causes a 75-fold increase in K_D. When the polar residues of Lys7, Asp25, Arg56 and Gln59, which are all involved in direct polar interactions with the ligand, were replaced by alanines, only slight increases in K_D were observed (no increase to 12-fold increase in K_D) (Table 3.2). The replacement of Lys13, whose terminal amino group is about 6 Å distant from the ligand, by glutamine resulted in a 14-fold increase in K_D, suggesting that this sidechain, which is covered from the ligand backbone by a symmetry-related molecule in both complex structures, could be involved in charge interactions with the ligand backbone in solution.

The strong affinity decreases involving hydrophobic replacements and the considerably smaller decreases when polar residues were replaced agree with further experimental results [69], which demonstrate that complexes of *Bs*-CspB and thymine-rich oligonucleotides cannot be dissociated by high salt concentrations. Their mode of interaction thus appears to be based on hydrophobic rather than of polar interactions. This is in good agreement with the binding mode observed in the complex crystal structures.

Construct	K_D (nM)	T_M (°C)
wt *Bs*-CspB[a]	1.8 ± 0.4	53.6 ± 0.1
F15A[a]	135 ± 4	36.2 ± 0.1
F17A[a]	345 ± 63	41.2 ± 0.3
F27A[a]	286 ± 7	47.9 ± 0.6
H29Q[a]	104 ± 6	48.7 ± 0.3
F30A[a]	208 ± 11	54.5 ± 0.1

Construct	K_D (nM)	T_M (°C)
F38A[a]	228 ± 9	56.8 ± 0.1
K7A	6.4 ± 0.6	44.8 ± 0.1
K13Q[a]	25.0 ± 3.0	55.3 ± 0.2
D25A	21.4 ± 0.5	51.3 ± 0.1
R56A	11.8 ± 1.4	59.2 ± 0.1
Q59A	2.7 ± 0.43	53.9 ± 0.1

Table 3.2: Equilibrium Dissociation constants (K_D) of *Bs*-CspB and mutant variants determined by fluorescence titrations with dT_7 (Chapter 2.3.3). *Bs*-CspB and mutant protein variants were titrated with increasing concentrations of dT_7 in a buffer containing 50 mM cacodylate / HCl, 100 mM KCl, pH 7.0, supplemented with 40 µM N-acetyl-tryptophanamide. The fluorescence signal of Trp8 recorded at 343 nm was used as a probe. The protein concentration was 25 nM for all K_D values determined, here. [a] data was kindly provided by Markus Zeeb [132, 136].

3.1.6 Preferential binding of heptanucleotides by CSP

Dissociation equilibrium constants (K_D) of various *Bs*-CspB- and *Bc*-Csp·heptanucleotide complexes were determined by further fluorescence titration experiments, in order to analyze the sequence-specific binding of ssDNA ligands by *Bs*-CspB and *Bc*-Csp. In addition, kinetic rate constants (k_{on}, k_{off}) describing complex formation and dissociation of some *Bs*-CspB·heptanucleotide complexes was determined by stopped-flow techniques.

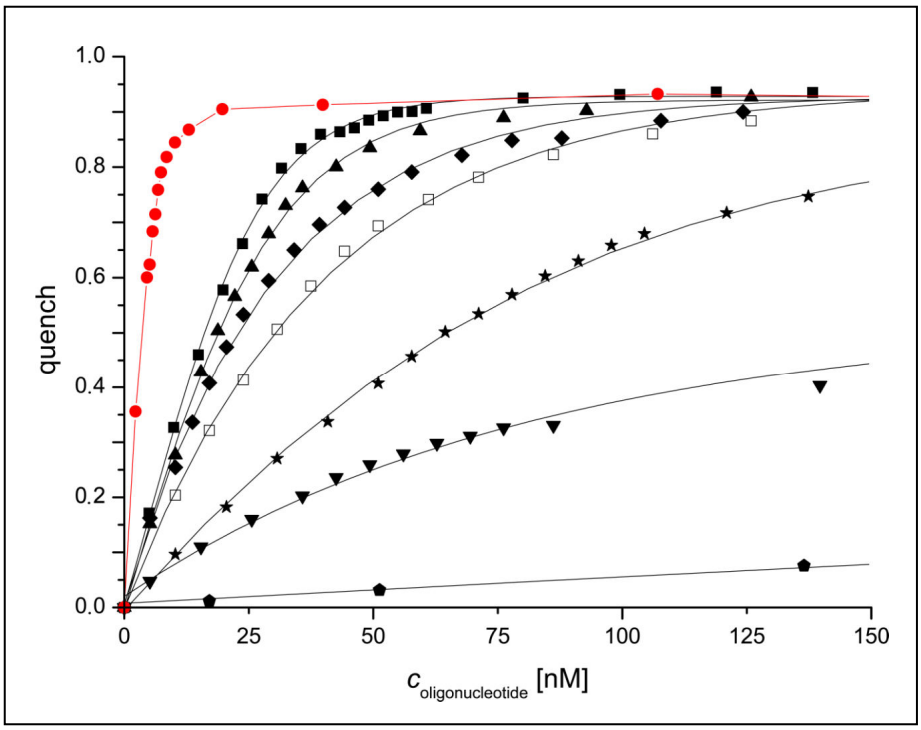


Figure 3.11: Determination of thermodynamic equilibrium constants of *Bs*-CspB·oligonucleotide complexes using fluorescence titrations. K_D values were determined according to Eftink and Lohman [12, 13] (Chapter 2.3.3)**.** Curves from representative experiments are shown, only. Experimental conditions and affinity data of all complexes from this figure are given in Table 3.3.
● - O1 (TTCTTTT), K_D = 0.2 nM ± 0.07 nM ; ■ - dT_7 (TTTTTTT), K_D = 1.8 nM ± 0.4 nM; ▲ - CT3 (CTCTTTC), K_D = 3.9 nM ± 0.2 nM; ◆ - CT1 (CTTTTTT), K_D = 5.9 nM ± 0.3 nM; □ - CT4 (CTCTCTC), K_D = 10.8 nM ± 0.9 nM; ★ - CT2 (CTTTTTC), K_D = 33.7nM ± 4.1 nM; ▼ - CT5 (CTCTTCC), K_D = 66.2 nM ± 4.8 nM; ⬟ - (CCCTTTC) CT7, K_D = 1808 nM ± 117 nM;

