Chapter 4 - The architecture of a *Bc*-Csp domain swap

4.1 Results

The crystal structure of the *Bc*-Csp·dT$_6$ complex features a domain swap. While structural and biochemical aspects of nucleic acid binding to *Bc*-Csp, which provide insight into CSP function on the molecular level, are dealt with in Chapter 3, the structural reorganization associated with the swapped subdomains provides further insight into the CSP architecture and will be the subject of this chapter. Some redundancy involving the results in Chapters 3 and 4 has been allowed in order to simplify understanding the line of the argument in both chapters. Solving the crystal structure of a *Bc*-Csp·dT$_6$ complex required somewhat more sophistication in terms of assigning the correct space-group and unit-cell parameters, indexing diffraction data, calculating molecular replacement, and revising structural models than for the other crystal structures solved in this work. Its structure-solution and refinement processes will therefore be described in detail.

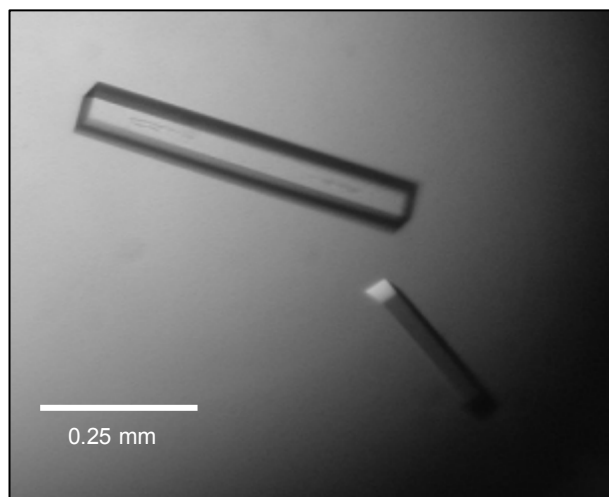| ***Crystallization*** |
| --- |
| **reservoir solution**
 35% 2-methyl-2,4-pentanediol
 0.1 M sodium acetate pH 4.6
 0.02 M CaCl$_2$ |
| **sample buffer**
 20 mM HEPES (pH 7.5)
 50 mM NaCl
  3 mM MgCl$_2$
 18 mg/ml *Bc*-Csp·dT$_6$ complex |
| **crystallization setup**
 hanging drop crystallization, <br> manual setup

 1 µl sample : 1 µl reservoir solution |

Figure 4.1: Protein crystallization of the *Bc*-Csp·dT$_6$ complex. The crystal shown on the right was crystallized according to the given conditions and used for data collection.

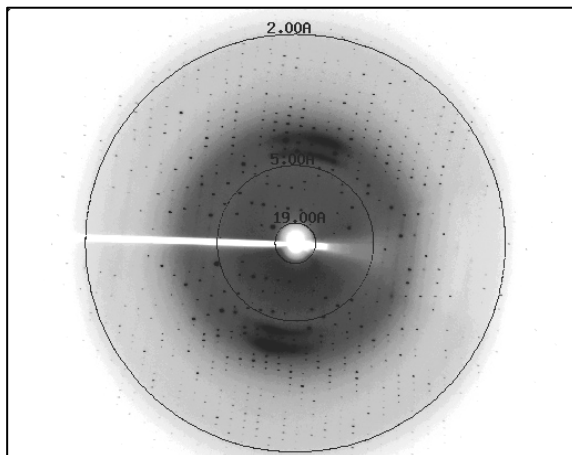| Data collection & processing	
wavelength (Å)	0.9184
resolution (Å)	20.00 - 1.29
resolution (Å) (last shell)	1.40 - 1.29
space group	$P2_12_12$
temperature (K)	100
X-ray source	BESSY BL 14.1
detector	mar345IP
unit-cell parameters a (Å) b (Å) c (Å)	74.34 64.89 31.20
unique reflections (last shell)	37,691 (7,838)
$I/\sigma(I)$ (last shell	14.8 (5.2)
% data completen. (last shell)	97.0 (94.4)
R_{meas} [a] (%) (last shell)	6.6 (32.0)

Figure 4.2: X-ray-diffraction-data statistics of a *Bc*-Csp·dT$_6$ complex crystal based on the final assignment of space-group and unit-cell parameters. One diffraction image is shown on the right. The angle of rotation is 1°. Resolution (Å) is indicated by rings.
[a] R_{meas} is a redundancy-independent R factor, which correlates intensities from symmetry-related reflections [3].
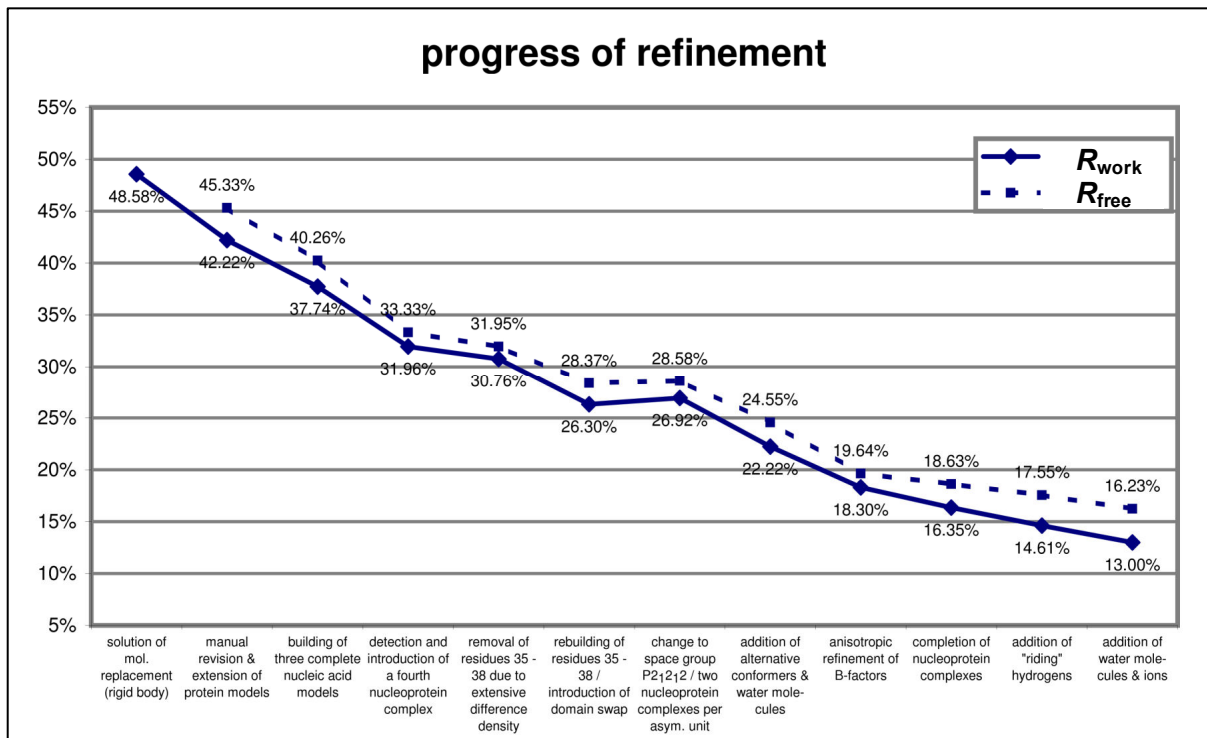

