Chapter 2 - Materials and methods

General cell-culture and molecular-biology methods used in this work refer to protocols from the text book Molecular Cloning [95]. Other methods will be described in detail if further references are incomplete or not available. All fine-chemicals were purchased from ROTH unless stated otherwise. All suppliers of materials used in this work are listed in section 2.7.

2.1 Molecular biology methods

2.1.1 Denaturing Schägger gels

Samples of bacterial cultures and aliquots of protein preparations were analyzed by discontinuous SDS-polyacrylamide gel electrophoresis (PAGE) according to Schägger & Jagow [96]. Under these conditions most proteins are unfolded and their charges are masked by detergent molecules. In this method proteins from the samples are first focussed in the top stacking gel, and then separated according to their apparent molecular sizes within a range of 1 to 100 kDa in the bottom separation gel. The Schägger method rather than the widely used Laemmli method [97] was used for separating protein samples containing CSP, because it provides a better resolution in the size range of 1 to 10 kDa.

Implementation

A cassette for minigels (80 mm × 60 mm × 5 mm) (BIO-RAD) was assembled and mounted according to the manifacturer's specifications. The separation gel was prepared by mixing:

6 ml H_2O,
8 ml gel buffer,
10 ml acrylamide / bisacrylamide solution,
200 µl 10% (w/v) ammonium persulfate ,
15µl TEMED.

The gel cassette was filled with liquid separation gel leaving a gap of 2 cm below the upper edge. The gel was overlaid with isopropanol. After polymerization the supernatant was removed, the gel was washed extensively with water, and overlaid with unpolymerized stacking gel. The stacking gel was prepared by mixing:

6.7 ml of H_2O,
2.75 ml of gel buffer,
1.9 ml of acrylamide / bisacrylamide solution,
100 µl 10% (w/v) ammonium persulfate ,
10 µl TEMED.

Pockets for samples were introduced into the liquid stacking gel by inserting a sample comb featuring 15 wells into the unpolymerized stacking gel. After polymerization the comb was removed. Gels were mounted in the electrophoresis chambers and the apparatus was filled with anode and cathode buffer. Samples consisting of 5 µl of protein solution were mixed with 5 µl of loading buffer, denatured at 95 °C for 5 min, and loaded into individual wells of the gel. The electrophoresis was started with a constant current of 20 mA per gel, until the bromophenol-blue band, used as a tracing component in the loading buffer, had migrated into the separation gel. Thereupon the current was changed to 40 mA per gel. Electrophoresis was terminated when the bromophenol-blue band had reached the lower end of the separation gel. Gels were removed from gel cassettes, stained with staining solution for 30 minutes, and de-stained until protein bands could clearly be distinguished from the background.

Solutions

6 × loading buffer	20% SDS, 50 mM TRIS pH 7.6, 1% bromophenol blue, 50% glycerol.
gel buffer	1.5 M Tris, 0.4% sodium dodecyl sulfate, pH 8.8.
cathode buffer	0.1 M TRIS, 0.1 M tricine, 0.1% (w/v) SDS, pH 8.23.
anode buffer	0.2 M TRIS, pH 8.9.
acrylamide / bisacrylamide sol.	37.5 : 1, 30% solution (Rotipherese Gel 30, ROTH).
staining solution	50% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.1% (w/v) Coomassie Brilliant Blue.
destaining solution	5% ethanol, 10% glacial acetic acid, 85% H_2O.

2.1.2 Native polyacrylamide gels

Continuous native gel electrophoreses using 17.5% polyacrylamide gels in TBE were performed to evaluate the stability of CSP·DNA complexes. The migrating speed of macromolecules in these gels depends on masses and charges of single components or complexes in the samples. All CSP evaluated in this work display pI values below pH 8.23, therefore the default polarity for denaturing gels (cathode towards the stacking gel, anode towards the separating gel) could be used for native gels. Upon binding of an oligonucleotide to a protein a complex of higher molecular mass is formed, which carries additional negative charges as compared to the free protein. The accelerating effect of additional charges has a greater influence on the electrophoretic mobility of the complexes than the decelerating effect of mass increase. As a consequence, protein·DNA complex bands are shifted towards the side of the anode (lower side).

Implementation

A cassette for minigels (80 mm × 60 mm × 5 mm) [BIO-RAD] was mounted in a stand according to the manifacturer's specifications and filled with separation gel, which was prepared by mixing:

3 ml H_2O,
5.8 ml acrylamide / bisacrylamide solution,
1 ml 10 × TBE,
200 µl 10% (w/v) ammonium persulfate,
15 µl TEMED.

A sample comb featuring 9 wells was inserted into the liquid gel. After polymerization, the comb was removed, the gel cassette was inserted into an electrophoresis chamber and the chamber was filled with cold 1 × TBE. Samples were prepared by mixing 5 µl of protein solution, 1 µl glycerol (80%), and 0.3 µl of loading buffer and filled into individual wells. Electrophoresis was carried out at 4 °C at a constant voltage of 50 V until the bromophenol-blue band used for tracing appeared in the middle of the gel. Gels were removed from the electrophoresis cassettes and silver-stained.

Solutions

10 × TBE	0.89 M TRIS, 0.89 M boric acid, 0.02 M EDTA.
acrylmide/bisacrylamide sol.	37.5:1, 30% solution (Rotipherese gel 30, ROTH).
Loading buffer	0.1% bromphenol blue, 20% ethanol.

2.1.3 Silver staining of polyacrylamide gels

A simple 30-min silver staining protocol based on [98] was used for staining proteins following electrophoresis in a native PAGE.

Implementation

Steps	Solution	Time
Fixation	60 ml acetone stock; 1.5 ml TCA stock; 25 µl 37% HCOH	5 min
Rinse	H_2O	3 × 5 s
Wash	H_2O	5 min
Rinse	H_2O	3 × 5 s
Pretreat	60 ml acetone stock	5 min
Pretreat	100 µl $Na_2S_2O_3$ stock	1 min
Rinse	H_2O	3 × 5 s
Impregnate	0.8 ml $AgNO_3$ stock; 0.6 ml 37% HCOH; 60 ml H_2O	8 min
Rinse	H_2O	3 × 5 s
Development	1.2 g Na_2CO_3; 25 µl HCOH; 25 µl $Na_2S_2O_3$ stock; 60 ml H_2O	10 - 20 s
Stop	1% glacial acetic acid in H_2O	30 s
Rinse	H_2O	10 s

Table 2.1: A short protocol for silver staining of polyacrylamide gels.

Solutions

acetone stock	50% acetone in H_2O.
TCA stock	50% TCA in H_2O.
$AgNO_3$ stock	20% $AgNO_3$ in H_2O (store in dark).
$Na_2S_2O_3$ stock	10% $Na_2S_2O_3$ in H_2O.

2.1.4 Generation of *Bs*-CspB mutant variants by QuikChange Mutagenesis

Variants of *Bs*-CspB featuring single amino-acid replacements were generated, in order to analyze the contribution of individual sidechains to ligand binding and protein stability. The mutants were made using the improved two-stage QuikChange Mutagenesis protocol [99]. In this reaction both strands of a template plasmid (vector pET11a *cspB* amp^R, supplied by Christine Magg, Lehrstuhl für Biochemie, University of Bayreuth) are fully amplified, originating from primers which contain the desired mutations. Separating the amplification of the two strands by using each primer in a separate reaction prevents the formation of primer dimers during the first rounds of amplification, which are expected to profoundly decrease the efficiency of mutagenesis. After the initial steps of amplification both reactions are mixed and some additional rounds of amplification are performed. Upon amplification, plasmids containing unmodified DNA strands are cleaved by adding *Dpn*-I restriction enzyme [NEW ENGLAND BIOLABS], which only cleaves methylated DNA. Mutated nicked plasmids generated in the amplification cycles do not carry methyl groups.
Upon cleavage, aliquots of the digests were transformed into *E. coli* Top10 cells, and bacteria which acquired functional plasmids were selected by appropriate antibiotics. All clones were checked for the correct introduction of the desired mutations and the absence of second-site mutations by gene sequencing (MWG BIOTECH) using vector specific primers (T7 / T7R).

Implementation

For each mutagenesis reaction two primers were generated, which cover 18 to 21 nucleotides upstream to the area of modification, the desired nucleotide changes, and 21 nucleotides downstream to the area of modification, for both strands of the template DNA (see materials, below).

25 pmol of one primer were added to a mixture of
10 ng template DNA,
5 µl 10 × PCR buffer (STRATAGENE),
4 µl 5 mM dNTP (each) (STRATAGENE),
the total volume was brought to 49 µl with H_2O in 0.2 ml PCR tubes.

Immediately before the mutagenesis reaction was started, 1 µl (5 u) of *Pfu* polymerase (STRATAGENE) was added to both tubes, the tubes were inserted into a thermocycler (PTC-200, BIO-RAD) and the following protocol was started:

Task	Time	Temperature	Comment
Initialization	30 s	95 °C	initial denaturation of DNA
Step 1: 10 cycles	30 s 30 s 300 s	95 °C 58 °C 86 °C	denaturation of DNA primer annealing primer elongation
Combination			after the first round of amplification, the content of both tubes was mixed.
Step 2: 20 cycles	30 s 30 s 300 s	95 °C 58 °C 86 °C	denaturation of DNA primer annealing primer elongation
Termination		4 °C	storage

Table 2.2: A protocol for site-directed mutagenesis of bacterial plasmids.

After the second round of amplification, 1 µl (5 u) of *Dpn*-I restriction enzyme was added to the mutagenesis setups and incubated for 4 h at 37 °C. Upon digestion 1.5 µl aliquots were used to transform chemo-competent cells of *E. coli* strain Top10. Transformed clones were selected on LB agar plates containing 100 µg / ml ampicillin.

Materials

DNA oligonucleotides (MWG BIOTECH, 0.2 µmol synthesis scale) for mutagenesis:

Mutation	rxn.	Sequence
K7A	1	ATGTTAGAAGGTAAAGTA**GCA**TGGTTCAACTCTGAAAAAGGT
K7A	2	ACCTTTTTCAGAGTTGAACCA**TGC**TACTTTACCTTCTAACAT
D25A	1	TCGAAGTAGAAGGTCAAGACG**C**TGTATTCGTTCATTTCTCTGC
D25A	2	GCAGAGAAATGAACGAATACA**G**CGTCTTGACCTTCTACTTCGA
R56A	1	TTTGAAATCGTTGAAGGAAAC**GC**CGGACCACAAGCTGCTAACGTT
R56A	2	AACGTTAGCAGCTTGTGGTCCG**GC**GTTTCCTTCAACGATTTCAAA
R59A	1	GTTGAAGGAAACCGCGGACCA**GC**AGCTGCTAACGTTACTAAAGA
R59A	2	TCTTTAGTAACGTTAGCAGCT**GC**TGGTCCGCGGTTTCCTTCAAC

Table 2.3: Primers used for the generation of mutant *Bs*-CspB variants by site-directed mutagenesis. For each construct two reactions (rxn.) were needed to introduce the desired mutation into the coding (1) and noncoding (2) strand. Changes from wildtype *Bs*-CspB are **indicated**.