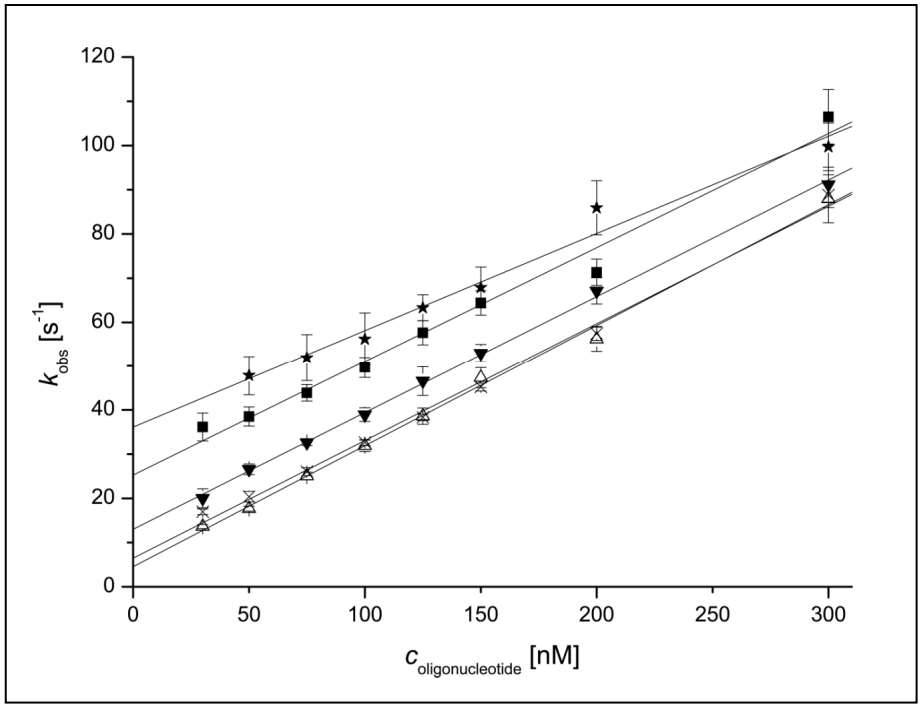


Figure 3.12: Determination of association kinetics of *Bs*-CspB·oligonucleotide complexes by stopped-flow analyses**.** Data were measured at least six-fold and averaged. Association and dissociation rate constants (k_{on}, k_{off}) were determined according to Bernasconi [1] (Chapter 2.3.4). Results and experimental conditions are given in Table 3.3. ■ - CT2, △ - CT3, × - CT4, ★ - CT5, ▼ - CT6.

Ligand	Sequence	CSP : ssDNA	K_D 10^{-9} (M)	k_{on} 10^8 ($M^{-1}s^{-1}$)	k_{off} (s^{-1})[a]	k_{off} (s^{-1})[b]
dT_7	TTTTTTT	1:1	1.8 ± 0.4	3.37 ± 0.11	1.7 ± 0.6	0.6 ± 0.6
CT1	CTTTTTT	1:1	5.9 ± 0.3	n.d.	n.d.	n.d.
CT2	CTTTTTC	1:1	33.7 ± 4.1	2.58 ± 0.01	25.3 ± 2.1	8.7±1.0
CT3	CTCTTTC	1:1	3.9 ± 0.2	2.74 ± 0.01	4.6 ± 1.1	1.1±0.1
CT4	CTCTCTC	1:1	10.8 ± 0.9	2.66 ± 0.01	6.5 ± 1.2	2.9±0.2
CT5	CTCTTCC	1:1	66.2 ± 4.8	2.20 ± 0.02	36.1±2.4	14.6±1.1
CT6	CTCCTTC	1:1	12.5 ± 0.6	2.64 ± 0.01	13.0±0.6	3.3±0.2
CT7	CCCTTTC	1:1	1808 ± 117	n.d.	n.d.	n.d.
CT9	TTTTTTC	1:1	5.5 ± 0.4	n.d.	n.d.	n.d.
dC_7	CCCCCCC	1:1	> 4000	n.d.	n.d.	n.d.
O1	TTCTTTT	1:1	0.2 ± 0.07	n.d.	n.d.	n.d.
CT10	CTCATTC	1:1	46. 8 ± 2.1	n.d.	n.d.	n.d.
CT12	CTCTATC	1:1	47.0 ± 0.7	n.d.	n.d.	n.d.
CT16	ATCTTTC	1:1	137 ± 16	n.d.	n.d.	n.d.
CT17	CACTTTC	1:1	256 ± 25	n.d.	n.d.	n.d.
CT18	CTATTTC	1:1	63.0 ± 2.9	n.d.	n.d.	n.d.
CT19	CTCTTAC	1:1	88.6 ± 5.7	n.d.	n.d.	n.d.
CT20	CTCTTTA	1:1	7.9 ± 2.3	n.d.	n.d.	n.d.

Table 3.3: Equilibrium dissociation constants (K_D) and kinetic association (k_{on}) and dissociation (k_{off}) rate constants of *Bs*-CspB·heptapyrimidine complexes (Chapter 2.3.3). All K_D values were determined by fluorescence titration of *Bs*-CspB with increasing concentrations of oligonucleotides at 15 °C in a buffer containing 50 mM cacodylate / HCl, 100 mM KCl, pH 7.0, supplemented with 40 µM N-acetyl-tryptophanamide. The fluorescence signal of Trp8 recorded at 343 nm was used as a probe. The protein concentration was 25 nM except for O1 (10 nM) and CT7 / dC_7 (100 nM). See Figure 3.11 for titration curves of representatives.
k_{on} and k_{off} rates were determined by measuring the decay of the Trp8 fluorescence signal using a stopped-flow device. [a] the dissociation rate k_{off} was determined from the offset of the linear fit of experimentally observed rate $k_{obs} = k_{on} \cdot [Bs\text{-CspB}] + k_{off}$ at various ssDNA concentrations. Plots are shown in Figure 3.12. The buffer and temperature conditions were identical as for determination of Equilibrium dissociation constants. The protein concentration was 30nM. [b] The dissociation rate k_{off} was calculated by $k_{off} = k_{on} \cdot K_D$. n.d. – not determined.

Oligonucleotide	Sequence	CSP : ssDNA	K_D (nM)
dT_7	TTTTTTT	1:1	0.9 ± 0.2
CT1	CTTTTTT	1:1	2.8 ± 0.9
CT2	CTTTTTC	1:1	8.3 ± 0.2
CT3	CTCTTTC	1:1	3.3 ± 0.2
CT4	CTCTCTC	1:1	5.2 ± 0.2
CT5	CTCTTCC	1:1	36.4 ± 3.4
CT6	CTCCTTC	1:1	3.5 ± 0.3
CT7	CCCTTTC	1:1	307 ± 33
CT9	TTTTTTC	1:1	1.5 ± 0.2
dC_7	CCCCCCC	1:1	> 4000

Table 3.4: Dissociation constants (K_D) of *Bc*-Csp·heptapyrimidine complexes. All K_D values were determined by fluorescence titration of *Bc*-Csp with increasing concentrations of oligonucleotides at 15 °C (Chapter 2.3.3). The same experimental conditions were used as for *Bs*-CspB (Table 3.3). The protein concentration was 25 nM except for CT7 and dC_7. Here the protein concentration was 100 nM.

For analyzing the preferential binding of pyrimidines, a set of heptanucleotides was used, whose sequences were altered gradually and progressively, starting with dT_7, then replacing the terminal thymines (oligonucleotides CT1 and CT9 in Table 3.3 and Table 3.4), followed by the thymine at position 3 (oligonucleotide CT3 in Table 3.3 and Table 3.4). This CTCTTTC scaffold, which turned out to bind with almost the same affinity as the dT_7 sequence, was then used to determine the preferences of the remaining sequence positions (oligonucleotides CT4 – CT7 in Table 3.3 and Table 3.4). By comparing K_D values of two ligands whose sequences differ only at the position of interest (Table 3.5), the relative preference for *Bs*-CspB for T or C at each position within a heptanucleotide was determined. These comparisons reveal small preferences for T over C at positions 1, 4, 5 and 7, associated with an approximately 3-fold increase in K_D as compared to dT_7. At positions 2 and 6 the K_D increase was 450 and 17-fold, respectively, when thymine was replaced by cytosine. In contrast, at position 3 cytosine was preferred over thymine as shown by a 9-fold decrease in K_D. The individual preferences at all 7 positions were combined to predict a new high affinity ligand (TTCTTTT, O1 in Table 3.3), which binds to *Bs*-CspB with a subnanomolar K_D. This is the highest affinity reported so far.