Figure 4.3: Refinement progress of the *Bc*-Csp·dT$_6$ complex crystal structure.

Model building & refinement	
PDB model for mol. replacement	1C9O
PDB access. of the refined structure	2HAX
reflections - working set	34221
reflections - free set (= 5.0%)	1908
R_{work}[a] (%)	16.23
R_{free}[a] (%)	13.00
mean B factor (Å^2)	10.84
number of non-hydrogen atoms	1,614
number of protein molecules	2
number of dT_6 molecules	2
number of water molecules	234

Model evaluation	
root-mean-square deviations	
bond lengths (Å)	0.016
bond angles (°)	1.605
torsion angles (°)	4.656
planarity (Å)	0.009
Ramachandran statistics	
residues in allowed regions	95.3%
residues in additionally allowed regions	4.7%
residues in generously allowed regions	0%
residues in forbidden regions	0%
real-space correlation coefficient[b]	95.8%
ESD coordinate uncertainty (Å)[b]	0.066
coord. uncert. from Luzzati plot (Å)[b]	0.145

Table 4.1: Building, refinement, and evaluation of the atomic model based on diffraction data from *Bc*-Csp·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].

4.1.1 Crystallization and structure solution of a *Bc*-Csp·dT_6 complex (2)

A complex of *Bc*-Csp and hexathymidine was crystallized (Figure 4.1). Diffraction data were collected up to a resolution of 1.29 Å at BESSY synchrotron beamline 14.1 [110] (Figure 4.2). An overview of the refinement process is given in Figure 4.3. The diffraction data is in agreement with an orthorhombic Bravais lattice and thus limits the crystal architecture to space groups P222, P2$_1$22, P2$_1$2$_1$2, or P2$_1$2$_1$2$_1$. Merging diffraction data accordingly using standard parameters was not successful, however, as indicated by a R_{meas} value of above 20% for the merged data. The data were therefore processed and merged in monoclinic space groups. In P2$_1$ three molecular replacement solutions were found using a single structural model of *Bc*-Csp (1C9O, chain A). Their translational and rotational parameters were applied to the search model and a trimeric structural model (protein molecule A, B, and C) of the crystal's asymmetric unit was generated.

After some initial rounds of rigid-body refinement to optimize molecular positions and some rounds of atomic positional refinement, the crystallographic R_{work} and R_{free} values of the struc-
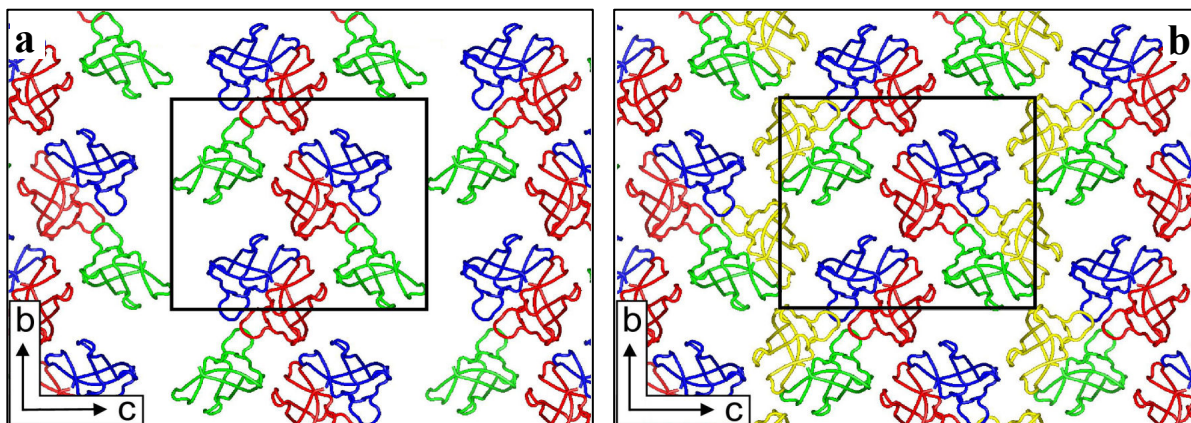
Figure 4.4: Initial (a) and optimized (b) molecular-replacement solutions for *Bc*-Csp·dT$_6$ in space group P2$_1$. View along the *a* axis. α-carbon traces of protein molecules are depicted as colored lines. The content of a unit cell is enclosed in a black box. a) Using the search model 1CSP, three protein molecules could be assigned inside the asymmetric unit (molecules A - red, B - green, C - blue). The packing of molecules is incomplete: Large gaps are visible in the molecular arrangement. At this stage no fourth molecule could be assigned. b) After some rounds of manual and automated refinement the anticipated fourth molecular replacement solution was found using the partially refined structure (molecule D – yellow) as a search model.