2.2 Cell biology and protein purification methods

All steps of protein purification were carried out at 4 °C unless stated otherwise. For fast protein liquid chromatography (FPLC) a PHARMACIA LKB FPLC system was used (controller: LCG-501 Plus, recorder: Rec 102, pump: P-500, photometer: Uvicord SII, fraction collector: FRAC 200). For purification at low pressure the PHARMACIA LKB GRADIFRAC system was used (controller: programmer GP-250 Plus, pump: P1, photometer: Control Unit UV-1, fraction collector: FRAC-100)

2.2.1 Recombinant overexpression of *Bs*-CspB and mutant variants in *Escherichia coli*

Protein for crystallization trials of *Bs*-CspB and *Bc*-Csp was generously supplied by Markus Zeeb and Christine Magg (both Lehrstuhl für Biochemie, University of Bayreuth). Protein for crystallization trials of *Bs*-CspB mutants M1R/E3K/K65I and A46K/S48R was provided by Michael Wunderlich (Lehrstuhl für Biochemie, University of Bayreuth). For biophysical studies of CSP·oligonucleotide complexes, *Bs*-CspB and variants of *Bs*-CspB were expressed in BL21(DE3) cells cotransfected with pLysS plasmid (NOVAGEN) in 2 × TY medium supplemented with 200 µg / ml ampicillin and 34 µg / ml chloramphenicol at 37 °C. For each construct two 2-l cultures were inoculated 1 : 10 from an overnight culture and induced with 1 mM IPTG (isopropylthiogalactoside, PHARMACIA) after a cell density of 0.8 - 1 $OD_{\lambda = 600\ nm}$ was reached. Two to three hours after induction the cells were pelleted in a cooled centrifuge (J2-MC, rotor: JA-10, BECKMAN) at 4000 × g / 4 °C, flash-frozen in liquid nitrogen, and stored at -70 °C.

2.2.2 Bacterial lysis

For protein purification a bacterial pellet was thawed at room temperature, resuspended in 35 ml ice-cold lysis buffer per 10 g of bacterial pellet and lysed using a French Press (French Pressure Cell Press, SMI-AMINCO) at 19,500 psi (internal chamber pressure). The viscous lysate was diluted 1 : 1 with lysis buffer and centrifuged at 65,000 × g for 30 min (J2-MC, rotor: JA-20, BECKMAN). The supernatant was aspirated and used for anion-exchange chromatography.

2.2.3 Anion-exchange chromatography

Anion-exchange chromatography (AEC) was the first step in the chromatographic purification strategy for CSP. Chromatographic purification was performed in order to remove most cellular components including most proteins from the protein of interest. Ion-exchange chromatography separates macromolecules according to their electrostatic properties using ion gradients. Anion exchangers bind proteins with anionic (negative) surface charge under low-salt conditions. Bound protein is eluted in high-salt buffer.

Implementation

All steps were carried out at a flow speed of 5 ml / min using a low-pressure GRADIFRAC device (PHARMACIA) at 4 °C. Cleared bacterial lysate was loaded on a column (Vantage L, MILLIPORE, manually packed with 200 ml DEAE sepharose from 33 g dry material, PHARMACIA) equilibrated in buffer A. This buffer was also used for washing until the $OD_{\lambda = 280nm}$ at the column outlet returned to base level. Proteins were eluted in a 600 ml gradient from 0% to 60% buffer B and collected in 15-ml fractions, which were analyzed by Schägger SDS page. Fractions containing protein of the correct size (~ 7.4 kDa) were pooled. Before loading, samples were concentrated 10-fold using a speedvac (SAVAN - SV200) and resuspended in 50% loading buffer.

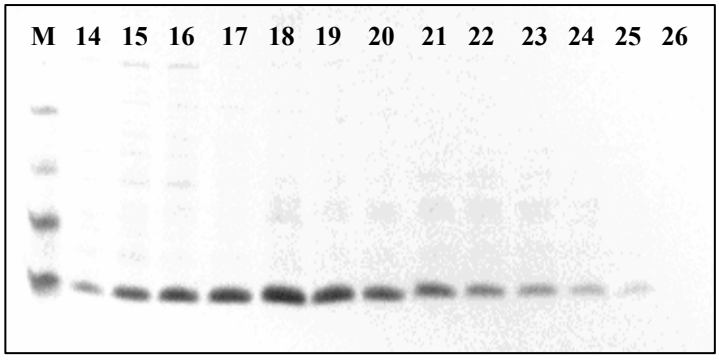

Figure 2.1: Anion-exchange chromatography of *Bs*-CspB variant R56A using a DEAE sepharose column. Fractions 14 to 26 (numbers) are shown. The fraction size was 15 ml. Equivalents of 100 µl eluate fraction were loaded in each lane. All fractions were pooled for hydrophobic interaction chromatography. M – marker, bands at 29,000 Da, 21,000 Da, 12,500 Da, 6,500 Da (top to bottom).

2.2.4 Ammonium-sulfate precipitation and butyl-sepharose hydrophobic-interaction chromatography (HIC)

In this step most bacterial proteins and nucleic acids were precipitated by adding powdered ammonium sulfate to a level of 50% saturation. Under these conditions all CSP used in this study (*Bs*-CspB and mutants as well as *Bc*-Csp) remained in solution. In the presence of high-salt concentrations these CSP bound to butyl-sepharose hydrophobic-interaction-chromatography columns and could be eluted in low salt buffer.

Implementation

Chromatography steps were carried out at flow rates of 2 ml / min using a FPLC system (PHARMACIA) at 4 °C with a maximal pressure set to 2 MPa.

Fine-powdered ammonium sulfate was gradually added to pooled CSP containing fractions from AEC stirred on ice until 50% (132 g / 100 ml) of saturation was reached. The suspension was centrifuged at 40,000 / 4 °C (BECKMAN J2-MC, rotor JA-10), in order to spin down precipitated proteins. The supernatant was loaded on a column (XK16 from PHARMACIA, packed with 25 ml of butyl-sepharose slurry, PHARMACIA), equilibrated with 37.5% buffer C / 62.5% buffer D (50% saturation). After loading, the column was washed with 37.5% buffer C / 62.5% buffer D until the $OD_{\lambda = 280\ nm}$ at the column outlet returned to base level. Elution was performed in a 20 ml gradient from 62.5% to 0% buffer D. During elution 2-ml fractions were collected and analyzed by Schägger SDS page. Fractions containing protein of the appropriate mass (~ 7.3 kDa) were tested by UV spectroscopy for nucleic-acid contamination, and fractions devoid of nucleic acids were pooled and concentrated to a final volume of 4 ml using spin concentrators (Vivaspin 15 - 3000 MWCO PES, SARTORIUS) at 3,500 g.

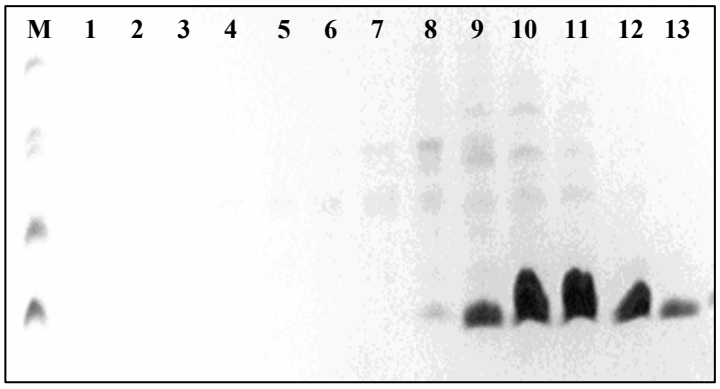

Figure 2.2: Hydrophobic interaction chromatography of *Bs*-CspB variant R56A. The size of the fractions collected was 3 ml. 5µl of the eluate fractions were loaded.
Fractions 9 – 13 (numbers) were pooled and used for gel filtration. M – marker, bands at 29,000 Da, 21,000 Da (double band), 12,500 Da, 6,500 Da (top to bottom).

2.2.5 Gel-filtration chromatography (GFC)

The last step of the chromatographic protein-purification strategy for CSP involved gel-filtration chromatography. Here, all molecules in the sample are separated according to their hydrodynamic radii. This step was required to remove some high molecular-weight impurities still present in the sample after HIC.

Implementation

Chromatography steps were carried out at flow rates of 2 ml / min using a FPLC system (PHARMACIA) at 4 °C with a maximal pressure set to 0.3 MPa. Concentrated CSP containing fractions from HIC (2 ml per run) were loaded on a HiLoad 16/60 Superdex 75 gel-filtration column (PHARMACIA) equilibrated with buffer E. The eluate was collected in 2-ml fractions and analyzed by Schägger SDS page. Fractions containing CSP were concentrated to approximately 20 mg / ml of protein using spin concentrators (Vivaspin 15 - 3000 MWCO PES, SARTORIUS) at 3,500 × g, flash-frozen in liquid nitrogen, and stored at -70 °C. Impure fractions were subjected to another round of gel-filtration chromatography.

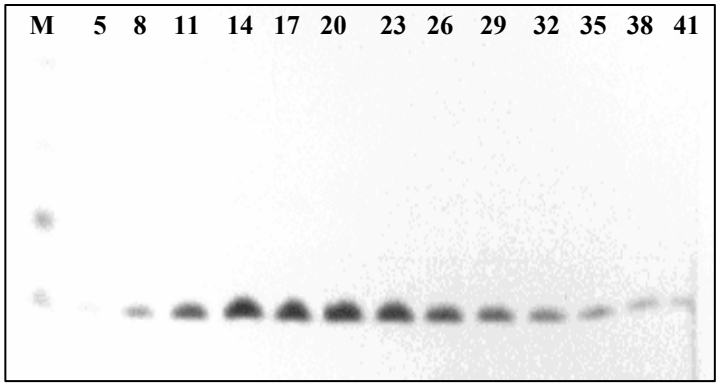

Figure 2.3: Gel-filtration chromatography of *Bs*-CspB R56A. Fraction size was 2 ml. 5 µl of sample were loaded. All fractions shown (numbers) as well as intermediate fractions were pooled and concentrated to a final concentration of 20 mg / ml protein. M – marker, bands at 12,500 Da and 6,500 Da (top to bottom).

