

3 Structures of Small Domains and Proteins

In the course of this thesis, the structures of three small domains and proteins were determined by solution NMR spectroscopy. All of these proteins were targets of the German structural genomics effort “protein structure factory” (psf). They were selected solely based on their suitability for NMR analysis, not on any common biological or molecular function. They will be discussed together in this chapter and the structure determination of each of them will be introduced by a short description of their respective biological roles. As discussed in chapter 1.1, one of the ultimate aims of structural genomics is to elucidate the structure to function relationship of proteins. Therefore, insights into the molecular roles were used in interpreting the structures wherever possible. The determination of these protein structures was used to develop a general strategy for assignment and structure calculation, which will be discussed in section 3.4.

3.1 SODD – BAG-Domain

3.1.1 SODD biology

‘Silencer of Death Domains’ (SODD) is a BAG-domain containing protein of 457 amino acids, which was identified as a binding partner for the intracellular domains of ‘Tumor Necrosis Factor Receptor 1’ (TNF-R1) and ‘Death Receptor 3’ (DR3) (Jiang et al., 1999). SODD also interacts with Bcl2 via its BAG-domain (residues 378 to 452) (Antoku et al., 2001). SODD was shown to act in an anti-apoptotic manner and was found to be overexpressed in pancreatic cancer cells (Ozawa et al., 2000). It was hypothesized that the BAG-domain also interacts with a predicted ATPase domain in TNF-R1 (Miki et al., 2002). In contradiction, the N-terminal region of SODD interacted in the original two-hybrid screens with the death domains of TNF-R1 and DR3 (Jiang et al., 1999). Neither the mechanism of the anti-apoptotic function of SODD nor the role of HSP70 in BAG-domain mediated apoptosis regulation and hence the role of the BAG/HSP70 interaction have been understood to date. Recent findings further add to

the complexity of SODD biology. Apparently, the loss SODD has no effect on TNF-R1 signalling as discovered in mice knock-outs (Endres et al., 2003). Another study reported the involvement of SODD in Apolipoprotein B mRNA editing (Lau et al., 2003).

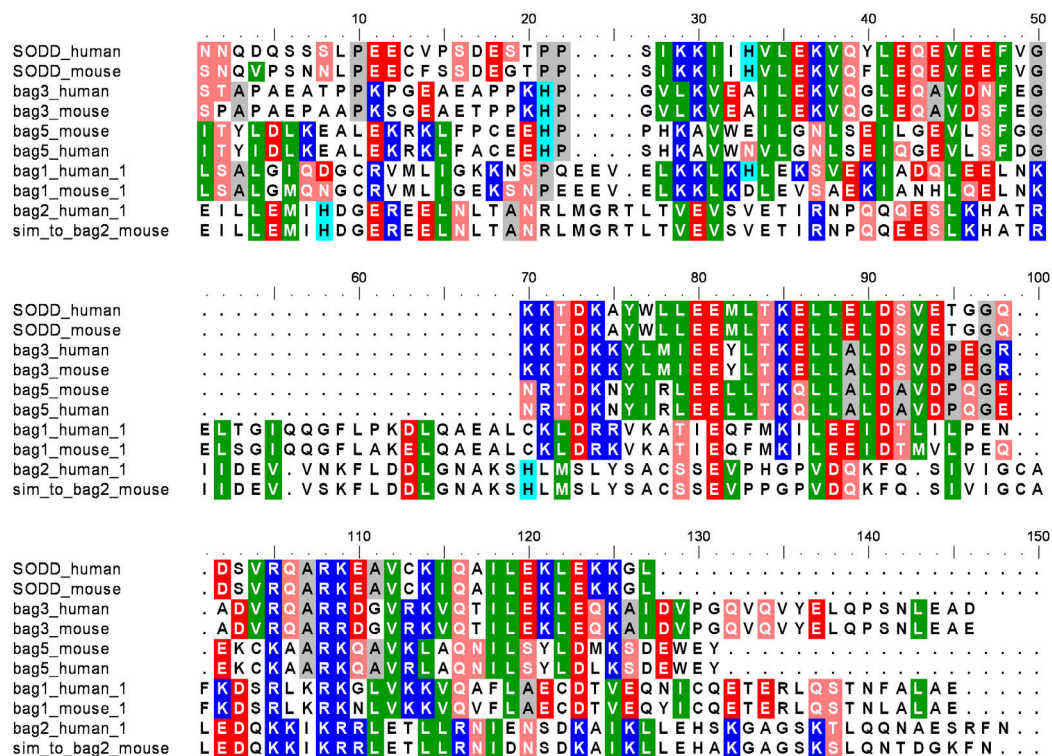


Figure 3.1 Sequence Alignment of BAG-domains.

The sequence alignment of selected BAG-domains shows the difference between the SODD- and the Bag1-subfamily. SODD subfamily proteins are characterized by a 19 amino acid gap and the conserved residues homologous to F400 and Y409 in SODD. Note that the alignment includes only a subset of known BAG-domains. Sequences were taken from the Swissprot database. The threshold for shading was set to 40% identity. Accession numbers from top: O95429, Q9CWG2, O95817, Q9JLV1, Q9DAU0, Q9UL15, Q99933, Q60739, O95816, Q91YN9.

BAG-domains were identified as protein-protein interaction domains with low sequence similarity (Figure 3.1) involved in apoptosis regulation (Bateman et al., 2002). They appear in a variety of vertebrate, insect, nematode, yeast and plant proteins (Takayama et al., 1999). Most of the known BAG-domains were found to interact with the ATPase domain of HSP70 and with the apoptosis regulator protein Bcl2 (Antoku et al., 2001; Takayama et al., 1999).

BAG-domains form three helix bundle structures (Briknarova et al., 2001; Briknarova et al., 2002; Sondermann et al., 2001). The interaction interface of the hBag1 BAG-domain in complex with HSP70 was analyzed by X-ray crystallography and is located at the interface of the second and third helix (Sondermann et al., 2001). NMR studies on the mBag1 BAG-domain with peptides derived from HSP70 were consistent with these results (Briknarova et al., 2001). The interaction of BAG-domains with Bcl2 has not been investigated in detail so far. Only a BH3 binding motif homologous to a motif in the Bcl2 interacting protein Bax was hypothesized to be present in BAG-domains (Antoku et al., 2001). Recent structural studies (Briknarova et al., 2002) showed that the SODD BAG-domain is smaller than its homologous domain in Bag1, defining a new subfamily of short SODD-like BAG-domains. However, it remained unclear whether the domains of this subfamily are really shorter or whether their domain boundaries are only shifted towards the N-terminus.

In this section, the NMR structure of an N-terminally extended SODD BAG-domain construct is reported together with relaxation rate measurements to clarify domain boundaries. A homology model of the SODD-BAG/HSP70 complex was constructed to analyze SODD-subfamily specific contacts in the binding interface and to investigate possible differences in HSP70 binding of the two BAG-domain subfamilies.

3.1.2 Structure of the SODD-BAG Domain

As expected from the sequence alignments and the 3D-structures of hBag1 (Sondermann et al., 2001) and mBag1 (Briknarova et al., 2001), the BAG-Domain of SODD is exclusively α -helical and forms a three helix bundle (Figure 3.2) as also found by Briknarova et al.. The ensemble of 20 lowest energy structures shows an average RMSD to the mean structure of 0.5 Å for the backbone atoms and 1.1 Å for the sidechain heavy atoms (Table 3.1). Residues I383, V386, V390, L393, V397, F400, Y409, L412, L416, L423, V426, A441, I445, I448, L449 and L452 form the hydrophobic core of the domain. Residues F400 and Y409, which are conserved in all short BAG-domains,

seem to play a key role in stabilizing the loop between the helices $\alpha 1$ and $\alpha 2$. The helix boundaries were found at residues S379 – F400, D406 – D424 and Q431 – K455 respectively.

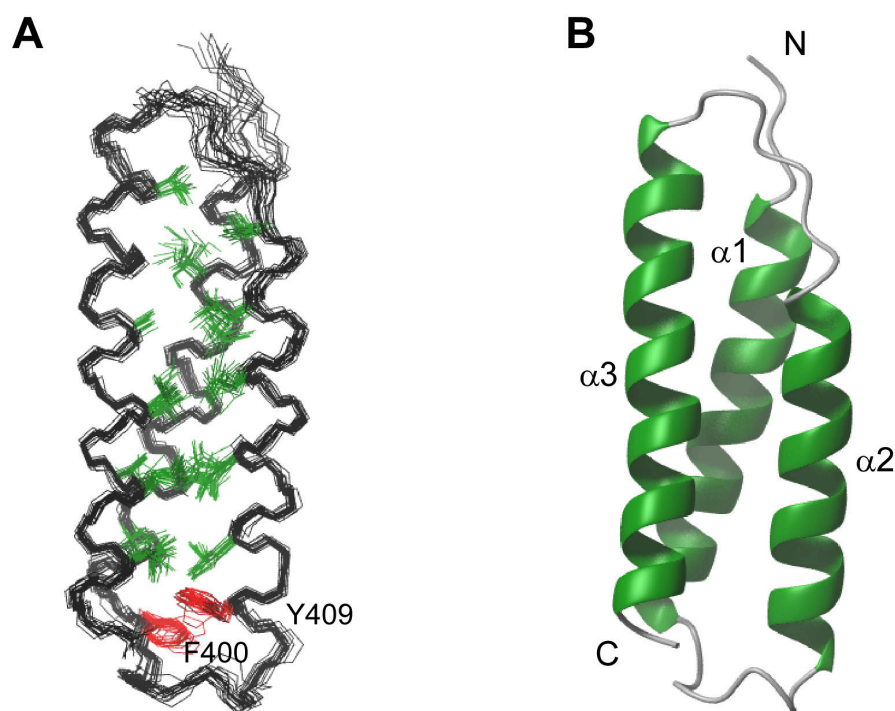


Figure 3.2 Structure of the SODD-BAG domain

(A) Ensemble of the 20 lowest energy structures. The backbone of the protein is shown as black lines. The heavy atoms of the hydrophobic residues making up the core of the protein are indicated in green. The conserved residues of the SODD-subfamily of BAG-domains F400 and Y409 are drawn in red. The residues 358 to 379 that were found to be unstructured are omitted for clarity. (B) Ribbon representation of the structure with the lowest RMSD to the mean structure of the ensemble. Figure prepared with MOLMOL.

The data derived from the assigned spectra indicates the presence of the expected three-helix bundle structure. Summarized in Figure 3.3 the T_1/T_2 ratio, the $\Delta\Delta\delta_i$ and the total number of NOE restraints per residue already show most features of the structure.

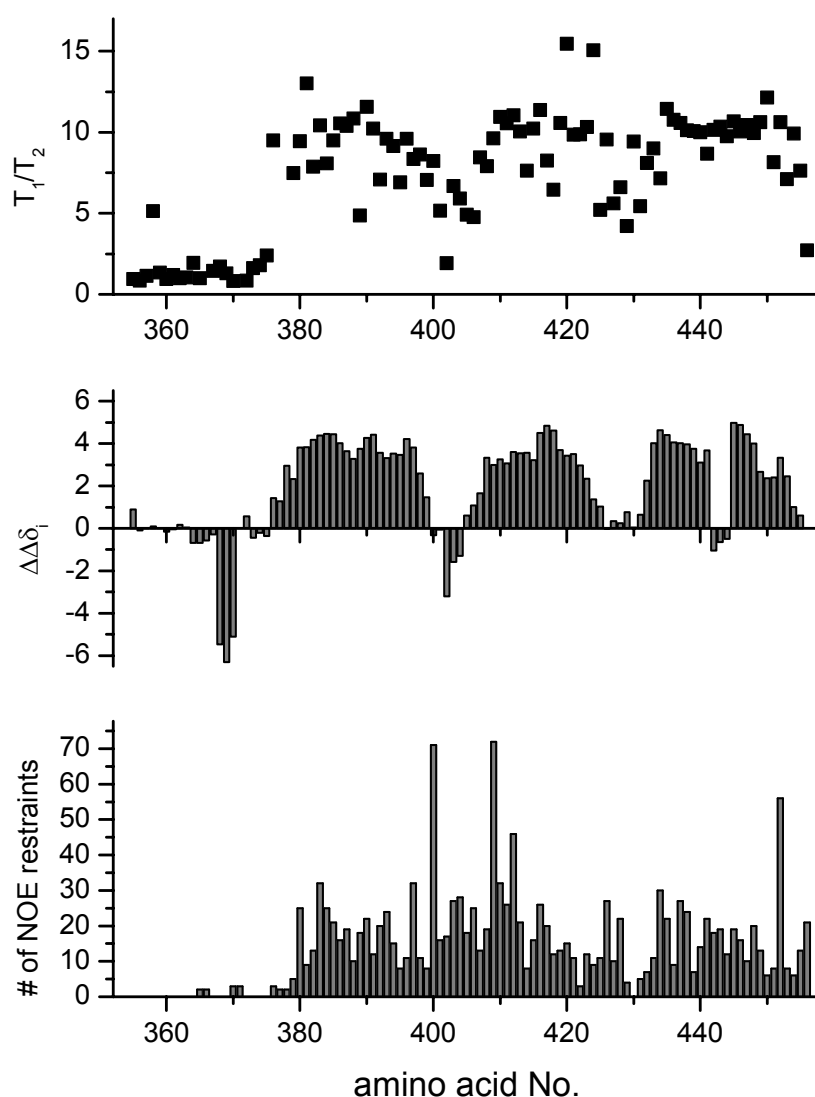


Figure 3.3 Summary of structural parameters for SODD-BAG

The T_1/T_2 ratio (**top**), the chemical shift difference from the random coil shifts for C_α and C_β as calculated from eq. 3.1 (**middle**) and the total number of NOE restraints per residue (**bottom**) are plotted against the number of the respective amino acid. Smaller T_1/T_2 ratios are indicative of faster motion, positive and negative $\Delta\Delta\delta_i$ indicate α -helices and β -sheets, respectively.

The first residues from 358 to 379 of the construct are unstructured as indicated by their small T_1/T_2 -ratio and the almost complete absence of inter-residue NOEs. In the structured region of the domain the three α -helices are clearly visible from their positive values in the $\Delta\Delta\delta_i$ plot. Interestingly, the loops connecting the three helices are slightly more mobile than the rest of the residues. The distribution of the NOE-

restraints on the amino acids is fairly uniform with the exception of F400 and Y409. These residues form the hydrophobic core at one end of the three-helix bundle and are therefore involved in many crucial sidechain interactions. The unique ^1H chemical shifts of their aromatic sidechains allow the identification of many contacts that would otherwise not be detected due to spectral overlap. The quality of the ensemble is summarized in Table 3.1.

Restraints:		Quality of the Ensemble:	
<i>NOEs:</i>		<i>RMSD (res 379-452): (MOLMOL)</i>	
Used for calculations	731	Backbone atoms	0.5 Å \pm 0.09 Å
Long range	219	Sidechain heavy-atoms	1.1 Å \pm 0.09 Å
Medium range	258		
Sequential	254	<i>Deviation from idealized geometry: (CNS)</i>	
		Bonds	0.0010 \pm 0.0001
<i>Other:</i>		Angles	0.2858 \pm 0.0029
H-bonds	52	Impropers	0.1651 \pm 0.0083
Dihedrals (TALOS)	140		
		<i>Violations:</i>	
		Distances > 0.2 Å	0
		Dihedrals > 5°	0
		<i>Ramachandran-Plot: (Procheck NMR)</i>	
		Most favoured regions	93.0 %
		Additionally allowed regions	7.0 %
		Generously allowed regions	0.0 %
		Forbidden regions	0.1 %

Table 3.1 Structural statistics for the SODD-BAG domain.