All oligonucleotides analyzed by stopped-flow techniques show similar association rate constants (k_{on}), which are in the range of 2.2 - 3.37 · 10^8 $M^{-1}s^{-1}$ (Table 3.3, Figure 3.12). In contrast, experimentally determined dissociation rate constants (k_{off}) show stronger variations and are correlated with experimentally determined K_D equilibrium constants (dT_7 - K_D = 1.8 ± 0.4 nM, k_{off} = 1.7 ± 0.6 · 10^{-8} $M^{-1}s^{-1}$; CT5 - K_D = 66.2 ± 4.8 nM, k_{off} = 36.1±2.4 · 10^{-8} $M^{-1}s^{-1}$,). This suggests, that the affinity of the *Bs*-CspB·oligonucleotide complexes depends only on the dissociation rates of the complexes.

pos.	oligo 1	oligo 2	***Bs*-CspB** K_D2/K_D1[a]	***Bc*-Csp** K_D2/K_D1[a]
1	dT_7 (TTTTTTT)	CT1 (CTTTTTT)	3.3 ± 0.6	3.1 ± 0.3
2	CT3 (CTCTTTC)	CT7 (CCCTTTC)	464 ± 6.2	93 ± 4.4
3 (a)	CT2 (CTTTTTC)	CT3 (CTCTTTC)	0.12 ± 0.01	0.40 ± 0.01
3 (b)	dT_7 (TTTTTTT)	O1 (TTCTTTT)	0.11 ± 0.01	n.d.
4	CT3 (CTCTTTC)	CT6 (CTCCTTC)	3.2 ± 0.1	1.1 ± 0.1
5	CT3 (CTCTTTC)	CT4 (CTCTCTC)	2.8 ± 0.1	1.6 ± 0.1
6	CT3 (CTCTTTC)	CT5 (CTCTTCC)	17 ± 0.4	11 ± 0.4
7	dT_7 (TTTTTTT)	CT9 (TTTTTTC)	3.0 ± 0.5	1.7± 0.1

Table 3.5: Binding preferences of *Bs*-CspB and *Bc*-Csp for heptanucleotides containing thymine and cytosine at individual positions. Relative preferences [(a)] for thymines / cytosines were determined, by comparing K_D values of two different CSP·heptapyrimidine complexes (see Table 3.3 & Table 3.4) whose the ligand sequences differed only at the position of interest (underlined). Their errors were estimated by Gaussian progression of mean errors. n.d. - not determined.

Due to limitations in the protein concentration range which could be studied in the stopped-flow device, k_{off} values could not be determined experimentally with high precision. To reduce the error, k_{off} rates were also calculated from k_{on} rates and K_D. Despite some differences in absolute values between both sets, they follow the same trend.

Fluorescence titrations involving *Bc*-Csp and the set of oligonucleotides used for studying the preferential binding of heptapyrimidines to *Bs*-CspB revealed small preferences for T over C associated with up to 3-fold increases in K_D for nucleobase positions 1, 5 and 7 (Table 3.4, Figure 3.7). At positions 2 (CT3, CT7) and 6 (CT3, CT5) the decrease in affinity was significantly stronger: When C was introduced here, the K_D increased 93- and 11-fold, respectively. At position 3 (CT2, CT3) C was slightly preferred over T, associated with a 2.5 fold decrease in K_D.

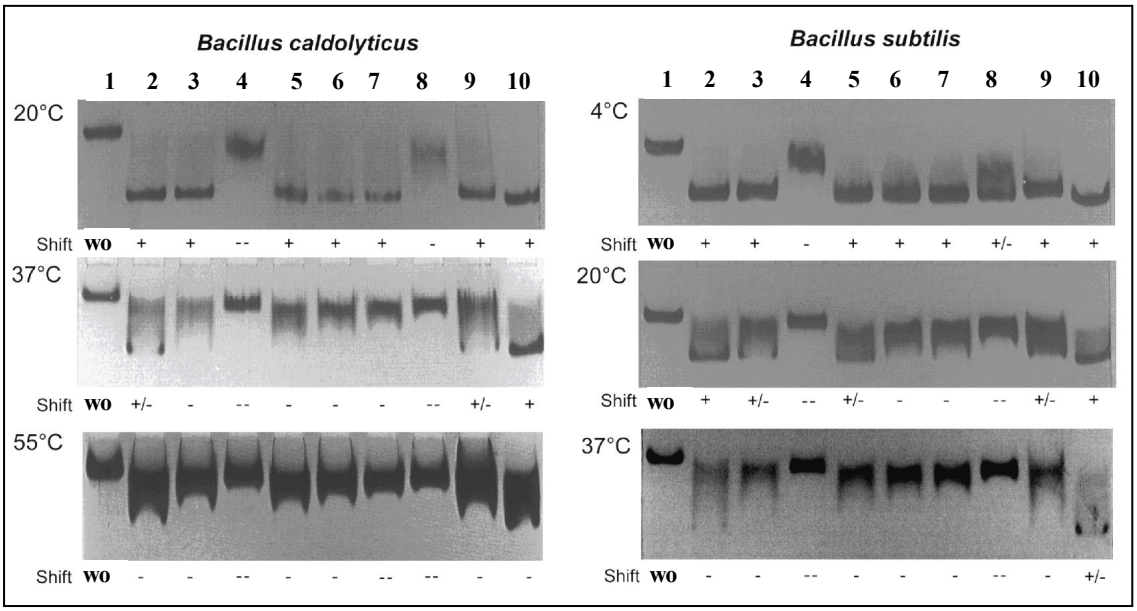
In addition to the preferential binding of pyrimidines, the effect of introducing purine nucleobases at individual sequence positions was analyzed for *Bs*-CspB, using the CTCTTTC sequence as a reference and another set of heptanucleotides, where individual pyrimidines were replaced by adenines (oligonucleotides CT3, CT10, CT12, CT16 – CT20 in Table 3.3 & Table 3.6). The presence of an adenine at sequence position 2 results in a 56-fold increase in K_D with respect to thymine. Although the increase in K_D (12.1 to 35-fold) is not as prominent at most other positions as observed for sequence position 2, all these sites show a significant preference for pyrimidines over adenine. Only for sequence position 7 the introduction of an adenine results in a small decrease in affinity, as indicated by a twofold increase in K_D.

In addition, the preferential binding of oligonucleotides by the two cold shock proteins and the effect of temperature on the stringency of binding were analyzed in qualitative terms. Therefore, CSP·heptapyrimidine complexes, previously used to determine positional binding preferences by fluorescence spectroscopy, were subjected to analytical gel electrophoreses at different temperatures (Figure 3.13).

Pos.	oligo 1	oligo 2	K_D2/K_D1 [a]
1	CT3 (CTCTTTC)	CT16 (ATCTTTC)	35 ± 2.3
2	CT3 (CTCTTTC)	CT17 (CACTTTC)	56 ± 3.0
3	CT3 (CTCTTTC)	CT18 (CTATTTC)	16 ± 0.1
4	CT3 (CTCTTTC)	CT10 (CTCATTC)	12.0 ± 0.1
5	CT3 (CTCTTTC)	CT12 (CTCTATC)	12.1 ± 0.4
6	CT3 (CTCTTTC)	CT19 (CTCTTAC)	23 ± 0.3
7	CT3 (CTCTTTC)	CT20 (CTCTTTA)	2.0 ± 0.5

Table 3.6: Binding preferences of *Bs*-CspB for heptanucleotides containing pyrimidines and purines at individual positions. Relative preferences [(a)] for pyrimidines / purines were determined, by comparing K_D values of two different CSP·heptanucleotide complexes (see Table 3.3) whose ligand sequences differed only at the position of interest (underlined). Their errors were estimated by Gaussian progression of mean errors.