2.2.6 Preparative formation of CSP·oligonucleotide complexes

CSP·oligonucleotide complexes were generated for crystallization setups. In order to remove excess single components the complexes were formed with an excess of oligonucleotides and isolated from unbound oligonucleotides by gel-filtration chromatography using a simple hand-packed gravity flow column.

Implementation

For complex formation 200 µl of oligonucleotide solution (oligonucleotides from MWG BIOTECH, c = 6 mM, diluted in buffer E) was added to 400 µl of CSP solution (c_m = 18 mg / ml, c ≈ 2.42 mM) and the mixture was incubated on ice for 30 min. A small self-made column (bed diameter: 1.0 cm, bed height: 14 cm, bed volume: ≈ 12 ml, produced by MDC workshop, Berlin) was packed by hand with Superdex G10 (PHARMACIA) bulk material hydrated in buffer E. The column outlet was connected to a valve to regulate the flow speed. The valve outlet was connected to an apex. The flow speed was set to one drop per two seconds (flow speed = 0.025 ml / s). Using 0.3 ml of 0.01% Coomassie Brillant Blue solution the exclusion volume of the column was determined. The molecular weight of the dye, like that of CSP·oligonucleotide complexes, is too high to enter the pores of the gel matrix. The tracer was observed in fractions of 48 to 75 drops ($V_{exclusion}$ ≈ 2.4 ml - 3.9 ml). For a preparative purification of the complex from excess oligonucleotides 0.3 ml of the complex were loaded on the column and the flowthrough was collected in 0.5-ml fractions, until the column volume was completely passed through. UV / vis spectra of all fractions were recorded (BECKMAN DU 7400) and those containing the complex showed a combination of the spectra of the single components (Figure 2.4). All fractions containing the complex were pooled and concentrated to a final concentration of approximately 20 mg / ml using spin concentrators (Vivaspin 2 – 3,000 MWCO PES, SARTORIUS) at 3500 × g in a cooled tabletop centrifuge. Protein concentration was determined using the Bradford method [100]. The protein concentration was adjusted to 20 mg / ml and the complex solution was used for crystallization trials.

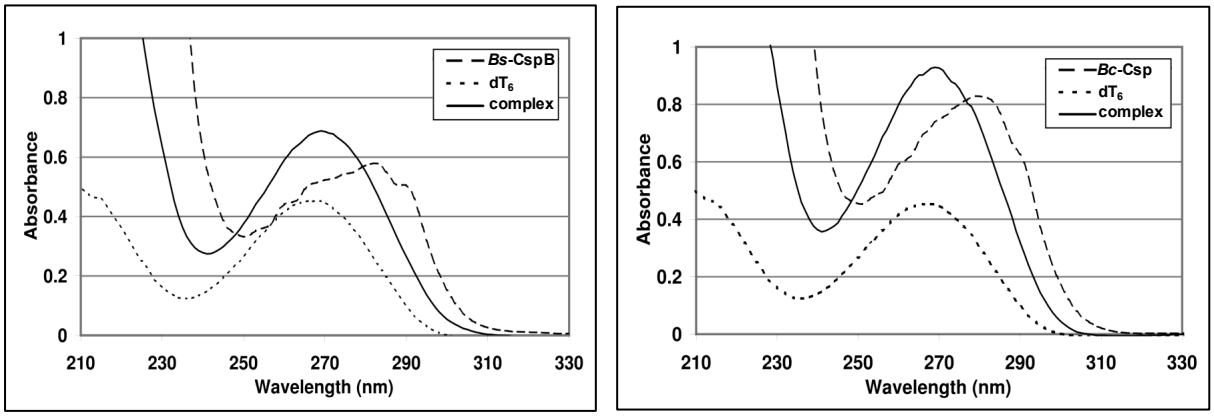

Figure 2.4 UV / vis spectra of single components and purified CSP·dT_6 complexes.

2.2.7 Solutions for protein purification

Lysis buffer	50 mM TRIS / HCl, pH 7.8, 1 mM EDTA, 1 mM DTT, pH 7.8.
Buffer A	50 mM TRIS / HCl, pH 7.8.
Buffer B	50 mM TRIS / HCl, 1 M NaCl, pH 7.8.
Buffer C	50 mM TRIS / HCl, pH 7.6.
Buffer D	50 mM TRIS / HCl, 80% ammonium sulfate, pH 7.6.
Buffer E	50 mM TRIS / HCl, 100 mM NaCl, pH 7.8.

2.3 Spectroscopic methods

2.3.1 Determination of protein and oligonucleotide concentrations using UV / vis spectroscopy

Concentrations of proteins and oligonucleotides were determined by using a UV / vis spectrometer (DU 7400, BECKMAN) and 50-µl microcuvettes with an optical pathlength of 1 cm. Spectra were recorded at room temperature and covered a range of 200 to 350 nm. Buffer without protein and nucleic acids was used as reference. Concentrations of protein and nucleic acids were calculated from the respective light absorptions at 280 and 260 nm according to Lambert-Beer's law:

$$c = \frac{E_\lambda}{\varepsilon_\lambda \cdot l}$$

c - molar concentration, E_λ - extinction at wavelength λ, ε_λ - molar extinction coefficient at wavelength λ, l - optical pathlength.

Equation 2.1

For *Bs*-CspB and *Bc*-Csp molar extinction coefficients at 280 nm (ε_{280}) (see Table 2.4) were computed from amino-acid sequences with the PROTPARAM program (www.expasy.ch, [101]), which uses the method of Gill and von Hippel [102] with some modifications [103]. Extinction coefficients for oligonucleotides at 260 nm (ε_{260}) (see Table 2.5) were calculated from the respective mononucleotides by adding extinction coefficients for the respective mononucleotides at 260 nm and pH 7.0. (dT : 8,400 M^{-1} cm^{-1}, dC: 7,050 M^{-1} cm^{-1}, dG: 12,010 M^{-1} cm^{-1}, dA: 15,200 M^{-1} cm^{-1}, [104]).

Protein	calculated ε_{280}
Bs-CspB	5,747
Bs-CspB, K7A	5,747
Bs-CspB, D25A	5,747
Bs-CspB, R56A	5,747
Bs-CspB, Q59A	5,747
Bs-CspB, M1R/E3K/K65I	5,747
Bs-CspB, A46K/S48R	5,747

Protein	calculated ε_{280}
Bc-Csp	6,990

Table 2.4: Calculated molar extinction coefficients of CSP.

Oligonucleotide	Sequence	calculated ε_{260}
dT_7	TTTTTTT	58,855
CT1	CTTTTTT	57,450
CT2	CTTTTTC	56,100
CT3	CTCTTTC	54,750
CT4	CTCTCTC	53,400
CT5	CTCTTCC	53,400
CT6	CTCCTTC	53,400
CT7	CCCTTTC	53,400
dC_7	CCCCCCC	49,350
CT9	TTTTTTC	57,450
CT10	CTCACCT	61,550
CT11	CTCGTTC	58,360
CT12	CTCTATC	61,550
CT13	CTCTGTC	58,360

Table 2.5: Calculated molar extinction coefficients of oligonucleotides.

2.3.2 Determination of melting temperatures from *Bs*-CspB and mutant variants using CD spectroscopy

CD (circular dichroism) spectroscopy is a sensitive method to analyze structural properties of biological macromolecules using circular polarized light. In all optically active substances the absorbance of left circularly polarized light differs from that of right circularly polarized light:

$$\Delta A_\lambda = A_{\lambda,\mathrm{l}} - A_{\lambda,\mathrm{r}}$$
Equation 2.2

A - absorbance, l and r, indices indicating left and right circular polarization at a given wavelength λ [nm].

For studying structural changes using CD spectroscopy, the molar dichroism ($\Delta\varepsilon_\lambda$) is used, which is concentration-independent :

$$\Delta A_\lambda = (\varepsilon_{\lambda,\mathrm{l}} - \varepsilon_{\lambda,\mathrm{r}})cl$$
Equation 2.3

ε - molar absorption coefficients, c – concentration [$\frac{\mathrm{mol}}{\mathrm{l}}$], l - optical path length [cm], see Equation 2.2.

$$\Delta\varepsilon_\lambda = \varepsilon_{\lambda,\mathrm{l}} - \varepsilon_{\lambda,\mathrm{r}} = \frac{\Delta A_\lambda}{cl}$$
Equation 2.4

The molar dichroism ($\Delta\varepsilon_\lambda = \varepsilon_{\lambda,\mathrm{l}} - \varepsilon_{\lambda,\mathrm{r}}$) is reported in degrees of ellipticity for historical reasons. Dichroism and ellipticity values can be interconverted:

$$\theta_\lambda = 3298 \cdot \Delta\varepsilon_\lambda$$
Equation 2.5

$\Delta\varepsilon_\lambda$ - molar dicroism, θ_λ molar ellipticity.

Secondary structure influences the CD spectra of macromolecules; therefore CD data can be used to trace structural changes of proteins and nucleic acids at varying temperatures, denaturant concentrations or other physicochemical parameters.

CD spectra covering a range of 200 to 250 nm (far UV) were recorded using a spectropolarimeter (J-720, JASCO) in 1 × 1 cm quartz cuvettes of 1 cm optical path length, filled with a total volume of 3.4 ml. The recording speed was 20 nm / min, the filter bandwidth was 1 nm at a resolution of 1 nm, the damping was set to 1 s. The protein concentration was adjusted to 100 µM in measuring buffer. Spectra were recorded at 15 °C and processed by subtracting buffer spectra recorded under the same conditions.