A structure induced sequence alignment (Figure 3.1) shows that all BAG-domains exhibit a high conservation of the binding interface to HSP70 in the C-terminal part of the domain. In this alignment, all BAG domains containing the motive F_i, Y_{i+9} as F400, Y409 in SODD-BAG were considered to belong to the shorter subfamily of BAG-domains with a characteristic 19 amino acid deletion. Mapping the conserved residues onto the surface of the SODD BAG-domain shows that all residues interacting specifically in the hBag1/HSP70 complex are conserved in SODD. This is shown in Figure 3.5 A – C, in which conserved residues found to be essential for the interaction to HSP70 in the crystal (Sondermann et al., 2001) or by mutagenesis experiments are

coloured in blue while other residues conserved throughout the whole family are coloured in cyan. SODD-subfamily-specific residues are highlighted in green. The conservation of the non-subfamily-specific residues extends over all other BAG-domains (Figure 3.1). In the new SODD subfamily two glutamic acids E413 and E427 are also remarkably conserved, and W410 and E419 appear at the edges of the common binding interface to HSP70 (Figure 3.5 A).

3.1.3 Computer Model of the SODD-BAG/HSP70 Complex

An homology model of the SODD-BAG/HSP70 complex (Sondermann et al., 2001) was built to address the question whether the differences in charge distribution between SODD and hBag1 BAG-domains could be compensated for in the complex or not. A common function for BAG domains is their interaction to the ATPase domain of HSP70 (Sondermann et al., 2001). It is unknown whether the interaction of the ‘large’ hBAG1-type domains and the ‘small’ SODD type domains have the same or different effects. Nevertheless, it can be expected that the two domains share a common interaction mechanism.

As a prerequisite for a conserved binding mode it would be expected that the electrostatic potential of the binding interface should be largely conserved. Indeed, many features of the charge distribution are similar (Figure 3.4 A, C). However, a remarkable difference is the larger negatively charged area due to E413 in the lower half of the interface. The differences in the electrostatic properties of the binding site are conserved in the SODD-subfamily, raising the question of whether the binding mode of this subfamily may indeed be the same as that found for hBag1.

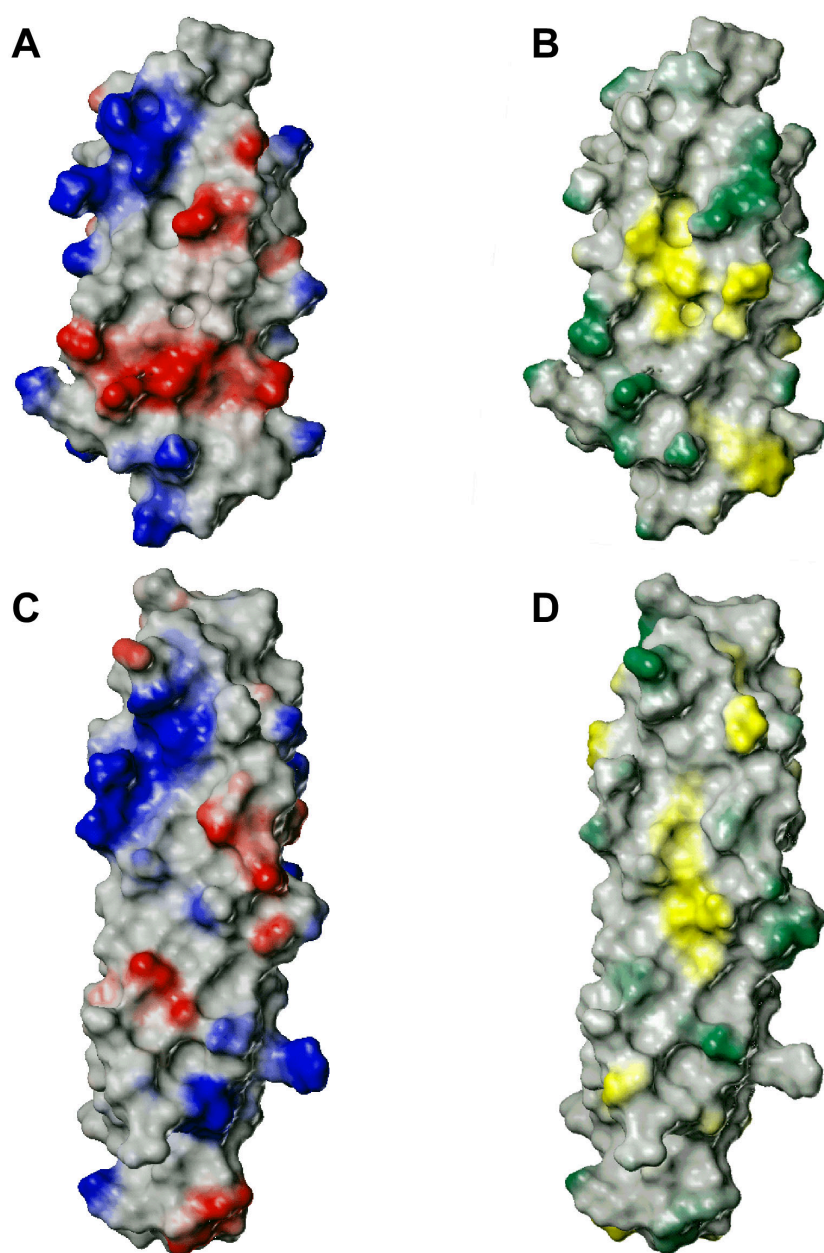


Figure 3.4 Comparison of the surface properties of the hBag1- and SODD-BAG domains
(A,B) Surface representations of the SODD-BAG domain. (C,D) Surface representations of the hBag-domain in complex with HSP70 (PDB: 1HX1) (A,C) Electrostatic surface. Positively charged areas are coloured in blue, negatively charged areas in red. (B,D) Hydrophobic surface. Hydrophobic areas are coloured in yellow, hydrophilic areas in green.

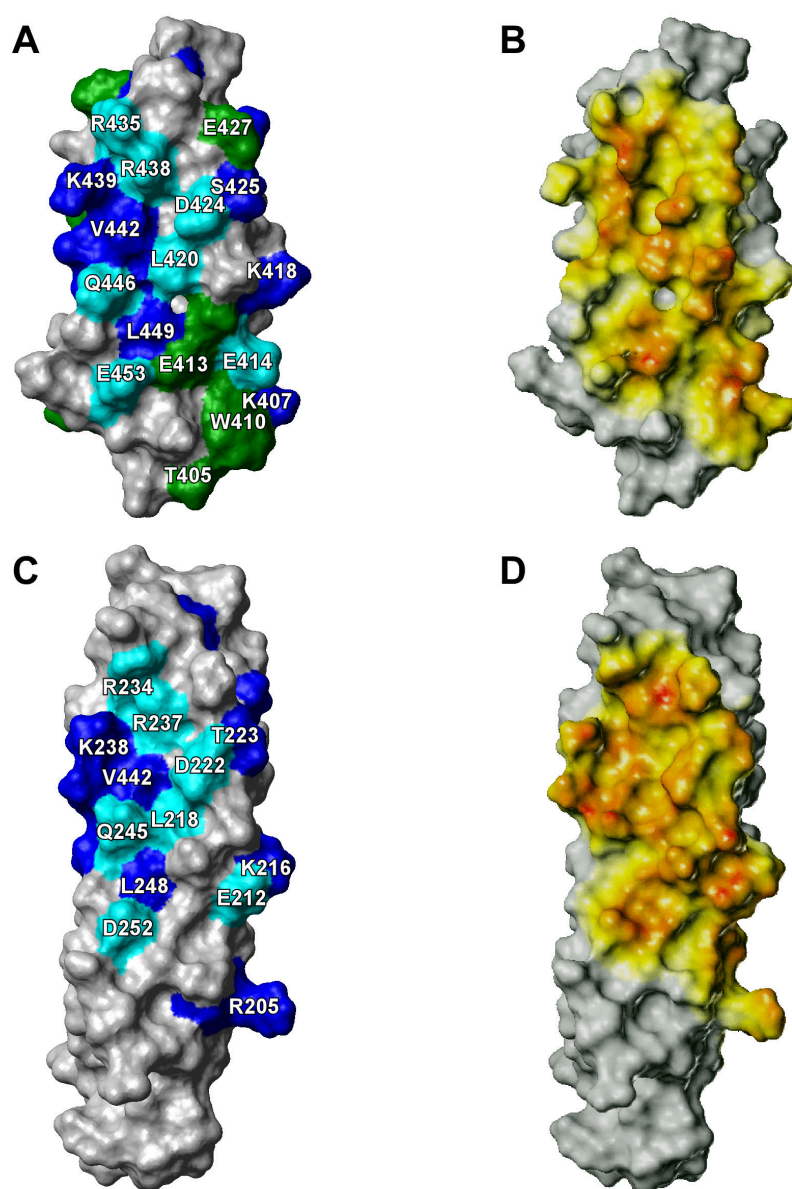


Figure 3.5 Comparison of the interaction surface of the hBag1- and SODD-BAG domains

Surface representations of the SODD-BAG domain **(A)** The conservation profile mapped on the surface of the BAG-domains. Conserved residues that are found to be essential for the interaction to HSP70 in the crystal or by mutagenesis experiments (Briknarova et al., 2001; Sondermann et al., 2001) are colored in blue while other residues conserved throughout the whole family are colored in cyan. SODD-subfamily-specific residues are highlighted in green. **(B)** the contact area of SODD to HSP70 in the modelled complex based on the crystal structure (1HX1) is depicted in orange. **(C,D)** Surface representations of the hBag1-BAG domain (from 1HX1) shown in the same orientation and colouring as SODD. **(A)** and **(C)** were prepared with MOLMOL

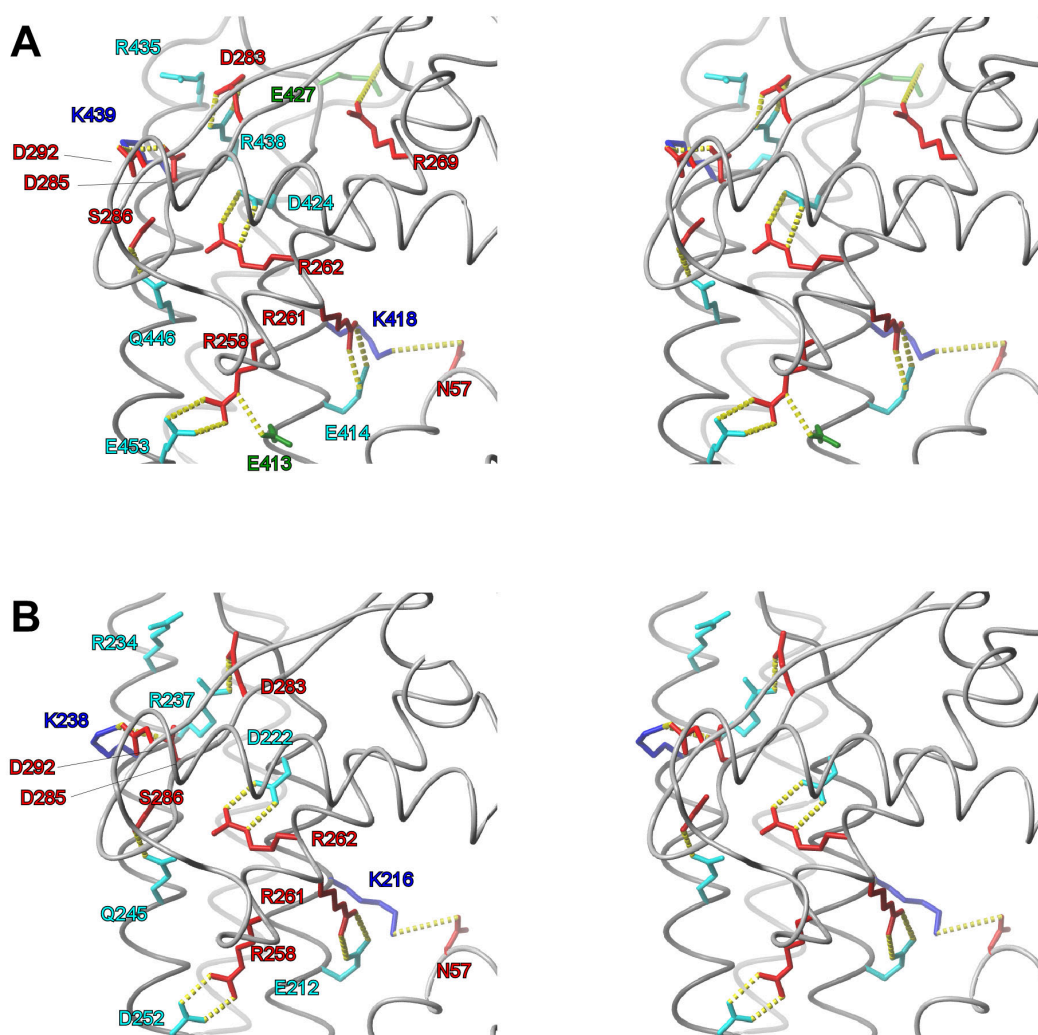


Figure 3.6 Interactions in the HSP70 complexes of SODD and hBag1

The stereo figures show the electrostatic interactions in the model of the SODD-BAG/HSP70 complex (A) and the crystal structure of the hBag1 complex (1HX1) (B). The residues in the BAG-domains are coloured as in Figure 3.4, residues of HSP70 are coloured in red. The protein backbones are indicated as C α -traces in dark and light grey for the back domains and HSP70, respectively. The orientation of the figure is roughly the same as in Figure 3.4. Figure prepared with MOLMOL

In our homology model, which does not differ considerably from the hBag1/HSP70 complex, the two subfamily-specific negatively charged amino acids E413 and E427 in SODD are found to be able to form H-bonds/salt bridges to R258 and R269 in HSP70, respectively. All other non-conserved residues are involved in hydrophobic interactions. The network of H-bonds found in the complex structure of hBag1 is extended by these additional interactions (Figure 3.6). While E413 is within the protein

interface, E427 is close to its edge and its interactions are thus interrupted during the dynamics simulation by the insertion of water molecules. On the other hand the C $^{\alpha}$ -C $^{\alpha}$ distance between the residues E427 and R269 remains within 13 Å allowing the formation of an H-bond or salt-bridge between residues E427/R269. Thus, our model shows that the additional charges in the binding interface of the SODD-subfamily can be compensated or delocalized in the complex without changing its overall geometry. It is therefore likely that the geometry of the complex is conserved for all BAG-domains. These charge differences might therefore modulate binding affinities of the two BAG-domain subfamilies to HSP70, as observed experimentally for mutants of IFN α 2 in its complex with ifnar2 (Piehler et al., 2003). In case the interface is also used for other interactions, they may ensure the binding of a different spectrum of interaction partners.

3.1.4 Discussion

The structure of the BAG-domain from human SODD revealed a second subfamily of BAG-domains differing in length of the three-helix bundle. In the structures of the short SODD-like subfamily that contains the F-X_s-Y motif as a common sequence feature, all helices are three to nine residues shorter than in the longer Bag1-like domains. Nevertheless, the interaction interface with the HSP70 ATPase domain seems to be conserved in both families of Bag-domains. Compared to hBAG1, in the BAG-domain of SODD all residues are conserved that are involved crucial with the HSP70 ATPase domain. In the model of the complex of SODD-BAG with HSP70 all of these contacts of hBAG1 are resembled. There are additional charged residues in SODD, which allow for additional contacts. Interestingly, these glutamic acids E413 and E427 were found to be conserved throughout the SODD-subfamily. The presence of these residues may ensure different binding affinities for SODD and hBag1 BAG-domains to HSP70 and in case the interface is used for other interactions to ensure the binding of a different spectrum of interaction partners.

3.2 Complex I – Subunit B8

3.2.1 Introduction to Complex I

NADH ubiquinone oxidoreductase (Complex I) is the first complex of the mitochondrial electron transport chain. It is one of the most complicated biological machineries with a large number of subunits and cofactors. Its basic function is the transfer of two electrons from NADH to ubiquinone resulting in proton translocation across the membrane (Vinogradov, 2001; Yano, 2002). Although it is common to most life forms, there are large differences between bacterial and mitochondrial complex I.

