Stamler R.; Kappe G.; Boelens W and Slingsby C. (2005). Wrapping the alpha-crystallin domain fold in a chaperone assembly. *Journal of Molecular Biology* 353 (1) 68-79.

Wrapping the α-crystallin domain fold in a chaperone assembly

Robin Stamler [1], Guido Kappé [2], Wilbert Boelens [2] and Christine Slingsby [1]

[1] Birkbeck College, Department of Crystallography, Malet Street, London WC1E 7HX, UK.

[2] Department of Biochemistry, Radboud University, Nijmegen, The Netherlands.

Small heat shock proteins (sHsps) are oligomers that perform a protective function by binding denatured proteins. Although ubiquitous, they are sequence variable except for a C-terminal ~90-residue "α-crystallin domain". Unlike larger stress response chaperones, sHsps are ATP-independent and generally form polydisperse assemblies. One proposed mechanism of action involves these assemblies breaking into smaller subunits in response to stress, before binding unfolding substrate and reforming into larger complexes. Two previously solved non-metazoan sHsp multimers are built from dimers formed by domain swapping between the α-crystallin domains, adding to evidence that the smaller subunits are dimers. Here, the 2.5 Å resolution structure of an sHsp from the parasitic flatworm *Taenia saginata* Tsp36, the first metazoan crystal structure, shows a new mode of dimerization involving N-terminal regions, which differs from that seen for non-metazoan sHsps. Sequence differences in the α-crystallin domains between metazoans and non-metazoans are critical to the different mechanism of dimerization suggesting that some structural features seen for Tsp36 may be generalized to other metazoan sHsps. The structure also indicates

scope for flexible assembly of subunits, supporting the proposed process of oligomer breakdown, substrate binding and reassembly as the chaperone mechanism. It further shows how sHsps can bind coil and secondary structural elements by wrapping them around the α-crystallin domain. The structure also illustrates possible roles for conserved residues associated with disease, and suggests a mechanism for the sHsp-related pathogenicity of some flatworm infections. Tsp36, like other flatworm sHsps, possesses two divergent sHsp repeats per monomer. Together with the two previously solved structures, a total of 4 α-crystallin domain structures are now available giving a better definition of domain boundaries for sHsps.

Keywords α-crystallin; *Echinococcus multilocularis;* molecular chaperone; *Schistosoma mansoni;* Small heat shock protein; *Taenia saginata*

Abbreviations used: ACD, α-crystallin domain; DTT, dithiothreitol; sHsp, small heat shock protein; MAD, multi-wavelength anomalous dispersion; RMSD, Root mean square deviation.

Introduction

Small heat shock proteins (sHsps) are ubiquitous, ATP-independent stress response chaperones. They have the smallest monomeric masses of the heat-shock (or chaperone) protein classes, ranging from 12 - 42 kDa, yet they usually associate into large polydisperse oligomers.[1] Their best-characterized role *in vitro* is to prevent the irreversible aggregation of non-native proteins and to deliver them to the ATP-dependent quality control chaperone systems.[2] *In vivo*, their role is believed to be a protective function, maintaining other proteins that are liable to unfold or disassemble.[2] sHsps are found in almost all organisms, with metazoans possessing ~10-16 sHsp genes.[3,4] For human disease, significant members include αA-crystallin, found in the lens, and αB-crystallin and several other sHsps that are distributed in a

variety of tissues, notably muscle and the nervous system.[2] sHsps interact with both myofibrils [5] and cytoskeletal elements.[6,7] Mutations in conserved residues cause cataract, myopathy [2] and neuropathy.[8,9] sHsps are associated with processes involving longevity [10] and cell survival. Consistent with this, sHsps accumulate in human degenerative diseases, particularly in long-lived cells[2], where abnormal protein aggregation is involved. This accumulation in aggregation-associated diseases includes central nervous system degenerative disorders such as Alexander's disease [11], Alzheimer's disease [12], and Parkinson's disease, where sHsps may protect against damaging effects mediated by alpha-synuclein. [13,14] sHsps have also been implicated in apoptosis, tumour growth and development.[15]

In vitro, the ability of sHsps to carry out their role in binding denatured protein is activated by elevated temperature. In addition, there may be higher level control mechanisms that regulate sHsp function. At least for mammalian systems, sHsps are targets for phosphorylation via stress sensitive signalling pathways involving MAP kinase p38 at N-terminal serine residues, which modulates their chaperone functions.[16]

sHsps normally exist as oligomers which are usually polydisperse and change size and organisation on exposure to stress and when interacting with substrate. [17,18,19,20,21] Three possible modes of chaperone action have been suggested: i) The large oligomeric form remains intact as it binds substrate on its surface.[22] i) The large oligomer breaks down into smaller subunits (that may be monomers, dimers or larger assemblies), which exposes hydrophobic surfaces enabling binding of unfolded substrate, and subsequently reassembles into large soluble complexes [2,23] that are then handled by ATP-dependent refolding machinery. iii) sHsp molecules are intercalated

into large insoluble protein aggregates which enables the subsequent disaggregation and refolding by the ATP-dependent refolding machinery. [24,25,26,27]

sHsp monomers have a conserved α-crystallin domain (ACD) of ~ 90 residues, with a variable N-terminus and a short C-terminal extension (Figure 1a).[1] Here the ACD domain is defined on the basis of its conserved tertiary fold β-strands 2-9, whilst β-strand 10 is considered part of the C-terminal extension (Figure 1b). Both N- and C-termini are thought to have a role in oligomerization.[2,28,29,30,31,32]

Two non-metazoan sHsp X-ray structures have been solved to date, Hsp16.5 from the archaeal hyperthermophile *Methanococcus jannaschii* (MJHsp16.5) [33] and Hsp16.9 from the wheat plant *Triticum aestivum* (WHsp16.9).[34] The ACD in both structures was shown to fold into a β-sandwich structure with hydrophobic grooves down each edge, termed the N-terminal and C-terminal edges (Figure 2a). Both structures formed a similarly-arranged dimer through a strand-exchange mechanism involving a loop protruding from the β-sandwich (Figure 2b). In both cases, the dimer is assembled into larger oligomeric forms, MJHsp16.5 forming a spherical 24-mer, whilst WHsp16.9 forms two six-membered rings stacked face to face to form a dodecamer. This higher-order assembly of the dimers was achieved through an articulated C-terminal extension from the ACD bearing a short I/V-X-I/V conserved motif (Figure 1b), which serves to patch the groove on the C-terminal edge of the sandwich. This groove has also been shown to be a site of substrate binding [18]. Higher-order assembly was also mediated by patching the groove on the N-terminal edge of the sandwich through the dimer strand-exchange mechanism. The variable N-terminus proximal to the ACD was not resolved in the MJHsp16.5 structure, but for WHsp16.9, six of the twelve N-termini were ordered. These served to stabilise the

oligomer by forming a loose knot between monomers from different rings in the two-ring stack that makes up the dodecamer.

The position of one proposed site of substrate binding in the C-terminal groove of the ACD β-sandwich suggests that the oligomeric structure may dissociate into subunits, exposing this binding site, and that there is flexibility in the mode of reassembly with bound substrate, supporting the hypothesis that the mode of action involves dissociation into subunits and reassembly. The observation that a dimer is a sub-assembly species complements the evidence that the unit of subunit exchange between closely related oligomers for at least some sHsps is a dimer. [23,35]

No metazoan sHsp to date has proved amenable to crystallization, because of their marked polydispersity. However, we now have available for study a monodisperse metazoan sHsp, Tsp36.[36] Tsp36 is a 314-residue sHsp from *Taenia saginata*, the beef tapeworm, and is thought to be located on the surface of the larval stage (oncosphere).[37] Homologous sHsps are found in other parasitic flatworms, including *Schistosoma mansoni* [38] and *Echinococcus multilocularis*.[39] The lifecycle of these parasitic flatworms involves a definitive host in which the adult reproduces and liberates eggs, and an intermediate host where the ingested eggs hatch and develop into the invasive larval form, where the surface-expressed sHsp may have a role in survival [39] or pathogenesis.[40]

For Tsp36, and other homologous parasitic flatworm sHsps, each polypeptide chain possesses two α-crystallin domains, ACD1 and ACD2. Unlike most other sHsps, Tsp36 exists as two distinct oligomeric species - a dimer in reducing conditions and a tetramer in oxidising conditions.[36] This is in marked contrast to the large assemblies seen for most sHsps. Furthermore, the size of the oligomer is smaller than that seen for other sHsps. Although its physiological role is unclear, Tsp36

demonstrates the *in vitro* chaperone properties of other sHsps as assessed using two different substrates (citrate synthase and insulin).[36] For reduction-induced aggregation of insulin, the dimeric form of Tsp36 was able to fully prevent aggregation of a 15 times molar excess of insulin, comparable to the chaperone-like activity of αB-crystallin. With temperature-induced aggregation of citrate synthase, the tetrameric form showed better chaperone-like activity than the dimeric form. Here we describe the X-ray structure of dimeric Tsp36, the first solved structure of a metazoan sHsp.

Results and Discussion

Sequence Conservation and Domain Boundaries

An important aspect of structure/function studies of sHsps is the definition of domain boundaries. Two repeat regions for the Tsp36 sequence can be discerned using RADAR [41] (Figure 1a and 1c). In addition to two α-crystallin domains, ACD1 and ACD2, two repeat segments on the N-terminal side of each ACD can be identified (Repeat 1 N-terminus - R1N, Repeat 2 N-terminus - R2N), with a linker region between ACD1 and R2N. There is a long (80 residue) N-terminal segment prior to R1N, but no identifiable C-terminus following either ACD, unless the linker is judged to represent the C-terminal region of ACD1.

A structural alignment (see Methods) of the two ACDs of Tsp36 with the α-crystallin domains of other parasitic flatworm sHsps and a selection of metazoan and non-metazoan sHsps, with the secondary structure of Tsp36 ACD1 indicated, is shown in Figure 1b. The alignment indicates sites of conservation across all sHsps, and also some distinctions between metazoan and non-metazoan sHsps, including a conserved proline-glycine motif seen in the region between β-strands 3 and 4, and a longer segment between β-strands 5 and 7, both specific for non-metazoans.

3D Structural Analysis

The structure of Tsp36 was determined using multiple-wavelength anomalous dispersion (MAD) method at 2.5 Å (Table 1). Crystallization conditions required the addition of dithiothreitol (DTT), resulting in a dimeric form of the protein.[36] Two molecules (A and B) were found in the asymmetric unit of a $P2_12_12_1$ cell. The electron density map was readily traceable except for gaps at residues 267 - 278 for chain A and residues 144 - 152 and 266 - 279 for chain B. The two molecules are related by a non-crystallographic 2-fold axis, and are thus taken to represent the solution dimer. We will describe the structure hierarchically and with reference to the non-metazoan structures.