In this assay differences in mobility cannot be directly linked to nucleobase alterations at individual sequence positions due to only qualitative measures of affinity (shift, incomplete shift, smear, no shift). Still, the results reflect the previously determined K_D values and binding preferences: While at low temperatures (at low stringency conditions) most complexes remained intact, those featuring cytosines at oligonucleotide positions 2 or 6 disassembled during electrophoreses. This indicates that the complex affinities do not solely depend on the number of thymines as previously suggested [71], because other oligonucleotides of the same



Lane	Oligo	Sequence	*Bc*-Csp K_D 10^{-y} (M)	*gel shift* low	*gel shift* med	*gel shift* high	*Bs*-CspB K_D 10^{-y} (M)	*gel shift* low	*gel shift* med	*gel shift* high
1	without									
2	dT_7	TTTTTTT	0.9 ± 0.2	+	+/-	-	1.8 ± 0.4	+	+	-
3	CT1	**C**TTTTTT	2.8 ± 0.9	+	-	-	5.9 ± 0.3	+	+/-	-
4	CT7	**CCC**TTT**C**	307 ± 33	--	--	--	1808 ± 117	-	--	--
5	CT3	**C**T**C**TTT**C**	3.3 ± 0.2	+	-	-	3.9 ± 0.2	+	+/-	-
6	CT6	**C**T**CC**TT**C**	3.5 ± 0.3	+	-	-	12.5 ± 0.6	+	-	-
7	CT4	**C**T**C**T**C**T**C**	5.2 ± 0.2	+	-	--	10.8 ± 0.9	+	-	-
8	CT5	**C**T**C**TT**CC**	36.4 ± 3.4	-	--	--	66.2 ± 4.8	+/-	--	--
9	CT9	TTTTTT**C**	1.5 ± 0.2	+	+/-	-	5.5 ± 0.4	+	+/-	-
10	O1	TT**C**TTTT	n.d.	+	+	-	0.2 ± 0.07	+	+	+/-

Figure 3.13: Temperature dependence on preferential binding of heptapyrimidines by CSP. Shift gels of *Bc*-Csp·heptapyrimidine and *Bs*-CspB·heptapyrimidine complexes (top). All complexes carry 6 additional negative charges and show an increased electrophoretic mobility as compared to CSP alone. Complexes which persisted during electrophoresis show prominent gel shifts (+), whereas dissociating complexes show tailed and discontious shifts (+/-), smears (-), or run like controls without (wo) oligonucleotide (--), depending on the CSP·oligonucleotide affinities. The results from of the shift gels are summarized in the table (bottom). The stringency conditions (low, med, high) refer to experimental temperatures (20 °C, 37, °C, 55 °C for *Bc*-Csp), (4 °C, 20, °C, 37 °C for *Bs*-CspB). K_D values were taken from Table 3.3 and Table 3.4.

base composition not carrying cytosines at position 2 or 6 do shift under this condition. The optimized ligand O1, which contains a cytosine at position 3, appears to be bound with highest affinity by both CSP under limiting stringencies (medium and high temperatures), followed by dT_7, which contains only thymines.

3.2 Discussion

Crystal structures of *Bs*-CspB and *Bc*-Csp in complex with hexathymidine were solved by molecular replacement. The arrangement of ligand and protein molecules differs significantly in the two structures. In the *Bs*-CspB·dT_6 crystal 5 nucleotides of the DNA ligand are revealed by electron density and an interspersed arrangement of protein and DNA is formed, which traverses throughout the crystal (Figure 3.8, Figure 3.15a, c, Figure 3.16a). In the *Bc*-Csp·dT_6 crystal two protein chains form a swapped dimer, whose functional units individually interact with single hexathymidines (Figure 3.15b, d, Figure 3.16b). Both structures feature a common DNA binding site, which interacts with ligand nucleobases through stacking interactions and some polar contacts. The molecular composition of the interaction surface in both structures would allow the binding of uridine-ribonucleotides without providing additional interactions.

Solution binding studies of *Bs*-CspB involving fluorescence spectroscopy revealed a ~200-fold higher K_D for heptathymidine than for hexathymidine, indicating that seven rather than six nucleotides fully occupy the CSP ligand binding site [132]. However, until now suitable crystallization conditions for well diffracting crystals of CSP in complex with deoxyheptanucleotides or with ribonucleotides have not been found. Further data from fluorescence titrations (Table 3.3, Table 3.4) and analytical gel electrophoresis (Figure 3.13) demonstrate distinctive binding preferences for individual positions within a heptanucleotide. Both CSP show a strong preference for thymine at positions 2 and significant preferences for cytosine and thymine at the sequence positions 3 and 6 (Table 3.5). At all other positions the distinction between pyrimidines was not as prominent, although thymines were slightly preferred over cytosines. By combining individual binding preferences an optimized binder featuring the sequence “TTCTTTT” was obtained, which is bound with highest affinities by both CSP. These results agree well with further binding studies of CSP binding based on fluorescence titrations [69, 71]and with studies involving a *Bs*-CspB and a DNA microarray featuring all 4096 sequences of DNA hexanucleotides (Malcom Walkinshaw and Hugh Morgan, University of Edinburgh, personal communication): When the top 42 high-affinity hits were aligned, their consensus corresponded to a heptanucleotide sequence, which strongly supports the idea

that a heptanucleotide has the optimal size to completely fill the binding site of *Bs*-CspB. A strong prevalence of thymine was found at sequence position two and significant prevalences of cytosine and thymine were observed for sequence positions 3 and 6, respectively. The overall consensus sequence of the top 42 binders was GTCTTT(G/T).

3.2.1 Assignment of seven common nucleobase-binding subsites on the CSP surface

In total 7 sites of protein-nucleobase interaction were found by comparing the crystal structures of both complexes (Figure 3.15), a finding that is in good agreement with observations that a heptanucleotide appears to be the optimal ligand size in solution [69, 132]. Four of these seven subsites are occupied by nucleobases in both structures, two of the three remaining positions are located at the borders of the binding site and therefore should not be occupied simultaneously by a hexanucleotide. Following the 5' to 3' polarity of the ligand molecules the seven nucleobase-binding subsites were indexed (Figure 3.15):

The first interaction subsite is only occupied by a nucleobase in the *Bs*-CspB·dT_6 structure: It involves an edge-on stack of the T5 base against Phe38. At subsite 2 the T6 / T1 nucleobase forms a double stack with Phe30 and Phe38. Its nucleobase headgroups are specifically contacted by Lys39 backbone groups through hydrogen bonds reminiscent of Watson-Crick basepairing (Figure 3.14). The third interaction subsite is only occupied in the *Bc*-Csp·dT_6 structure. Here, the T2 nucleobase makes an edge-on contact with Phe30. In the *Bs*-CspB·dT_6 structure the T1 nucleotide, for which no electron density was observed, is expected to occupy this space but its interaction with the protein is too weak to define its conformation in the crystal. A non-terminal nucleotide of a ssDNA ligand (as observed in the *Bc*-Csp complex structure) is expected to bind this third subsite in a more ordered fashion. At contact subsite 4

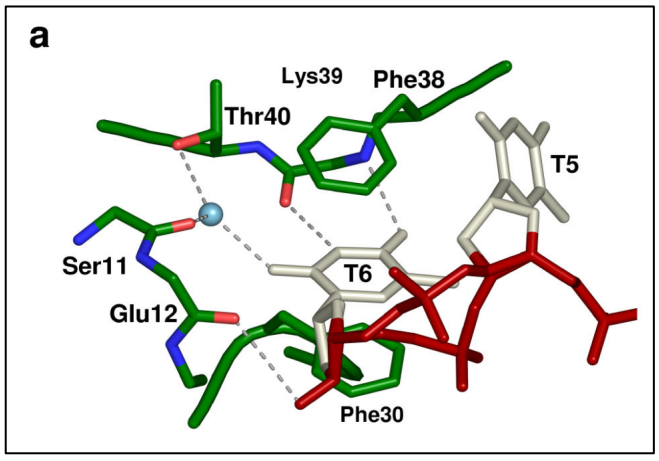


Figure 3.14: Nucleobase headgroups interact with protein groups from subsite 2 through hydrogen bonds reminiscent of Watson-Crick basepairs. The layout of molecules from these figures follows that of Figure 3.5. This figure is based on the *Bs*-CspB·dT_6 complex.