Thermal folding transitions of *Bs*-CspB mutants were recorded at 222.6 nm using a spectropolarimeter (J-720, JASCO). The temperature was increased from 5 °C to 90 °C at a heating rate of 30 °C / h. Absorbance data was corrected by subtracting a buffer control and used for the calculation of θ_λ. Plotting θ_λ plotted against temperature resulted in sigmoidal melting curves which feature single inflection points and support a simple two-state model of protein folding. The data was fitted [16, 17] and melting temperatures were determined.

measuring buffer : 50 mM cacodylate, 100 mM KCl, pH 7.0

2.3.3 Determination of equilibrium dissociation constants (K_D) of CSP·oligonucleotide complexes by fluorescence spectroscopy

The binding of single-stranded nucleic acids was monitored by measuring the fluorescence of a single tryptophane (Trp8), which is quenched upon binding of a ligand [69, 84]. The quench relates to the fraction of ligand-bound protein in solution and can be used to determine affinities (thermodynamic equilibrium constants) and stoichiometric relations of protein·ligand complexes. The equilibrium between association and dissociation of protein and a ligand can be described by the law of mass action:

$$\mathrm{P} + \mathrm{L} \rightleftarrows \mathrm{PL}$$

Equilibrium reaction: association of protein (P) and ligand (L) molecules *vs.* dissociation of protein-ligand complexes (PL).

$$K_\mathrm{D} = \frac{[\mathrm{P}]\cdot[\mathrm{L}]}{[\mathrm{PL}]}$$

Equation 2.6

K_D - dissociation equilibrium constant, K_A - association equilibrium constant, [ ] - concentrations in $\frac{\mathrm{mol}}{\mathrm{l}}$.

$$K_\mathrm{A} = \frac{[\mathrm{PL}]}{[\mathrm{P}]\cdot[\mathrm{L}]} = \frac{1}{K_\mathrm{D}}$$

Equation 2.7

For the determination of the K_D, *Bs*-CspB und *Bc*-Csp were titrated with various oligonucleotides while measuring their Trp8 fluorescence using a spectrofluorimeter equipped with a thermostate cuvette holder and a stirrer (FP 6500, JASCO). The excitation and emission wavelengths were set to 280 nm and 343 nm, and their filter bandwidths were set to 5 nm and 10 nm. The photomultiplier voltage was set to 450 and 500 V, depending on the required sensitivity of the spectrofluorimeter.

For the correct determination of the equilibrium constants protein concentrations had to be below or close to the K_D of the respective nucleoprotein complex and were therefore adjusted within a range of 10 – 100 nM, depending on a previous K_D determination. For the titration of most ligands a protein concentration of 25 nM was used. For very good binders the concentration was lowered to 10 nM. At lower concentrations the fluorescence signal contained too much noise for a valid evaluation. For very weak binders the protein concentration was set to 100 nM in order to detect a stronger fluorescence signal of the protein. This was necessary in order to improve the accuracy of the measurements as the oligonucleotides turned out to increase the fluorescence signal at high concentrations. This additional fluorescence signal was suspected to originate from impurities (faints of fluorescence labels) in the synthesis device (personal communication with Prof. Dr. Welfle, MDC Berlin).

The titrations were carried out in 1 × 1 cm quartz cuvettes, filled with a total volume of 1,750 µl of measuring buffer. During the measurements the sample temperature was set to 15 °C and the sample was constantly stirred. At the beginning of each titration the cuvette was filled with buffer, preincubated for several minutes to establish thermal equilibrium, and the background fluorescence in the absence of the protein was determined. Protein was then added to the solution from a 600 nM stock solution and, after 3 min of equilibration, the fluorescence in the absence of a ligand was determined. Oligonucleotides were added progressively from 0.6, 3, 6 and 60 µM stock solutions. At each step of titration, the fluorescence was recorded 6 times over a period of 10 s. Average values of fluorescence were calculated, which were corrected for the dilution caused by addition of ligand from stocks. Fluorescence data were corrected by subtracting buffer fluorescence and fluorescence of oligonucleotides, which were determined in separate experiments in the absence of the protein. Corrected fluorescence data were normalized to initial protein fluorescence and designated as quench:

$$Q = \frac{F_0 - F}{F_0}$$

F - corrected fluorescence signal from protein, F_0 - protein fluorescence signal in the absence of a ligand.

Equation 2.8

After all fluorescence data of a complete titration experiment was recorded, the concentration of the oligonucleotide used was calculated for each titration step. By plotting quench against concentration hyperbolic binding curves were obtained. The data was fitted according to Eftink and Lohman [12, 13] using the program GraFit 3.0 (Erithacus software, www.erithacus.com):

$$Q = Q_{\max} \cdot \frac{A - \sqrt{A^2 - 4n \cdot [P]_0 \cdot [L]_0}}{2 \cdot [P]_0}, \quad A = K_D + [P]_0 + n \cdot [N]_0$$

Equation 2.9

Q - quench of fluorescence at each step of titration. Q_{max} is reached when no free protein is present (at infinite ligand concentration), n - stoichiometric relation of protein and ligand (number of ligand molecules bound to one protein molecule), $[P]_0$, $[L]_0$ initial protein and ligand concentration, K_D - dissociation equilibrium constant.

measuring buffer : 50 mM cacodylate, 100 mM KCl, pH 7.0, supplemented with 40 µM N-acetyl-tryptophanamide to reduce protein adsorption on the glass surface of the cuvettes.

2.3.4 Determination of kinetic rate constants for CSP·nucleotide complex formation and dissociation using stopped-flow techniques

Following the principles described in 2.3.3 the quenching of the fluorescence signal of *Bs*-CspB was used to determine kinetic rate constants (k_{on}, k_{off}) of *Bs*-CspB·oligonucleotide complex formation and dissociation. The allocation of equilibria between complexed and free state from isolated components was determined using a stopped-flow device (DX.17MV, APPLIED PHOTOPHYSICS connected to a BIO-LOGIC SFM3 spectrofluorimeter equipped with a PMS400 detection unit). The dimensions of the measuring chamber were 10 mm × 2 mm × 1 mm. The fluorophore was excited at 280 nm. To exclude the detection of light below 300 nm, a cuvette filled with acetone was placed between the measuring chamber and the photomultiplier. In this setup fluorescence quenching of Trp8 ($\lambda \geq 320$ nm) upon oligonucleotide binding was detected.

The formation of chemical equilibria depend on the initial protein and ligand concentrations as well as on kinetic rate constants k_{obs}:

$$P + L \xrightarrow{k_{obs}} [\, P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL \,]$$

k_{obs} - rate constant describing equilibrium formation, association, k_{on} - association rate constant, k_{off} dissociation rate constant.

k_{obs} can be determined experimentally by fitting the monoexponential decay of the fluorescence signal from a stopped-flow experiment to the equation:

$$F(t) = A \cdot e^{-k_{obs} \cdot t} + b$$
Equation 2.10

k_{obs} - rate constant describing equilibrium formation, t - time, b - offset (fluorescence after equilibrium has formed; at t = ∞), A - pre-exponential factor.

By varying the ligand concentration while keeping the protein concentration constant, a pseudo-first order relationship is established, which allows to determine kinetic rate constants of complex formation and dissociation by linear fitting of k_{obs} values to ligand concentrations [1]:

$$k_{obs} = k_{on} \cdot [L] + k_{off}$$
Equation 2.11

The association rate constant k_{on} is a kinetic rate constant of a second order reaction, which depends on protein and ligand concentration (see Equation 2.10) while the dissociation rate constant k_{off} is a rate constant from a first order reaction and depends on the concentration of the protein·ligand complex, only. The conditions used for this kinetic analysis (k_{on}·[P] >> k_{off}) allowed a reliable determination of k_{on}. For a reliable determination of k_{off} the protein concentration should be much lower (k_{off} >> k_{on}·[P]). Due to high-affinity binding of oligonucleotides used in this study by *Bs*-CspB suitable protein concentrations would be below 1 nM. The resulting change in the fluorescence signal would be too weak to be detected using the specified equipment. For a better estimation of the dissociation rates, k_{off} was therefore calculated from equilibrium constants K_D:

$$k_{off} = K_D \cdot k_{on}$$
Equation 2.12

measuring buffer : 50 mM cacodylate, 100 mM KCl, pH 7.0

2.4 Determination of molecular masses by analytical equilibrium ultracentrifugation

Analytical equilibrium ultracentrifugation was performed to determine the mass of the *Bs*-CspB·dT$_6$ complex. The sedimentation rate of particles depends on many parameters, which include their molecular masses and specific volumes, the acceleration of the system, the density of the solvent and the temperature. Apart from sedimentation the motion of particles is influenced by diffusion. In order to keep the diffusion and sedimentation rates of protein particles in the same range, strong acceleration is required. Ultracentrifugation is performed until sedimentation and diffusion rates are in equilibrium (Figure 2.5).

Diluted solutions of *Bs*-CspB·dT$_6$ were analyzed using an analytical ultracentrifuge (XL-A, BECKMAN) equipped with absorbance optics. Samples of about 70 µl containing different concentration of complex in the range of 85 - 200 mg / l were centrifuged in externally loaded 6-channel cells against buffer (20 mM TRIS-HCl, 50 mM NaCl, 3 mM $MgCl_2$, pH 7.5) for 2 h at 32,000 rpm (over-speed) followed by 26 h at 28,000 rpm (equilibrium) at 10 °C. The radial concentration distributions of each sample in sedimentation equilibrium were recorded at three different wavelengths between 270 and 220 nm and fitted globally to Equation 2.13 using the program POLYMOLE [105].

$$A_r = A_{rm} e^{MF}$$

Equation 2.13

A_r - radial absorbance at position r, A_{rm} radial absorbance at meniscus position, M - molecular weight.

with $$F = \frac{\left(1 - \rho\bar{\nu}\right)\omega^2\left(r^2 - r_m^2\right)}{2RT}$$

Equation 2.14

ρ is the solvent density, $\bar{\nu}$ is the partial specific volume of the particle, ω is the angular velocity, r_m is the position of the meniscus, R is the gas constant, and T is the absolute temperature.

Because the K_D of the complex is sufficiently low (< 10^{-8} M), the mass of the complexes determined from the distribution of absorbance in equilibrium did not need to be corrected for dissociation effects. Several mass determinations were performed at different complex concentrations. Their results were extrapolated to infinite dilution (Figure 2.6).