The α-crystallin domains

The X-ray structure shows that the two ACDs are β-sandwich structures with a 7-strand arrangement (Figure 3) similar to the previously solved sHsps (see Table 2 for RMSDs). The structure (Figure 4) shows the large protruding loop between β-strands 5 and 7, termed the B5-B7 loop (Figures 2b and 3), which in the previously-solved non-metazoan sHsp crystal structures is involved in dimer formation through strand exchange. The spatial relationship between the connected ACDs does not resemble that seen for the ACDs comprising the dimers of the previously solved structures (Figure 5a and 5b), the two domains within the monomer being orientated such that the B5-B7 loops point away from each other (Figure 4).

The N-terminal regions

In sHsps, it is the N-terminal region that is highly variable and for which 3D structural information is sought, especially in relation to oligomer assembly. In Tsp36

there is a global N-terminal region in front of the connected ACDs as well as short segments (R1N and R2N) local to each ACD (Figure 1a), and all are visible in the X-ray structure. The R1N and R2N segments that are N-terminal to ACD1 and ACD2 respectively are largely coil but are orientated differently, R1N pointing away from the adjoining ACD1 to lie on one side of ACD2, whilst R2N also associates with ACD2, lying on its other side (Figure 4). The 80-residue global N-terminus consists of a series of helices with sharp turns between them marked by prolines (Figure 4 and Figure 1c).

The Tsp36 dimer

The assembly of dimeric Tsp36 does not resemble that seen in the two previously solved structures. This metazoan sHsp does not assemble via strand exchange between ACDs (Figure 5a and 5b), as already suggested by the shorter length of the B5-B7 loop in metazoans. The dimeric assembly is instead mediated mostly by the well-resolved helical packing of the N-termini which cross over the dimeric interface (Figure 4). In addition, the initial coiled region of the N-terminus associates with the edges of ACD1. The dimer buries 3200 $Å^2$ surface area of each monomer.

The Tsp36 dimeric structure provides some clues as to the impact on assembly of the sequence divergence of the metazoan sequences from non-metazoan. The dimerization of WHsp16.9 and MJHsp16.5 involved both strand exchange and a ring stacking interaction of prolines close to the dyad axis (Figure 5c). Comparison of the WHsp16.9 dimer (Figure 5c) shows this proline is absent in ACD1, and the substantially smaller B5-B7 loop is displaced from the strand-exchange position with the aid of a phenylalanine residue. The proline and phenylalanine are conserved in

non-metazoan and metazoans sHsps respectively (Figure 1b). The differences seen for the Tsp36 sequence are conserved across metazoan sHsps and suggest that metazoan sHsps do not assemble into dimers via B5-B7 loop strand exchange and instead have some alternative mode of assembly. For ACD2, whilst missing density prevents the assignment of the B5-B7 loop, the two domains from each partner molecule are arranged in a fashion that precludes the previously seen dimeric conformation (Figure 5b). Together, these findings in Tsp36 suggest a different mode of assembly for metazoan sHsps from that seen previously. Structural and sequence comparisons of the ACDs for Tsp36, WHsp16.9 and MJHsp16.5 are indicated in Table 2.

Implications for Function

The β-sandwich structures of both ACDs create 2 edges, termed here the N-terminal edge (comprised of β-strands 2, 3 and 7) and C-terminal edge (comprised of β-strands 4 and 8) (see also Figure 2a), that are closed off by short stretches of polypeptide in a fashion that is referred to here as patching.

The N-terminal edge of ACD1 is patched intramolecularly by hydrophobic residues from the N-terminus which insert into the hydrophobic groove (Figures 4 and 6). In addition, a conserved arginine abutting this groove (here Arg158) forms a polar interaction with an N-terminal serine residue (Figure 6). Mutations of this conserved arginine in human sHsps are associated with cataract, myopathy[2], and neuropathy[8,9],whilst phosphorylation of N-terminal serines alters sHsp function.[16] The relationship seen here between an edge-located conserved arginine which is disease-related and a serine which is potentially a functionally important phosphorylation site is striking. An edge location may represent either a substrate binding site or a site for higher-order assembly. The disease association indicates the site is functionally

important, whilst its patching by the N-terminal serine suggests a mechanism by which phosphorylation could exert its effect. Thus for human sHsps the interaction and conformation of the β-sandwich edge and the N-terminus may affect either higher order assembly or substrate binding, and that mutation in this region may lead to chaperone failure.

The N-terminal edge of ACD2 is patched less extensively, a short antiparallel β-strand from the R1N segment serving to close the edge.

For the C-terminal edge of ACD1, the first 5 residues of the N-terminus of the partner monomer patches the hydrophobic groove in a fashion that closely resembles the interaction of the C-terminal extension with non-dimer partner related monomers seen in the previously-solved non-metazoan structures (Figures 4 and 7). Tsp36 monomers, as predicted by alignment (Figure 1b), possess no C terminal extension. Thus the N-terminus fulfils the role of the ordered C-terminal extension seen in the previously solved structures in creating higher-order assembly. This interchangeability of role in higher-order assembly between the N- and C-termini may hint at a general device in sHps for flexibility in the mode of reassembly of complex once substrate is bound. As described above, one model of the sHsp chaperone mechanism requires oligomer breakdown, substrate binding at oligomer interfaces and reassembly into complexes. If so, it is likely that the oligomer interfaces that present themselves for higher-order assembly may vary depending on the presence or absence of substrate, and that both N- and C-termini will be required to interact with these interfaces.

The C-terminal edge of ACD2 is also patched by hydrophobic residues, but these come from the linker region and R2N segments of the same monomer rather than from the dimer partner.

A striking feature of the Tsp36 assembly is that ACD2 is "wrapped" by bordering segments R1N and R2N, which form irregular coils. These are in turn enclosed by helices from the long N-terminus, which also encircle ACD1 (Figure 8). This double wrapping around the ACDs by both polypeptide regular secondary structural elements and non-regular coil structure, indicates how sHsps may serve as high-capacity chaperones for unfolding protein. The presence of hydrophobic patches, particularly at the N- and C-terminal edges of the β-sandwich structure, provide the anchoring points by which these extended lengths of unfolding protein may be stabilized (Figure 8). The structure supports the model for sHsp function in which the large oligomer breaks down into smaller subunits, exposing hydrophobic surfaces in the α-crystallin domain which enables binding of unfolded substrate, followed by reassembly into large soluble complexes aided by sequence extensions.

The Tsp36 dimer forms a tetramer under oxidising conditions.[36] Each dimer contains two cysteines located in the B5-B7 loop of ACD1 (Figure 4), the only cysteines present, and which are therefore the sites for tetramer formation by disulphide bonding. The significance of this redox-dependent structural change is not clear, although *in vitro* the tetrameric form is a better chaperone than the dimer. [36]

A model of the tetramer is shown in Figure 9, using the structure of the B5-B7 loop seen in ACD1 for one monomer to model other missing B5-B7 loops. The demonstrated site of interaction between the dimers at the protruding cysteines of the B5-B7 loops creates a covalently linked assembly with a large hole in the centre.

Implications for Pathogenicity

The physiological role of Tsp36 in *T. saginata* larvae is unknown, but could involve self-protection and pathogenesis. It may resemble the single-domain

Mycobacterium tuberculosis sHsp 16.3, which is present in abundance on the external surface of the cell wall and is thought to have a role in infection and also to act as a chaperone in long-term survival. [42] Further light on Tsp36 function may be shed by the sHsps of several parasitic flatworms related to *T. saginata*. Some of these sHsps are major antigens, including a double-domain sHsp from *Schistosoma mansoni* [38], secreted by the flatworm's eggs, which lodge in tissues and cause the destructive chronic granulomatous inflammation that leads to the huge morbidity and mortality of schistosomiasis. The immunodominant T-cell epitope of the sHsp [40] corresponds by sequence alignment to the exposed R2N segment of random coil in the Tsp36 structure (Figure 4).

The oncosphere of *Echinococcus multilocularis*, a tapeworm with a highly invasive larval stage, is an increasingly significant cause of human disease.[43] It also expresses on its surface [39] a double-domain sHsp highly homologous (87% identity) to Tsp36. Uniquely it possesses a putative integrin-binding RGD domain (Figure 1b) on the protruding B5-B7 loop at the site of the cysteine residue seen in Tsp36 (Figure 4). This raises the question as to whether the unusually invasive character of *E. multilocularis* is related to its surface-located sHsp and its RGD domain. The location of the RGD domain on a protruding loop may suggest a possible mechanism of attachment by integrin binding and an area for therapeutic exploration using synthetic RGD peptides, which could act as competitive inhibitors.

Comparison of the metazoan Tsp36 sHsp quaternary structure with single-domain sHsps thus shows how embellishments of the α-crystallin domain fold can control variable assembly that supports protective binding functions as well as creating a potentially pathogenic device harmful to mankind.

Materials and Methods

Protein Expression

The cDNA containing the coding sequence of *T. saginata* Tsp36 [37] was cloned in the pET3a expression vector (Novagen) as described before.[36] Selenomethionine-substituted Tsp36 was expressed in B834(DE3)pLysS cells (Novagen). The transformed cells were grown in Luria broth (LB) medium to an OD_{600} of 0.7, then harvested and resuspended in the same volume of M9 medium.[44] Cells were further cultured for 1 hour at 37°C before protein expression was induced by adding 0.4 mM IPTG with the addition of 50 mg/L seleno-methionine (Sigma-Aldrich). The cells were collected and lyzed in TEN_{50} buffer pH 7.4 (20 mM Tris-HCl, 1 mM EDTA and 50 mM NaCl) by freeze-thawing and subsequently incubated for 30 minutes at 4°C with 0.5 mg/ml lysozyme and another 30 minutes at 4°C with 0.05 mg/ml DNaseI in the presence of 12 mM $MgCl_2$. The lysate was dialyzed against TEN_{50} buffer and the Tsp36 protein was purified by a fast flow DEAE sepharose column (Amersham Pharmacia Biotech) using a gradient from 50 to 300 mM NaCl.