tural model decreased of to 42.2% and 45.3%, respectively. The packing of protein molecules inside the crystal appeared to be incomplete, however: the molecular arrangement contained gaps, which were big enough to accommodate another search model (Figure 4.4a). At this stage of the refinement it was not possible to find a fourth *Bc*-Csp position by molecular replacement, however. F_o-F_c maps revealed large patches of positive difference density featuring several nucleobase groups and sugar rings of the ligand. Into this density three dT$_6$ molecules (chains E, F and G) were built. In addition, protein groups were revised and parts from the model surrounded by negative difference density were excluded from the model.

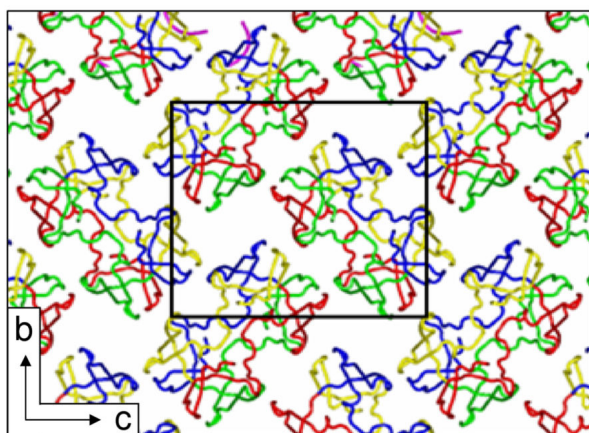After the next round of automated refinement, which from now on also included an isotropic refinement of the atomic displacement (mean atomic *B* factors), the quality of the atomic model improved significantly, as indicated by a 5% decrease in crystallographic *R* factors. The F_o-F_c map based on the improved model contained further patches of interpretable F_o-F_c difference density in gapped regions of the crystal lattice. Using the partially refined model, another round of molecular replacement was calculated, which provided the anticipated fourth solution and decreased the

Figure 4.5: Formation of domain-swapped tetramers in the *Bc*-Csp·dT$_6$ crystal. The layout follows Figure 4.4. The content of a unit cell is enclosed by a black box. View along the *a* axis.

crystallographic *R* factors by approximately 6% (Figure 4.4b, protein molecule D). The F_o-F_c difference density of the resulting tetrameric model indicated a conflicting progression of residues 35 to 39. This region was excluded from the model and another round of refinement was performed, which revealed a domain-swapped architecture and required the reallocation of residues 37 to 67 of the four structural models (Figure 4.5, Figure 4.9): In contrast to all other CSP structures reported so far, whose single globular domains are composed of single protein chains, in the *Bc*-Csp·dT$_6$ structure the corresponding globular entities are formed by N-terminal (residues 1-36) and C-terminal (residues 37-66) subdomains of two different protein chains. Rebuilding the region of transition (residues 35 to 39) and reassigning the molecular architecture accordingly decreased the crystallographic *R* factors by approximately 5%.

A thorough assessment of the molecular arrangement revealed two apparently non-crystallographic symmetry elements in addition to the P2$_1$ screw axis along the *b* axis, which is part of the P2$_1$ space group: 2-fold rotation axes were observed parallel to the *a* axis, and 2$_1$ screw axes were observed parallel to the *c* axis of the unit cell (Figure 4.6). In order to analyze symmetry relations in reciprocal space, 0- and 1-layer precession images were generated from diffraction data processed in space group P1 (Figure 4.7). Symmetry relationships between reflections included 2-fold rotations along the *h*, *k*, and *l* directions and systematic absences of odd reflections along the 0*k*0, 00*l* layer lines, and hence supported the assignment of two 2$_1$ screw axes parallel to *b* and *c*, and one 2-fold rotation axes parallel to *a*. This assignment of Miller indices and crystallographic axes does not comply with the prerequisites of P2$_1$2$_1$2 ,
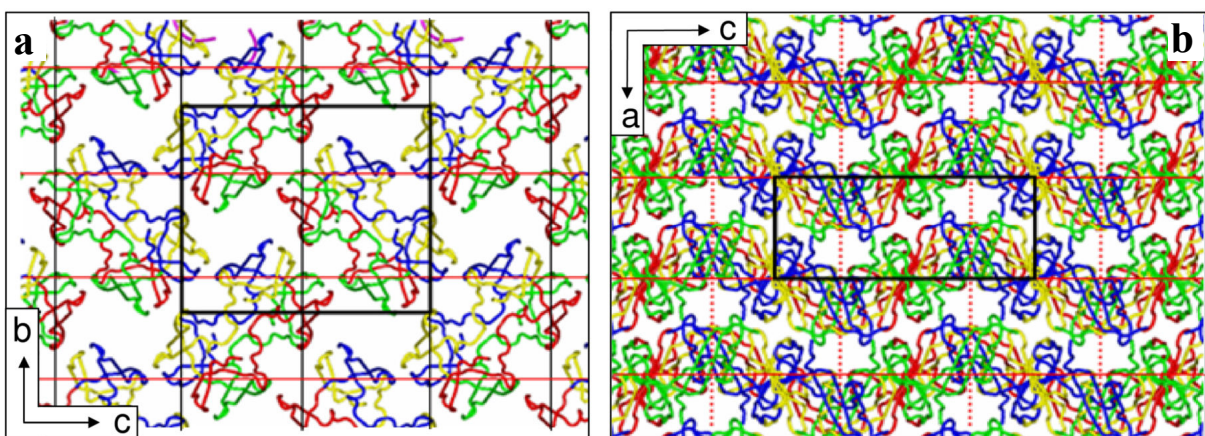