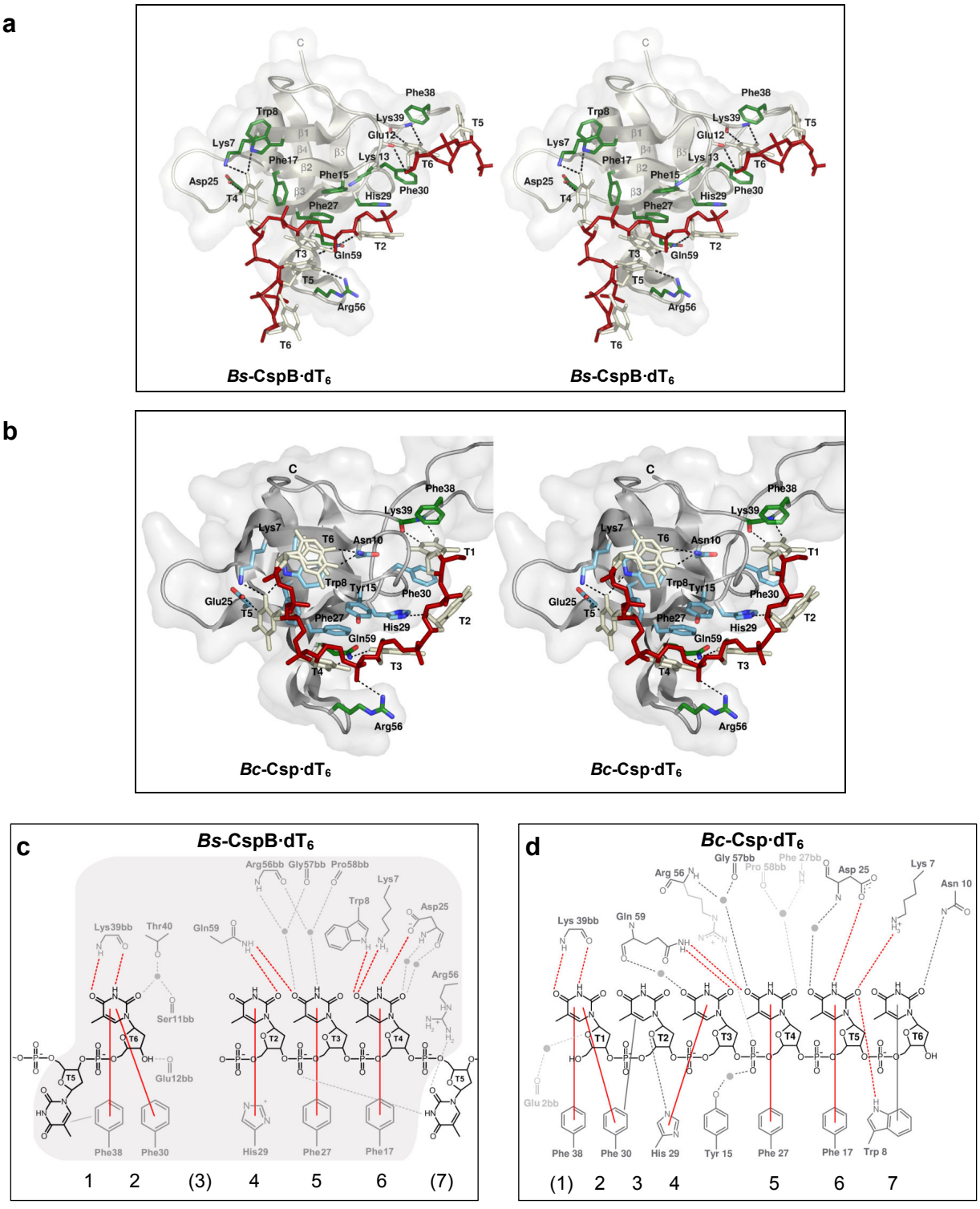


Figure 3.15: Intermolecular interactions between CSP and hexathymidine.
a) and b) Stereo projections of *Bs*-CspB·dT_6 (a) and *Bc*-Csp·dT_6 (b) complexes. Hydrogen bonds are indicated as dashed lines. The layout of molecules from these figures follows Figure 3.5. c) & d) Schematic overview of intermolecular interactions: DNA (black) and protein groups (gray) interact through stacking interactions (solid lines) and hydrogen bonds (dashed lines). Some contacts are mediated by water molecules (circles). A common set of interactions observed in both crystal structures are highlighted in red. For the nucleobases binding subsites were assigned (numbers below the schemes), which are discussed in Chapter 3.2.1 . Subsites, which are not occupied by nucleobases, are parenthesized.
c) Representative section from a continuous arrangement of *Bs*-CspB and dT_6 molecules. The interaction surface of a single protein molecule is highlighted in gray.
d) Interaction between dT_6 molecules and functional units of the protein swapped dimer. Interactions observed in only one functional unit of the structure are colored in light gray.

the T2 / T3 nucleobase stacks with His29, which extends the double stack of Phe38, T6 / T1 and Phe30 from contact subsite 1. At contact site 5 the nucleobase of T3 / T4 is stacked against Phe27. The headgroups of the nucleobases at subsites 4 and 5 are contacted by the sidechain amino group from Gln59. At contact site 6 the T4 / T5 base is specifically contacted by Lys7, Trp8, and Asp25 and forms a stack with Phe17. This nucleobase, however, shows different orientations in both complex crystal structures, which can be superimposed by 180° flips alongside an axis through nucleobase atoms N^3 - C^6. As a consequence, the nucleobase groups addressed through direct hydrogen bonds by the protein are either O^4 and N^3 (in the *Bs*-CspB complex) or O^2 and N^3 (in the *Bc*-Csp structure). These differences in nucleobase orientation are associated with different progressions of the DNA sugar phosphate backbone: In the *Bc*-Csp·dT_6 structural models the 3' terminal nucleobase form a stack with Trp8 and its O^2 group is contacted by the sidechain aminogroup of Asn10 at the designated seventh subsite. In the *Bs*-CspB complex structure, the equivalent T5 nucleotide bends away from the protein, its O^2 still being contacted by the guanidinium group of Arg56, while its aromatic nucleobase ring associates with the first binding subsite of a different protein molecule through an offset stack with Phe38. The Trp8 sidechain in the *Bs*-CspB complex structure forms a crystal contact with its symmetry-related equivalent as previously discussed (Figure 3.6), and may be seen as one example of differences in the crystal packing environments of the two structures, which are expected to cause the diverse architectures of the two complexes. As the orientation of this sidechain strongly deviates from that of all other CSP structures (Figure 3.18), an arrangement of the terminal two nucleobases as observed in the *Bc*-Csp·dT_6 probably reflects the binding of a thymine-based ligand in solution.