Crystallization and data collection

Purified protein was concentrated to 3.5 mg/ml in 10mM Tris-HCl (pH 7.5), 1mM EDTA, 150 mM NaCl, 1mM DTT. Prior to crystallization, protein was incubated at 37 ˚C for 15 minutes with 5mM DTT. Crystallization was performed at room temperature using the hanging drop method by mixing the protein with an equal volume of reservoir solution consisting of 1.7 M ammonium sulphate, 10mM DTT, 100 mM ADA (pH 6.5), 90 mM sodium potassium tartrate. Rod-like crystals (0.4 mm x 0.07 mm x 0.07 mm) were flash-frozen in liquid nitrogen using a cryoprotectant solution of glycerol 30% v/v, 2M ammonium sulphate, 100mM ADA (pH 6.5), 100

mM sodium potassium tartrate. A MAD dataset were collected at the ESRF on beamline 14.4, and processed using MOSFLM [45], SCALA [46], and TRUNCATE.[47] Crystals diffracted to 2.5 Å, showing a cell size of a = 51.6 Å, b = 102.1 Å, c = 139.5 Å, a space group of $P2_12_12_1$ (Table1), and two molecules in the asymmetric unit (V_M= 2.6 $Å^3$ Da^{-1} for a solvent content of 52%).

Structure determination and refinement

The MAD dataset was processed in SHELXD [48] to determine the sites of the 14 selenium atoms present in the asymmetric unit, and these were refined and phases then calculated in AUTOSHARP [49] (Table 1). The resulting electron density map was improved through multiple cycles of solvent-flipping and density modification.

Model building was carried out using XFIT [50] and refinement was performed using CNS.[51] After each cycle of refinement, the model was manually adjusted against $2F_o$-F_c and F_o-F_c maps. Residues 267-278 of chain A, and residues 144-152 and 266-279 of chain B could not be assigned. Refinement statistics (Table 1) were determined with CNS and PROCHECK.[52]

Structural Alignments

Structural alignment of Tsp36 ACD1, Tsp36 ACD2, WHsp16.9 and MJHsp16.5 was performed using CORA.[53] This alignment was used to generate a Hidden Markov model using HMMER.[54] This HMM profile was then used to align other sequences to it.

Acknowledgements

We are grateful to our colleagues in the Birkbeck Crystallography department and to Wilfried de Jong of Radboud University Biochemistry department for helpful advice and discussions. Data were collected at the ESRF and Daresbury SRS facilities. RS and CS are grateful to the Medical Research Council (London) for financial support.

Co-ordinates

The co-ordinates and structure factors have been deposited in the PDB with accession codes 2BOL and R2BOLSF respectively.

References

1. De Jong, W. W., Caspers G. J. & Leunissen, J. A. (1998). Genealogy of the alpha-crystallin - small heat-shock protein superfamily. *Int. J. Biol. Macromol.* **22**, 151-162.
2. Van Montfort, R., Slingsby, C. & Vierling, E. (2001). Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv. Protein Chem.* **59**, 105-156.
3. Kappé, G., Franck, E., Verschuure, P., Boelens, W. C., Leunissen, J. A. & de Jong, W. W. (2003).The human genome encodes 10 alpha-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress Chaperones,* **8**, 53-61.
4. Ding, L. & Candido, E. P. (2000). HSP25, a small heat shock protein associated with dense bodies and M-lines of body wall muscle in Caenorhabditis elegans. *J Biol Chem.* **275**, 9510-9517.
5. Bullard, B., Ferguson, C., Minajeva, A., Leake, M. C., Gautel, M., Labeit, D., Ding, L., Labeit, S., Horwitz, J., Leonard, K. R. & Linke, W. A. (2004). Association of the chaperone alphaB-crystallin with titin in heart muscle. *J. Biol. Chem.* **279**, 7917-7924.
6. Perng, M. D., Wen, S. F., van den, I. P., Prescott, A. R. & Quinlan, R. A. (2004). Desmin aggregate formation by R120G alphaB-crystallin is caused by altered filament interactions and is dependent upon network status in cells. *Mol. Biol. Cell,* **15**, 2335-2346.
7. Mounier, N. & Arrigo, A. P. (2002). Actin cytoskeleton and small heat shock proteins: how do they interact? *Cell Stress Chaperones,* **7**, 167-176.
8. Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., Michalik, A., De Vriendt, E., Jacobs, A., Van Gerwen, V., Vennekens, K., Mazanec, R., Tournev, I., Hilton-Jones, D., Talbot, K., Kremensky, I., Van Den Bosch, L., Robberecht, W., Van Vandekerckhove, J., Broeckhoven, C., Gettemans, J., De Jonghe, P. & Timmerman, V. (2004). Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat. Genet.* **36**, 597-601.
9. Benndorf, R. & Welsh, M. J. (2004). Shocking degeneration. *Nat. Genet.* **36**, 547-548.
10. Hsu, A. L., Murphy, C. T. & Kenyon, C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science,* **300**, 1142-1145.
11. Head, M. W. & Goldman, J. E. (2000). Small heat shock proteins, the cytoskeleton, and inclusion body formation. *Neuropathol. Appl. Neurobiol.* **26**, 304-312.
12. Mao, J. J., Katayama, S., Watanabe, C., Harada, Y., Noda, K., Yamamura, Y., Nakamura, S. (2001). The relationship between alphaB-crystallin and neurofibrillary tangles in Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **27**, 180-188.
13. Rekas, A., Adda, C. G., Andrew Aquilina, J., Barnham, K. J., Sunde, M., Galatis, D., Williamson, N. A., Masters, C. L., Anders, R. F., Robinson, C. V., Cappai, R. & Carver, J. A. (2004). Interaction of the molecular chaperone alphaB-crystallin with alpha-synuclein: effects on amyloid fibril formation and chaperone activity. *J. Mol. Biol.* **340**, 1167-1183, doi:10.1016/j.jmb.2004.05.054

14. Zourlidou, A., Payne Smith, M. D. & Latchman, D. S. (2004). HSP27 but not HSP70 has a potent protective effect against alpha-synuclein-induced cell death in mammalian neuronal cells. *J. Neurochem.* **88**, 1439-1448.
15. MacRae, T. H. (2000). Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas. *Cell. Mol. Life. Sci.* **57**, 899-913.
16. Theriault, J. R., Lambert, H., Chavez-Zobel, A. T., Charest, G., Lavigne, P. & Landry, J. (2004). Essential role of the NH2-terminal WD/EPF motif in the phosphorylation-activated protective function of mammalian Hsp27. *J. Biol. Chem.* **279**, 23463-23471.
17. Basha E., Lee G.J., Demeler B. & Vierling E. (2004). Chaperone activity of cytosolic small heat shock proteins from wheat. *Eur J Biochem.* **271**, 1426-1436.
18. Lee, G. J., Roseman, A. M., Saibil, H. R. & Vierling, E. (1997). A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**, 659-671.
19. Horwitz, J. (1992). Alpha-crystallin can function as a molecular chaperone. *Proc. Natl. Acad. Sci. USA*, **89**, 10449-10453.
20. Ehrnsperger, M., Graber, S., Gaestel, M., Buchner, J. (1997). Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J.* **16**, 221-229.
21. Studer, S., Narberhaus, F. (2000). Chaperone activity and homo- and hetero-oligomer formation of bacterial small heat shock proteins. *J. Biol. Chem*. **275**, 37212-37218.
22. Kim, R., Lai, L., Lee, H. H., Cheong, G. W., Kim, K. K., Wu, Z., Yokota, H., Marqusee, S. & Kim S. H. (2003). On the mechanism of chaperone activity of the small heat-shock protein of Methanococcus jannaschii. *Proc. Natl. Acad. Sci. USA,* **100**, 8151-8155.
23. Haslbeck, M., Walke, S., Stromer, T., Ehrnsperger, M., White, H. E., Chen, S., Saibil, H. R. & Buchner, J. (1999). Hsp26: a temperature-regulated chaperone. *EMBO J.* **18**, 6744-6751.
24. Basha, E., Lee, G. J., Breci, L. A., Hausrath, A. C., Buan, N. R., Giese, K. C. & Vierling, E. (2004). The identity of proteins associated with a small heat shock protein during heat stress in vivo indicates that these chaperones protect a wide range of cellular functions. *J. Biol. Chem.* **279**, 7566-7575.
25. Mogk, A., Schlieker, C., Friedrich, K. L., Schonfeld, H. J., Vierling, E. & Bukau, B. (2003). Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. *J. Biol. Chem.* **278**, 31033-31042.
26. Haslbeck, M., Miess, A., Stromer, T., Walter, S., Buchner, J. (2005). Disassembling Protein Aggregates in the Yeast Cytosol: The cooperation of Hsp26 with Ssa1 and Hsp104. *J Biol Chem*. **280**, 23861-23868.
27. Cashikar, A.G., Duennwald, M., Lindquist, S.L (2005). A Chaperone Pathway in Protein Disaggregation: Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. *J Biol Chem*. **280**, 23869-23875.
28. Giese K.C. & Vierling E. J. (2004). Mutants in a small heat shock protein that affect the oligomeric state - Analysis and allele-specific suppression. *J. Biol. Chem.* **279**, 32674-32683.