Figure 4.6: Symmetry elements parallel to *b* and *c* axes indicate a higher-symmetry space group than P2$_1$. The layout follows Figure 4.4. 2$_1$ screw axes parallel to the *b* axis present in P2$_1$ are displayed by black lines. Additional 2$_1$ screw and 2-fold rotation axes not present in P2$_1$ are displayed by red solid and dashed lines, respectively. These relate molecule A (red) to C (yellow), and molecule B (green) to D (blue). All symmetry elements are compatible with space group symmetry P2$_1$2$_1$2. a) view along the *a* axis, b) view along the *b* axis. Axes of the real-space crystallographic coordinate system are indicated by arrows.
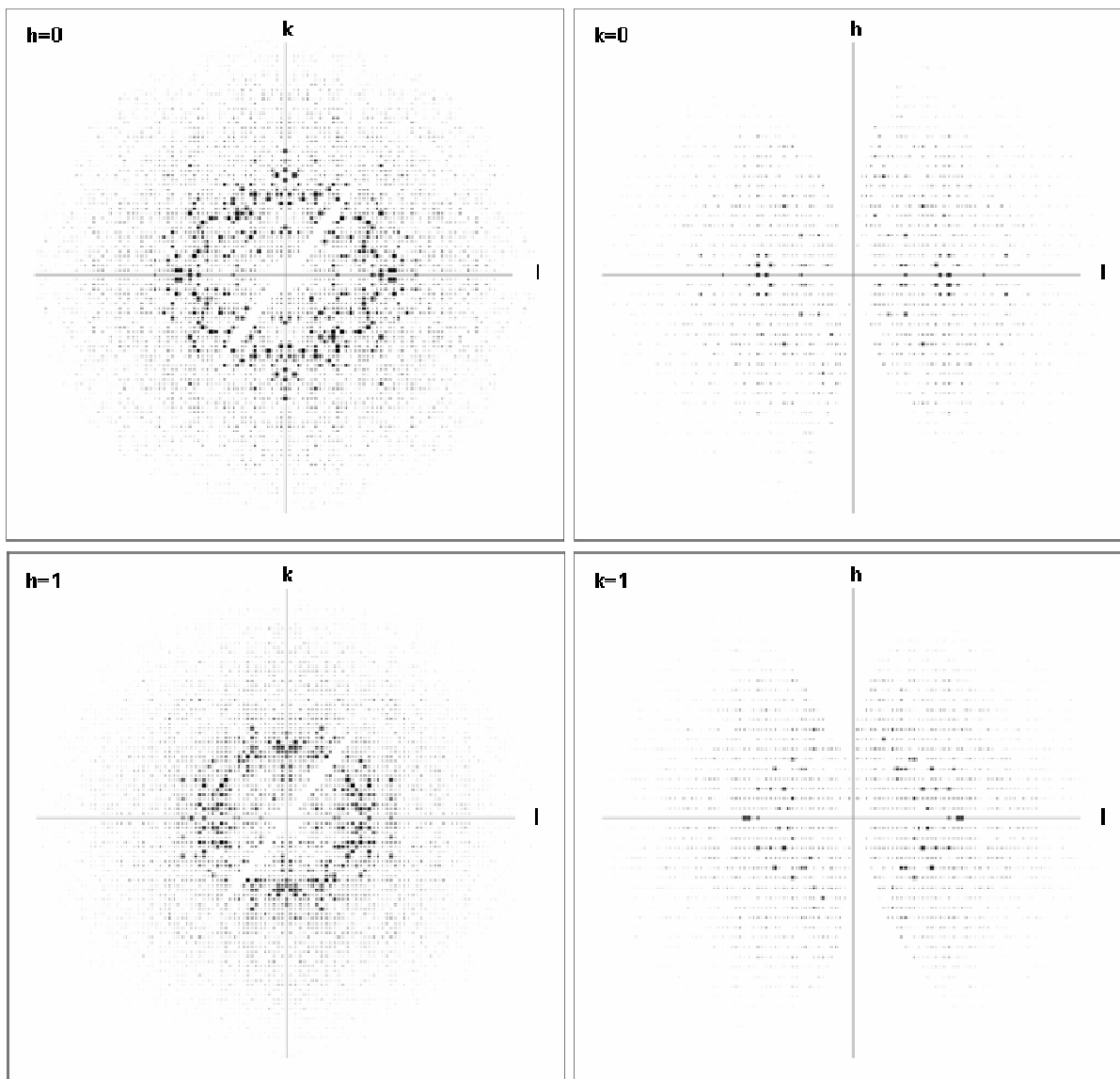
Figure 4.7: Zero-layer (top) and first-layer (bottom) precession images calculated from diffraction data from a *Bc*-Csp·dT$_6$ dataset processed in space group P1. Some diffraction data are missing due to a limited rotation angle of 132° along the φ axis and limitations in the crystal's orientation during data collection. 2-fold rotation symmetries along the reciprocal axes (*h*, *k*, *l*) in all layers indicate a 2/*mmm* (orthorhombic) Laue class. Systematic extinctions along the respective 0*k*0 and 00*l* crystallographic axes support space group P2$_1$2$_1$2. In order to comply with the symmetry elements of this space group, the reciprocal *h* and *l* indices of the measured reflections had to be exchanged.