Shortly after its first isolation (Hatefi et al., 1961), it was shown that mitochondrial complex I can be decomposed into several fragments that partially retain some activity (Hatefi et al., 1969). These can be further resolved into individual subunits, most of which were sequenced in recent years (Anderson et al., 1982; Fearnley et al., 1992; Hirst et al., 2003). Similar studies on bacterial systems revealed significant differences in subunit composition between the mitochondrial complex and that of bacteria. While there are 46 subunits reported for the bovine complex (Carroll et al., 2003), the bacterial enzyme consists of only 14 (Yagi et al., 1998), all of which have homologues in the mitochondrial complex. It is widely assumed that the homologues of the bacterial subunits also form the minimal functional complex of higher organisms (Fearnley et al., 1992), which account also for all known iron-sulfur cluster binding sites (Friedrich et al., 2000). Recently, it was shown that the *Neurospora Crassa* complex contains subunits that are exclusive to fungi (Abdrakhmanova et al., 2004).

Seven of the 14 conserved subunits are integral membrane proteins. These are encoded in mitochondrial genome, while all other subunits are encoded in the nucleus. A clear picture about the role of the so-called supernumary (also sometimes called 'accessory') subunits is only about to emerge. The supernumary MWFE subunit was shown to be required for complex I activity in mammals (Au et al., 1999). Recent findings suggest that it plays a role in complex I assembly (Yadava et al., 2004).

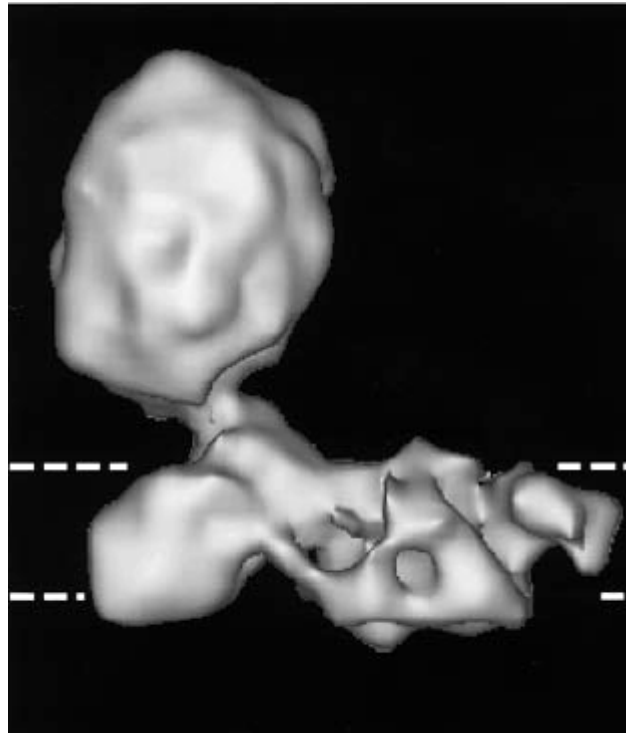


Figure 3.7 Structure of bovine Complex I determined electronmicroscopy

The figure shows the structure of Complex I at a resolution of 100Å. The peripheral arm that protrudes into the mitochondrial matrix is pointing upwards. The membrane bi-layer of the inner mitochondrial membrane is indicated by dashed lines. Figure reprinted from Grigorieff, 1998 with permission from Elsevier

The tight linkage of complex I to the production of so-called reactive oxygen species (ROS) and thus to possible cell damage connects it to a number of diseases. Mutations in complex I proteins and the above mentioned generation of ROS are assumed to be at least in part responsible for Leigh's (Benit et al., 2004), Parkinson's (Turrens, 2003) and Alzheimer's (Kim et al., 2002) diseases. ROS are formed if the electron transport chain is blocked or if the NAD(P)⁺-pool is almost completely reduced (Kushnareva et al., 2002). Even decreased ADP-levels in resting cells lead to a higher production of ROS. All this demands for a regulation of complex I, for example by mild uncoupling (Starkov, 1997). In this context, it was found that the activity of the complex is at least partially modulated by thiol-reactants interacting with thiol groups in complex I subunits (Taylor et al., 2003), increasing its superoxide production.

All available information on complex I evolution and differences between species was compiled, to construct this putative interaction map of all complex I subunits. Figure reprinted from Gabaldon et al., 2005 with permission from Elsevier.

From a bioinformatic study gathering all available data on the interaction and evolution of Complex I subunits it was possible to suggest an interaction map (Figure 3.8) in which the relative positions of the subunits are hypothesised (Gabaldon et al.,

2005). Structure determination of complex I by X-ray crystallography has so far been hindered by its size and the homogeneity problems caused by its large number of subunits. Until now, the lack of high resolution structural information, both of atomic coordinates for individual subunits as well as of tertiary contacts between subunits, has hindered the elucidation of the exact mechanism of electron transport. Recently, the structure of the *Thermus thermophilus* complex I was solved to a resolution of 4 Å by X-ray crystallography (Hinchliffe et al., 2005). Although only the largest features of the complex are visible at this resolution, the authors could determine the exact positions of all iron sulfur clusters in the complex. With a tentative assignment based on the electro negativity of the clusters, this allowed for the first time to describe the path of the electrons within the complex. At this stage, it is impossible to assign the individual subunits to parts of the electron density because too little information on the organisation of the subunits is known to guide a fit of the proteins into the electron density map. A structure of the protein part of the complex therefore depends on crystals that diffract to higher resolutions. Although the resolution with 4Å is not sufficient to model the proteins, the positions of the FeS-clusters are clearly detectable. This allows to assign the identities of the clusters based on their EPR-characteristics.

B8 is one of the most conserved subunits of complex I with 94% identity between mammals and 40% identity between all instances of occurrence (Figure 3.9). Interestingly, there are two other homologous proteins of about 100 amino acids in size, namely the mitochondrial ribosomal protein L51 and the mitochondrial 28S ribosomal protein S25. These proteins form together the L51-S25-CIB8 domain family in the Pfam Database (Bateman et al., 2002). Besides their common sequence features they seem all to be involved in large protein complexes.

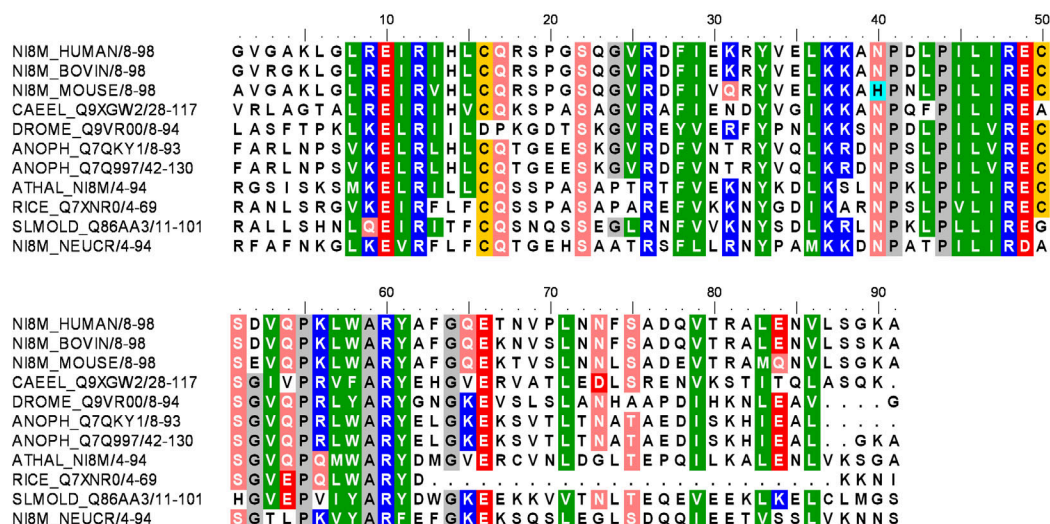


Figure 3.9 Alignment of B8 subunits

Alignment of all sequence homologs of CI-B8 from different species containing the motif PxxPILIRE. Although not all of them are annotated as Complex I subunits in the protein databases, these sequences can be assumed to be the B8 subunits from the respective organisms. The shading threshold was set to 70% identity. Accession numbers (from top): O43678, Q02370, Q9CQ75, XP_214570, CAE71812, Q9XGW2, Q9VR00, Q7QKY1, ATHAL_NI8M, Q7XNR0, AAQ63699, EAA60556, NI8M_NEUCR, EAA47426.1, EAA71427

In this chapter, the NMR solution structure of the oxidized subunit B8 from human complex I will be described.

3.2.2 Structure of CI-B8

Overall a well defined structural ensemble was obtained. The structure shows a four stranded mixed parallel/anti-parallel β -sheet with three α -helices packed against one side (Figure 3.10 B). The ensemble of 19 final structures (Figure 3.10 A) shows an average RMSD to the mean structure of 0.43 Å for the backbone atoms and 0.92 Å for sidechain heavy atoms in the ordered regions (Table 3.2). The helices are composed of residues Q31-K39 (α 1), Y41-A47 (α 2) and A84-G97 (α 3), the strands range from E18-H22 (β 1), I53-K56 (β 2), K64-R68 (β 3) and E74-V77 (β 4).

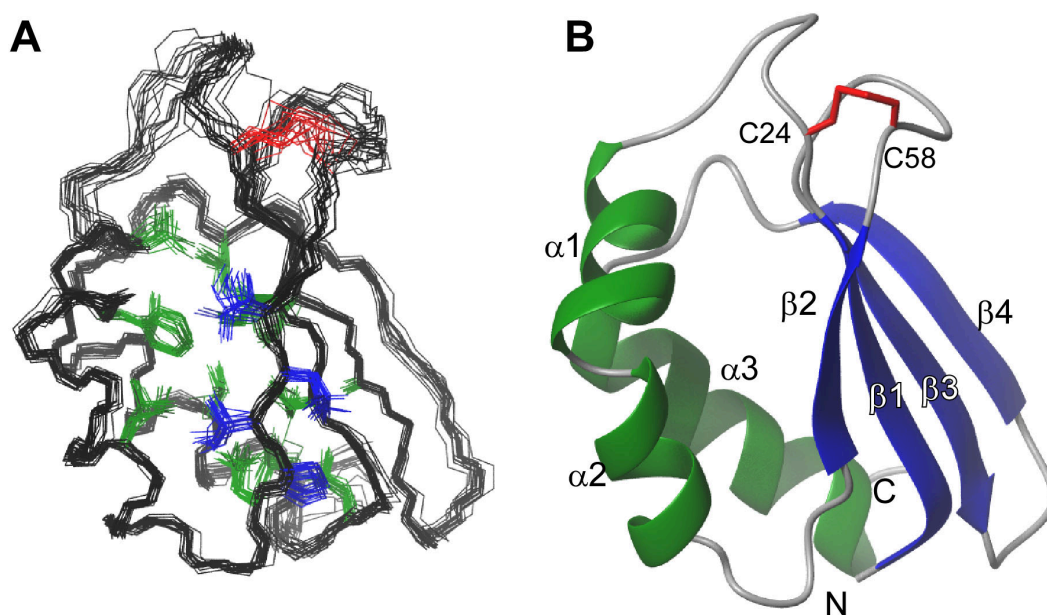


Figure 3.10 Structure of the subunit B8 from Complex I

(A) Ensemble of the 19 lowest-energy structures. The backbone of the protein is shown as black lines. The sidechain heavy atoms of the hydrophobic residues making up the core of the protein are indicated in green. The cysteines C24 and C58 that form the disulfide bridge are drawn in red. The conserved motif PILI P52-I55 is drawn in blue. The residues M1 to L14 that were found to be unstructured are omitted for clarity. **(B)** Ribbon representation of the structure with the lowest RMSD to the mean structure of the ensemble indicating the elements of secondary structure. This figure was prepared with MOLMOL

The hydrophobic core of the Protein is formed by residues L16, I19, I21, V33, F36, L44, L51, I53, L65, A67, L79, V87, L91 and L95. These residues are strongly conserved throughout all proteins of this domain family (Figure 3.9). The kink between $\alpha 1$ and $\alpha 2$ was verified by careful analysis of the 3D-NOESY-spectra. While H^α of residues Q31-F36 show the expected H^α_i to $H^{\beta_{i+3}}$ NOE-crosspeaks, a crosspeak for I37- C^α to Y41- H^β was found. The H^α of E38 and R39 are not involved in contacts to $H^{\beta_{i+3}}$. For residues Y40 to K45 the H^α to $H^{\beta_{i+3}}$ pattern can be found. Furthermore, TALOS-predicted backbone dihedral angles and chemical shift deviations from random coil values also do not support an α -helical structure in the region K39 to Y41.

Restraints:		Quality of the Ensemble:	
<i>NOEs:</i>		<i>RMSD (res. 16-22, 30-56, 63-98): (MOLMOL)</i>	
Used for calculations	1166	Backbone atoms	0.43 Å ± 0.06 Å
Long range	378	Sidechain heavy-atoms	0.92 Å ± 0.07 Å
Medium range	271	<i>Deviations from idealized geometry: (XPLOR-NIH)</i>	
Sequential	517	Bonds	0,0156 ± 0,0003
<i>Other:</i>		Angles	1,9755 ± 0,0451
H-bonds	37	Impropers	2,3166 ± 0,0957
Dihedrals (TALOS)	108	<i>Violations:</i>	
		Distances > 0.5 Å	0
		Dihedrals > 5°	0
		<i>Ramachandran-Plot: (Procheck NMR)</i>	
		Most favoured regions	76.08 %
		Additionally allowed regions	20.41 %
		Generously allowed regions	2.74 %
		Forbidden regions	0.82 %

Table 3.2 Structural statistics for the CI-B8

The regions L23-G29 and E57-Q62 show larger deviations from the mean structure due to a severe lack of NOE-signals involving these residues. Similarly, the amino acids in the region M1 to L14 are not involved in any inter-residue NOE contacts that contain structural information (Figure 3.11 C). Since their signals also show much lower T_1/T_2 -ratios than those signals from the ordered regions of the protein (Figure 3.11 A), these residues were omitted from the structure calculations. Those residues in regions L23-G29 and E57-Q62 of which H-N crosspeaks were observed, however, do not show significantly different dynamics than the ordered part of the protein, as indicated by the T_1/T_2 -ratios. We assume that the lack of signal here is rather due to exchange broadening than to internal disorder.

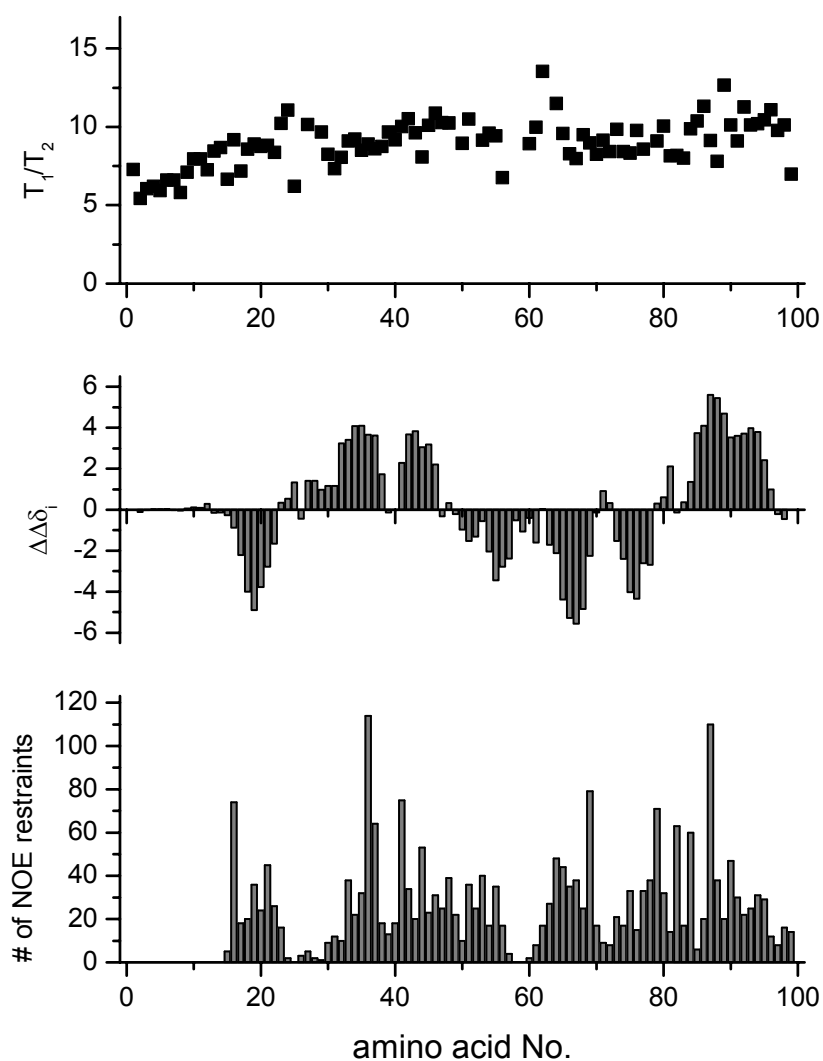


Figure 3.11 Summary of structural parameters for CI-B8

The T_1/T_2 ratio (**top**), the chemical shift difference from the random coil shifts for C_α and C_β as calculated from eq. 3.1 (**middle**) and the total number of NOE restraints per residue (**bottom**) are plotted against the number of the respective amino acid. Smaller T_1/T_2 ratios are indicative of faster motion, positive and negative $\Delta\Delta\delta_i$ indicate α -helices and β -sheets, respectively.