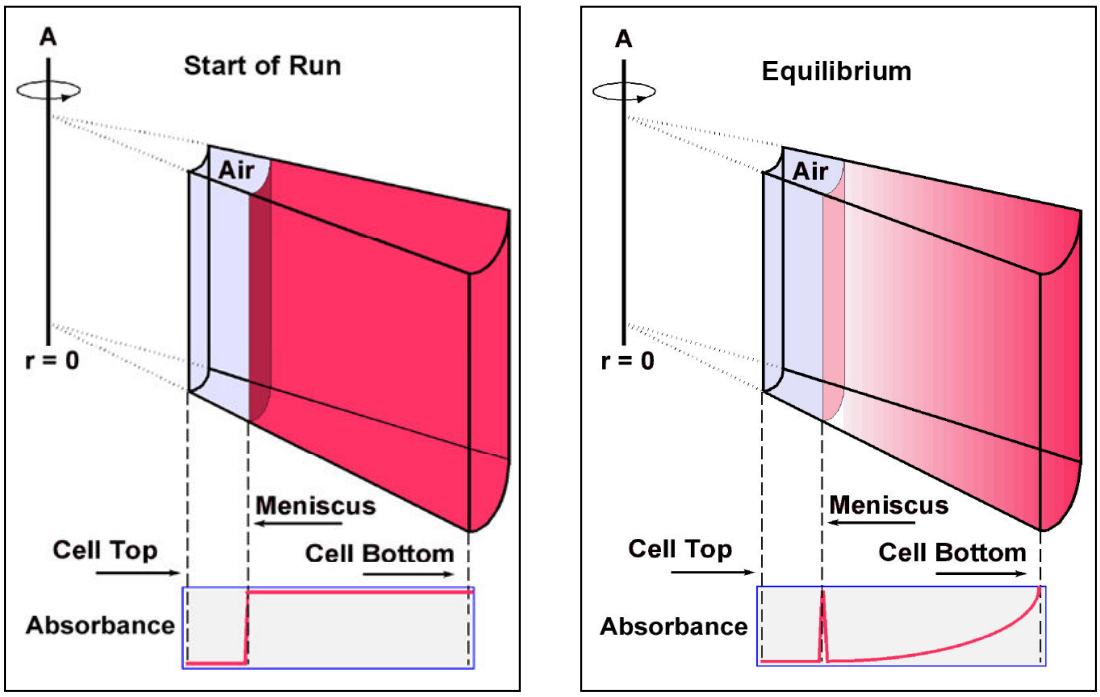

Figure 2.5: Principles of analytical equilibrium ultracentrifugation. Left - at the beginning of an experiment, right - after an equilibrium between diffusion and sedimentation rates has formed.

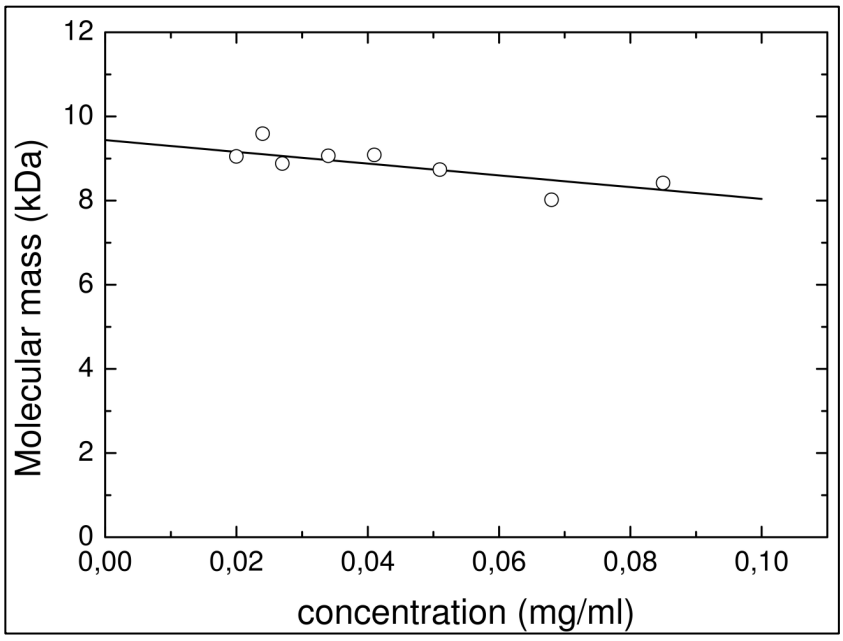

Figure 2.6: Mass determination of the *Bs*-Csp·dT_6 complex in solution. The slow decrease of mass with increasing concentration is typical for a non-interacting (monomer) component and can be addressed to virial effects (Personal communication with Prof. Joachim Behlke, MDC Berlin). The molecular mass extrapolated to infinite dilution is approximately 9.3 kDa, which is close to the sum (9.131 kDa) of the protein (7.368 kDa) and the ligand molecule (1.763 kDa). This supports a 1:1 stoichiometry in solution.

2.5 Structural biology methods

All procedures used in this part are only described in brief, in order to illustrate fundamental tasks required to solve the crystal structures in this work. For further information refer to common textbooks on X-raycrystallography [106, 107].

2.5.1 Growth of macromolecular crystals

All structures of proteins and protein·DNA complexes in this study were determined by X-ray diffraction techniques. A prerequisite for this method is the availability of single crystals of these compounds, as the intensity and design of current X-ray sources is too weak to use diffracted X-ray beams of single molecules for high-resolution structure determination at present. A crystal acts as a signal amplifier due to positive interference of equivalent X-ray beams diffracted at individual molecules throughout the highly ordered crystalline lattice. Crystallization is performed from a solution containing the molecules of interest at very high concentrations. Several parameters influence the formation of well-ordered crystals of reasonable size: These include purity (absence of other molecules which impair crystal-lattice formation) and homogeneity of the materials to be crystallized, as well as solvent conditions and temperature. The initial step in crystallization is the formation of crystal nuclei, which grow out through the addition of further molecules (Figure 2.7). If the concentration of the macromolecules is sufficiently high (supersaturated), individual crystal nuclei can form from interacting molecules until the solution is no longer supersaturated. If the molecule concentration is too high the protein forms an amorphous precipitate, or microcrystals, which originate from (too) many nuclei. If the macromolecule concentration is too low, no nuclei will form. Ideally, only a limited number of nuclei form and crystal growth occurs over a period of a few days. The conditions for crystallization vary for each macromolecule and hence have to be found empirically. Important parameters that influence crystallizations of macromolecules are: the pI of the macromolecule, the pH, ionic strength and polarity of the solvent, and in

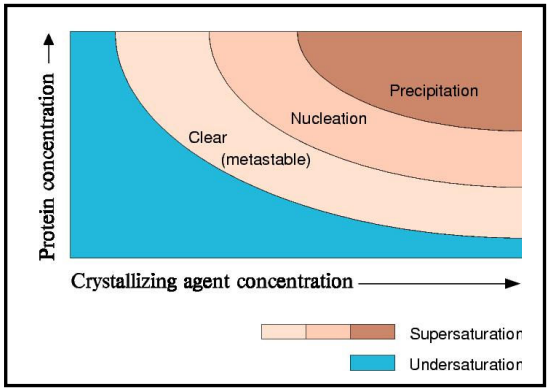

Figure 2.7: Phase diagram of protein crystallization. © 1999-2005 Airlie J McCoy, University of Cambridge. All rights reserved.

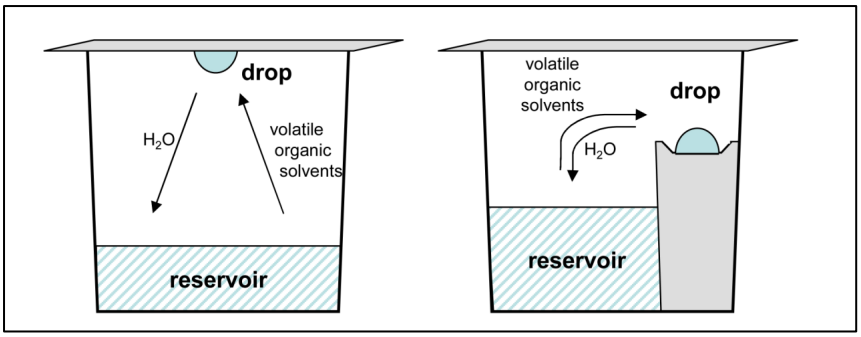

Figure 2.8: Protein crystallization setups using hanging-drop (left) and sitting drop-drop (right) vapour diffusion methods.

some cases the presence of cofactors. Some examples for cofactors are ions, peptides, nucleosides, sugars, lipids, or other macromolecules including proteins or nucleic acids. Cofactors stabilize certain conformations of the macromolecules or they participate in intermolecular packing.

Two methods, the sitting-drop and hanging-drop crystallization, are frequently used for protein crystallization. In both methods drops (crystallization setups with a volume range of range of 200 nl to 10 µl) are formed by mixing aliquots of a protein stock solution with crystallization (trial) solutions. The drops are equilibrated in individual closed containers against a reservoir filled with crystallization solution. The difference between the two methods mainly affects the position of the drop (Figure 2.8). As there is no physical contact between the drop and the reservoir, diffusion of volatile compounds such as water and organic solvents through the gas phase drives the formation of an equilibrium between individual vapor pressures of drop and reservoir. Initially, the vapor pressure of water in the drop is higher than that of the reservoir due to its lower concentration of solvated substances (salts, PEG, organic molecules). As a consequence, water evaporates from the drop and condenses in the reservoir and hence the concentration of the macromolecules and precipitating agents increase in the drop. In crystallization trials which involve crystallization buffer containing volatile organic solvents such as ethanol and isopropanol, these components follow a concentration gradient of the opposite direction, thereby diffusing from the reservoir to the drop, decreasing its polarity, and hence lowering the solubility of the macromolecules therein.

Implementation

The *Bs*-CspB·dT_6 and *Bc*-Csp·dT_6 complexes were crystallized according to the hanging-drop vapor-diffusion method. For crystallization trials, 1 µl of a protein solution was pipetted on a siliconized coverslip (Menzel Gläser, 22 × 22 mm, NUNC, siliconized with Aquasil, HAMPTON RESEARCH) and 1 µl of a precipitant solution was added carefully to its center. The coverslip was then attached to the greased rim (BaySilone, BAYER) of a well in a Linbro plate (ICN FLOW), which was filled with 800 µl precipitant solution. For initial screening the Crystal

Screen I™ and Crystal Screen II™ were used (HAMPTON RESEARCH), which follow the concept of factorial crystal setups [108] and refer to the original sparse matrix screen by Jancarik & Kim [109]. Small crystals of the two CSP·dT_6 complexes grew within a period of three weeks. The initial crystallization conditions were optimized by varying the protein concentration, the pH of the crystallization buffer, and its precipitant concentration, until crystals of larger size and more ordered morphology were obtained.

Mutants of *Bs*-CspB were crystallized using automated crystal setups featuring the sitting-drop vapor-diffusion method. For crystallization trials, 0.2 µl of a protein solution was added to a drop of 0.2 µl precipitant solution previously applied to the drop holder platform of a Crystal Quick microtitre plate (low layout, GREINER BIO-ONE), using a Hydra-plus-One robot (MATRIX TECHNOLOGIES). Crystal formation and growth were monitored using an automated imaging system (based on camera TM-1010, progressive scan from PULNIX). For initial screening, the Crystal Screen I™ and Crystal Screen II™, the PEG-Ion screen, the Salt-HT, and the Index-HT screen (all based on formulations from HAMPTON RESEARCH) were used. Screens were produced using a LISSY XXL robot (ZINSSER). Initial protein crystals appeared within one day. Crystallization conditions were optimized by screening pH, precipitant and protein concentration based on the original screening conditions. In these setups 0.4 µl of the protein stock solution was mixed with 0.4 µl of the appropriate precipitant solution. Crystals of appropriate size (at least 100 × 100 × 50 μm^3) were tested for diffraction and used for data collection.