3.2.2 How CSP may bind to heptanucleotide motifs and functional implications

As previously shown by mutational analysis (Table 3.2) and structural data (Figure 3.7, Figure 3.15), hydrophobic interactions predominantly contribute to ligand binding by CSP. Nevertheless, a very strong preference for thymine at positions 2 and significant preferences, at positions 3 for cytosine and at position 6 for thymine (Table 3.5, Table 3.6), suggest, that specific contacts between protein groups and polar nucleobase headgroups may help to discriminate between pyrimidines. This information may be used to relate the ligand molecule structures of both complexes and get an idea how heptapyrimidine motifs interact with the CSP binding site in solution: The highest preference for a thymine is expected for the contact

subsite 2 (Figure 3.15) in both complexes: Here the O^4 and N^3 functions are contacted through hydrogen bonds by the backbone of Lys39. The deprotonated N^3 function of a cytosine located at this site could not act as a hydrogen bond donor and would be repelled by the Lys39 carboxyl group (d = 2.88 Å), while its N^4 would be unfavorably lined up against the Lys39 amide group. The distortion of the protein-nucleobase interaction at subsite 2 is expected to cause a larger increase in K_D than those at the other subsites, because two stacking interactions with protein sidechains would be disturbed, whereas all other interactions involve only single stacks or edge-on contacts. A second site which is expected show a strong preference for thymine is subsite 6, which specifically addresses the protonated N^3 group of a thymine through a high-affinity hydrogen bond by the carboxylate group of Asp25 (d = 2.85 Å). The deprotonated N^3 of a cytosine would unfavorably interact with the sidechain's carboxylate group (d = 2.85 nm) and destabilize the ligand binding at this subsite.

A stabilizing effect of a cytosine to subsite 3 is not immediately evident from the crystal structure, because no nucleobase specific hydrogen bonds are observed in this site. The introduction of a cytosine may stabilize the edge-on contact between Phe30 and the base due to improved stacking interactions: The delocalized π-electrons from the nucleobase ring obtain a more prominent negative potential as compared to thymine due to additional positive inductive and mesomeric effects of its N^4 group [137].

The agreement of experimentally determined binding preferences and those derived from structural data support an assignment of seven nucleobase-binding subsites with distinctive binding properties. In addition, 1:1 stoichiometries of CSP·hexanucleotide and heptanucleotide complexes based on NMR data and analytical ultracentrifugation (Figure 2.6 & [132]) finally suggests that in solution complexes of a single protein molecule and a single heptanu-

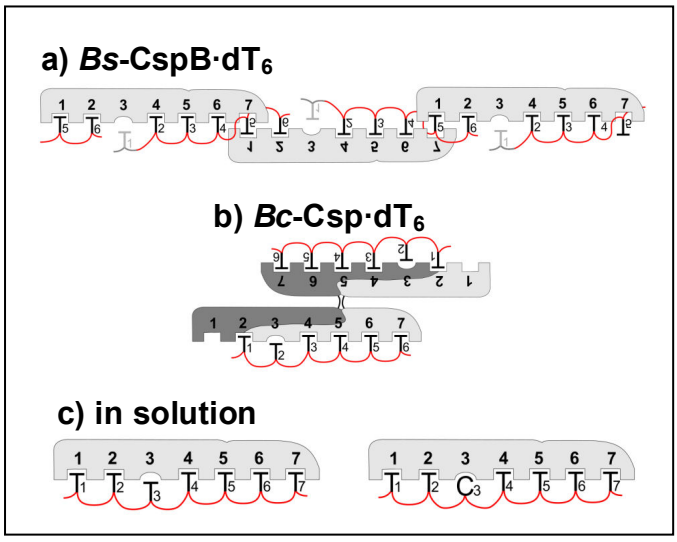


Figure 3.16: Schematic overview of CSP·oligonucleotide interactions. Protein molecules (gray objects) interact with bases from oligonucleotides at distinct binding subsites (cavities).
a) in the *Bs*-CspB·dT_6 crystal, a continous arrangement of protein and DNA molecules is formed. A gap between the 3' nucleotide (bound to subsite 2) and the first structured 5' nucleotide (bound to subsite 4) exists, which is expected to bind the unstructured T1 nucleotide (gray character).
b) In the *Bc*-Csp·dT_6 crystal structure, two protein chains (light and dark gray) form two functional units. Each of them binds a DNA molecule.
c) In solution, all seven subsites are occupied by a single oligonucleotide molecule.

cleotide motif are formed in such a way, that the 5' nucleobase interacts with subsite 1 and the following nucleotides proceed towards the seventh subsite (Figure 3.16c, Figure 3.19a). The biological functions of CSP are still under investigation. Certain CSP have initially been reported to function as transcriptional activators of cold induced genes such as *hns*, encoding the major nucleoid protein [68], and *gyrA*, encoding a subunit of DNA gyrase [138]. The ability of CSP to bind to single-stranded nucleic acids and prevent their association to double strands *in vitro* led to the assumption that these proteins may function as RNA chaperones [72], which may prevent the formation of mRNA double strands at low temperatures. Whether mRNA basepairing indeed impairs gene expression in the cold and whether CSP binding can prevent this, remains to be shown. It is quite clear, however, that the CSP are required for cold adaptation: Strains of *B. subtilis*, with two of three CSP genes inactivated show severe restrictions in growth at low temperatures [63]. Likewise, an *E. coli* strain with four (of nine) CSP genes inactivated is cold-sensitive and impaired in upregulation of many cold-induced genes [8]. For RNA chaperone activity, a low-affinity sequence-unspecific binding of mRNA by CSP may be sufficient to prevent mispaired mRNA double-strand formation in the cold. It remains to be seen, however, which role the preferential binding of CSP to uracil-rich sequences may have in this context.

In addition to the presumed RNA chaperone activity there is good evidence that some *E. coli* CSP function as transcriptional antiterminators of cold-regulated polycistronic genes [74]. Some post-terminator cistrons have been attributed to cold regulation, and the amount of their mRNA was shown to increase upon either a down-shift in temperature or the overexpression of certain CSP genes [59, 139]. Transcription termination occurs at terminator sites on the 3' end of an mRNA transcript which contains two inverted, guanine and cytosine-rich sequence motifs, which can form a stem-loop. This stem-loop is followed on the downstream side by

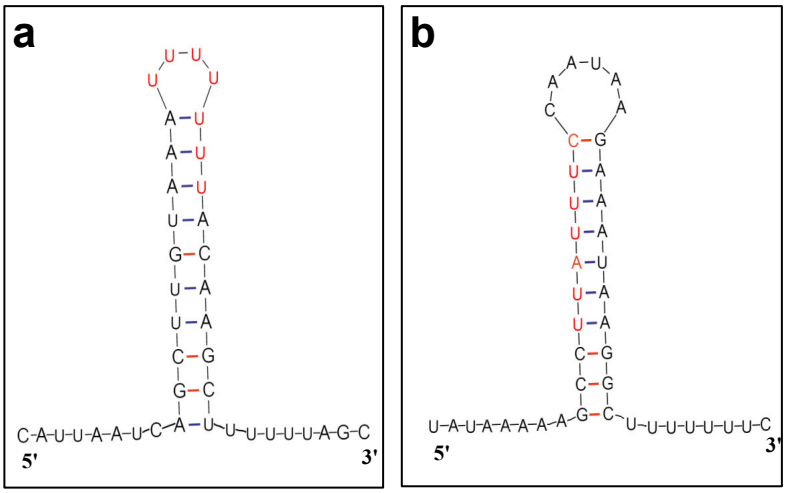


Figure 3.17: Two examples of predicted terminator sites which contain heptanucleotide sequence motifs in agreement with presumed binding preferences of CspB (highlighted in red). Terminator sites were predicted using the TransTerm program [9] and depicted using the mfold program [14]. Base pairing of U-rich downstream sequences, which may participate in the dissociation of the transcript is not shown.
a) predicted terminator site ~1200 nt upstream of the *des* gene (fatty-acid desaturase. b) predicted terminator site ~2500nt upstream of the *gyrA* gene , which encodes for a subunit of gyrase.

five or more uridines. A current model of transcription termination assumes the dissociation of a nascent transcript from the DNA upon stem-loop formation at the terminator site. When the adjacent uracil-rich sequence is being transcribed, the affinity of the dA-U basepairs is insufficient to stabilize the DNA-RNA duplex, and the transcript, together with the RNA polymerase, disassembles from the template. It is expected that the obstruction of stem-loop formation prevents termination and leads to transcription of cistrons on the 3' side of termination sites. The ability of the CSP to prevent double-strand formation in RNA may destabilize terminator stem-loops and promote transcription antitermination [74]. The preference of CSP binding to uracil-rich sequences may direct the CSP to terminator sites. Interestingly, some predicted terminator sites in cold-induced genes of *Bacillus subtilis* appear to contain sequence motifs, which could form high-affinity interactions with *Bs*-CspB sequences (Figure 3.17). Bound to CSP the basepairing of a heptanucleotide would be prevented, while the strong ligand curvature would complicate basepairing of adjacent nucleotides. Whether these sites indeed regulate gene-expression involving sequence-specific interactions with CSP has to await further experimental validation.