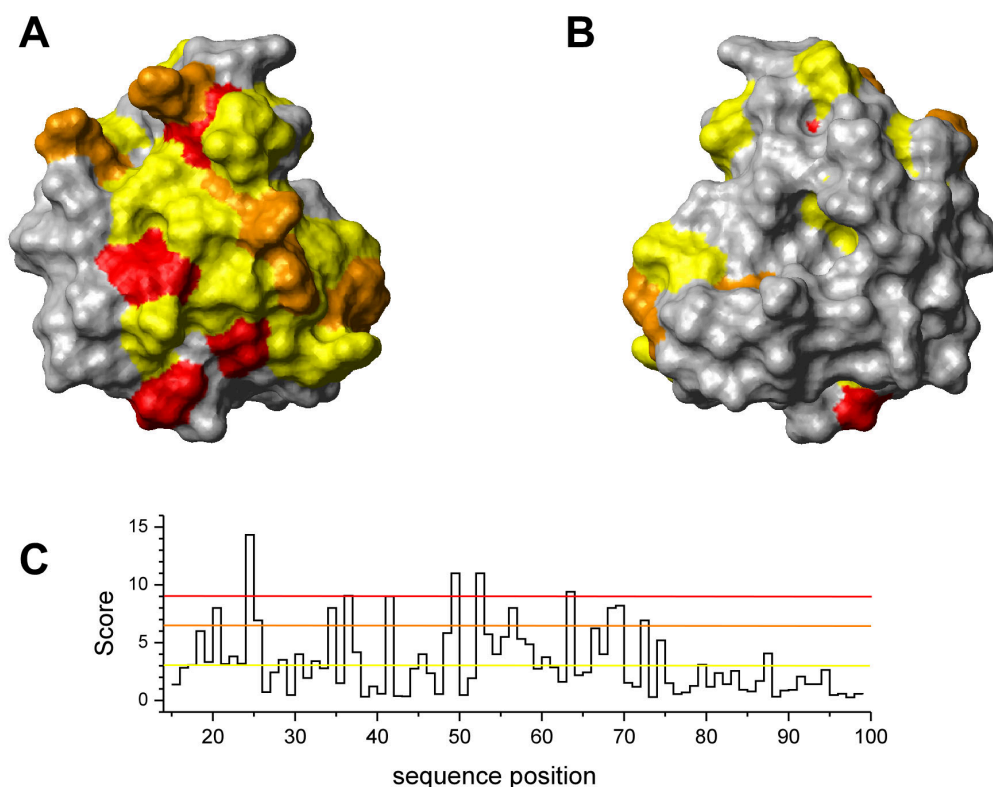


Figure 3.12 Conservation of surface residues in CI-B8

(A,B) Surface of CI-B8 coloured according to the Blossum32 similarity score given in (C). residues with a score bigger than 9 are coloured red, those with a score bigger than 6.5 orange. Residues with scores between 3 and 6.5 are coloured in yellow. In (A) the protein is shown in the same orientation as in Figure 3.10. (B) orientation turned by 180° relative to (A). (C) Blossum32 similarity scores for all positions in the alignment given in Figure 3.9.

Mapping the degree of conservation onto the surface of human CI-B8 reveals that the highly conserved residues cluster all in one area (Figure 3.12 A). While one face of the protein contains almost all conserved residues, the degree of conservation is much lower on the other side of the protein (Figure 3.12 B). The motif 'PILIRE', P52-E57, is in the center of the most conserved area on the surface and the most conserved feature in all B8 subunits (Figure 3.9). It forms a large and shallow hydrophobic patch together with I37, F41 and V42 from $\alpha 1$ and $\alpha 2$ (Figure 3.13 C). Around this central motif, a set of very polar residues is found, namely R34, Q25, E57, R56, R20, R68 (Figure 3.13 C,D). This dominantly positive charge pattern is highly conserved for all known CI-B8 subunits (Figure 3.13 A,B). Worth of note, analysis of the hydrophobic surface potential

revealed another hydrophobic cleft near F82 which is located in the less conserved surface region (Figure 3.12 B).

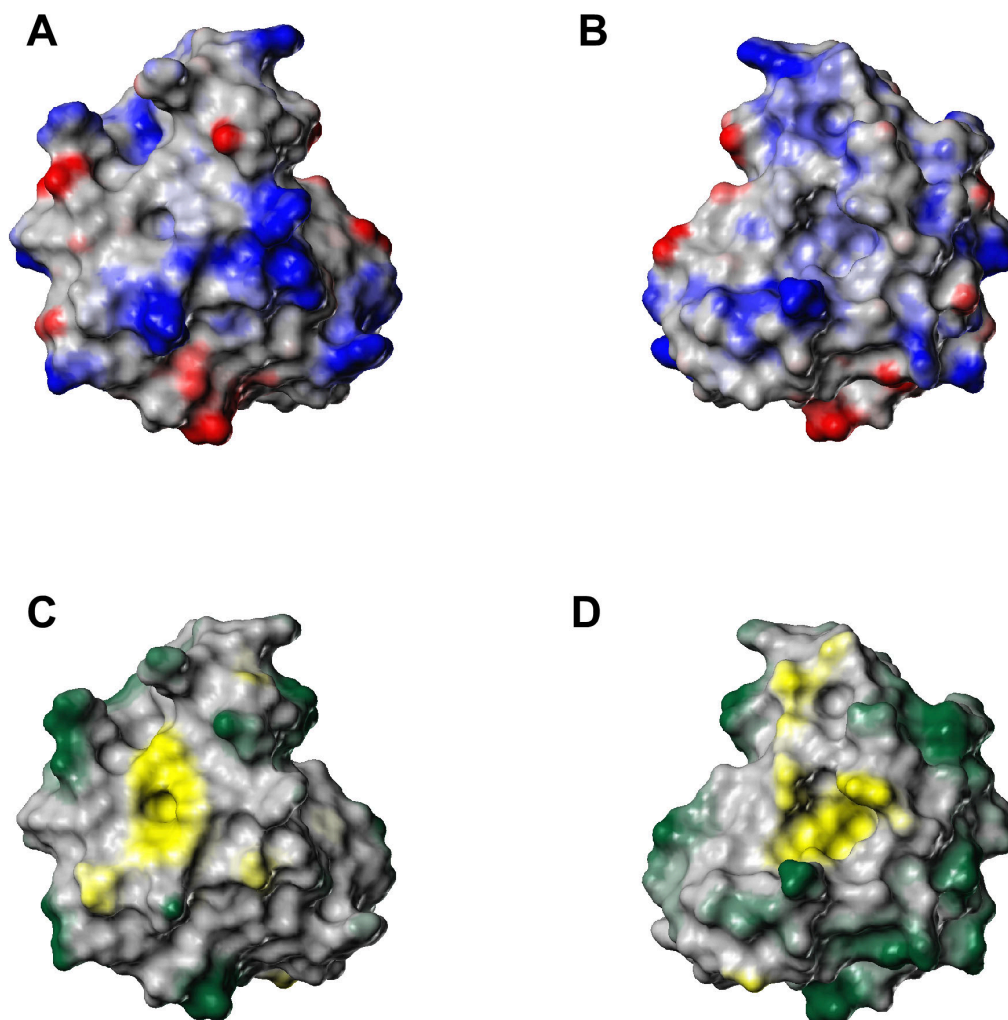


Figure 3.13 Surface properties of CI-B8

(A,B) Electrostatic surface. Positively charged areas are coloured in blue, negatively charged areas in red. (B) orientation turned by 180° relative to (A). (C,D) Hydrophobic surface. Hydrophobic areas are coloured in yellow, hydrophilic areas in green. (D) has the same orientation as (B). In (A) and (C) the protein is shown in the same orientation as in Figure 3.10

3.2.3 Comparison to other Structures

The DALI search showed CI-B8 to be similar to a wide class of thioredoxin-related proteins such as thioredoxins, thioredoxin-reductases and Fe₂S₂-ferredoxins (Table 3.3). The highest Z-scores were obtained for a human C73S thioredoxin-mutant (5.7) and thioredoxin 2 from *Anabaena sp* (5.6) (Table 3.3). Furthermore, a thioredoxin-like Fe₂S₂-ferredoxin from *Aquifex aeolicus* (PDB: 1F37 (Yeh et al., 2002)) showed a Z-score of 4.9. Interestingly, all found thioredoxin-related proteins contain active sites in regions, which can be superimposed with the loops 23-29 and 57-62 in CI-B8.

PDB	Z	%IDE	Protein
1erv	5,7	12	Thioredoxin Mutant (Homo sapiens)
1thx	5,6	9	Thioredoxin (thioredoxin 2) (Anabaena sp.)
1b9y-C	5,4	13	Transducin fragment (gt beta) fragment (gt g) (Bos taurus)
1f37-A	4,9	12	Ferredoxin [2fe-2s] (Aquifex aeolicus)
1qgv-A	4,5	4	Spliceosomal protein u5-15kd fragment (Homo sapiens)
1a8y	4,2	8	Calsequestrin (Oryctolagus cuniculus)
1gp1-A	4	10	Glutathione peroxidase (Bos taurus)
1hyu-A	3,7	12	alkyl hydroperoxide reductase subunit f (ahpf) (Salmonella typhimurium)
1h75-A	3,6	11	Glutaredoxin-like protein nrh (Escherichia coli)
1g7e-A	3,6	11	Endoplasmic reticulum protein erp29 fragment (Rattus norvegicus)
1qmh-A	3,5	8	RNA 3'-terminal phosphate cyclase (Escherichia coli)
1nhy-A	3,4	7	Elongation factor 1-gamma 1 fragment (Saccharomyces cerevisiae)
1eej-A	3,4	9	thiol:disulfide interchange protein (Escherichia coli)
1jfu-A	3,3	9	thiol:disulfide interchange protein ttpa fragment (Bradyrhizobium japonicum)
1g7o-A	3,3	12	Glutaredoxin 2 (grx2) (Escherichia coli)
1psq-A	3,2	7	Probable thiol peroxidase (Streptococcus pneumoniae)
1kte	3,2	8	Thioltransferase (Sus scrofa)
1hd2-A	3,2	6	Peroxiredoxin 5 residues 54-214 (Homo sapiens)
1gwc-A	3,2	12	Glutathione s-transferase tsi-1 (tagstu4-4 glutathione) (Triticum tauschii l.)
1bjx	3,2	11	Protein disulfide isomerase fragment (Homo sapiens)

Table 3.3 Results from the DALI search with CI-B8

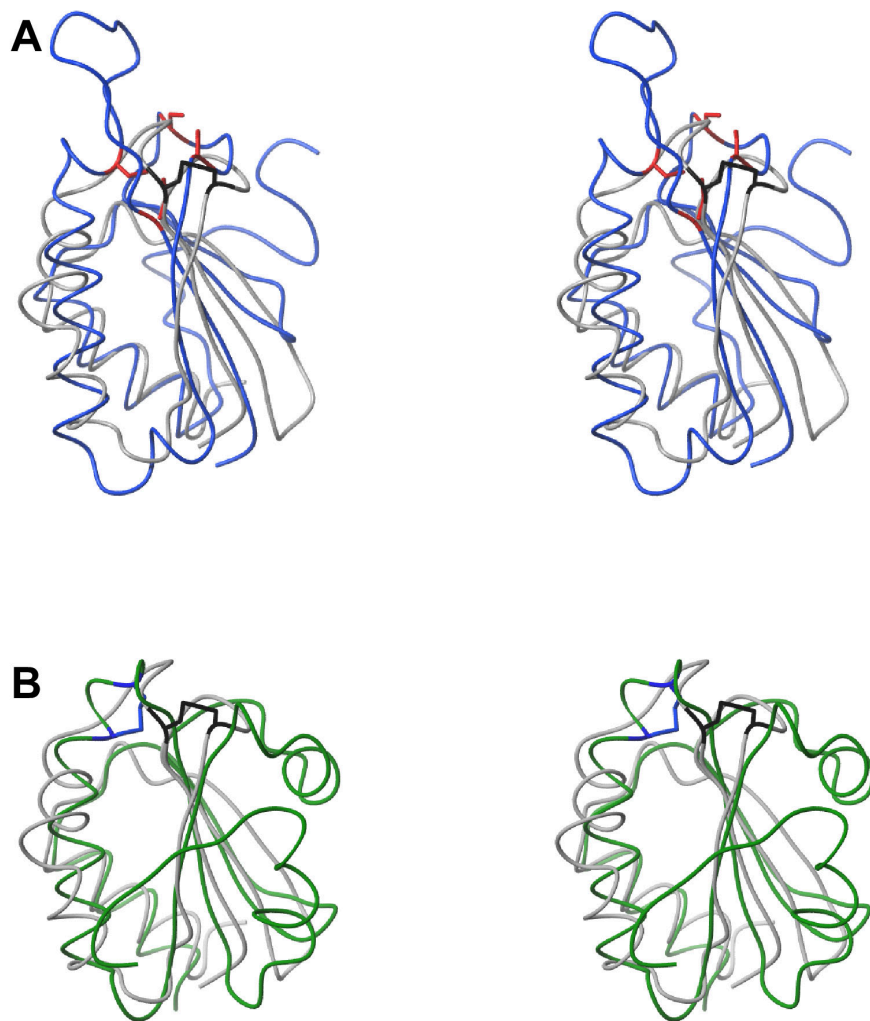


Figure 3.14 Structural comparison to a human thioredoxin and the bacterial ferredoxin from *Aquifex Aeolicus*.

(A) Superposition of CI-B8 (gray), and of the *A. aeolicus* ferredoxin (PDB: 1F37)(blue). The positions of the cysteine residues are indicated in black for CI-B8 and in red for the ferredoxin. The fit resulted from a superposition of the C $^{\alpha}$ of residues 15-26,33-38,40-51,63-68,74-80,82-94,95-98 of CI-B8 onto the ones of residues 1-12,27-32,33-44,64-69,70-76,77-89,91-94 in the ferredoxin, respectively. (B) The C $^{\alpha}$ trace of CI-B8 (gray)and the trace of the *Anabaena sp.* thioredoxin (PDB: 1THX) (green). Positions of the disulfide bridges are given for CI-B8 in black, for thioredoxin in blue. The fit was achieved by a superposition of the C $^{\alpha}$ of residues 17-31,33-36,41-49,50-58,59-78,80-87,89-97 of CI-B8 onto the ones of residues 22-36,38-41,42-50,52-60,72-91,92-99,100-108 in thioredoxin, respectively

The comparison to thioredoxins, e.g. from *Anabaena sp* (PDB: 1THX (Saarinen et al., 1995)), shows that their first strand of the β -sheet and the helices $\alpha 1$ and $\alpha 3$ are not present in CI-B8 (Figure 3.14 B) but the disulfide bridge of CI-B8 is located in roughly the same area as the one in thioredoxin. However, in CI-B8 it connects two different loops which is in contrast to the well known CxxC-motif of thioredoxins.