29. Fu, X., Zhang, H., Zhang, X., Cao, Y., Jiao, W., Liu, C., Song, Y., Abulimiti, A., & Chang, Z. (2005). A dual role for the N-terminal region of Mycobacterium tuberculosis Hsp16.3 in self-oligomerization and binding denaturing substrate proteins. *J. Biol. Chem*. **280**, 6337-6348.
30. Lindner, R.A., Carver, J.A., Ehrnsperger, M., Buchner, J., Esposito, G., Behlke, J., Lutsch, G., Kotlyarov, A., Gaestel, M. (2000). Mouse Hsp25, a small shock protein. The role of its C-terminal extension in oligomerization and chaperone action. *Eur J Biochem.* **267**, 1923-1932.
31. Stromer, T., Fischer, E., Richter, K., Haslbeck, M., Buchner, J. (2004). Analysis of the regulation of the molecular chaperone Hsp26 by temperature-induced dissociation: the N-terminal domail is important for oligomer assembly and the binding of unfolding proteins. *J. Biol. Chem.* **279**, 11222-11228.
32. Bova, M.P., McHaourab, H.S., Han, Y., Fung, B.K. (2000). Subunit exchange of small heat shock proteins. Analysis of oligomer formation of alphaA-crystallin and Hsp27 by fluorescence resonance energy transfer and site-directed truncations. *J. Biol. Chem.* **275**, 1035-1042.
33. Kim, K. K., Kim, R. & Kim, S. H. (1998). Crystal structure of a small heat-shock protein. *Nature,* **394**, 595-599.
34. van Montfort, R. L., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* **8**, 1025-1030.
35. Sobott, F., Benesch, J .L., Vierling, E., Robinson & C. V. (2002). Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277**, 38921-38929.
36. Kappe, G., Aquilina, J. A., Wunderink, L., Kamps, B., Robinson, C. V., Garate, T., Boelens, W. C. & de Jong, W. W. (2004). Tsp36, a tapeworm small heat-shock protein with a duplicated alpha-crystallin domain, forms dimers and tetramers with good chaperone-like activity. *Proteins,* **57**, 109-117.
37. Benitez, L., Harrison, L. J., Parkhouse, R. M. & Garate, T. (1998). Sequence and preliminary characterisation of a Taenia saginata oncosphere gene homologue of the small heat-shock protein family. *Parasitol. Res.* **84**, 423-425.
38. Nene, V., Dunne, D. W., Johnson, K. S., Taylor, D. W. & Cordingley, J. S. (1986). Sequence and expression of a major egg antigen from Schistosoma mansoni. Homologies to heat shock proteins and alpha-crystallins. *Mol. Biochem. Parasitol.* **21**, 179-188.
39. Merckelbach, A., Wager, M. & Lucius, R. (2003). Analysis of cDNAs coding for immunologically dominant antigens from an oncosphere-specific cDNA library of Echinococcus multilocularis. *Parasitol. Res.* **90**, 493-501.
40. Chen, Y. & Boros, D. L. (1998). Identification of the immunodominant T cell epitope of p38, a major egg antigen, and characterization of the epitope-specific Th responsiveness during murine schistosomiasis mansoni. *J. Immunol.* **160**, 5420-5427.
41. Heger, A. & Holm, L. (2000). Rapid automatic detection and alignment of repeats in protein sequences. *Proteins,* **41**, 224-237.
42. Valdez, M. M., Clark, J. I., Wu, G. J. & Muchowski, P. J. (2002). Functional similarities between the small heat shock proteins Mycobacterium tuberculosis HSP 16.3 and human alphaB-crystallin. *Eur. J. Biochem.* **269**, 1806-1813.

43. McManus, D. P., Zhang, W., Li, J. & Bartley, P. B. (2003). Echinococcosis. *Lancet,* **362**, 1295-1304.
44. Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Meth. Enzymol.* **276,** 523-530.
45. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
46. Evans, P. R. (1997). Scala. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography,* **33**, 22-24.
47. French, G. S. & Wilson, K. S. (1978). On the treatment of negative intensity observations. *Acta Cryst.* **A34**, 517-525.
48. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Cryst.* **D58**, 1772-1779.
49. De La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood Heavy Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods Enzymol.* **276**, 472-494.
50. McRee, D. E. (1999). XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. *J. Structural Biology,* **125**, 156-165.
51. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. -S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Cryst*. **D54**, 905-921.
52. Laskowski, R.A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*. **26**, 283-291.
53 Orengo C.A. (1999). CORA - topological fingerprints for protein structural families. *Protein Sci.* **8**, 699-715.
54 Eddy S.R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755-763.

Table captions

Table 1. Crystallographic statistics

Table 2. Structural and sequence identity comparisons of the α-crystallin domains of Tsp36, WHsp16.9 and MJHsp16.5. Upper right values are RMSDs (Å), and lower left values are sequence identities (%). Divergent loop regions (between strands 2 and 3, and strands 5 and 7) are excluded.

Figure legends

Figure 1. Domain boundaries and alignments. (a) Domain boundaries for WHsp16.9, MJHsp16.5, and repeat units and regions of Tsp36. R1N - Repeat1 N-terminus, ACD1 - Repeat1 α-crystallin domain, L - Linker, R2N - Repeat2 N-terminus, ACD2 - Repeat2 α-crystallin domain. (b) Structural alignment of the ACD and C-terminal extension of metazoan (red) and non-metazoan (green) sHsps. Both sHsp domains of Tsp36, *Echinococcus multilocularis* and *Schistosoma mansoni* are aligned with the following single domain sHsps: human αA crystallin, human αB crystallin, human Hsp27, *Ciona intestinalis* (Q95P25), *Caenorhabditis elegans* (HSP11_CAEEL), *Drosophila melanogaster* (HSP23_DROME), *Arabidopsis thaliana* (HSP12_ARATH), *Pisum sativum* (HSP11_PEA), WHsp16.9, *Saccharomyces cerevisiae* (HSP26_YEAST), *Mycobacterium tuberculosis* (14KD_MYCTU), *Synechocystis* (P72977) and MJHsp16.5. Conserved and conservatively substituted residues are indicated in grey, except for the conserved disease-related arginine (blue) and C-terminal motif (purple). Residues that differ for non-metazoans (cyan) and metazoans (yellow) are indicated. The RGD domain of *E. multilocularis* is shown in magenta. The secondary structure of Tsp36 ACD1 is indicated as black arrows, whilst the C-terminal extension in WHsp16.9 and MJHsp16.5 contains an extra β-strand 10 comprised of the conserved C-terminal residues. (c) The full sequence of Tsp36, with domain boundaries coloured as indicated in (a). The secondary structure for the N-terminus is indicated by cylinders, and residues involved in patching the edges of the ACDs are indicated in red.

Figure 2. (a) An ACD from WHsp16.9, showing arrangement of the β-strands 2-9 and the N- and C-terminal edges of the β-sandwich. β-strands are numbered according to scheme derived from MJHsp16.5 (Kim et al., 1998). (b) The dimer of WHsp16.9 (yellow and green), indicating the strand exchange between domains of β-strand 6 (blue) from the loop between β-strands 5 and 7 which hydrogen bonds with β-strand 2 of the dimer partner.

Figure 3. The ACD topology of the α-crystallin domains of Tsp36 (a), WHsp16.9 (b) and MJHsp16.5 (c). For Tsp36, no β-strand 6 is present in the B5-B7 loop. β-strands L1-L2 from the linker region L are also indicated. The colour scheme is as for Figure 1a. Both WHsp16.9 and MJHsp16.5 show an additional β-strand 10 containing the conserved I/V-X-I/V motif in the C-terminus, whilst MJHsp16.5 shows an additional β-strand 1, part of the variable N-terminus, that packs alongside the ACD at the N-terminal edge.

Figure 4**.** The structure of dimeric Tsp36. Molecule A (coloured as for Figure 1a) forms a dimer with molecule B (yellow). In molecule A the single cysteine in the B5-B7 loop is indicated in CPK colours ball and stick format, the conserved disease-related arginine (blue) in stick format (Arg158). The black oval indicates interface between ACD2 domains with missing density, the blue oval indicates missing density in molecule B.

Figure 5. Differences in organization of ACDs of WHsp16.9 (a) and Tsp36 (b). The orientation of one monomer of WHsp16.9 (green) is shown the same as for ACD1 of Tsp36 (magenta). The relationship of the second monomer of WHsp16.9 (yellow) is different from that of Tsp36 ACD2 (cyan). (c) Differences between WHsp16.9 and Tsp36 B5-B7 loop. Superposition of Tsp36 ACD1 (magenta) on a dimer of WHsp16.9 (yellow and green). The divergence of the B5-B7 loops of Tsp36 and WHsp16.9 is highlighted. The conserved Pro62 from each dimer partner of WHsp16.9 is indicated in spacefill format.

Figure 6. The hydrophobic N-terminal edge of ACD1 (β-strand 2 coloured red, β-strand 3 coloured blue, β-strand 7 coloured yellow) patched by Phe99, and hydrophobic residues in the N-terminal region 57-63. The interaction between Ser59 (green, main chain and side chain oxygens in red) and Arg158 (stick format, nitrogens of guanidinium group in blue) is indicated.

Figure 7. Patching of C-terminal β-sandwich edge. (a) In WHsp16.9, the C-terminal extension of a partner molecule (cyan) patches the edge comprised of β-strands 4 and 8. Two conserved isoleucine residues (ball and stick format) insert into the hydrophobic groove. (b) In Tsp36, the N-terminus of the partner molecule (cyan) patches the equivalent site in ACD1. A proline and isoleucine (ball and stick format) insert into the hydrophobic groove.

Figure 8. Wrapping of ACDs (purple, cartoon format) of molecule A by R1N, linker and R2N neighbouring regions (coloured as in Figure 1a with surface representation) which are in turn enclosed by the N-terminus (cyan, cartoon format). The inset shows the ACDs alone, with a surface coloured red for hydrophobic residues to indicate the hydrophobic patches that can act as anchoring points for substrate.

Figure 9. A model of the tetramer of Tsp36 shown in spacefill. The missing B5-B7 loops were modeled based on the loop seen for ACD1 in molecule A. The dimer-dimer interaction was created by modeling a pair of disulphide bonds and assuming a 2-fold axis between dimers.

Table 1.

	Data Collection (ESRF 14.4)					
	Peak		**Inflection**		**Remote**	
Wavelength (Å)	0.97927		0.97942		0.9686	
	Overall	Outer shell	Overall	Outer shell	Overall	Outer shell
Resolution(Å)	2.5	2.64 - 2.5	2.5	2.64 - 2.5	2.5	2.64 - 2.5
R_{merge}	0.061	0.18	0.06	0.159	0.059	0.22
I/ó (I)	16.9	4.5	18.3	5.2	16	3.4
Completeness (%)	96.9	81.4	96.6	79.5	97.2	82.1
Multiplicity	5.0	3.6	5.2	3.8	5.4	4.0

	Phasing Power		
Acentric (Iso/Ano)	0 / 1.509	0.533 / 1.121	0.748 / 1.087
Centric (Iso)	0	0.446	0.66

	Refinement
R/R_{free}	20.7 / 26.4
RMS deviations bond lengths (Å)	0.008
RMS deviations bond angles (°)	1.3

Table 2.