2.5.2 Freezing of protein crystals

For data collection all crystals were frozen in liquid nitrogen and mounted in a setup with cryogenic cooling. Collecting diffraction data below room temperature is advantageous for the following reasons: 1) a decrease in temperature also decreases the atomic motion and results in a better signal-to-noise ratio of the diffraction data; 2) at lower temperatures the speed of damaging reactions (e.g. decarboxylations of glutamates and aspartates) inside the crystal are decreased, which arise from ionizations of chemical groups and radical formation by X-rays. This results in a better consistency of diffraction data from different times (images). Before the crystals were flash frozen in liquid nitrogen, their mother liquors had to be tested for and optimized against ice-crystal formation upon freezing, which may destroy macromolecular crystals. In addition, ice crystals give rise to strong diffraction spots or rings, which may overlap with the protein diffraction pattern and reduce the amount of usable data.

Implementation

The crystallization conditions for the *Bs*-CspB·dT$_6$ and *Bc*-Csp·dT$_6$ complexes and those for the *Bs*-CspB stability mutant M1R/E3K/K65I turned out to be cryo-conditions, which allowed cryogenic freezing without forming ice crystals. In order to prevent ice formation 15% glycerol were added to the reservoir solution of variant A46K/S48R. For freezing, crystals were removed from the crystallization drop using nylon loops on copper pins (Crystal Cap Copper, MAR RESEARCH), drawn through a 10-µl aliquot of cryosolution (if required), and submerged in liquid nitrogen.

2.5.3 Principles of X-ray diffraction / collection of X-ray diffraction data

When an electromagnetic wave interacts with an object of a similar size as its wavelength the wave is diffracted by this object. Electrons from atomic orbitals act as diffractors for the electromagnetic waves within a wavelength range of about 1 to 10 nm (X-rays). The waves of two or more diffractors add up if they have the same orientation and are in-phase. For the collection X-ray diffraction data in this work the BESSY synchrotron was used [110]. Synchrotrons provide a very brilliant radiation (high intensity and very low beam divergence) and their X-ray wavelength can be tuned.

The crystalline lattice is composed of elementary cells. These are the smallest and simplest repeating units uniquely and completely defining the crystal, which can be generated by using translation operations along the three crystallographic unit cell axes (*a*, *b*, and *c*), only.

In order to describe and predict conditions of positive interference arising from diffraction events on the crystalline lattice, a geometrical model system has been introduced, which regards diffraction conditions according to Bragg's Law (Equation 2.15) as geometrical reflections on grid-planes: The crystal lattice is divided in sets of parallel planes, which are anchored on lattice points from each of the three crystallographic unit axes (*a* - along *x*, *b* - along *y*, *c* - along *z*), of the crystal's coordinate system. Positive interference occurs between

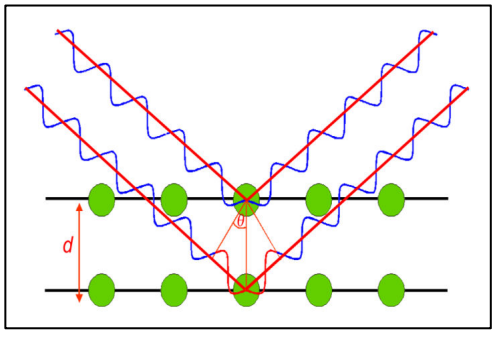

Figure 2.9: Bragg's law visualized schematically: X-rays (blue waves) are diffracted by the crystalline lattice (lattice points displayed as green circles). Individual reflection conditions can be described as reflections of X-ray beams by planes (black lines), which are anchored on lattice points. Positive interference occurs, if Bragg's law is true: Then the path difference between the diffracted waves equals λ or multiples of it, and hence the diffracted beams are in-phase. Whether diffraction conditions are fulfilled depends on the angular orientation (θ) of the crystal to the X-ray beam and the distance between the planes of reflection (d).

X-rays diffracted at layers of one set, if their path difference equals their wavelength, or multiples of it (Figure 2.9). In order to fulfill this condition, primary beam and crystal have to be oriented in a certain way with respect to each other (at a certain angle of reflection).This condition is fulfilled if Bragg's law is true:

$$2d \sin\theta = n\lambda$$
Equation 2.15

θ is the angle of reflection, λ is X-rays wavelength, d is the distance between lattice planes of one set, $n \in \mathrm{N} \setminus \{0\}$.

Each reflection is non-ambiguously defined by the axis intercepts (Miller indices) describing the lattice planes of reflection from which it "originates". The three Miller indices h, k, l, define the axis intercepts of a lattice plane along the crystallographic unit axes (a, b, c) of the crystal's elementary cell. The maximal resolution (d_{max}) is defined by the outermost measurable reflections, which originate from sets of planes with the smallest inter-lattice spacing. The highest theoretical resolution (smallest inter-lattice spacing) equals $0.5 \times \lambda$. The quality of the crystal packing limits d_{max}, however. The highest-resolution protein crystal structure deposited in the Protein Data Bank (PDB) [111] until now was that of crambin at a resolution of 0.54 Å [112]. Typical values for d_{max} are in the range of 1.5 – 2.5 Å, and typically lower than 3 Å for large macromolecular assemblies. d_{max} also defines the minimal distance of molecular features that can be discriminated in the electron density, which equals about $0.75 \cdot d_{max}$. Higher resolutions result in finer resolved electron density maps, which reveal more molecular details. In addition, higher resolutions provide more diffraction data, which can be used to refine further parameters (additional water molecules, anisotropic B factors, hydrogen atoms). The accuracy of structural models based on diffraction data therefore increases with resolution.

Diffraction conditions for a crystal can be described using Bragg's law and the concept of the lattice planes, and only depend on lattice- and orientational parameters and the X-ray wavelength (Figure 2.9). The intensity of each diffracted beam and its phase shift with respect to the primary beam, however, depend on all atom positions and element types in the crystal's unit cell and on their orientation with respect to the planes of reflection (defined by the reflection condition):

$$\varphi_a = 2\pi \cdot (hx_a + ky_a + lz_a)$$
Equation 2.16

The phase shift φ (denoted in radians ($0 \leq \varphi \leq 2\pi$)) caused by a single atom a, depends on the reflection condition (h,k,l), and its atomic position (expressed in fractional coordinates) in the unit cell.

Individual phase shifts caused by atoms in the unit cell add up and define the overall intensity of the diffracted beams. Each diffracted beam can be fully described by its structure factor, a complex vectorial entity that contains information about the amplitude of the beam as well as its phase shift:

$$\vec{F}_{hkl} = |\vec{F}_{hkl}| \cdot e^{i\varphi_{hkl}}$$

Equation 2.17

$\vec{F}_{hkl}$ - structure factor of a diffracted beam, originating from grid plane *hkl,* which defines its reflection condition.
$|\vec{F}_{hkl}|$ - structure factor amplitude (amplitude of that X-ray beam), φ - phase shift (denoted in radians: $0 \leq \varphi \leq 2\pi$).

Structure factors can also be calculated from the atomic model of the crystallographic unit cell:

$$\vec{F}_{hkl} = \sum_{j=1}^{n} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)}$$

Equation 2.18

n - number of atoms in the unit cell of a crystal, f_j atomic scattering factor of atom j, x_j,y_j,z_j - fractional atomic coordinates of atom j within the unit cell, *hkl* - grid plane.

Using CCD-detectors or image plates, the intensities of the diffracted X-rays are being recorded and can be used to calculate structure factor amplitudes:

$$|\vec{F}_{hkl}| \cong \sqrt{I_{hkl}}$$

Equation 2.19

I_{hkl} - Intensity of a beam, diffracted by grid plane *hkl*.
$|\vec{F}_{hkl}|$ its structure-factor amplitude.

In order to enable different diffraction conditions, the crystal is rotated, while diffraction images are being recorded. Each picture depicts the diffractions of a certain angle of rotation (typically 1°). Information about phase shifts cannot be determined, however, using this direct approach.

Implementation

Single-wavelength X-ray data collection was performed at the ID14.1 and ID14.2 beamlines at the BESSY synchrotron, Berlin [110] at 110 K using a mar165CCD camera or a mar345 image-plate detector (MAR RESEARCH).

2.5.4 Data indexing and integration

For the prediction of reflection conditions and the assignment of reflections recorded in the data, the (theoretical) concept of the reciprocal space is used, in which all reflection conditions are defined as a three dimensional lattice. This reciprocal lattic has lattice spacings of reciprocal elementary cell dimensions.

In primitive crystal lattices, unit cells and elementary cells are synonymous, in surface-centered and inner-centered lattices, elementary cells contain two or more unit cells. Depending on the molecular symmetry in the unit cell of the crystal, its molecular structure is fully described by an even smaller entity, the crystal's asymmetric unit: If areas within the unit cell are related by a crystallographic symmetry (2-fold, 3-fold, 4-fold, 6-fold rotation and / or screw axes), the crystal's asymmetric unit is limited to this fraction of the unit cell ($\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{6}$), and the intensities of some reflections are identical and do not need to be measured for a second time. If the crystal is perfectly aligned and no anomalous data are collected (*e.g.* when structures are determined by molecular replacement), a recording range of about 360° / *n* / 2 is sufficient (where *n* is the highest crystallographic symmetry observed). Some redundancy, however, increases the accuracy of the data and allows the estimation of data quality by comparing intensities of reflections (the Bijvoet pairs) that should have the same intensities using R_{sym}:

$$R_{sym} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{\left|\bar{I}_j - I_{j_k}\right|}{\left|\bar{I}_j\right|}$$

Equation 2.20

N is the number of all uniquely measured reflections, *M* is the number of multiple measurements of an unique reflection *j*, $\bar{I}_j$ is its mean intensity based on *M* multiple measurements, and I_{j_k} is the intensity of an individual measurement.

When all reflections have been assigned and all parameters describing the crystal architecture and unit-cell dimensions are in correspondence with the measured data (reciprocal cell axes, point group symmetries, systematically absent reflections), a set of structure factor amplitudes is generated, which can be used for crystal structure determination.

While processing the diffraction images of a dataset, the intensities of all reflections are scaled in resolution- and rotation-angle dependent manner. This is done to compensate for effects arising from X-ray absorption and X-ray scattering (Compton effect) by the crystal and adherent materials (nylon loop, cryosolution, precipitates), as well as for crystal damage and anisotropic diffraction (caused by differences in crystalline order). At the end, a unique set of structure factor amplitudes is generated.