The comparison with the ferredoxin from *A. aeolicus* (Figure 3.14 A) does not only show considerable similarity in topology but its active center comprising four cysteines ligating the Fe_2S_2 -cluster is located in exactly the same area as the disulfide bridge in CI-B8. Two of the cysteines match positions with C24 and C58 as they are located at the end of $\beta 1$ and $\beta 2$.

Searches for active site-patterns using the program PINTS against a database of active sites (3A) (Stark et al., 2003) yielded fragments of a variety of binding sites of e.g. alcohol-dehydrogenases or NiFe-hydrogenases (Data not shown). However, these results are highly dependent on the conformation of the regions L23-G29 and E57-Q62 for which we do not know the exact structure. The majority of the found active sites covers the two cysteines C24 and C58.

3.2.4 Biophysical Experiments

The appearance of a large conserved hydrophobic patch on the surface of a protein raises the question, whether these residues are still solvated with water or buried by oligomerization of the protein. To test the surface accessibility of the PILI motif at the surface of CI-B8, the distance between individual the methyl groups and the solvent was measured. This can be achieved by observing the changes in the individual relaxation times upon addition of a relaxation enhancing agent such as Gd^{3+} -DTPA-BMA to the protein solution.

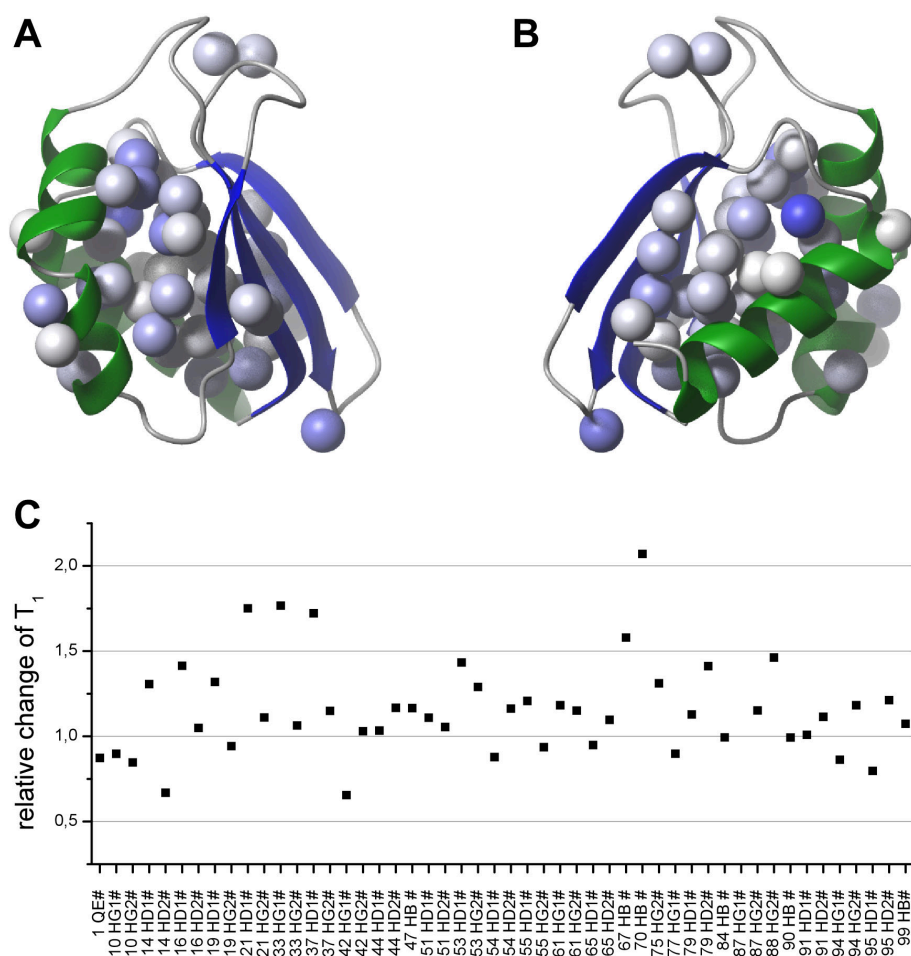


Figure 3.15 Changes of the T₁-relaxation by Gd³⁺-DTPA-BMA

(A,B) Changes in the T₁-relaxation rates of methyl groups upon addition of Gd³⁺-DTPA-BMA. Methyl groups, for those T₁-relaxation rates could be measured are painted as balls. The backbone of the protein is indicated as in (Figure 3.10 B). The intensity of the change is indicated by the level of blue. In (A) the protein is shown in the same orientation as in (Figure 3.10 B). (B) orientation turned by 180° relative to (A). (C) Relative changes of T₁ of the individual methyl groups. Figures (A) and (B) were prepared with MOLMOL.

The changes in the T₁-relaxation rates (Figure 3.15 C) occur in all sequence positions. The larger changes between the Gd³⁺-sample and the unperturbed sample occur, as expected, with a strong tendency towards the surface of the protein (Figure 3.15 A, B). Since some of the methyl groups in the hydrophobic patch exhibit larger changes than other methyl groups, it can be safely assumed that they are as solvent accessible as any other residue on the protein surface.

The structural homology of CI-B8 to thioredoxin-type proteins raised the question if CI-B8 could also have a thioredoxin-like function. Since the disulfide bond in CI-B8 is located in a position similar to that of the catalytically active disulfide bond in thioredoxin-like proteins, the redox-potential of CI-B8 was measured.

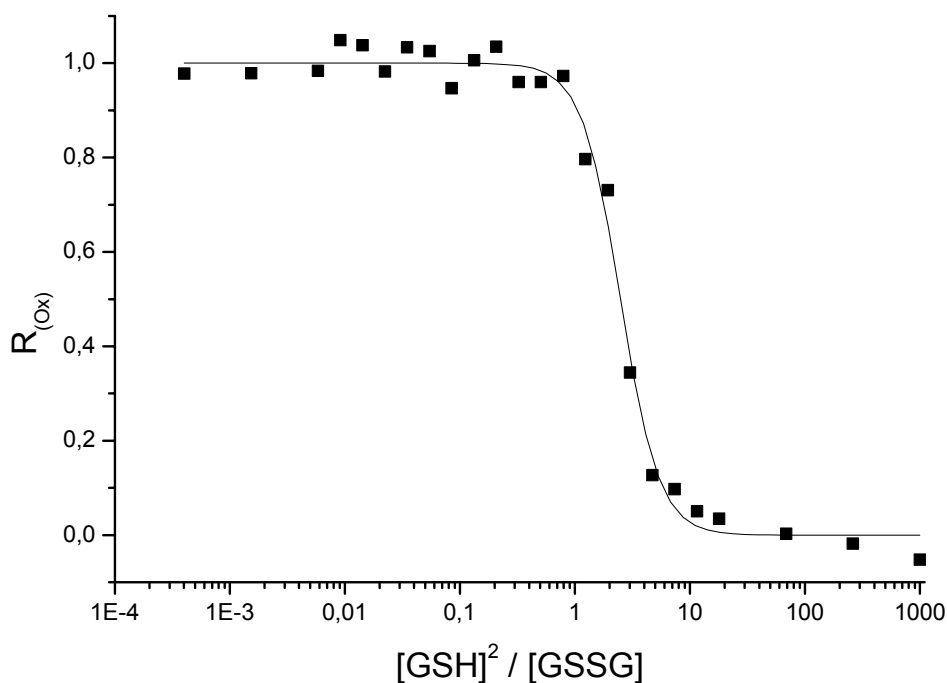


Figure 3.16 Determination of the redox potential

The ratio of oxidized protein was derived from tryptophane fluorescence of CI-B8 determined at various ratios of $[GSH]^2$ to $[GSSG]$. An average from three measurements is plotted as black squares. The fit to the modified model from Mossner et al. (Mossner et al., 1998) is given as a line.

Using a GSH/GSSG redox buffer system a redox potential of -251,6 mV was measured (Figure 3.16), which is very similar to that of other thioredoxin-like proteins.

3.2.5 Discussion

The structure of CI-B8 shows remarkable similarity to thioredoxin related proteins. This structural homology revealed a similar position of the two cysteines (C24 and

C58) in CI-B8 as in both thioredoxins and ferredoxins (Figure 3.14). However, a function based on both cysteines would be limited to mammals, insects and plants as C58 is not conserved in nematodes or fungi (Figure 3.9). One possibility is that the disulfide-bridge forming capability of these cysteines is of functional relevance. The redox potential of $-251,6\text{ mV}$ of the CI-B8 falls well into the range covered by catalytically active disulfide bridges in thioredoxins or thioredoxin-reductases. Thus, it is possible that both cysteines could play a role in redox-processes. In contrast to all thioredoxin-like proteins, though, the topology of the disulfide bridge is different. Furthermore, while the cysteines of the active disulfide-bridge in thioredoxins are located in the same loop forming the well known CxxC-motif, the disulfide bridge in CI-B8 connects two different loops.

However, complexation of metal ions by the two cysteines in CI-B8 seems also possible. Together with H22 which is conserved in most higher organisms (Figure 3.9), they may form a binding site for divalent ions. Such sites can be found, for example, in Sco1 which binds Cu^{2+} (Balatri et al., 2003) or alcohol-dehydrogenases, which bind Zn^{2+} (Li et al., 1994). In both cases, the metal ion is bound between the two cysteines and the histidine in an almost trigonal geometry. In alcohol-dehydrogenases (PDB: 1ADF) a fourth position is occupied by a water molecule. Since we do not know the exact structure of this region in CI-B8, and because H22 is sufficiently close to the cysteines, a metal ion binding site cannot be excluded.

Although also iron-sulphur proteins are found to be similar to CI-B8 by the DALI search, a ferredoxin-like Fe_2S_2 cluster in CI-B8 is not likely. In contrast to ferredoxins, CI-B8 contains only two cysteines which are not sufficient to accommodate an Fe_2S_2 -cluster. Although some thioredoxin-related Fe_2S_2 -ferredoxins with a Cys₃Ser-ligation of their cluster are known (Yeh et al., 2002), there are so far no proteins with a Cys₂Ser₂-type ligation of a Fe_2S_2 -cluster. In CI-B8, however, only S27 and S59 would be able to adopt a geometry sufficient for FeS-cluster ligation together with C24 and C58. These serines are both conserved in mammals and plants but not in insects or fungi.

Subunit B8 is among the most conserved complex I proteins in mammals (94%) (Hirst et al., 2003) and thus likely to play an important role in the function of the complex. Its structure shows a thioredoxin-related fold with residues that could potentially contribute to its function. One possibility is the involvement of the disulfide bridge in redox-processes, since the in-vitro redox potential of CI-B8 (-251,6 mV) lies within the -370 mV to -220 mV range covered by FMN and the FeS-clusters known in complex I (Friedrich et al., 2000). The cysteines may alternatively contribute to metal ion binding or least likely to a FeS-cluster ligation site. From this and the structural similarity to other proteins one can speculate that CI-B8 is either involved in regulating Complex I activity or its assembly via assistance in redox-processes.

3.3 Zinc Finger Domain from the Human Hypothetical Protein BC018415

3.3.1 Available Biological Information

Perhaps the most challenging structures to determine in the course of a structural genomics project are those that deviate from the common protein characteristics. The second An1-like zinc finger from the human hypothetical protein BC018415 (or Q8WV99) is such a case. On a length of 47 amino acids this domain contains two zinc centres that can be predicted from the alignment (Figure 3.17). Therefore, most of the internal space of the protein is occupied by the zinc centres leaving little space for a hydrophobic core.

Zinc fingers in general are a very diverse class of protein domains (Laity et al., 2001). Since their first discovery as a repeated DNA binding motif with a zinc centre coordinated by two cysteines and two histidines (short CCHH) and a $\beta\beta\alpha$ -framework, many additional zinc finger domains with different coordination and fold patterns have been found extending the classical zinc-finger domain definition to all small proteins that rely on zinc ions to define their tertiary structure. The functions of known zinc fingers include DNA and RNA binding as well as apoptosis regulation, protein

folding and assembly and lipid binding (Laity et al., 2001). It is therefore difficult to generate functional hypotheses for an unknown domain solely from the fact that is a zinc finger-domain.

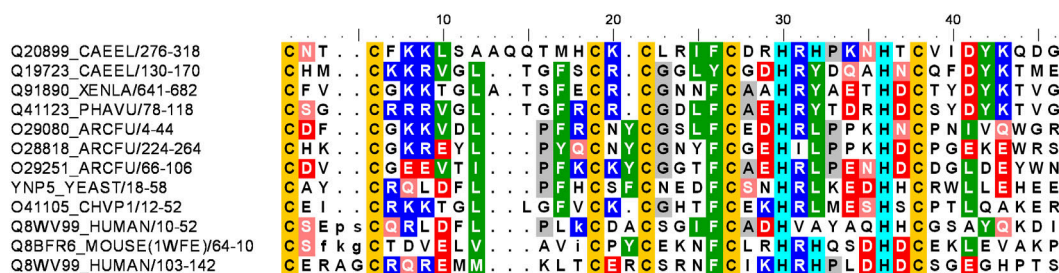


Figure 3.17 Alignment of the An1 like Zinc Domains

The Allingment shows some of the sequences with a An1-like zinc finger present in the Pfam database. The domain discussed here is the second domain of Q8WV99 shown in the last row. The zinc coordinating residues can be easily identified from this alignment. The threshold for shading was set to 40% identity.

The hypothetical protein BC018415(or Q8WV99) contains two An1-like zinc finger domains and two ubiquitin interaction motifs (UMI) (Schultz et al., 2000). Although other proteins with this domain architecture are known, none of them has a known function assigned to it yet. Blast searches for similar zinc finger domains (Altschul et al., 1990) leads to domains that are contained in nucleotide associated proteins such as the “Arsenite inducible RNA associated protein”. However, for these proteins, a detailed molecular function is also unclear.

3.3.2 Strategy for Structure Determination

Although the identities of the Zinc coordinating residues can be inferred from the alignment, the coordination of the two zinc centres is unknown. The two complexes could either be symmetrical (both CCCH) or different (one CCCC and the other CCHH). Recently, the Riken structural genomics consortium published a series of similar zinc finger structures in the PDB database of which the sequence in 1WFE showed the largest similarity to BC018415 (45% identical residues). In this structure, a solution for the assignment of the zinc coordinating residues was included although it

was not reported how this assignment was achieved. In the structure 1WFE, the residues that correspond to C5, C9, C28 and H31 in the construct analysed here coordinate one of the zinc-ions, while C20, C22, H37 and C39 coordinate the other. Interestingly, the N^δ of H31 is reported to coordinate the zinc-ion in the first complex while N^ε of H37 is believed to coordinate the other. All interpretation of our structural data was aided by this reported structure.