where 2$_1$ screw axes are oriented parallel to *a*, and *b*. The crystallographic data were thus manually reindexed by exchanging the *h* and *l* indices, and could then be integrated and merged successfully in the higher-symmetry space group P2$_1$2$_1$2, as indicated by an R_{meas} of 5.3%. The reassignment of the *h* and *l* indices also resulted in a relocation of the *a* and *c* axes in real space. Molecular replacement of a single swapped protein chain associated with a single dT$_6$ molecule in space group P2$_1$2$_1$2, yielded two solutions (Figure 4.8). The resulting model of the crystal's asymmetric unit contains two protein molecules, which form a swapped dimer and associate with two dT$_6$ molecules. In space group P2$_1$2$_1$2$_1$ no molecular
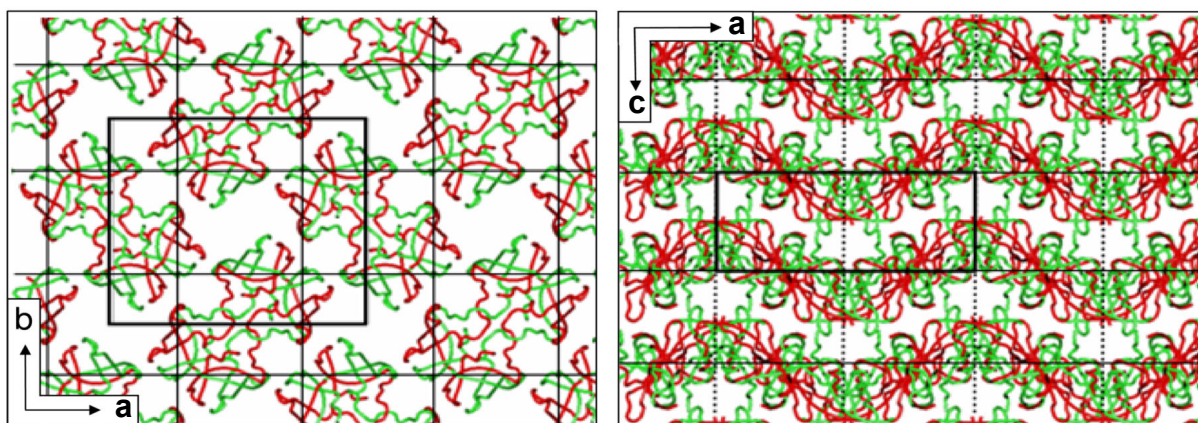
Figure 4.8: Molecular replacement solutions for *Bc*-Csp·dT$_6$ in the (correct) space group P2$_1$2$_1$2. The layout follows Figure 4.4. The unit-cell content is enclosed in a black box. 2$_1$ screw and 2-fold rotation axes derived from space-group symmetry elements are displayed by solid and dotted lines. The two independent positions of protein chains (red, green) in the crystal are displayed by α-carbon traces. a) view along the *a* axis. b) view along the *b* axis.

replacement solution was found. This space-group would also be in agreement with the diffraction data processed in P1 due to missing 00*l* reflections.

The swapped dimer model was gradually expanded and solvent molecules associated with protein and DNA groups through hydrogen bridges were included. In the last refinement steps the *B* factors were refined anisotropically. When the refinement of the structural model was completed, the R_{work} and R_{free} values corresponded to 13.0% and 16.2%.

4.1.2 Structural features of the *Bc*-Csp·dT$_6$ domain swap

The architecture of the functional units closely resembles that of all other structural models of CSP, featuring 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* (Figure 4.9). 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, as will be discussed later. The swapped chains can be related by a non-crystallographic twofold rotation axis. Both functional units superimpose with a RMSD of 0.11 Å. The formation of a swapped dimer reduces the solvent accessible surface per subunit by 5.4% (493 Å^2).
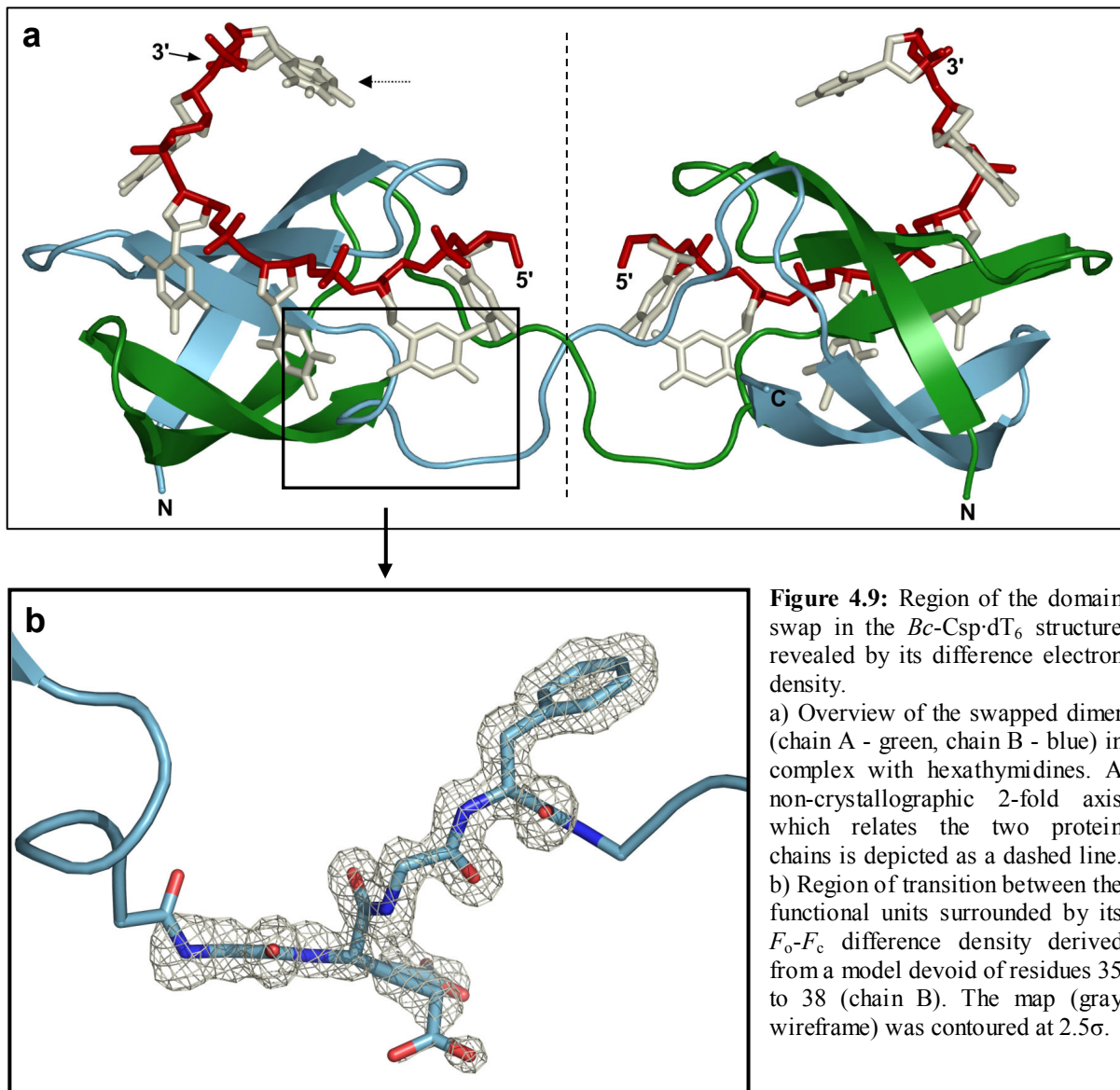
Figure 4.9: Region of the domain swap in the *Bc*-Csp·dT$_6$ structure revealed by its difference electron density.