3.2.3 Structural superpositions of CSP structures reveal a preformed nucleotide binding site

Quite remarkably, most hydrophobic sidechains (Trp8, Phe15 / Tyr15, Phe17, Phe27, His29, Phe30, and Phe38) which participate in ligand binding in both complex structures (Figure 3.7, Figure 3.15), are located and oriented in almost invariant ways when comparing all CSP crystal structures from *Bacilli* currently deposited in the PDB (Figure 3.18). Only Phe38, which shows two alternative conformation in five ligand-free structures, and Trp8, whose reorientation in the *Bs*-CspB·dT_6 complex may be explained by crystal packing effects as previously discussed (Figure 3.6), show some conformational variability. Similar conformational restrictions apply to polar groups, which are expected to establish specificity at the designated subsites 2 and 6 in both complex structures. These involve the amino and carbonyl groups of Lys39 (subsite 2), which are restricted by being part of the backbone, and the sidechains of Asp25 and Lys7 (subsite 6), which restrict their flexibility by forming a salt bridge. The sidechain of Gln59 in contrast, which interacts with O^2 and O^4 nucleobase groups at subsites 4 and 5 but apparently does not contribute to nucleobase-specific binding at these sites, as indicated by only slight preferences of thymines over cytosines at these positions, adopts various conformations. The overall contribution of this sidechain to the ligand binding appears to be

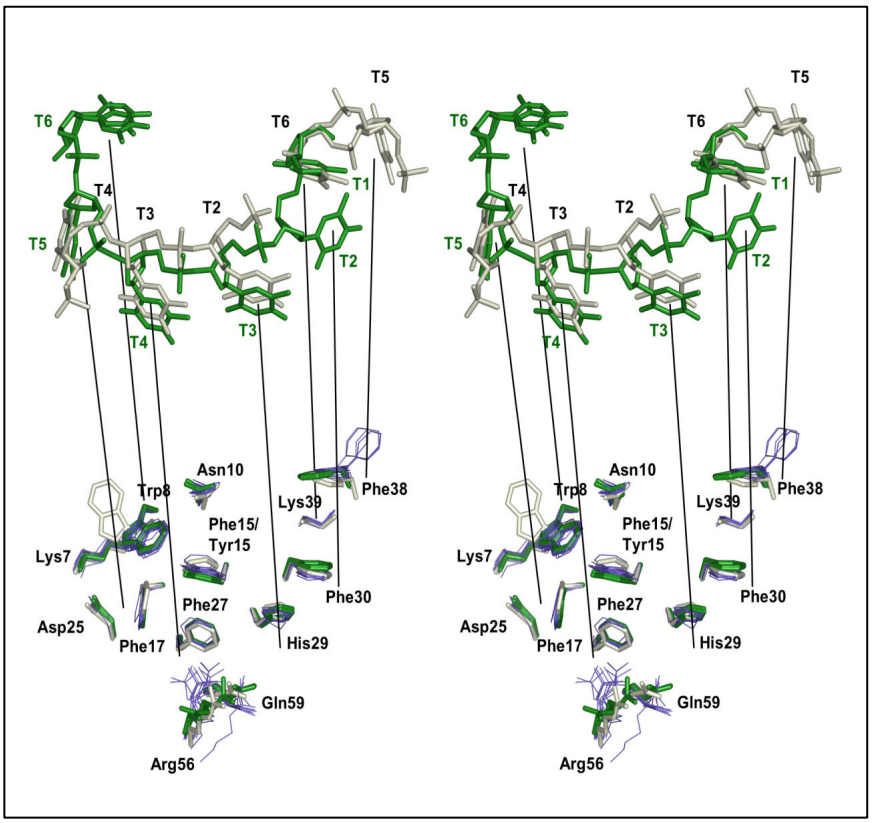


Figure 3.18: Superimposition of *Bacillus* CSP crystal structures from the PDB (1CSP, 1CSQ, 2ES2, 2HAX, 1HZ9, 1HZA, 1HZB, 1HZC, 1I5F). Structures of nucleobase ligands from the *Bs*-CspB·dT_6, gray, and *Bc*-Csp·dT_6, green, complexes have been shifted (top) to allow a better view on the CSP binding site (bottom). Lines schematically relate individual nucleobases to their position in the complex structures. Protein groups involved in nucleobase interactions in the complex structures are shown as sticks in corresponding colors. Their equivalents from CSP structures without an oligonucleotide ligand are displayed as blue lines.

small; its replacement by alanine does not significantly increase the K_D for dT_7 (Table 3.2). Arg56, which is associated with O^2 from the nucleobase T5 in the *Bs*-CspB·dT_6 complex structure, does contribute to ligand binding as shown by a 6.5-fold increase in K_D associated with its replacement to alanine. In accordance with this, its interaction with the phosphodiester group linking T3 and T4 in one functional unit of the *Bc*-Csp·dT_6 complex and its prominent conformational variability in all other CSP structures suggests, that it also interacts with other nucleotides in a sequence-unspecific way. The restricted conformational flexibility of most protein groups interacting with the ligand is expected to decrease the entropic penalty, which would arise if a flexible site of interaction becomes fixed upon ligand binding. In addition, a preformed binding site is in agreement with the observation that kinetic rate constant (k_{ass}) for the formation of CSP·oligonucleotide complexes are hardly influenced by the ligand sequence (Table 3.3): If the binding site was gradually formed upon interaction with a ligand molecule, unfavorable interactions between ligand and protein groups should delay the formation and its k_{ass} would depend on the K_D of the complexes.