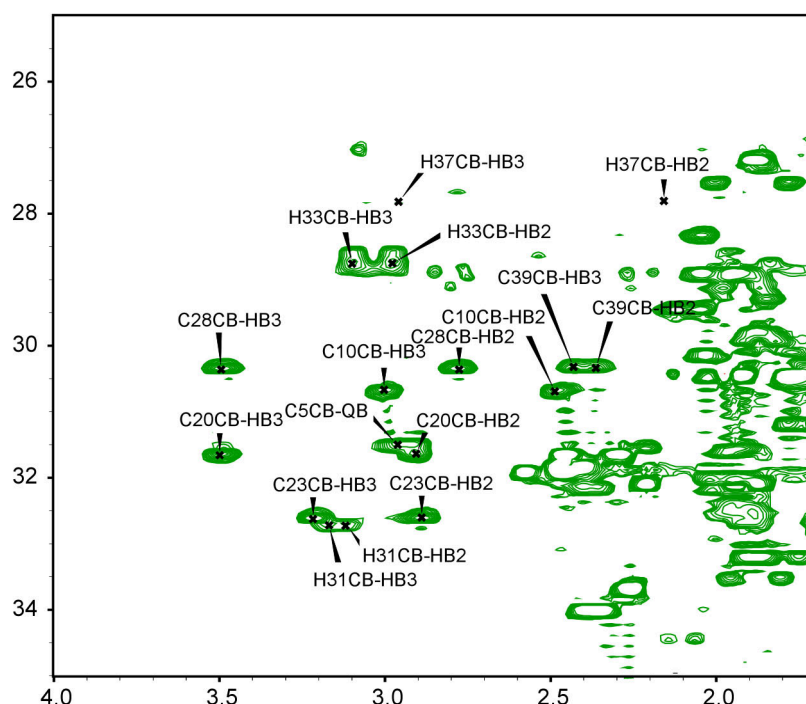


Figure 3.18 Section of the ^{13}C -HSCQ spectrum of the An1-like zinc finger domain

The Figure shows the region of the spectrum that contains all the C^β-H^β cross peaks of the histidines and cysteines. Since these represent the majority of the protons present in the core of the protein, NOE assignment is difficult and alternative interpretations might be possible.

The assignment of the resonances was performed independently as described in chapter 4.2.2. The crucial assignment of the histidine sidechains was achieved using the information from the ^{13}C -NOESY spectra. The H^β-protons of these residues are involved in most of the long range contacts within the protein and are thus important for the correct determination of the structure. Interestingly, all cysteine and histidine

residues that are coordinating the zinc centres show $H^\beta-C^\beta$ chemical shifts in the same chemical shift range (Figure 3.18).

A model of the structure of the second An1-like zinc-finger domain was built from its sequence based on the structure of 1WFE. The interpretation of the NOE data was guided by this model. The NMR-structures obtained this way are in good agreement with unrestrained CANDID/DYANA runs. Since the data did not contain information about the coordinating residues of the zinc centres, this information was also taken from the model. To obtain good ramachandran-characteristics for the backbone, it was necessary to restrict the backbone dihedral angles to region ± 30 degrees to that of the model. These restrictions did not contradict the NOE-based restraints.

3.3.3 Structure of the An1-like Zinc Finger

The structure of the An1-like zinc finger domain of BC018415 shows two short stretches of antiparallel β -sheet the first comprising residues S2-K4 and R14-M26, the second K17-E21 and S24-F27. The residues I29-R32 and P34-H37 both form a single turn of 3_{10} -helix (Figure 3.19).

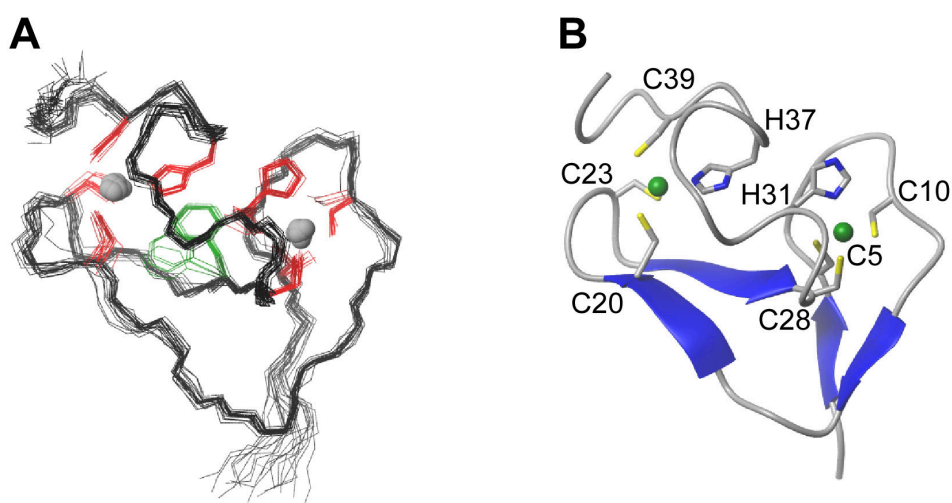


Figure 3.19 Structure of the second Zinc finger of the hypothetical protein BC018415

(A) Ensemble of the 19 lowest energy structures. The backbone of the protein is shown as black lines. The heavy atoms of F27 making up the core of the protein are indicated in green. The sidechains of the cysteine and histidine residues that coordinate the zinc ions (grey) are drawn in red. (B) Ribbon representation of the structure with the lowest RMSD to the mean structure of the ensemble indicating the elements of secondary structure. The sidechains of the zinc coordinating residues are represented as sticks the zinc ions as green spheres. Figure prepared with MOLMOL.

The structure displays some remarkable features. All positively charged residues cluster at one end of the protein (Figure 3.20). Considering the fact that the protein shows similarities to RNA associated proteins (Section 3.3.1) this might constitute a possible interaction site with nucleic acids.

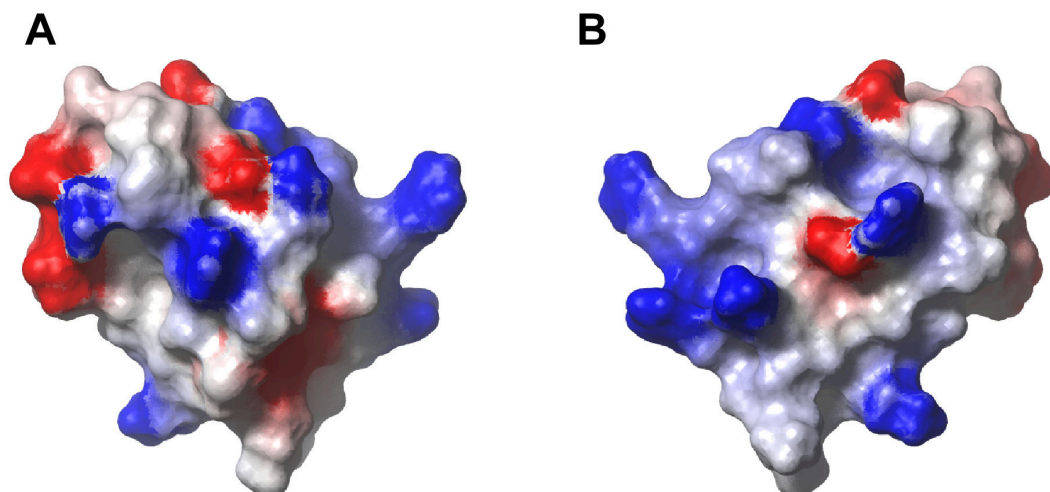


Figure 3.20 Surface properties of the zinc finger domain of BC018415

(A,B) Electrostatic surface. Positively charged areas are coloured in blue, negatively charged areas in red. In (A) the protein is shown in the same orientation as in Figure 3.19 (B) orientation turned by 180° relative to (A). Figure prepared with MOLMOL.

3.3.4 Quality of the Obtained Ensemble

The characterisation of the An1-like zinc finger of BC018415 based on the chemical shift and the relaxation times shows two interesting features. The first is the very uniform distribution of relaxation times (Figure 3.21 A): only a few amide groups seem to relax faster than the average. The other most prominent feature is the complete lack

of systematic secondary chemical shift deviations indicating secondary structure (Figure 3.21 B). Also the backbone angle prediction software TALOS (Cornilescu et al., 1999) or the analysis of the NOESY spectra do not indicate any regular secondary structure. The failure of the prediction methods might be due to the fact that the six zinc coordinating cysteines, which show very unusual C^β chemical shifts, make up almost 14% of the protein. Since the analysis in TALOS is usually done for triples of amino acids, almost every triple is obscured by a zinc coordinating cysteine residue. Nevertheless, also the NOE analysis does not reveal many NOE-pattern indicative of regular secondary structure. Compared to other zinc finger domains, the An1-like zinc finger seems to have an unusually low content of secondary structure. Also the structure of 1WFE does not show much secondary structure.

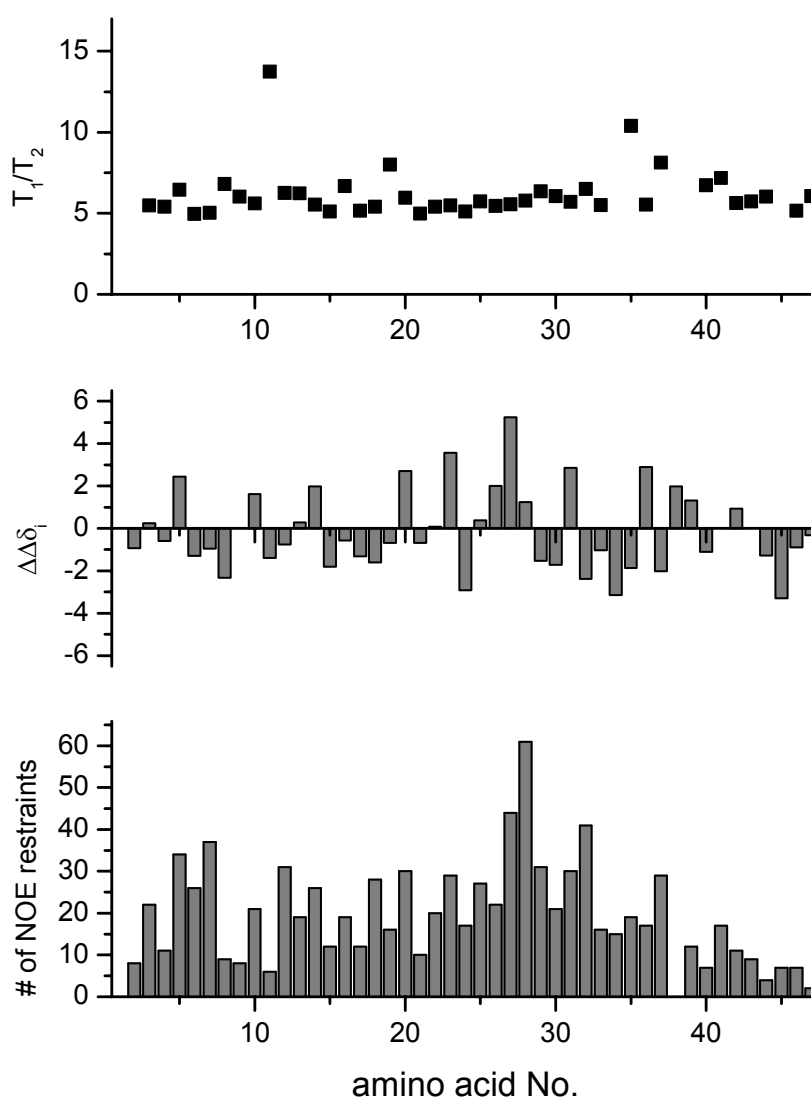


Figure 3.21 Summary of structural parameters for the An1-like Zinc Finger

The T_1/T_2 ratio (**top**), the chemical shift difference from the random coil shifts for C_α and C_β as calculated from eq. 3.1 (**middle**) and the total number of NOE restraints per residue (**bottom**) are plotted against the number of the respective amino acid. Smaller T_1/T_2 ratios are indicative of faster motion, positive and negative $\Delta\Delta\delta_i$ indicate α -helices and β -sheets, respectively.

Although there is a sufficient amount of NOE restraints (see Figure 3.21), they do not define the backbone of the protein in a way that conforms to the favoured regions of the ramachandran plot. Only if the dihedral angles are loosely constrained to the angles of the structure of 1WFE, a structural ensemble can be obtained showing very good agreement of the coordinates of backbone and sidechain heavy atoms (Table 3.4).

The distribution of the backbone dihedral angles is worse than expected for a high quality structure but within the range that is usually obtained for NMR structures (Nabuurs et al., 2004). If the angle restraints constructed from the model are not used, this distribution is much worse (residues in most favoured regions 50.0%, additional allowed regions 40.6%, generously allowed regions 6.0%, forbidden regions: 3.4%), although there is virtually no difference in the amount of restraint violations.

Restraints:		Quality of the Ensemble:	
<i>NOEs:</i>		<i>RMSD (res. 3-42): (MOLMOL)</i>	
Used for calculations	450	Backbone atoms	0,40 ± 0,098
Long range	168	Sidechain heavy-atoms	1,16 ± 0,171
Medium range	77		
Sequential	205	<i>Deviations from idealized geometry: (XPLOR-NIH)</i>	
		Bonds	0,0131 ± 0,0004
<i>Other:</i>		Angles	1,6706 ± 0,0673
H-bonds	0	Impropers	1,6447 ± 0,0470
Dihedrals (Model)	38		
		<i>Violations:</i>	
		Distances > 0.5 Å	1
		Dihedrals > 5°	0
		<i>Ramachandran-Plot: (Procheck NMR)</i>	
		Most favoured regions	74,30%
		Additionally allowed regions	25,40%
		Generously allowed regions	0,30%
		Forbidden regions	0,10%

Table 3.4 Structural statistics for the hypothetical protein BC018415

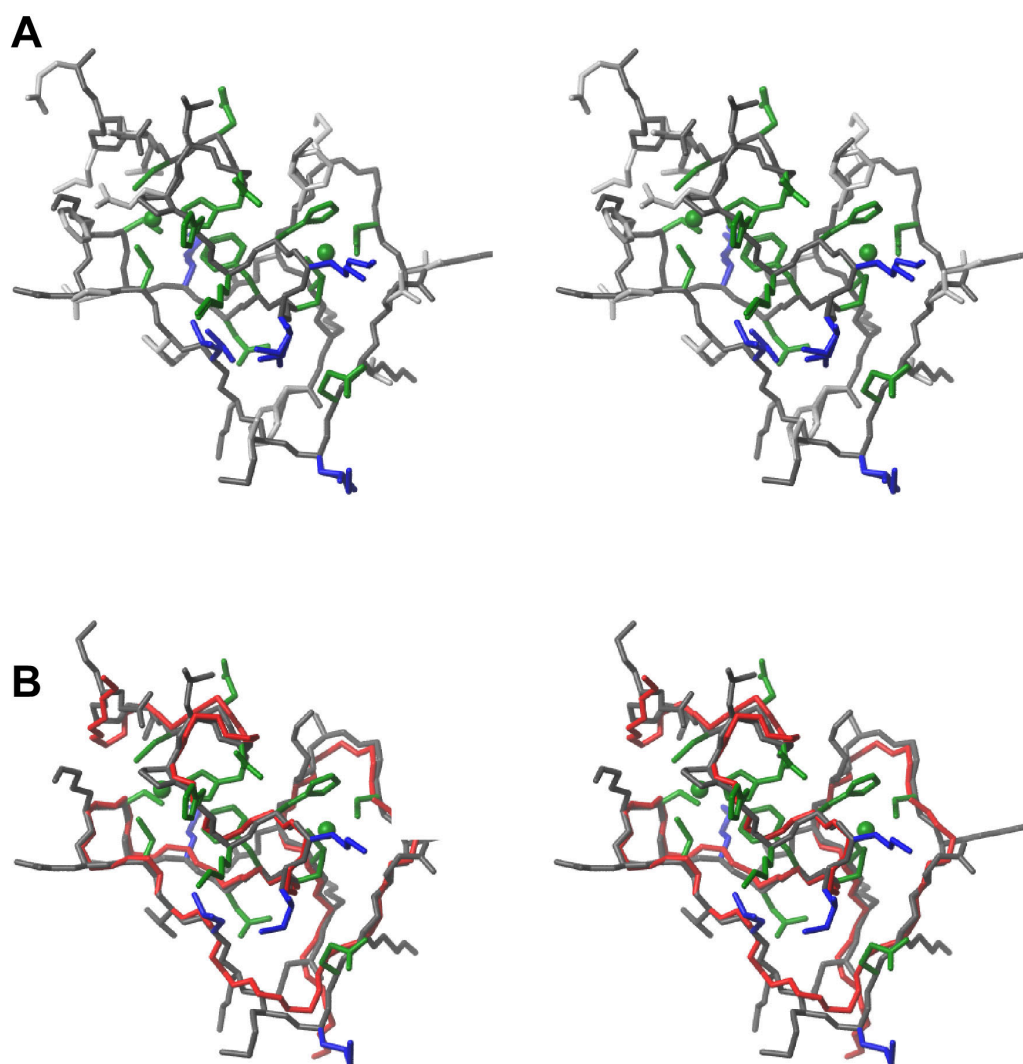


Figure 3.22 Comparison of the initial model with the structure of 1WFE and the final structure

(A) Comparison of the initial model (dark grey) with the structure of 1WFE (light grey). Identical residues are plotted in green similar residues in blue. All other residues have the corresponding shade of grey. The backbone is identical for both molecules. **(B)** Backbone of the representative structure of the final NMR ensemble compared to the initial model on that the assignments were based. The model is plotted as in (A), the backbone of the final structure is traced in red. Figure prepared with MOLMOL.