	Tsp36ACD1	Tsp36ACD2	WHsp16.9	MJHsp16.5
Tsp36 ACD1		1.9	1.9	2.2
Tsp36 ACD2	24		1.5	1.6
WHsp 16.9	19	21		1.5
MJHsp 16.5	29	18	34	

Fig1a

Click here to download high resolution image

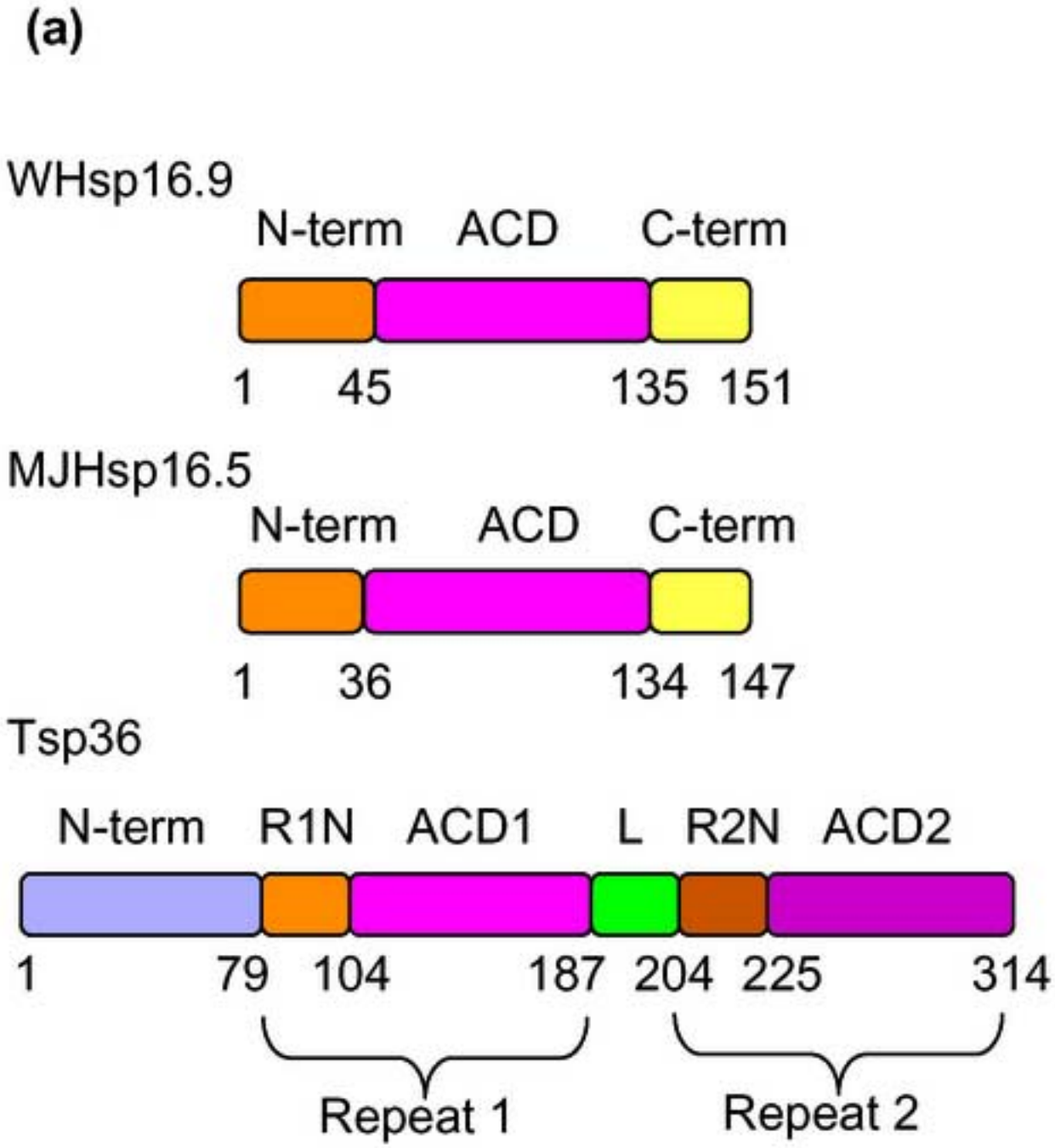

Fig1b

Click here to download high resolution image

(b)