Implementation

During data collection, space groups and unit-cell parameters were analyzed and strategies for obtaining complete data were developed using the XDS software [113]. The final assignment of space groups and unit-cell parameters was performed using the XDS package in the home workstation pool for the complete datasets. Averaging of intensities from related reflections and creation of datasets with a set of unique reflections were performed using the XSCALE program [113]. The raw data files were converted into the binary MTZ format for further processing by using the XDSCONV program [113], and the CCP4 suite [114].

2.5.5 Solution of the crystallographic phase problem by molecular replacement

The structure factor of a diffracted beam can be constructed from its amplitude if its phase shift is known (Equation 2.17). The availability of crystallographic models therefore allows phase determination by molecular replacement. In this method, atomic coordinates of crystallographic models from similar macromolecules are used to calculate the missing experimental phases, whereas structure-factor amplitudes are based on experimental data. For molecular replacement, the structural model must have an architecture (arrangement and spacing of macromolecules) which is equivalent to that in the respective crystal on which the dataset is based: Both model and crystal architecture must be isomorphous. In cases of disagreeing architectures, the replacement model has to be reoriented and shifted in the unit cell appropriately, until an isomorphous model is obtained [115]. Model architecture is tested for isomorphism using the Patterson function, which can be calculated using structure-factor amplitudes from experimental data (Equation 2.17) as well as from atomic coordinates (Equation 2.18). Phase information is not required:

$$P(u,v,w) = \frac{1}{V}\sum_{hkl}\left|\vec{F}_{hkl}\right|^2 e^{-2\pi(hu+kv+lw)}$$

Equation 2.21

The Patterson function is based on structure-factor amplitudes. In illustrative terms, a Patterson function represents the convolution of two identical electron densities. Both densities are shifted with respect to each other by u (along x), v (along y), and w (along z). Patterson functions hence reveal interatomic distance vectors of all molecules in and between unit cells, shifted into the origin of the coordinate system. At the end of each vector a maximum is observed in the Patterson function.

For finding the isomorphous replacement model, the Patterson function based on experimental data is multiplied with the Patterson function of different model architectures. When the isomorphous model architecture is found, both densities overlap and their product it maximized:

$$S = \int_{r=r_1}^{r_2} P_{\text{experimental}} \cdot P_{\text{model}} dr$$

Equation 2.22

A scoring function used in the search for a correct molecular replacement model. Patterson densities based on experimental data ($P_{\text{experimental}}$) and a model (P_{model}) are multiplied. The product density is integrated within a centrosymmetric sphere shell in the interatomic distance-vector space. The inner and outer borders of this sphere are defined by r_1 and r_2.

The computational effort of this search can be decreased, if the model's orientation and position are investigated independently by using intramolecular and intermolecular vectors for rotational and translational searches in consecutive order. This is possible by choosing the shell of integration appropriately. The search process thus involves three steps: Rotational and translational searches followed by a simple rigid-body refinement (see also Chapter 2.5.7). At the end of each step, crystallographic *R* factors are calculated in order to evaluate the agreement of structure factor amplitudes from model and experimental data. Based on the Patterson scoring function (Equation 2.22) and the crystallographic *R* factor (Equation 2.27), the best solutions were selected for further processing. When the searches were finished, the solutions were checked for intermolecular overlaps (clashes of atoms). Correct molecular replacement solutions tend to show the best agreement of experimental and calculated structure factor amplitudes, as indicated by lowest crystallographic *R* factors, and the highest Patterson scoring function values without causing extensive intermolecular overlaps.

Implementation

The phase problem could be solved *via* molecular replacement using the AMoRe software [116]. Structures of *Bs*-CspB (1CSP & 1CSQ, Protein Data Bank) were used as templates for *Bs*-CspB based structures. A structure of *Bc*-Csp (1C9O) was used as a template for *Bc*-Csp based structures. When parameters for space-group and unit-cell parameters could not uniquely be assigned in the indexing step, all potential solutions were used for molecular replacement. Rotational and translational parameters of the best solutions were fed into the PDBSET program of the CCP4 suite to orient the search models and generate starting models for the refinement process. Prior to automated refinement these models were visually inspected using the O software [117].

2.5.6 Calculation of electron density maps

Electron density maps show the distribution of electrons within a crystal. Peaks in the electron density maps indicate atom positions. The electron density at any given position within the unit cell can be calculated from the structure factors by performing a Fourier transform:

$$\rho_{x,y,z} = \frac{1}{V}\sum_h\sum_k\sum_l \vec{F}_{hkl} \cdot e^{-2\pi i(hx+ky+lz)}$$

Equation 2.23

$\rho_{x,y,z}$ is the electron density (probability of occurrence) at position *x, y, z*. *V* is the volume of the unit cell. $\vec{F}_{hkl}$ is the structure factor of reflection condition *hkl*. The accuracy of ρ increases with the number of available structure factors.

Structure factors, can be calculated from a Fourier backtransform of the electron density in the unit cell:

$$\vec{F}_{hkl} = V\sum_x\sum_y\sum_z \rho \cdot e^{2\pi i(hx+ky+lz)}$$

Equation 2.24

See Equation 2.23. The accuracy of $\vec{F}_{hkl}$ increases with the completeness of the atomic model.

As previously denoted, the information about phase shifts of diffracted beams is lost in a typical diffraction experiment and has to be determined using other experimental methods, such as isomorphous replacement, multiple anomalous diffraction, single wavelength anomalous diffraction, or molecular replacement.

Difference electron density maps are use to compare electron densities based on atomic models with those based on experimental data in real space, in order to detect disagreements:

$2F_o$-F_c density map

$$\rho_{x,y,z} = \frac{1}{V}\sum_{hkl}\left(2\left|\vec{F}_{hkl,obs}\right| - \left|\vec{F}_{hkl,calc}\right|\right) \cdot e^{-2\pi i(hx+ky+lz)+i\varphi_{calc}}$$

Equation 2.25

F_o-F_c density map

$$\rho_{x,y,z} = \frac{1}{V}\sum_{hkl}\left(\left|\vec{F}_{hkl,obs}\right| - \left|\vec{F}_{hkl,calc}\right|\right) \cdot e^{-2\pi i(hx+ky+lz)+i\varphi_{calc}}$$

Equation 2.26

Difference electron density at a given position x, y, z ($\rho_{x,y,z}$) is calculated to visualize differences between experimental data (crystal structure factor amplitudes $\left|F_{hkl,obs}\right|$) and their calculated counterparts from a structural model ($\left|F_{hkl,calc}\right|$). Phases ($\varphi$) are derived from the model, only.

Electron density is visualized as an isosurface map contoured at a certain threshold, which is expressed by multiples of the map's standard deviation σ. F_o-F_c maps are usually contoured either at +3σ (positive difference density) or at -3σ (negative difference density). These maps

display differences between electron densities based on structural models and experimentally determined crystallographic data. Patches of positive difference density indicate missing parts in the model, whereas patches of negative difference density indicate too much density in the model, which may be caused by false or incorrectly positioned groups. $2F_o$-F_c maps, usually contoured at 1σ, display electron density based on the experimental data and highlight missing density in the model.

Implementation

$2F_o$-F_c and F_o-F_c difference electron density maps were calculated based on atomic coordinates and X-ray diffraction data using the FFT program [114]. Maps were extended using the the Extend program [114] and converted into the BRIX map format using the mapman program [118]. The O software [117] was used to display density contour maps and atomic models in real space.

2.5.7 Structure refinement

Using difference electron density maps it is possible to refine a structural model in real space. This is done by adjusting the structural model by adding, removing, and reorienting groups until the model corresponds to the observed difference densities in real space. Manual optimization is followed by computational optimization of atomic positions in reciprocal space. In this approach, atomic positions of the model are varied systematically. At the end of each refinement step structure factors are calculated for the refined model, which are compared to experimental data using the crystallographic *R*-factors. Refinement is carried out, until after some rounds of equilibration no further improvement of *R*-factors is achieved.

$$R_{\text{model}} = \sum_{j=1}^{N} \frac{\left\| \left| \vec{F}_{obs,j} \right| - \left| \vec{F}_{calc,j} \right| \right\|}{\left| \vec{F}_{obs,j} \right|}$$

Equation 2.27

N is the total number of uniquely measured reflections, and $\vec{F}_{obs,j}$ is the structure factor of reflection j, determined experimentally, $\vec{F}_{calc,j}$ is its calculated counterpart.

In the structure refinement the cross-validation method [2] was used to prevent overfitting of structural data: A small set of data (5-10% of the unique reflections) was usually set aside and not used for refinement, to have an independent measure (R_{free}) of the fit of the refined structural model to the experimental data.

In addition to experimental data, additional parameters specifying chemical bond lengths, bond angles, torsion angles, and planar groups were restrained using force-field methods

[119] to prevent adjustments in the refinement which would violate the general stereochemistry of macromolecules:

$$E_{\text{total}} = E_{\text{X-ray}} + n_1\left(n_2 \cdot E_{\text{bond angles}} + n_3 \cdot E_{\text{bond distances}} + n_4 \cdot E_{\text{torsion angles}} + n_5 \cdot E_{\text{planar groups}}\right)$$

Equation 2.28

E-terms represent energy terms based on structural data (X-ray) and force field parameters. $n_1 - n_5$ are weighting factors.

Each refinement step was accepted, if an overall decrease in energy (Equation 2.28) was observed. In the first round of automated refinement, a rigid-body refinement was performed. Here, the starting model generated by molecular replacement was fitted to the experimental data by fine-tuning the position of the whole structural model within the unit cell without allowing individual atoms to move. In the consecutive steps of automated refinement, optimization of individual atomic positions was performed. This required three parameters (x, y, z) per atom. Positional refinement was followed by refining atomic displacement. This term summarizes individual atomic motion as well as static or dynamic disorder (movement) of atomic groups in the crystal. Depending on the resolution of data, atomic displacement was either refined isotropically, using spherical models, or anisotropically, using ellipsoid models for describing the displacement of individual atoms. Isotropic description of atomic displacement requires one additional parameter, the *B* factor (mean atomic displacement or Debye-Waller correction) for structure factor calculation:

$$\vec{F}(h,k,l) = \sum_{j=1}^{n} f_j \, e^{-\frac{B_j \sin(\theta)^2}{\lambda^2}} e^{2\pi \mathrm{i}\left(\mathrm{h}x_j + k y_j + l z_j\right)}$$

Equation 2.29

An additional term is introduced into the structure factor equation (Equation 2.20), which corrects the atomic scattering factor using an isotropic model of displacement. B_j, the "B-factor" is the surface area from a sphere describing the mean displacement of an atom j from its coordinates.