An analysis of the structure in light of the NOE constraint list reveals cases, where one would expect clear NOE-crosspeaks such as H^{α} - H^{α} contacts in the β -sheets, but which are absent. Also in regions, where the few hydrophobic contacts occur in the

structure, a number of NOE peaks is missing. This leads to the assumption that there are either severe problems with the sensitivity of the NOE spectra, and indeed especially the ^{13}N OESY-HMQC for the aromatic residues shows only a very few peaks, or that the NMR ensemble obtained is not accurate enough. The lack of a clear measure for the precision of the structure determination (Section 1.4.4) makes it impossible to determine the agreement of the calculated structure with the data.

Comparing the structure to the initial model that was used to assign the NOE-crosspeaks (Figure 3.22) shows that the NOEs support a structural ensemble that is very close to the initial model. The near perfect homology model that can be build on the basis of 1WFE (Figure 3.22 A) does already show all of the features that are contained in the final ensemble. The fact that our structure calculations in CANDID / DYANA as well as from manually assigned NOEs lead to a structure that is very similar to that independently obtained on a homologous protein shows that the two proteins show at least similar NMR features.

3.4 Towards an Efficient NMR Structure Calculation Strategy

In this section the structure determination process, and in particular the calculation strategies, of the three projects will be compared and discussed. From this analysis, a strategy will be developed that should allow obtaining high quality structures faster than through traditional manual assignment strategies. As a last topic the outcome of the experiments related to the function of the proteins will be discussed.

3.4.1 Comparison of the Structure Determination Processes

Although all three proteins show reasonable resolution in their ^{15}N -HSQC and ^{13}C -HMQC experiments, the NMR structure determination process was different in many respects. The SODD-BAG domain is a completely α -helical protein and has a rather repetitive amino acid composition (Table 3.5). The fact that all similar amino acids in the helices tend to be located in very similar structural environments leads to massive

problems in the assignment of ^1H -resonances in the NOE-spectra. Although the secondary structure is well defined through chemical shift analysis and medium range NOE contacts, only very few long range contacts can be assigned unambiguously without a structural model. This lack of unambiguous tertiary structure-determining NOE-cross-peaks leads to severe difficulties for the automated assignment strategy as discussed in Section 3.4.2. Rather than an automated approach, careful NOE-based model building and iterative manual assignment lead to a well-defined structural ensemble for the SODD-BAG domain. Another difficulty for the automated procedures is the unstructured N-terminus that contains almost 20 residues contributing to the assignment possibilities without taking part in the structure.

		SODD-BAG			CI-B8			An1-Zinc Finger			Average
		Nr.	%	D%	Nr.	%	D%	Nr.	%	D%	%
Ala	(A)	4	3,85	-3,98	12	11,88	4,05	1	2,13	-5,70	7,83
Cys	(C)	2	1,92	0,38	2	1,98	0,44	6	12,77	11,23	1,54
Asp	(D)	5	4,81	-0,50	4	3,96	-1,35	2	4,26	-1,05	5,31
Glu	(E)	16	15,38	8,76	6	5,94	-0,68	4	8,51	1,89	6,62
Phe	(F)	1	0,96	-3,04	3	2,97	-1,03	1	2,13	-1,87	4
Gly	(G)	6	5,77	-1,17	8	7,92	0,98	4	8,51	1,57	6,94
His	(H)	1	0,96	-1,32	1	0,99	-1,29	4	8,51	6,23	2,28
Ile	(I)	5	4,81	-1,10	5	4,95	-0,96	1	2,13	-3,78	5,91
Lys	(K)	12	11,54	5,63	7	6,93	1,02	3	6,38	0,47	5,91
Leu	(L)	12	11,54	1,90	10	9,90	0,26	2	4,26	-5,38	9,64
Met	(M)	1	0,96	-1,41	1	0,99	-1,38	2	4,26	1,89	2,37
Asn	(N)	1	0,96	-3,21	5	4,95	0,78	2	4,26	0,09	4,17
Pro	(P)	5	4,81	-0,03	5	4,95	0,11	2	4,26	-0,58	4,84
Gln	(Q)	7	6,73	2,79	5	4,95	1,01	1	2,13	-1,81	3,94
Arg	(R)	2	1,92	-3,43	9	8,91	3,56	6	12,77	7,42	5,35
Ser	(S)	9	8,65	1,79	6	5,94	-0,92	4	8,51	1,65	6,86
Thr	(T)	4	3,85	-1,58	2	1,98	-3,45	2	4,26	-1,17	5,43
Val	(V)	8	7,69	0,97	7	6,93	0,21	0	0,00	-6,72	6,72
Trp	(W)	1	0,96	-0,19	1	0,99	-0,16	0	0,00	-1,15	1,15
Tyr	(Y)	2	1,92	-1,14	2	1,98	-1,08	0	0,00	-3,06	3,06

Table 3.5 Comparison of the amino acid composition

The table shows the occurrence of each amino acid in all three proteins, the percentage that this particular amino acid contributes to the protein and the difference from the average contribution ($\Delta\%$) to the SwissProt Database. Differences greater than 5% are highlighted in green and blue.

This is very different in the case of the Complex I subunit B8. It displays a broad mixture of structural features in terms of secondary structure and amino acid

composition. The well-defined secondary structure elements already define most of the topology of the protein since the single β -sheet is interleaved with the three α -helices. In the case of β -sheets NOEs that define secondary structure also define long range contacts and thus help to achieve a well defined structure in the first steps of the automated assignment process. The composition of the amino acids which is fairly similar to the average of all proteins (Table 3.5) further facilitates the automated process since the number of ambiguities is drastically reduced when compared to the SODD-BAG domain. For CI-B8 CANDID results in a near perfect structure in the first completely unconstrained run. This protein is therefore the most successful example of automated structure determination discussed in this chapter.

The An1-like zinc finger domain proves more difficult for the application of automated procedures. Due to its small size and very unusual amino acid composition (Table 3.5) which is caused by the existence of two zinc centres in only 46 amino acids, there are not many strong long range contacts defining the packing of the sidechains within the protein. This is also reflected by the very low overall content of hydrophobic amino acids. Although CANDID produces a structure that is very similar to the homologous structure 1WFE (which itself was calculated completely automatically in CANDID), it is very difficult to find a consistent network of contacts that support this structure. The severe lack of secondary structure does not allow to constrain the backbone of the protein to facilitate a better packing of the sidechains. If the structural calculation is conducted without dihedral restraints extracted from the homology model, the NOE based restraints even fail to restrict the backbone to conformations with favourable Ramachandran characteristics. In this case, where only little experimental data is available, the lack of a reliable correctness measure is most noticeable. The structure retrieved by the process as described in Chapter 3.3 might be anything from the true structure of the domain, a partially unfolded state of a domain without its ligands held together by the zinc ions to an artefact based on errors either the resonance and/or the NOE assignment.

3.4.2 A Practical Strategy for Structure Calculation

For this work, two of the automated NOE-assignment methods that were presented in Section 1.3.3 were available. Both ARIA (Linge et al., 2003) and CANDID/DYANA (Herrmann et al., 2002) are able to produce mostly correct assignments and structures from reasonably clean NOE-peak-lists and resonance assignments. The typical procedure for these programs that is common to most of the NOE-assignment software available is presented in Figure 3.23.

From the resonance assignment and the NOE-peak-list an initial restraint list is built. Using ambiguous distant restraints (ADR) (Linge, 2000), a first ensemble of structures is calculated. From these structures, the contribution of the individual restraints to the ambiguous restraint is evaluated and restraints that have only little contribution are removed. Based on that new restraint set a new ensemble is calculated. Usually this procedure is able to eliminate all incorrect assignments to a peak within seven to eight cycles of structure calculation. Systematic investigations on the starting conditions have shown that it is crucial for the success of these approaches that the correct assignment is contained in the ambiguous restraint and that the number of contributions to a single ADR is not too high (Fossi et al., 2005). The structure calculation algorithms do not succeed if the correct fold is not obtained in the first cycle of calculation.

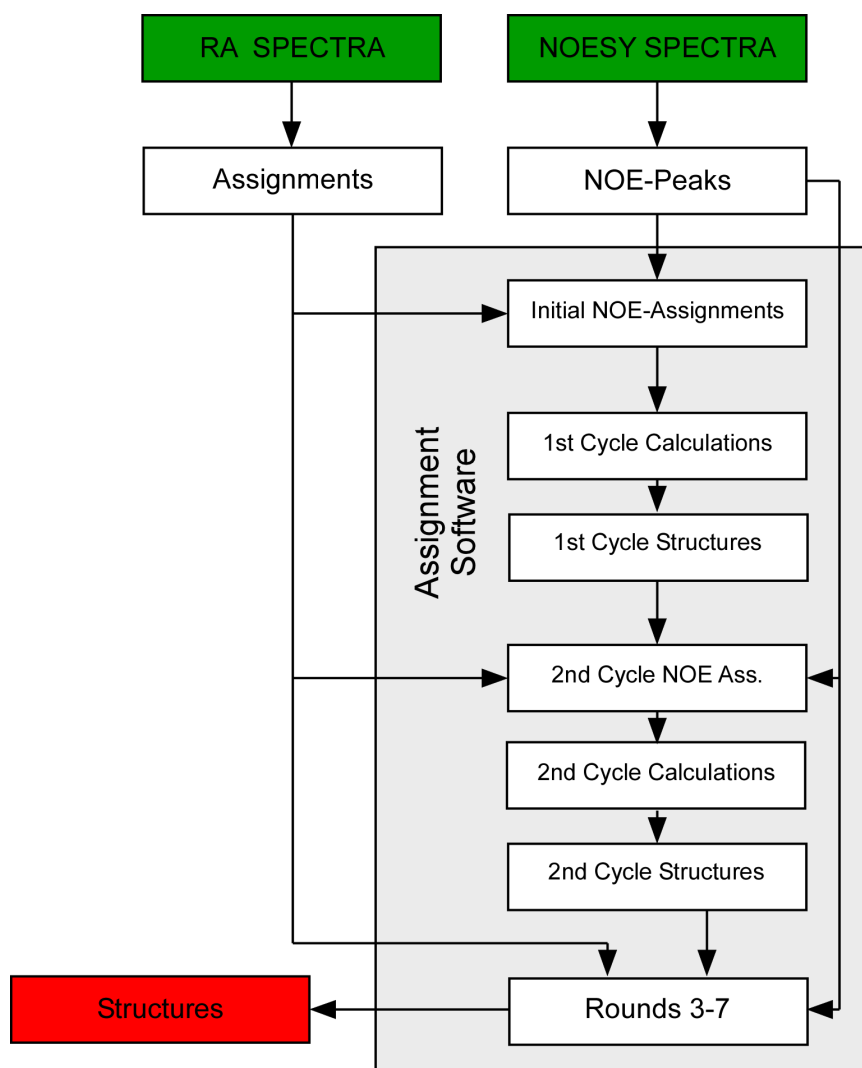


Figure 3.23 Flowchart of a typical automated NOE-assignment/ structure calculation

The figure shows the steps that are implemented in a typical automated NOE-assignment / structure calculation software such as CANDID/DYANA or ARIA. The number of iterative cycles of structure calculation and re-assignment is usually set to seven or eight. The initial NOE-assignment step also includes calibration of the peak volumes.

The development of automated NOE assignment/structure calculation software was initially aimed at complete automation. Thus apart from preparing the input files, these programs require no user input. Also, it is generally assumed that if the software is able to determine a structure, this structure will be the correct one and there will be nothing to improve. However, the requirements for such flawless structure determination are difficult to meet. For example, the resonance assignment should be complete and correct. While the latter is possible to achieve, usually only about 95% of all resonances are assigned. Another requirement for a successful automated structure

determination is that the lists of NOE peaks that are submitted to the procedure are virtually noise-free. Although there are developments for intelligent peak-picking no automated picking procedure is able to generate such noise free peak-lists, yet. Even by manual picking it is sometimes difficult to decide, whether a peak is noise or a real peak.

Thus, in reality automated procedures tend to generate errors. These range from wrongly assigned peaks that lead to violations of correct distance restraints up to major distortions in the structure. Because of the iterative design of the automation procedures it is a very difficult task. The only parameter that can be controlled by the user is the input for the software. But since it is very difficult to predict what the software will return or why errors occurred, modifying the input in a way that minimizes the errors present at the end of the NOE assignment / structure calculation is nearly impossible. Finally, there might be constellations where correct input leads to wrong structures because of chemical shift overlap or regions with only a very few restraints.

Still most of the assignments generated by the software are usually correct thousands of assignments become available after only a few hours. Thus an optimal strategy would combine the improved speed of the automated structure calculation and an error analysis and correction similar to classical manual NOE-assignment.

Tests with ARIA and CANDID on CI-B8 showed that additional mechanisms of CANDID/DYANA in the first few rounds of structure calculation improved the robustness of the algorithm: "restraint combination" and "network anchoring" (Section 1.3.3). These features make the CANDID/DYANA strategy more suitable for the initial assignment from automatic picked NOE-peak-lists, since a greater number of wrong "noise" peaks can be tolerated. In this work the software CANDID/DYANA has been used, although in principle every NOE-assignment/structure calculation software can be used to obtain the initial assignments.

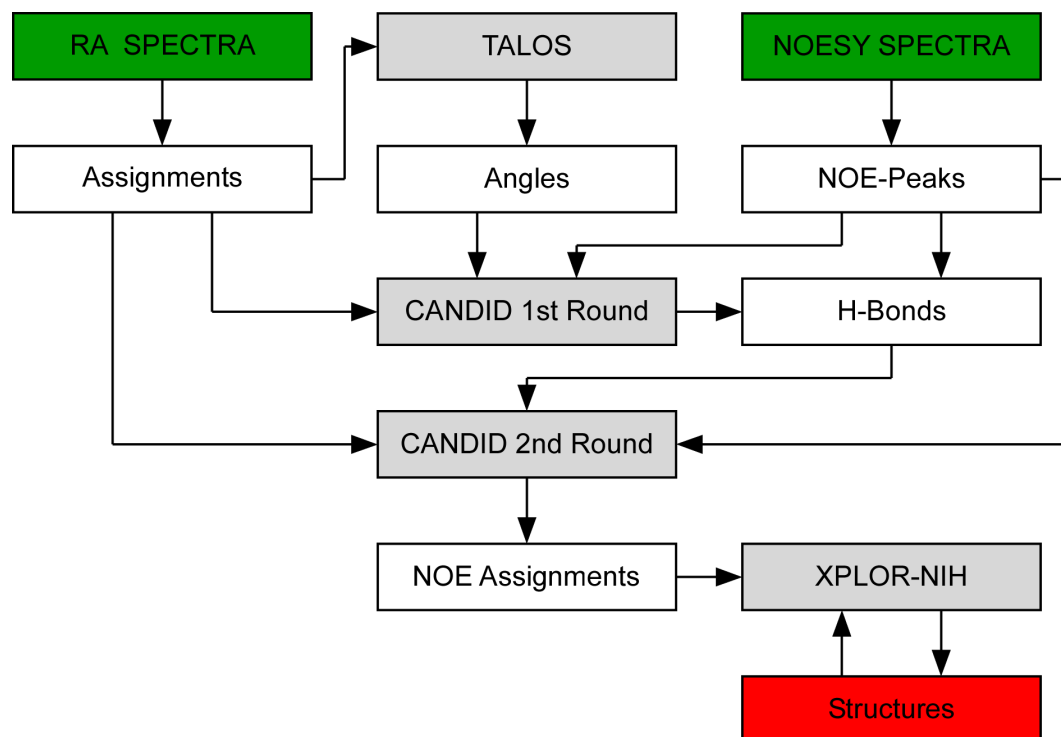


Figure 3.24 Flowchart of the assignment strategy

This flowchart outlines the assignment strategy as used in chapter 3.2. The refinement of the restraint lists was achieved in iterative cycles of XPLOR-NIH structure calculation and correction of the restraint lists.