a) Overview of the swapped dimer (chain A - green, chain B - blue) in complex with hexathymidines. A non-crystallographic 2-fold axis which relates the two protein chains is depicted as a dashed line.
b) Region of transition between the functional units surrounded by its F_o-F_c difference density derived from a model devoid of residues 35 to 38 (chain B). The map (gray wireframe) was contoured at 2.5σ.

4.2 Discussion

4.2.1 Backbone torsions, which mediate domain swapping in *Bc*-Csp, confer a two-state structural variability in closed monomeric CSP

In the *Bc*-Csp·dT$_6$ crystal structure, two protein chains form a swapped dimer, whose functional units associate with single dT$_6$ molecules (Figure 4.9). The domain swap involves a unique combination of the torsion angles ψ of Glu36 and ϕ of Gly37 located within loop L$_{34}$ and separates the two β-sheets which form the globular β-barrel. Interestingly, at the position of the domain swap the torsions in crystal structures of monomeric *Bc*-Csp show a two-state conformational variability (Figure 4.10): Out of ten protein models from six different *Bc*-Csp crystal structures five models display Glu36 ψ / Gly37 ϕ mean torsion angle combinations of 162° ± 14° and 99° ± 5°, designated as state 1. The other models show mean torsion angle combinations of -15° ± 15° and -65° ± 8°, designated as state 2 (Figure 4.10a, b). In all crystal structures featuring two protein molecules (1HZ9, 1C9O, 1HZB, 1HZC), one molecule maintains Glu36 ψ / Gly37 ϕ torsion angle combinations according to state 1 whereas the other molecule adopts a conformation according to state 2. In monomeric *Bc*-Csp (1I5F, 1HZA) and *Bs*-CspB crystal structures (1CSP, 1CSQ, 2ES2, 2HAX) Glu36 ψ / Gly37 ϕ torsion angle combinations according to state 1 are observed, only. All other torsions in *Bc*-Csp and *Bs*-CspB represent single states with standard deviations of less than 20°, except for the terminal residues. The angular differences between states 1 and 2, -177° for Glu36 ψ and -164° for Gly37 ϕ, roughly compensate each other and do not result in noticeable tertiary structural deviations between monomers in state 1 and state 2 (Figure 4.10c). The segregation into the two conformational substates is caused by differences in the crystal packing environment from sidechains of Glu36, Phe38, and Lys 37, which significantly influence the backbone conformation of loop L$_{34}$.