The radius of displacement can be calculated from the B-factor:

$$r = \sqrt{\frac{B}{8\pi^2}}$$

Equation 2.30

radius *r* defines a sphere of mean displacement, see Equation 2.29.

For refining the atomic displacement anisotropically, six additional parameters are required to define a thermal ellipsoid: 3 shape parameters (half axes) and 3 orientational parameters. This method requires more than twice as many data as the description of isotropic displacement and can therefore only be used if the resolution is sufficiently high (typically better than 1.4 Å). The number of refinement parameters can be reduced by using collaborative rather

than individual atomic displacements. This is achieved by defining parts of the structural model as a TLS group, whose translational, librational, and screw displacement parameters are refined as a pseudo rigid body [120]. For the refinement of displacement parameters of each TLS group 20 parameters are required.

Each cycle of automated refinement is followed by the calculation of difference density maps and a manual inspection and optimization of the refined model in real space. Improved structural models obtained from automated refinement procedures allow the calculation of better phases and hence provide better electron density maps, which reveal more molecular details of the model in real space. In this way, automated refinement greatly contributes to the fast adjustment of model parameters to the experimental data, whereas real-space optimization helps to overcome local scoring function minima in automated refinement, highlights inconsistencies between model and data, and thus enables the completion of the model.

Implementation

For the structural refinement approximately 95% of the reflections, the work set, were used in all structures. The remaining reflections, the free set, were kept separate and used for evaluating the refinement process by a cross-validation method [2]. The refinement included automated positional and temperature factor refinement based on maximum likelihood methods implemented in REFMAC5 [121], automated molecular building routines for water molecules using ARP/wARP [122], and manual model building and optimization of atomic positional parameters using the O software [117]. The structural models were gradually altered and extended in such a way, that structure factor amplitudes derived from them increasingly resembled their experimentally determined counterparts, as indicated by decreasing R_{free} and R_{work} factors (Equation 2.20). In the CSP·dT_6 complex structures large patches of positive electron density could be gradually filled with atomic models of nucleobases, deoxyribose sugars, and phosphate groups. The models of the *Bs*-CspB stability mutants, which differed from the wildtype protein only by up to three amino acids, caused big model bias, and even after removing differing groups the difference density was insufficient to allow the introduction of alternative sidechains. Consequently, refinements were (re)started using poly-alanine models and all sidechains were gradually built into the electron density.

When the refinement of macromolecular structures could not be advanced any further, automated and manual searches of water molecules were initiated. Water positions were checked for consistency and maintained in the refinement model only, if the following criteria were fulfilled: 1) spherical electron density maxima are present in both the $2F_o$-F_c map contoured at 1σ and the F_o-F_c map contoured at 3σ at each potential water position; 2) each water molecule

is connected to other solvent molecules or to polar groups from macromolecules by hydrogen bonds within a distance of 2.4 - 3.6 Å, and 3) their B-factors were not exceedingly high. Some molecules initially identified as waters could be assigned as ions, based on distance criteria, geometrical coordination of surrounding groups, and occurrence in the crystallization setups. In the structure of *Bc*-Csp two 2-methyl-2,4-pentanediol (MPD) molecules were found in the protein-solvent interface. Each building cycle of solvent molecules was followed by further cycles of positional and temperature-factor refinement. When the solvent structure was complete, virtual hydrogen atoms were included in the atomic model, which were constrained to associated heavy atoms. These "riding" hydrogens were used for improved geometrical and interatomic refinement procedures.

For the *Bc*-Csp·dT_6 structure, which is based on diffraction data with a maximum resolution of 1.29 Å, the displacement parameters of all atoms were refined anisotropically in the final steps of refinement, using ellipsoids for the description of the atomic displacement. For the *Bs*-CspB·dT_6 structure based on diffraction data with a maximum resolution of 1.78 Å, three groups were defined, whose combined displacement was refined by TLS refinement, in addition to refining isotropic displacement of individual atoms:

TLS group 1:	the complete protein molecule
TLS group 2:	nucleotides 1 to 5
TLS group 3:	nucleotide 6

For the evaluation of atomic displacement in this structure, anisotropic displacement parameters were calculated for individual atoms by combining isotropic displacement parameters with tensors from TLS groups using the TLSANAL program from the CCP4 suite [114]. Likewise, the displacement parameters of the of *Bs*-CspB mutant A46/S48R were refined by combining isotropic atomic displacement with TLS refinement of the complete molecule being defined as a single TLS group, using diffraction data with a maximum resolution of 2.30 Å. Atomic displacement in the structure of *Bs*-CspB mutant M1R/E3K/K65I was refined isotropically. Due to the limited resolution (d_{max} = 2.55 Å) of the diffraction data no TLS refinement was performed.

2.5.8 Structure evaluation and validation

Various methods developed for the assessment of crystallographic data and structural models were used in this work: An important indicator for the quality of the diffraction data is the symmetry *R* factor (R_{sym}, Equation 2.20, or R_{meas} [3]), which gives an overall estimation of the consistency of intensities derived from symmetry-related reflections that should be identical. Good values for symmetry *R* factors are below 7%. Values above 12% may indicate problems during data collection (e.g. anisotropic diffraction or radiation damage). High values (above 30%), however, indicate that a wrong space group may have been used for indexing and merging the data.

The crystallographic *R* factors (Equation 2.27) were used as measures for the increasing correlation of experimental and model data throughout the refinement. The R_{free} [2] is especially relevant, because it is based on data which have not been used in the refinement process. As a rule of thumb, structures at a resolution of 2.5 Å or better with crystallographic *R* factors below 25% are largely correct. By plotting crystallographic *R*-factors as a function of resolution (Luzzati plots, [123]) the maximal level of uncertainty in the atomic coordinates could be estimated.

Another quality indicator is the real-space correlation coefficient, which describes the agreement between electron density maps based on the model and experimental data [124]. Real-space correlation plots were used to detect regions in the model which are either poorly defined in the electron density or incorrectly built in the model.

In addition to consistency checks involving model and experimental data, the models were checked for plausibility by comparing some of their structural properties to those of other well refined crystal structures and other types of structure based information, using statistical methods: This involves individual parameters such as bond lengths, bond angles [125], planarities of certain groups [121], two dimensional analyses of backbone torsion angles ("Ramachandran plots", [126]), and analyses of torsions from sidechains and ligand molecules [127]. Further important parameters for evaluating plausibilities of structural models are non-covalent interatomic distances and orientations of functional groups with respect to each other [128]. In particular, the models were checked for distance violations (bumps) and stabilizing effects typically observed in protein structures (van-der-Waals contacts, hydrophobic interactions, hydrogen bonds, salt bridges). Apart from macromolecules (proteins, DNA, ligands) the latter analyses also include solvent molecules (water, ions) which are ordered in the structure. In addition, information about the location of functional groups was considered: In the inside

of a protein charged groups rarely occur, whereas hydrophobic groups tend to be located here but rarely on the protein surface.

Implementation

R_{meas} values were taken from the report of the XSCALE program [113] using all data of a diffraction dataset. Crystallographic *R* factors and RMS values of basic stereochemical parameters were reported by the REFMAC5 structure refinement program [121]. Backbone and sidechain torsion angles as well as model geometries were analyzed using the Procheck [129] program. The WHATCHECK [130] program was used to check the model for further structural inconsistencies (short interatomic distances, unsatisfied hydrogen bonding, improper geometries of functional groups, overall packing quality, backbone torsion distribution and quality) . Real-space correlation coefficients between structural models and experimental data and maximal coordinate errors were calculated using the SFCHECK software [15].

2.5.9 Molecular visualization and evaluation

All molecular graphics in this publication were generated using the Pymol software [131]. Electrostatic surface potentials were calculated using the adaptive Poisson-Solver (APBS) [10]. Molecular properties such as torsion angles, interatomic distances, hydrogen-bonding patterns, van-der-Waals contacts, symmetry-related interactions, *etc*., were evaluated using the WHAT IF program [128] and the CCP4 program suite [114].

2.6 List of companies

APPLIED PHOTOPHYSICS LTD., 203/205 Kingston Road, Leatherhead Surrey, KT22 7PB, United Kingdom.

BAYER MaterialScience AG, Communications, Geb. W1, Kaiser-Wilhelm-Allee 1, 51368 Leverkusen, Germany.

BECKMAN Coulter GmbH - Biomedical Research, Europark Fichtenhain B13, 47807 Krefeld, Germany.

BIO-LOGIC, 1, rue de l' Europe, F-38640 Claix, France.

BIO-RAD Laboratories GmbH, Heidemannstrasse 164, 80939 München, Germany.

GREINER BIO-ONE GmbH, Rosenkamper Straße 6, 42719 Solingen-Wald, Germany.

MATRIX TECHNOLOGIES, 22 Friars Drive, Hudson, NH 03051, USA.

HAMPTON RESEARCH, 34 Journey, Aliso Viejo, CA 92656, USA.

ICN FLOW, 3300 Hyland Ave, Costa Mesa, CA 92626 USA.

JASCO Labor- und Datentechnik GmbH, Robert-Bosch-Straße 11, 64823 Groß-Umstadt, Germany.

NEW ENGLAND BIOLABS, 240 County Road, Ipswich, MA 01938-2723, USA.

NUNC GmbH & Co. KG, Rheingaustraße 32, 65201 Wiesbaden, Germany.

MAR RESEARCH GmbH, Hans-Böckler-Ring 17, 22851 Norderstedt, Germany.

MILLIPORE GmbH, Am Kronberger Hang 5, 65824 Schwalbach, Germany.

MWG BIOTECH AG, Anzingerstr. 7a, 85560 Ebersberg, Germany.

PHARMACIA currently belongs to GE Healthcare: Amersham Biosciences Europe GmbH, Munzinger Straße 9, 79111 Freiburg, Germany.

PULNIX Inc, 625 River Oaks Parkway, San Jose, CA 95134, USA.

ROTH, Schoemperlenstraße 3-5, 76185 Karlsruhe, Germany.

SAVAN Group, PO Box 2061, Steyning, Sussex BN5 0AQ, United Kingdom.

SARTORIUS, Weender Landstrasse 94-108, 37075 Goettingen, Germany.

SMI-AMINCO, Abcoude, the Netherlands.

STRATAGENE, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA.

ZINSSER Analytic GmbH, Eschborner Landstrasse 135, 60489 Frankfurt, Germany.