The final strategy for semi-automated structure determination is outlined in Figure 3.24. Two types of information are retrieved from the software in subsequent cycles. In the first round of structure calculation the only information used is the chemical shift assignment and a NOE-peak-list. If this structure calculation converges, it will result in structures that might have severe flaws in one or more aspects but will have an approximately correct topology in terms of secondary structure and packing. From these initial structures it is easy to confirm the secondary structure with predictions based on chemical shift information (for example from the software TALOS (Cornilescu et al., 1999)) and the characteristic pattern in the NOE-spectra. All secondary structure that can be confirmed from the spectra can now be used in the calculations as H-bond restraints. The second round of automatic structure calculation now additionally includes those H-bond restraints and will result in greatly improved structures and a mostly self consistent NOE-assignment. This assignment can now be refined in a more traditional of structure determination for example in XPLOR-NIH

(Schwieters et al., 2003), in which the assignment is not changed by the structure calculation software. This allows to track the impact of changes in the assignments on the structure and to eliminate erroneous restraints in iterative cycles of correction and structure calculation. This evaluation can be repeated until the structural ensemble fulfils the quality criteria. If the calculation itself is fast enough so that the correction is the time limiting step, it is possible to obtain high quality NMR structures from the resonance assignment in a few days.

3.4.3 Considerations on the Additional Experiments

The additional experiments described in this chapter, although they do not strictly fit into the scope of a structural genomics project, display the range of possibilities that structure determination by NMR opens to investigate protein properties. In the case of the SODD-Bag domain there was detailed knowledge on molecular interaction partners such as HSP70 or BCL-2. The biological function of these interactions, though, remains unclear. With the NMR-structure of the SODD-BAG domain it was possible to further investigate the structural details of the interaction to BCL-2 and HSP70. Peptide titration studies with peptides from BCL-2 selected from its surface using *in silico* docking studies did not yield any results (data not shown in this thesis) thus the focus was shifted towards the interaction with HSP70, where structural information was available from the hBag1/HSP70 complex. Based on information from the complex of this homologous protein it was possible to model the SODD-BAG/HSP70 complex and to discover new interactions specific to the SODD-subfamily of BAG-domains that are not found in the complex of hBag1.

Though the overall function of complex I is very clear and has been studied for decades, no detailed function can be assigned to most of the 46 (in case of the mammalian complex) subunits. Thus the NMR structure determined here was used to generate hypotheses about the possible function using bioinformatical methods. A fold-similarity to thioredoxin-like proteins was discovered and a putative metal binding site was proposed instead of the found disulfide bridge. To further elucidate

the thioredoxin like function, the redox-potential of the disulfide bridge was subsequently measured to verify if thioredoxin like function is possible. All these findings lead to the conclusion that Ci-B8 might have a catalytical function of its own that is most likely involved in redox-processes.

Least was known about a possible the function of the An1-like Zinc-finger domain from the hypothetical protein BC018415. Here only a homology to an “Arsenite inducible RNA associated protein” is found. The occurrence of a large number of arginine residues (Table 3.5) forming a cluster of positive charges on the surface of the protein leads to the conclusion that the domain might interact with nucleic acids, a function commonly found for zinc fingers domains.

3.4.4 Conclusion

In summary, the protein structures discussed here show both the possibilities as well as the limitations of NMR structure determination in a structural genomics approach. The example of CI-B8 shows how NMR structure determination can work in an efficient and time-effective manner. Both other proteins highlight the problems that may occur if unfortunate features such as low variations in the amino acid composition and a mostly α -helical secondary structure come together. In such cases manual intervention is required to obtain high quality high resolution NMR structures, which may require months of expert time. Nevertheless, the general approach, in which manual refinement of the NOE-assignment is combined with automated methods, leads to faster high quality NMR structure determination.

3.5 References for Chapter 3

Abdrakhmanova,A., Zickermann,V., Bostina,M., Radermacher,M., Schagger,H., Kerscher,S., and Brandt,U. (2004). Subunit composition of mitochondrial complex I from the yeast *Yarrowia lipolytica*. *Biochim. Biophys. Acta* 1658, 148-156.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.

- Anderson,S., de Bruijn,M.H., Coulson,A.R., Eperon,I.C., Sanger,F., and Young,I.G. (1982). Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156, 683-717.
- Antoku,K., Maser,R.S., Scully,W.J., Jr., Delach,S.M., and Johnson,D.E. (2001). Isolation of Bcl-2 binding proteins that exhibit homology with BAG-1 and suppressor of death domains protein. *Biochem. Biophys. Res. Commun.* 286, 1003-1010.
- Au,H.C., Seo,B.B., Matsuno-Yagi,A., Yagi,T., and Scheffler,I.E. (1999). The NDUFA1 gene product (MWFE protein) is essential for activity of complex I in mammalian mitochondria. *Proc. Natl. Acad. Sci. U. S. A* 96, 4354-4359.
- Balatri,E., Banci,L., Bertini,I., Cantini,F., and Ciofi-Baffoni,S. (2003). Solution structure of Sco1: a thioredoxin-like protein Involved in cytochrome c oxidase assembly. *Structure. (Camb.)* 11, 1431-1443.
- Bateman,A., Birney,E., Cerruti,L., Durbin,R., Eddy,S.R., Griffiths-Jones,S., Howe,K.L., Marshall,M., and Sonnhammer,E.L. (2002). The Pfam protein families database. *Nucleic Acids Res.* 30, 276-280.
- Benit,P., Slama,A., Cartault,F., Giurgea,I., Chretien,D., Lebon,S., Marsac,C., Munnich,A., Rotig,A., and Rustin,P. (2004). Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome. *J. Med. Genet.* 41, 14-17.
- Berman,H.M., Bhat,T.N., Bourne,P.E., Feng,Z., Gilliland,G., Weissig,H., and Westbrook,J. (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* 7 Suppl, 957-959.
- Briknarova,K., Takayama,S., Brive,L., Havert,M.L., Knee,D.A., Velasco,J., Homma,S., Cabezas,E., Stuart,J., Hoyt,D.W., Satterthwait,A.C., Llinas,M., Reed,J.C., and Ely,K.R. (2001). Structural analysis of BAG1 cochaperone and its interactions with Hsc70 heat shock protein. *Nat. Struct. Biol.* 8, 349-352.
- Briknarova,K., Takayama,S., Homma,S., Baker,K., Cabezas,E., Hoyt,D.W., Li,Z., Satterthwait,A.C., and Ely,K.R. (2002). BAG4/SODD protein contains a short BAG domain. *J. Biol. Chem.* 277, 1528-1532.
- Carroll,J., Fearnley,I.M., Shannon,R.J., Hirst,J., and Walker,J.E. (2003). Analysis of the subunit composition of complex I from bovine heart mitochondria. *Mol. Cell Proteomics.* 2, 117-126.
- Cornilescu,G., Delaglio,F., and Bax,A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289-302.
- Endres,R., Hacker,G., Brosch,I., and Pfeffer,K. (2003). Apparently normal tumor necrosis factor receptor 1 signaling in the absence of the silencer of death domains. *Mol. Cell Biol.* 23, 6609-6617.
- Fearnley,I.M., and Walker,J.E. (1992). Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim. Biophys. Acta* 1140, 105-134.

- Fecke, W., Sled, V.D., Ohnishi, T., and Weiss, H. (1994). Disruption of the gene encoding the NADH-binding subunit of NADH: ubiquinone oxidoreductase in *Neurospora crassa*. Formation of a partially assembled enzyme without FMN and the iron-sulphur cluster N-3. *Eur. J. Biochem.* 220, 551-558.
- Fossi, M., Linge, J., Labudde, D., Leitner, D., Nilges, M., and Oschkinat, H. (2005). Influence of chemical shift tolerances on NMR structure calculations using ARIA protocols for assigning NOE data. *J. Biomol. NMR* 31, 21-34.
- Friedrich, T., and Böttcher, B. (2004). The gross structure of the respiratory complex I: a Lego System. *Biochim. Biophys. Acta* 1608, 1-9.
- Friedrich, T., Brors, B., Hellwig, P., Kintscher, L., Rasmussen, T., Scheide, D., Schulte, U., Mantele, W., and Weiss, H. (2000). Characterization of two novel redox groups in the respiratory NADH:ubiquinone oxidoreductase (complex I). *Biochim. Biophys. Acta* 1459, 305-309.
- Gabaldon, T., Rainey, D., and Huynen, M.A. (2005). Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* 348, 857-870.
- Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *J. Mol. Biol.* 277, 1033-1046.
- Hatefi, Y., HAAVIK, A.G., and GRIFFITHS, D.E. (1961). Reconstitution of the electron transport system. I. Preparation and properties of the interacting enzyme complexes. *Biochem. Biophys. Res. Commun.* 4, 441-446.
- Hatefi, Y., and Stempel, K.E. (1969). Isolation and enzymatic properties of the mitochondrial reduced diphosphopyridine nucleotide dehydrogenase. *J. Biol. Chem.* 244, 2350-2357.
- Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319, 209-227.
- Hinchliffe, P., and Sazanov, L.A. (2005). Organization of iron-sulfur clusters in respiratory complex I. *Science* 309, 771-774.
- Hirst, J., Carroll, J., Fearnley, I.M., Shannon, R.J., and Walker, J.E. (2003). The nuclear encoded subunits of complex I from bovine heart mitochondria. *Biochim. Biophys. Acta* 1604, 135-150.
- Jiang, Y., Woronicz, J.D., Liu, W., and Goeddel, D.V. (1999). Prevention of constitutive TNF receptor 1 signaling by silencer of death domains. *Science* 283, 543-546.
- Kim, S.H., Shim, K.S., and Lubec, G. (2002). Human brain nascent polypeptide-associated complex alpha subunit is decreased in patients with Alzheimer's disease and Down syndrome. *J. Invest. Med.* 50, 293-301.
- Kushnareva, Y., Murphy, A.N., and Andreyev, A. (2002). Complex I-mediated reactive oxygen species generation: modulation by cytochrome c and NAD(P)⁺ oxidation-reduction state. *Biochem. J.* 368, 545-553.

- Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11, 39-46.
- Lau, P.P., and Chan, L. (2003). Involvement of a chaperone regulator, Bcl2-associated athanogene-4, in apolipoprotein B mRNA editing. *J. Biol. Chem.* 278, 52988-52996.
- Li, H., Hallows, W.H., Punzi, J.S., Marquez, V.E., Carrell, H.L., Pankiewicz, K.W., Watanabe, K.A., and Goldstein, B.M. (1994). Crystallographic studies of two alcohol dehydrogenase-bound analogues of thiazole-4-carboxamide adenine dinucleotide (TAD), the active anabolite of the antitumor agent tiazofurin. *Biochemistry* 33, 23-32.
- Linge, J.P. (2000). New methods for automated NOE assignment and NMR structure calculation. Books on Demand).
- Linge, J.P., Habeck, M., Rieping, W., and Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315-316.
- Miki, K., and Eddy, E.M. (2002). Tumor necrosis factor receptor 1 is an ATPase regulated by silencer of death domain. *Mol. Cell Biol.* 22, 2536-2543.
- Mossner, E., Huber-Wunderlich, M., and Glockshuber, R. (1998). Characterization of *Escherichia coli* thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. *Protein Sci.* 7, 1233-1244.
- Nabuurs, S.B., Nederveen, A.J., Vranken, W., Doreleijers, J.F., Bonvin, A.M., Vuister, G.W., Vriend, G., and Spronk, C.A. (2004). DRESS: a database of REfined solution NMR structures. *Proteins* 55, 483-486.
- Ozawa, F., Friess, H., Zimmermann, A., Kleeff, J., and Buchler, M.W. (2000). Enhanced expression of Silencer of death domains (SODD/BAG-4) in pancreatic cancer. *Biochem. Biophys. Res. Commun.* 271, 409-413.
- Piehler, J., Roisman, L.C., and Schreiber, G. (2003). New structural and functional aspects of the type I interferon-receptor interaction revealed by comprehensive mutational analysis of the binding interface. *J Biol Chem* 275, 40425-40433.
- Saarinen, M., Gleason, F.K., and Eklund, H. (1995). Crystal structure of thioredoxin-2 from *Anabaena*. *Structure*. 3, 1097-1108.
- Sazanov, L.A., and Walker, J.E. (2000). Cryo-electron crystallography of two sub-complexes of bovine complex I reveals the relationship between the membrane and peripheral arms. *J. Mol. Biol.* 302, 455-464.
- Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28, 231-234.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Marius, C.G. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn Reson.* 160, 65-73.

- Sondermann,H., Scheufler,C., Schneider,C., Hohfeld,J., Hartl,F.U., and Moarefi,I. (2001). Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* 291, 1553-1557.
- Stark,A., and Russell,R.B. (2003). Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. *Nucleic Acids Res.* 31, 3341-3344.
- Starkov,A.A. (1997). "Mild" uncoupling of mitochondria. *Biosci. Rep.* 17, 273-279.
- Takayama,S., Xie,Z., and Reed,J.C. (1999). An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J. Biol. Chem.* 274, 781-786.
- Taylor,E.R., Hurrell,F., Shannon,R.J., Lin,T.K., Hirst,J., and Murphy,M.P. (2003). Reversible glutathionylation of complex I increases mitochondrial superoxide formation. *J. Biol. Chem.* 278, 19603-19610.
- Turrens,J.F. (2003). Mitochondrial formation of reactive oxygen species. *J. Physiol* 552, 335-344.
- Vinogradov,A.D. (2001). Respiratory complex I: structure, redox components, and possible mechanisms of energy transduction. *Biochemistry (Mosc.)* 66, 1086-1097.
- Xia,D., Yu,C.A., Kim,H., Xia,J.Z., Kachurin,A.M., Zhang,L., Yu,L., and Deisenhofer,J. (1997). Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. *Science* 277, 60-66.
- Yadava,N., Houchens,T., Potluri,P., and Scheffler,I.E. (2004). Development and characterization of a conditional mitochondrial complex I assembly system. *J. Biol. Chem.* 279, 12406-12413.
- Yagi,T., Yano,T., Di Bernardo,S., and Matsuno-Yagi,A. (1998). Procaryotic complex I (NDH-1), an overview. *Biochim. Biophys. Acta* 1364, 125-133.
- Yankovskaya,V., Horsefield,R., Tornroth,S., Luna-Chavez,C., Miyoshi,H., Leger,C., Byrne,B., Cecchini,G., and Iwata,S. (2003). Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* 299, 700-704.
- Yano,T. (2002). The energy-transducing NADH: quinone oxidoreductase, complex I. *Mol. Aspects Med.* 23, 345-368.
- Yeh,A.P., Ambroggio,X.I., Andrade,S.L., Einsle,O., Chatelet,C., Meyer,J., and Rees,D.C. (2002). High resolution crystal structures of the wild type and Cys-55-->Ser and Cys-59-->Ser variants of the thioredoxin-like [2Fe-2S] ferredoxin from *Aquifex aeolicus*. *J. Biol. Chem.* 277, 34499-34507.