In contrast to closed *Bc*-Csp structures, the two polypeptide chains of the swapped dimer feature a unique combination of torsion angles at the site of the domain swap, a Glu36 ψ torsions angle of 141° ± 4° (similar to state 1) and a Gly37 ϕ torsions angle of -85 ± 1° (similar to state 2). This results in effective ~180° rotations of their mainchains at the C-terminus of Glu36 ψ, causes the opening of their globular fold and allows them to re-associate as domain-swapped dimers. Apart from differences in the progression of the protein backbone at the point of transition, the overall structure of the functional units is not altered significantly
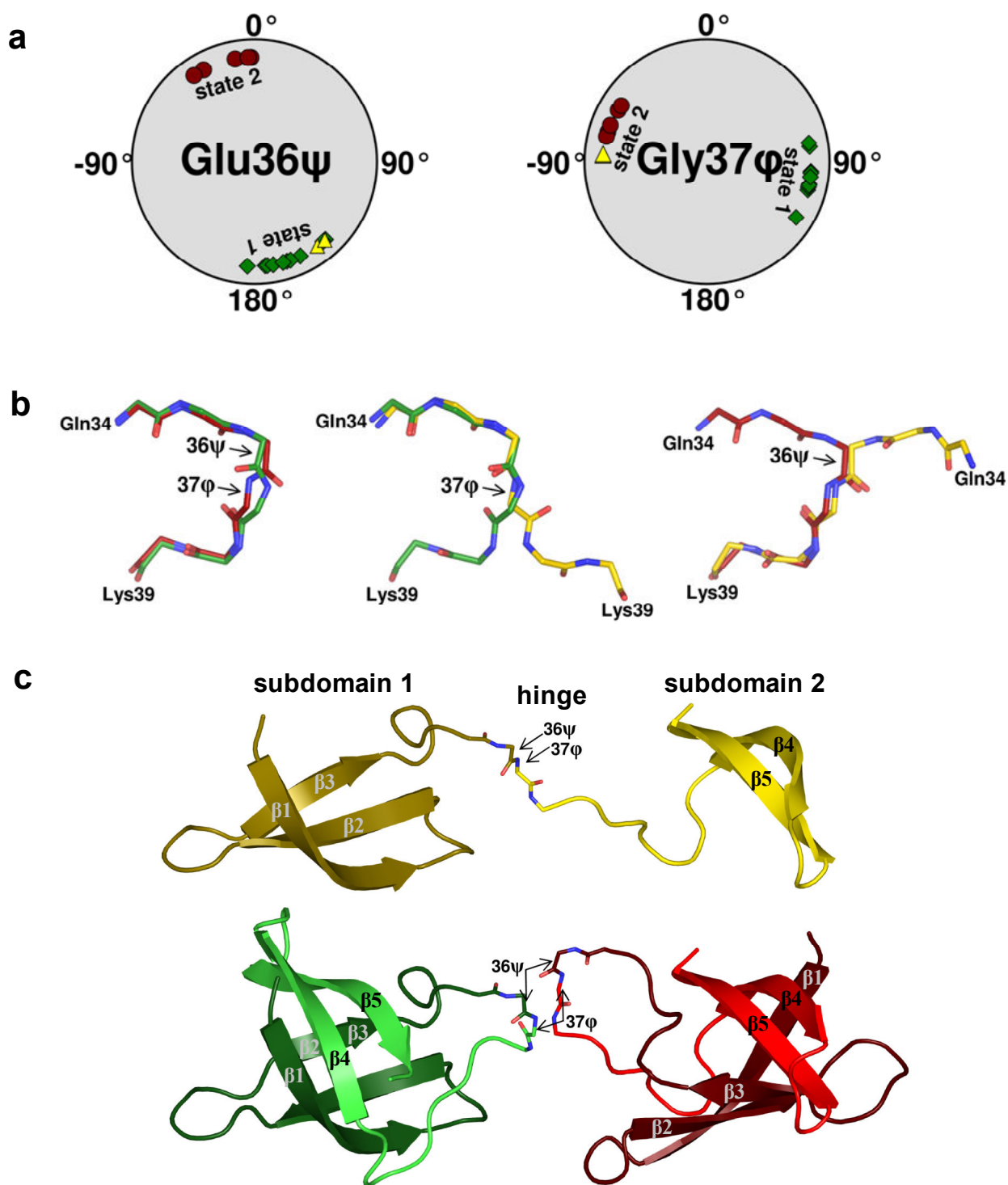
Figure 4.10: Comparison of open (domain-swapped) and closed states of *Bc*-Csp. The structural models of 2HZA represent closed monomeric CSP in b and c.

a) Torsion angles of Glu36 ψ (left) and Gly37 φ (right) from 14 closed models of *Bc*-Csp and *Bs*-CspB and two domain-swapped *Bc*-Csp molecules (yellow triangles). The closed structures feature a two-state structural variability in torsion angles of either 162° ± 14° and 99° ± 5° (state 1, green squares) or -15° ± 15° and -65° ± 8° (state 2, red circles). The domain-swapped structures feature a unique combination of torsions: 141° ± 4° for Glu36 ψ (similar to state 1) and -85 ± 1° for Gly37 φ (similar to state 2).

b) Comparisons of loop L$_{34}$ (Gln34 to Lys39) backbone structures displaying different conformational states based on Glu36 ψ/ Gly37 φ torsion angles. Left: Superimposition of two models corresponding to state 1 (green) and state 2 (red). Center: Superimposition of a model corresponding to state 1 (green) and a domain-swapped structure (yellow). Residues 34-37 were superimposed. Right: Superimposition of a model corresponding to state 2 (red) and a domain-swapped structure (yellow). Residues 36 to 39 were superimposed.

c) Comparison of an open monomer CSP (yellow) with two closed monomeric CSP structures corresponding to state 1 (green) and state 2 (red).

compared to the closed monomers in state 1 and 2 as indicated by RMSD values below 0.50 Å for equivalent α-carbon positions (see also Figure 3.5b). The Glu36 ψ / Gly37 φ torsion thus apparently constitute two pivot points, through which the structure can be converted between swapped (open) and closed monomeric states by a single ~180° rotation along either torsion (Figure 4.10b, c).