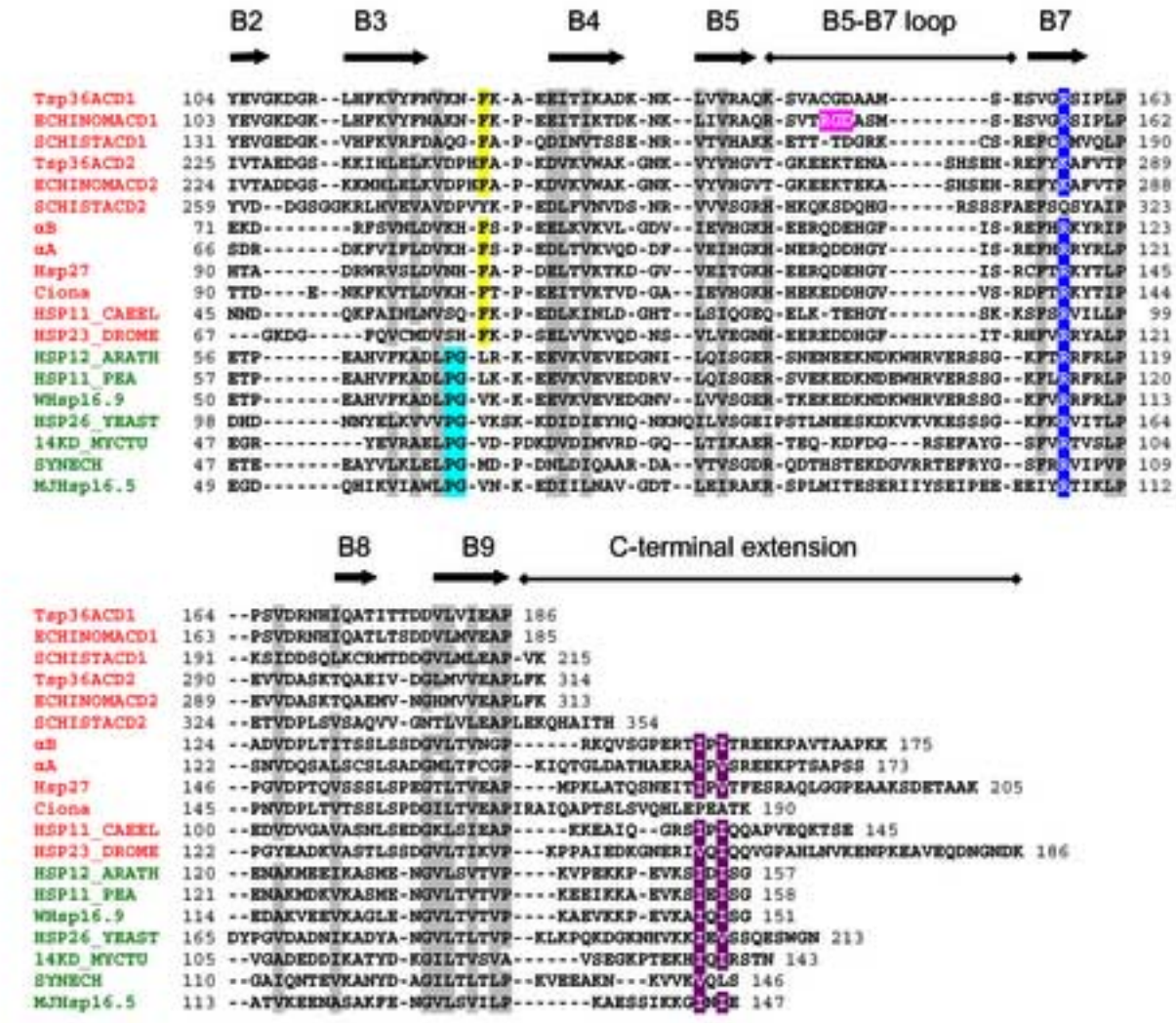

Fig1c

Click here to download high resolution image

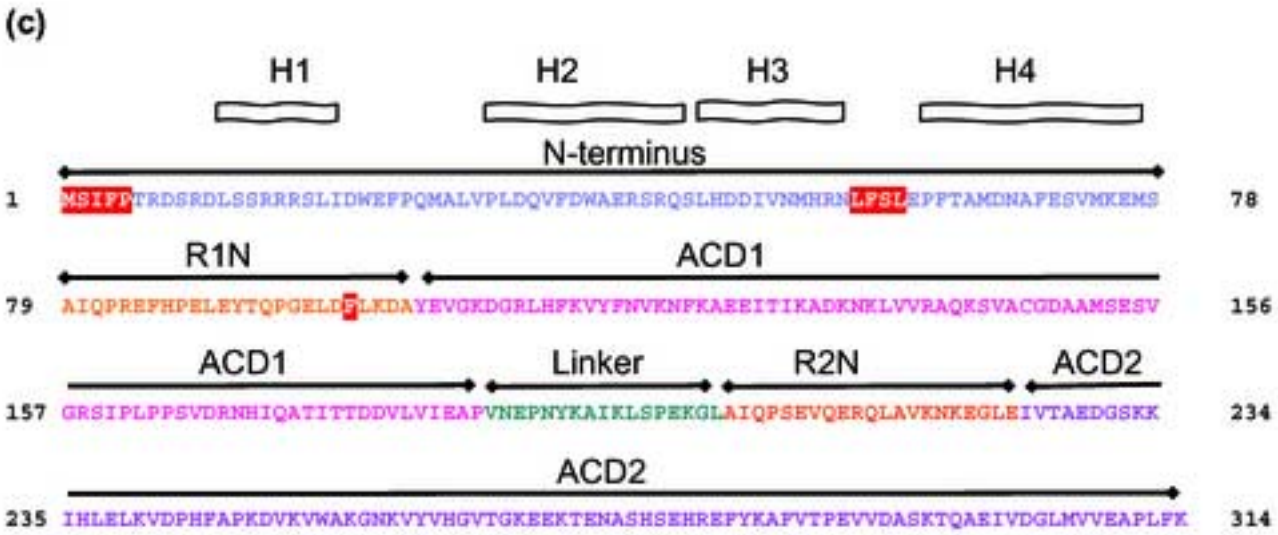

Fig2a

Click here to download high resolution image

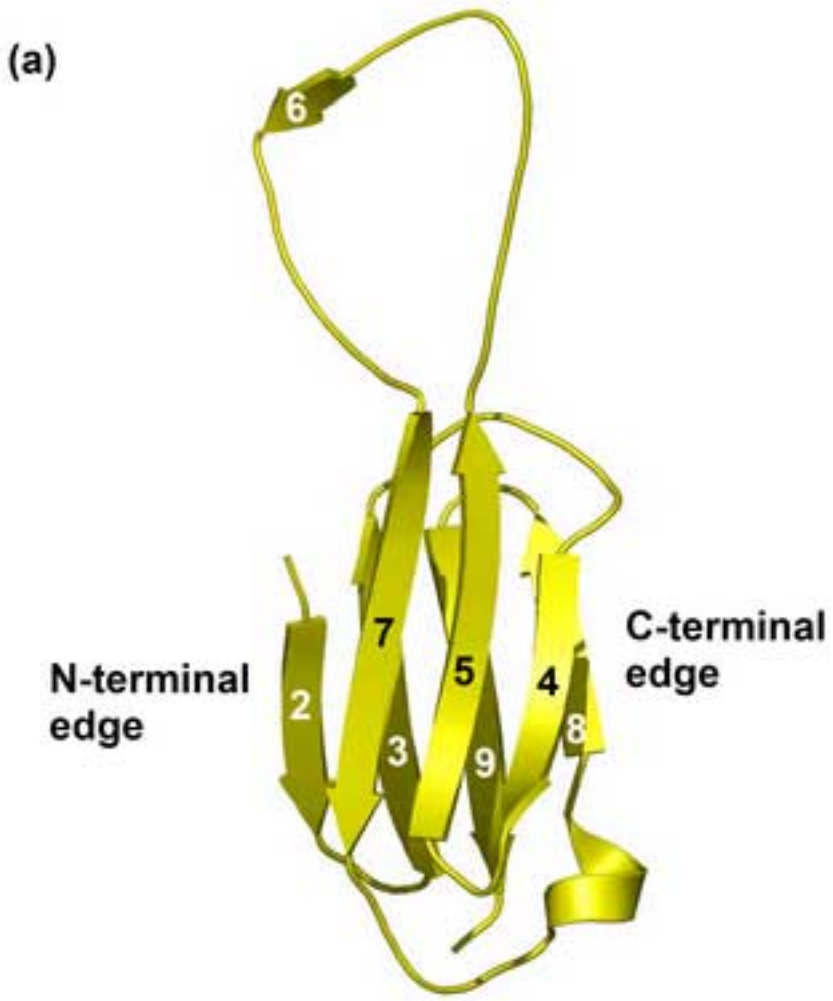

Fig2b

Click here to download high resolution image

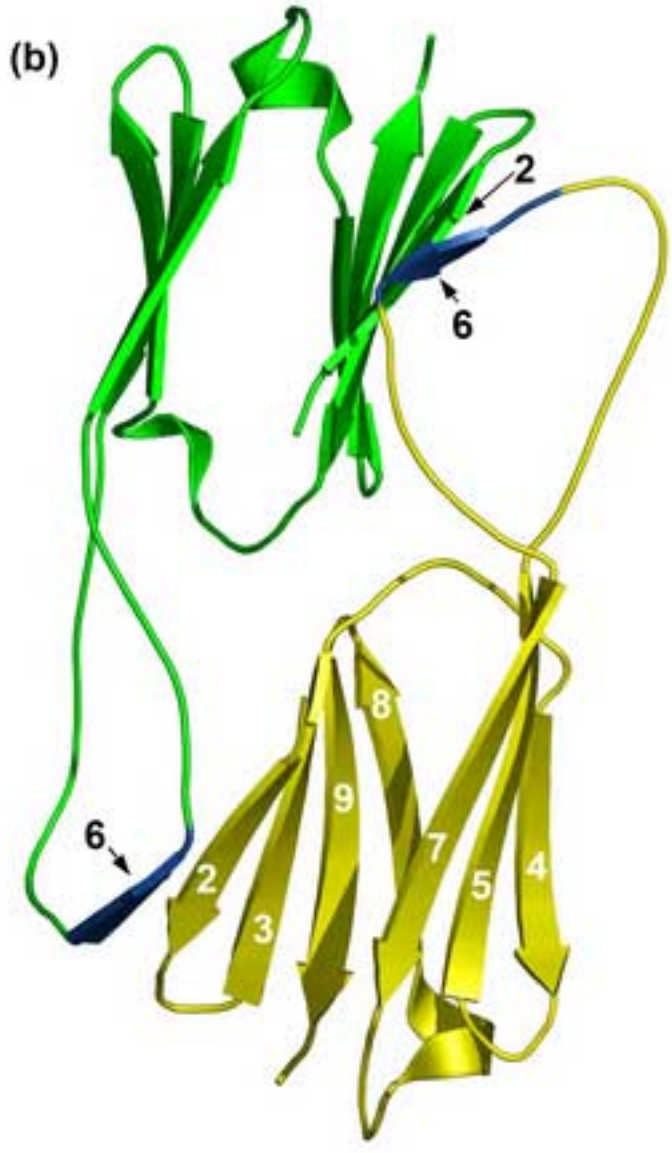

Fig3

Click here to download high resolution image