3.2.4 Ligand binding interfaces from CSP and Y-box proteins are highly conserved

Apart from *Eubacteria*, the CSP and homologs can be found in most kingdoms of life, except *Cyanobacteria* and *Archaea*. The sequence conservation of bacterial CSP orthologs is usually higher than 50% identity. Similar degrees of conservations are between bacterial CSP paralogs: The lowest level of homology between CSP paralogs in *E. coli* is observed for CspA and CspH, which share 48% sequence identity. In *B. subtilis* CspC and CspD share 68% sequence identity. Conserved residues are not equally distributed over the protein surface: Residues from the binding interface - defined on the basis of the two CSP·dT_6 structures - are typically conserved on a level above 75% sequence identity (Figure 3.19). In cases of diverging residues within the binding interface similar amino acids are observed (Tyr / Phe / Trp, Ser / Thr). Residues from other surface areas show a much lower level of conservation (less than 30% sequence identity). This suggests that binding modes and preferences of most CSP in bacteria are very similar, and may be attributed to similar biological function(s) - a finding that was already demonstrated for some CSP paralogs of *B. subtilis* and *E. coli*, which were shown to complement each other [63, 140]. The systematic analysis of *Bs*-CspB and *Bc*-Csp binding preferences by fluorescence titration confirms these observations (Table 3.5): Although the preferences of *Bc*-Csp for individual nucleotides at most positions was not as pronounced as for *Bs*-CspB, they clearly followed the same trend. The smaller base discrimination by *Bc*-Csp may have been caused by the experimental parameters: All fluorescence titration measurements were performed at 15 °C in order to allow good comparisons with previously published data. While for *B. subtilis* this temperature is close to its growth conditions, the temperature optimum for *B. caldolyticus* is more than 40 °C higher. The analyses of CSP·heptapyrimidine complexes by analytical gel electrophoresis reveal, that indeed both proteins display very similar binding preferences at three different levels of stringency, based on different temperatures (Figure 3.13). For *Bs*-CspB the temperature of each level (4 °C, 20 °C, 37 °C) was about 17 °C lower than for *Bc*-Csp (20 °C, 37 °C, 55 °C). A temperature-dependence on ligand binding is not unexpected as Gibbs free energies in general depend on temperature ($\Delta G = \Delta H - T \cdot \Delta S$). Their differences in affinity despite similar binding preferences may suggest that further intermolecular interactions between *Bc*-Csp and the ligand account for some additional affinity in the hyperthermophilic protein. The high degree of conservation in ligand-binding interfaces of *Bacillus* (Figure 3.15) and other CSP (Figure 3.19), however,

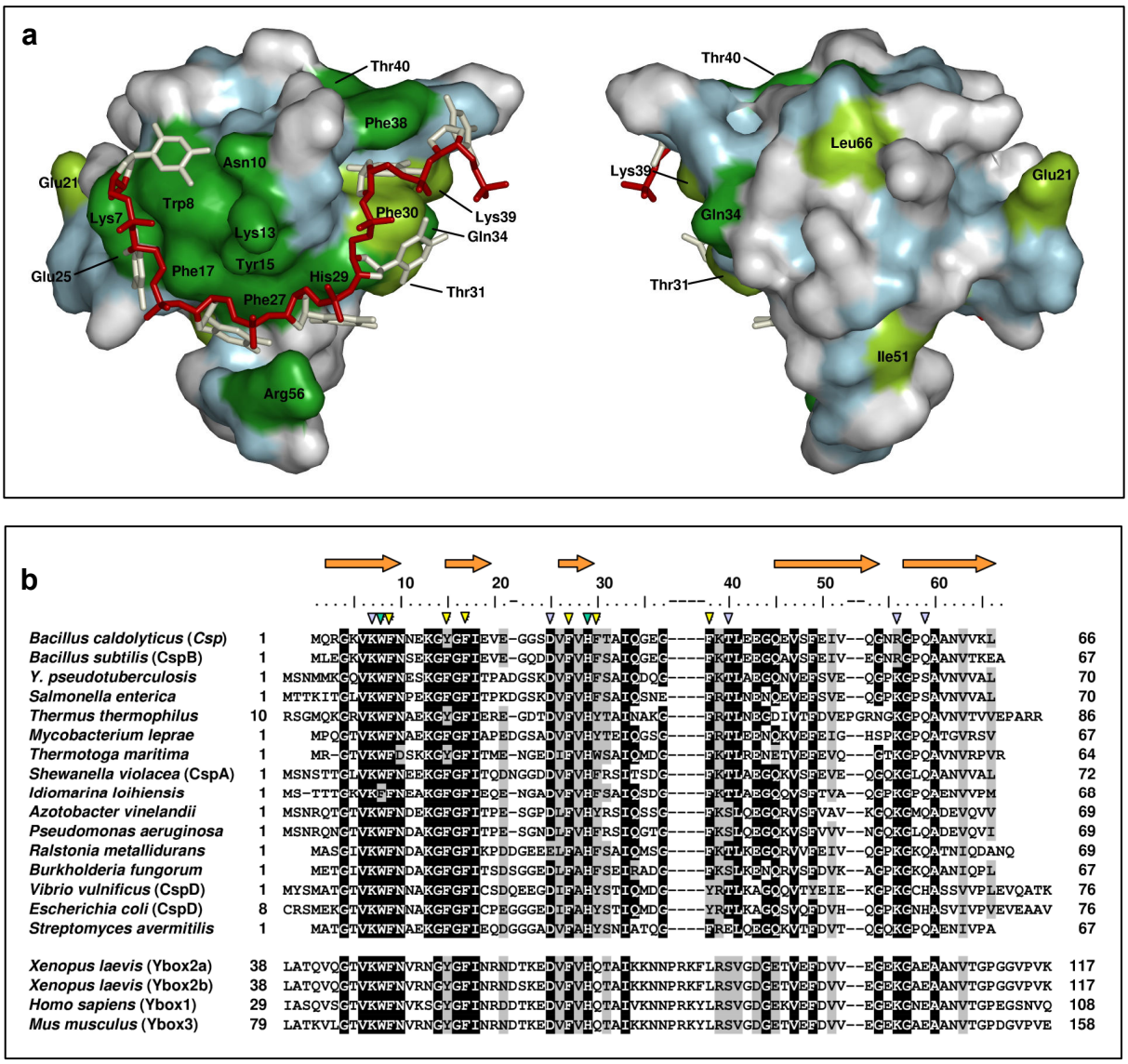


Figure 3.19: The binding site for nucleic acids is conserved throughout the CSP and Y-box proteins.
a) CSP sequence conservation mapped to the surface of *Bc*-Csp bound to dT_6. Most residues forming the ligand interaction site are conserved on the level of at least 75% identity (dark green) and similarity (light green). Other areas are conserved to a lesser extent. Invariant regions which originate from the protein backbone are colored light blue. The structure shown in this figure is based on a single functional unit of a swapped *Bc*-Csp·dT_6 dimer. The model of a heptanucleotide ligand (sticks) was assembled by combining structural information of the *Bs*-CspB·dT_6 and *Bc*-Csp·dT_6 complex structures.
b) Sequence alignment of CSP representatives (upper part) and Y-box proteins (lower part). Residues which are conserved at a level of at least 75% identity or similarity are highlighted in black or grey, respectively. Sidechains involved in DNA binding in the CSP·dT_6 complex crystal structures are marked by triangles (blue - polar contacts, yellow - hydrophobic contacts, green - both).

does not support the idea that additional interactions are responsible for the higher stabilities of *Bc*-Csp·heptapyrimidine complexes. Instead, greater protein stability appears to results in higher substrate affinities observed for *Bc*-Csp. Furthermore I presume that RNA binding preferences published for the *E. coli* CSP [141] and other CSP are incomplete because they consider absolute affinities and not the dependence of binding stringency on temperature. This dependence may also explain, why despite the high degree of identity within their binding

interfaces, for some (CspA, CspC, CspE) but not for all *Ec*-CSP (CspB) analyzed, transcription anti-termination activity could be demonstrated [74]. It remains to be shown for other CSP orthologs and paralogs, to what extent protein function depends on stability and is influenced by temperature, and what the biological significance of this may be.

Apart from *Eubacteria*, cold shock domains can also be found in *eukaryotic* proteins. A high degree of sequence identity (> 45%) was reported between the CSP and the nucleic acid-binding domains of the eukaryotic Y-box proteins [86]. Y-box proteins have been implicated in transcriptional activation and repression, regulation of alternative splicing, regulation of mRNA stability, translational activation or repression, and RNA packaging. It was also shown recently that a Y-box protein was required for cell growth at cold temperatures [88]. Most residues involved in preferential ligand binding in the two CSP·dT_6 crystal structures are conserved in the sequences of Y-box proteins, with the exception of sidechains forming nucleobase-binding subsites 1 and 2. I therefore expect that the Y-box and CSP proteins do not only have a common ancestor but also show common modes of nucleotide binding and similar binding preferences.