4.2.2 The domain-swapped architecture of *Bc*-Csp may provide insight in the process of CSP folding and misfolding

Structures of the CSP contain two subdomains of similar length, which are separated by a long loop: Subdomain 1 is a sheet including β-strands 1 to 3, subdomain 2 is a β-ladder comprising strands 4 and 5. In the closed monomers of *Bc*-Csp these two subdomains are stabilized by 26 backbone (local) hydrogen bonds, whereas the interface between them involves 8 backbone (non-local) hydrogen bonds (Figure 4.11). Almost all of these interactions can also be found in the swapped dimers. [15]N relaxation and H/D-exchange NMR experiments of *Bs*-CspB, which shares 86% sequence identity with *Bc*-Csp, suggest that loop L_{34}, which divides the subdomains, remains flexible even in the folded state of the protein [132]. It is therefore likely that the structural transition between closed monomeric, open monomeric and swapped dimeric state could occur without unfolding the two subdomains. A state similar to that of the open monomer may thus reflect a state in the folding pathway of *Bc*-Csp: In such an arrange-
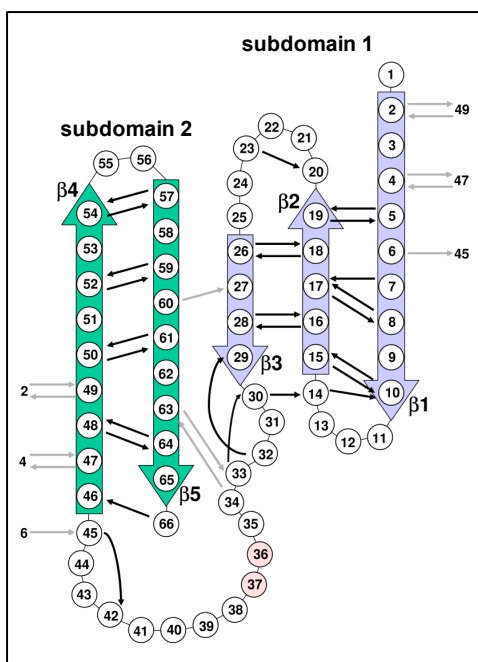

Figure 4.11: Topology plot of the *Bc*-Csp architecture. β-strands are depicted as blue (subdomain 1) and green (subdomain 2) large arrows. Intra-subdomain and inter-subdomain hydrogen bonds from the protein backbone are indicated as black and gray arrows, respectively, which follow the direction of the hydrogen bond (donor → acceptor). All hydrogen bonds can be observed in closed as well as in domain-swapped *Bc*-Csp structural models. In the latter case the intersubdomain hydrogen bonds are formed between different protein chains, however. In open monomers, as discussed in the text, only intrasubdomain (local) hydrogen bonds would be formed. Residues 36 and 37, whose backbone torsion angles display a two-state conformational variability in closed CSP structures and enable the domain swap in the *Bc*-Csp·dT$_6$ structure, are depicted in pink.

ment, the subdomains could independently form from an unfolded chain (Figure 4.10). The formation of the subdomains would already contribute two thirds of all backbone hydrogen bonds and stimulate the organization of a bipartite hydrophobic core, which would be solvent-exposed at this stage. The association of subdomains results in the formation of the closed β-barrel burying the hydrophobic core and provides the residual 8 backbone hydrogen bonds.

In the context of protein misfolding, domain swapping has recently been implicated in the formation of amyloids [34]. Interestingly, *Ec*-CspA [93], which shares 57% sequence identity with *Bc*-Csp, as well as fragments of *Bs*-CspB [142, 143], which share 86% sequence identity with *Bc*-Csp, were shown to form amyloid fibrils under acidic conditions. NMR Studies on amyloid formation of *Ec*-CspA have revealed time dependent changes in ^{15}N T2 relaxation accompanying the exponential phase of polymerization, which suggest that the first three β-strands may form an association interface that promote aggregate growth [93]. In the late stage of denaturation, signals of ^{1}H-^{15}N HSQC spectra from the N-terminal half of the molecules (equivalent to residues 5 to 36 in *Bc*-Csp) appear to be more severely broadened than those from the C-terminal half. This uncoupling of entities which refer to the subdomains suggests that a (re)organization of the protein chain as observed between swapped and un-swapped CSP structures could play a role in CSP misfolding. Domain swapping has recently been implicated in amyloidogenesis. This extends the cross-β model of amyloids, which is based on studies of amyloids by X-ray fiber diffraction [31, 32] and cryo-electron microscopy [33] (see also Chapter 1.1): For ribonuclease A, which has 30% β-strand content, amyloid formation involving intermolecular recombination of subdomains could be demonstrated [34], which apparently does not destroy enzymatic activity. It has therefore been suggested that amyloid cross-β spines consisting of extended β-sheets may be formed from domain-swapped protein assemblies with native-like architectures. This model of amyloid protofilament formation could also be relevant for the CSP, which have a high β-strand content (~56%), and it may explain why the bipartite organization, which allows the structural transition between closed monomers and swapped dimers, can be observed in the late stage of *Ec*-CspA amyloidogenesis. It must be noted, however, that amyloid fibers can also be formed from isolated peptide fragments of *Bs*-CspB which correspond to each of the two halves constituting the *Bc*-Csp swapped dimer, and from a smaller fragment encompassing the first 22 amino acids [142, 143].

4.2.3 Ligand binding in the context of domain swapping

The question remains, whether DNA binding to the CSP is responsible for domain swapping, or whether domain swapping favors ligand binding. So far, solution data supported the formation of complexes consisting of one CSP and one ligand molecule for ligands in the range of hexa- or heptanucleotides (Figure 2.6 & [132]). It cannot be excluded that an equilibrium between monomeric and swapped dimer state exists, which is predominantly on the side of the monomer. So far it was not possible, however, to separate two molecular species by analytical ultracentrifugation or by gel filtration, neither in the presence nor in the absence of a ligand. The *Bs*-CspB·dT$_6$ structure [144] clearly demonstrates that a domain swap is not required for ligand binding (Figure 3.15). In fact, the binding of a nucleobase to subsite one appears to be possible in the closed monomeric state, only: In the swapped-dimer state the backbone of Gly37 and Phe38 would be too close to allow a formation of an edge-on stack between a thymidine and the sidechain of Phe38. This may also explain, why in the *Bc*-Csp·dT$_6$ structure the hexathymidine ligands occupy subsites 2 to 7. Most protein-ligand contacts of the two complex structures are in good agreement, suggesting that domain swapping does not change the ligand interaction sites of the CSP. The transition of the closed-monomeric to the domain-swapped state is not possible when a ligand is bound to the binding site, which is formed by residues from both subdomains. Hence, ligand binding does not favor domain swapping.