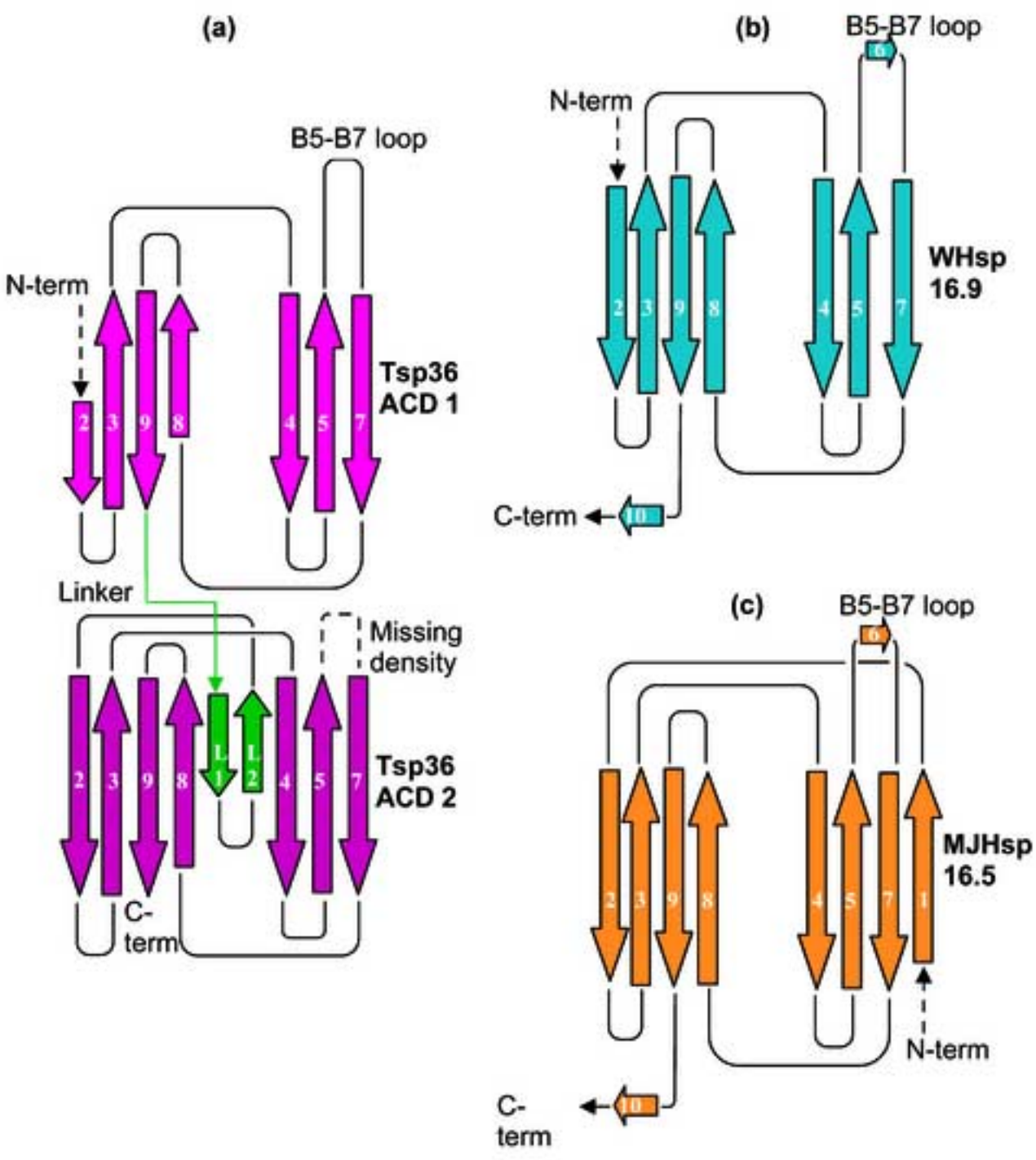

Fig4

Click here to download high resolution image

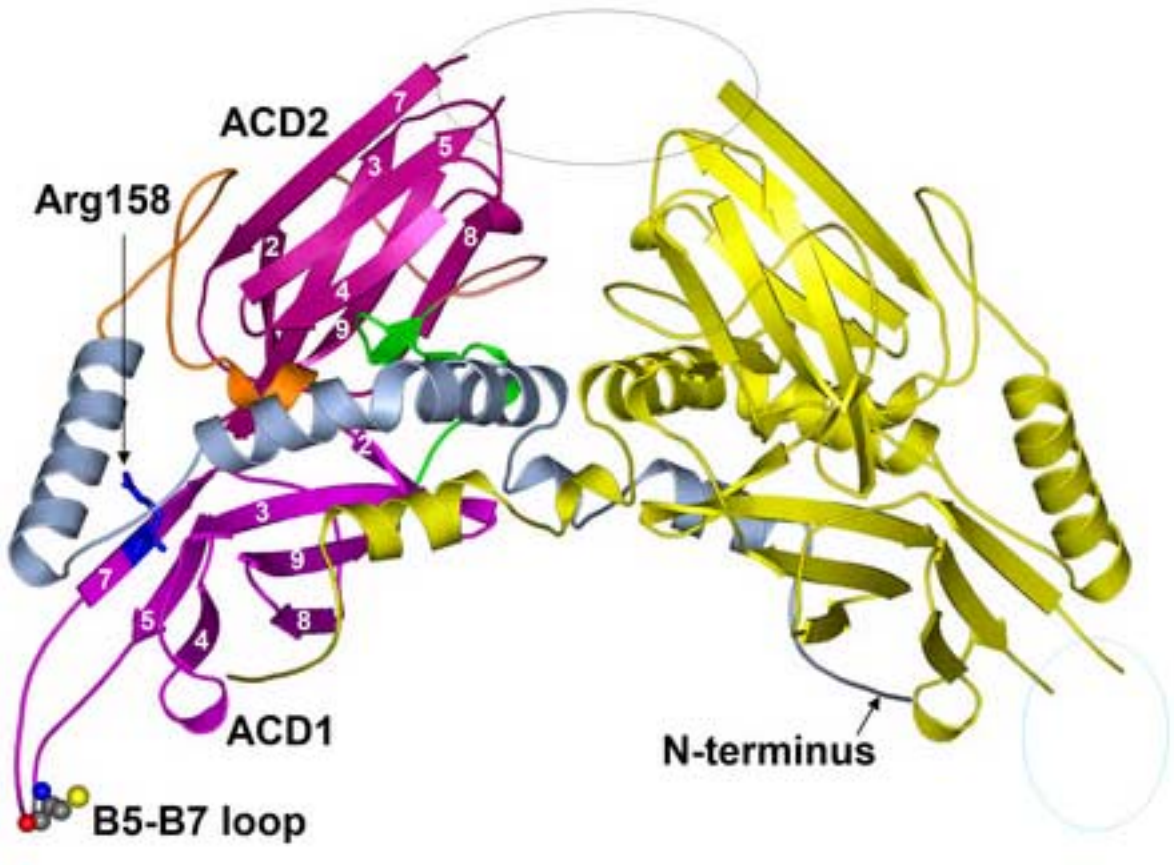

Fig5a-b

Click here to download high resolution image

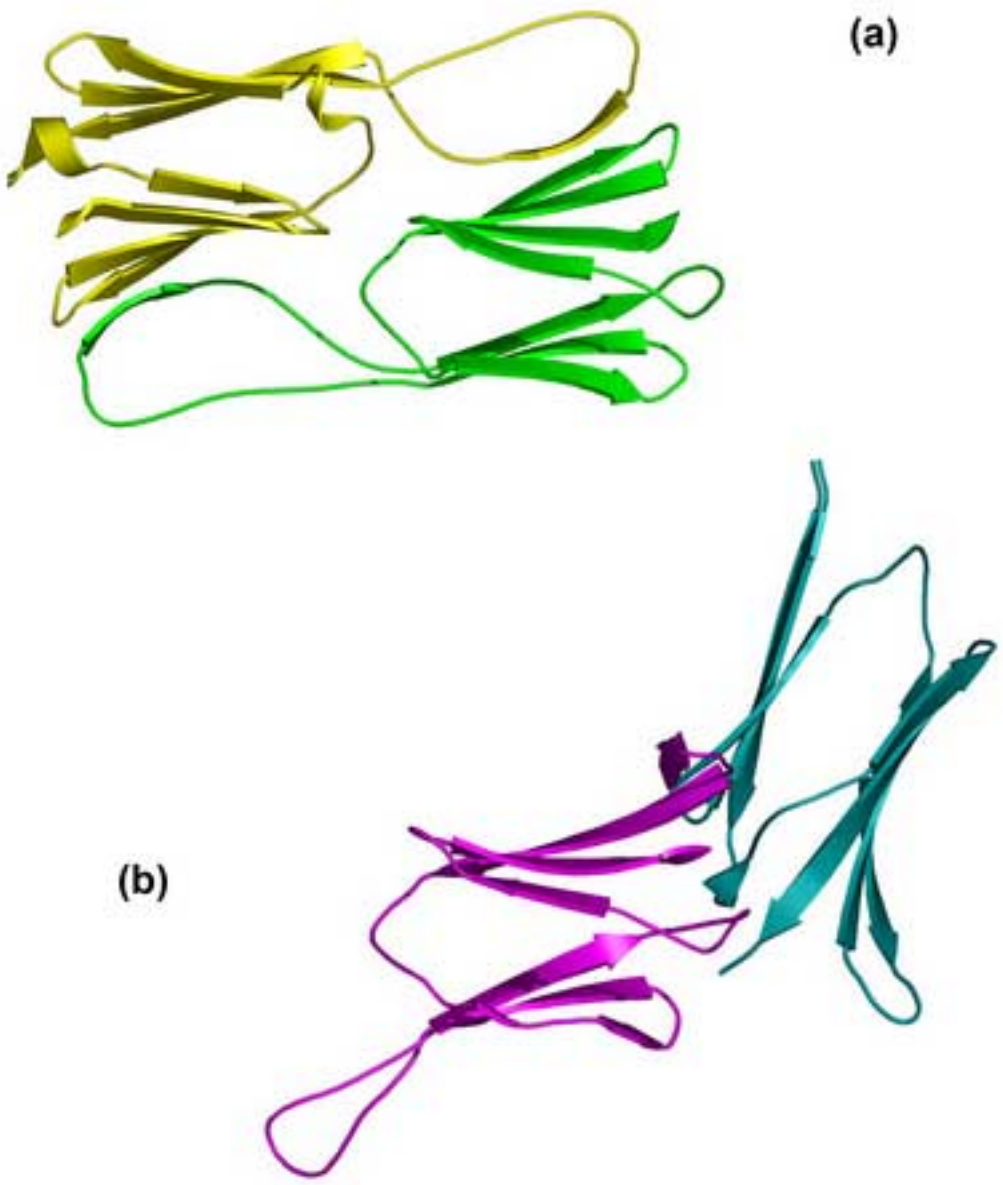

Fig5c
Click here to download high resolution image

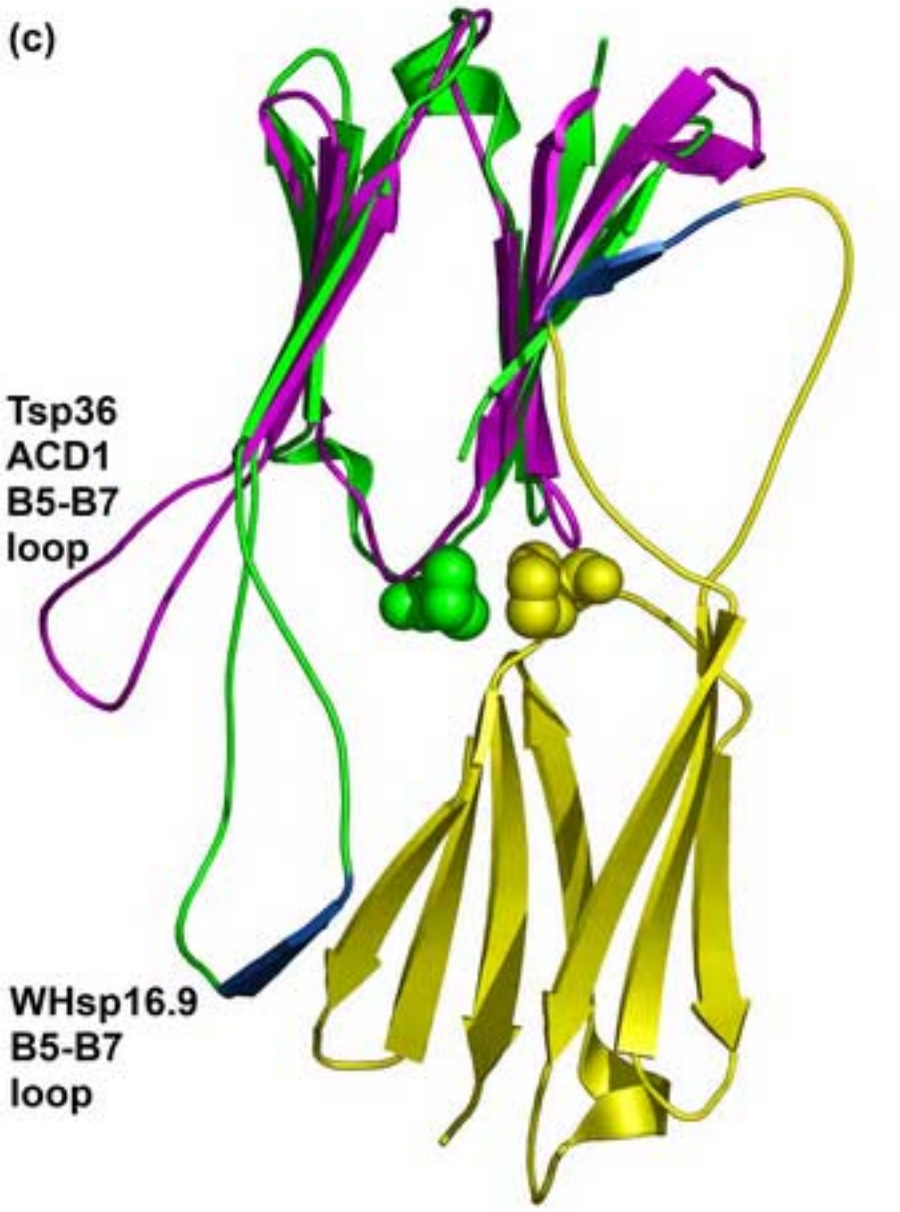

Fig6

Click here to download high resolution image

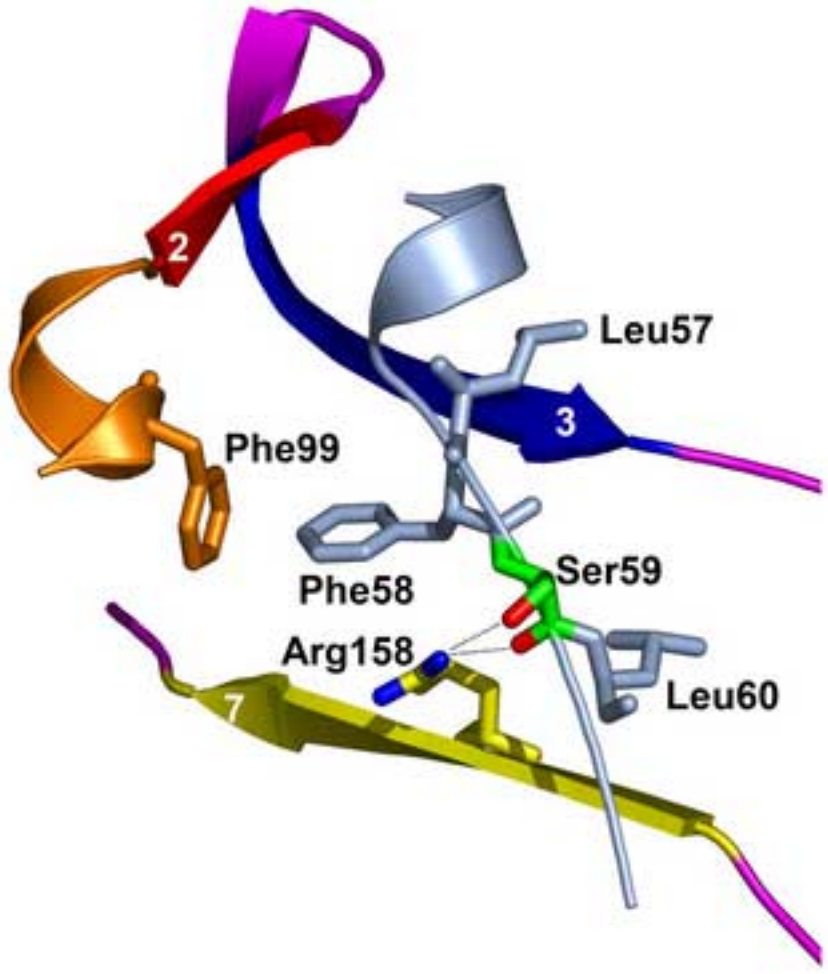

Fig7a
Click here to download high resolution image

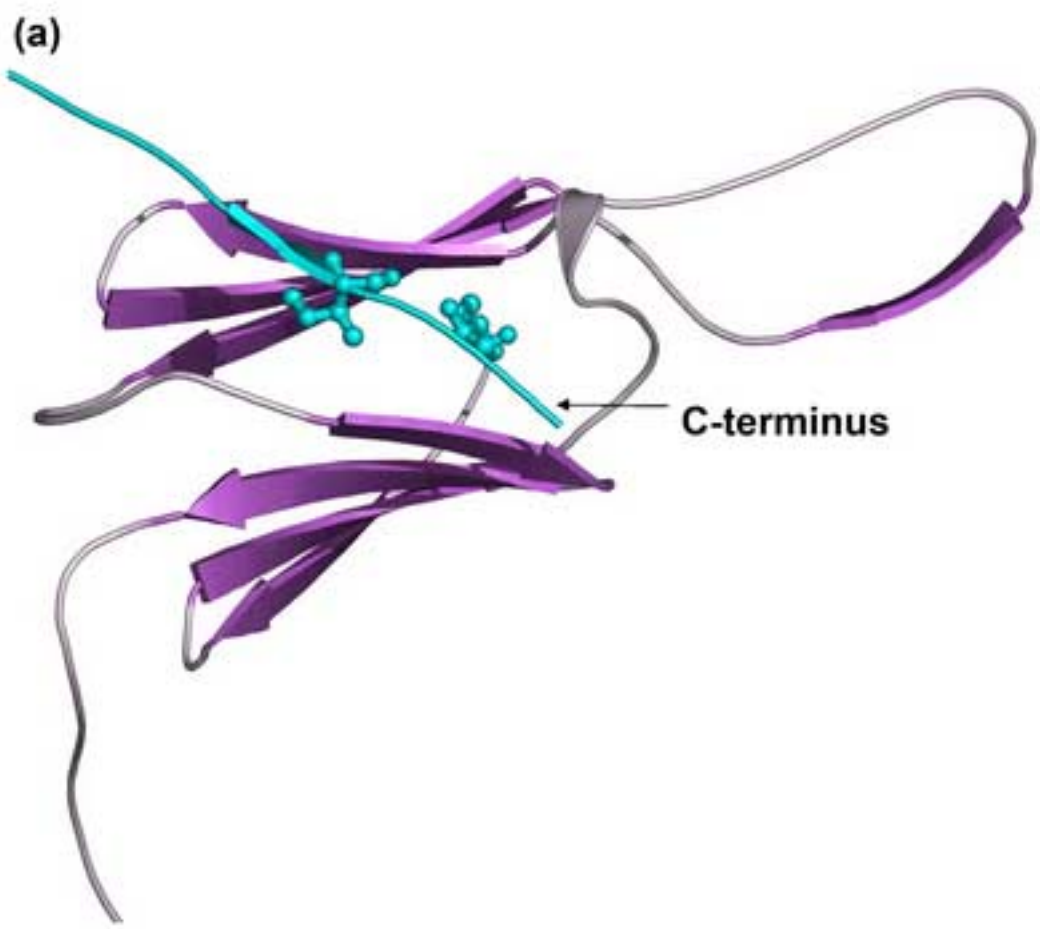

Fig7b

Click here to download high resolution image

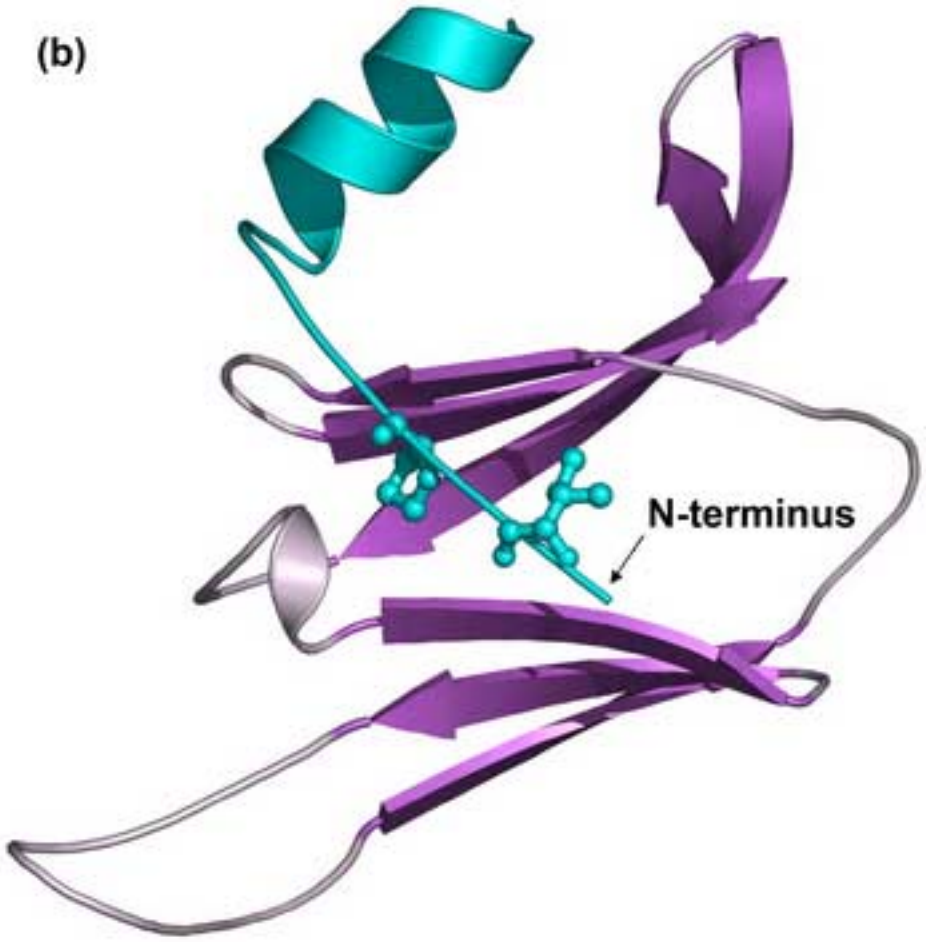

Fig8

Click here to download high resolution image

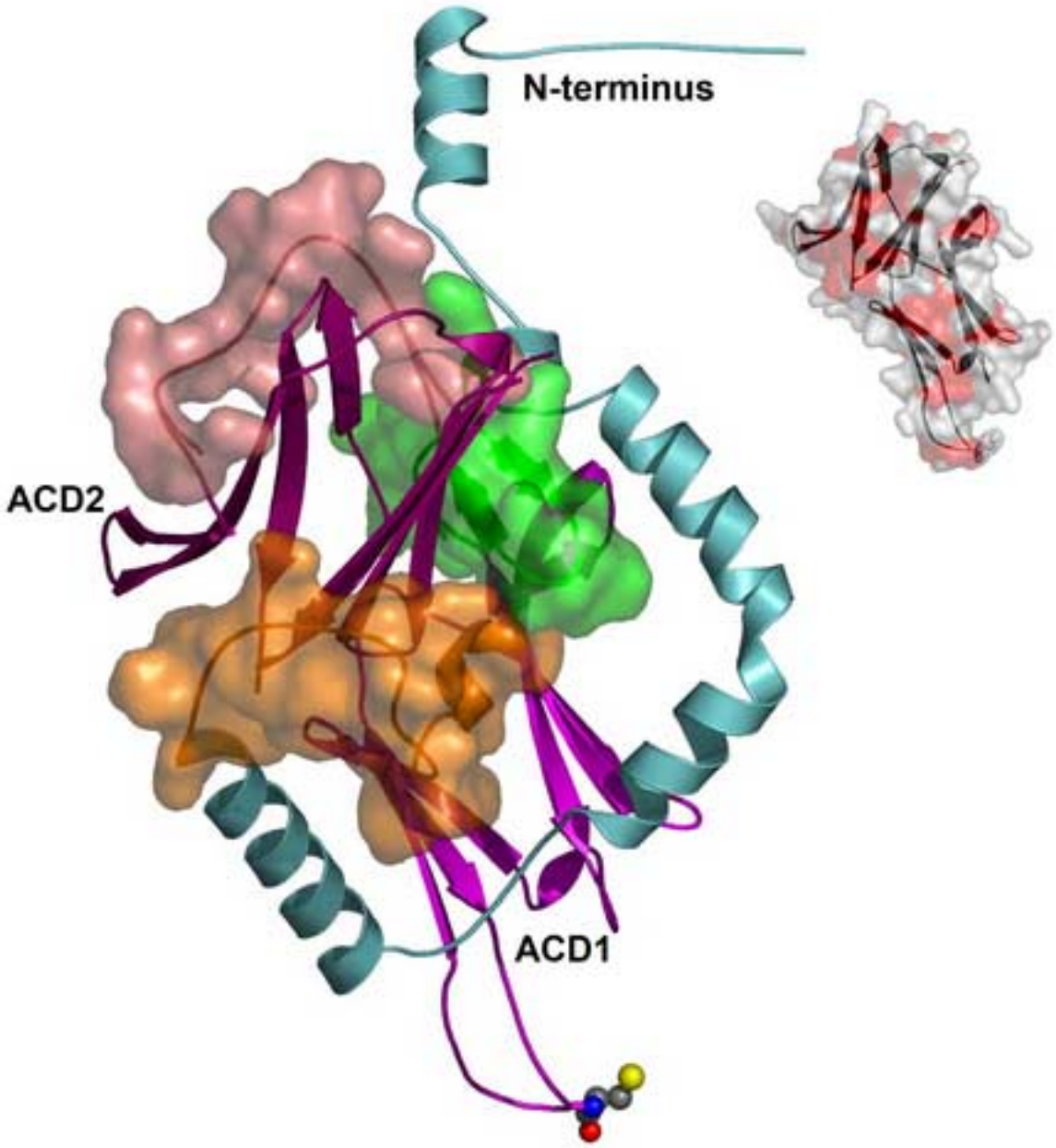

Fig9

Click here to download high resolution image