

# Homology of SMP domains to the TULIP superfamily of lipid-binding proteins provides a structural basis for lipid exchange between ER and mitochondria

Klaus O. Kopec, Vikram Alva and Andrei N. Lupas\*

Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstr. 35,  
72076 Tübingen, Germany

Associate Editor: Alfonso Valencia

## ABSTRACT

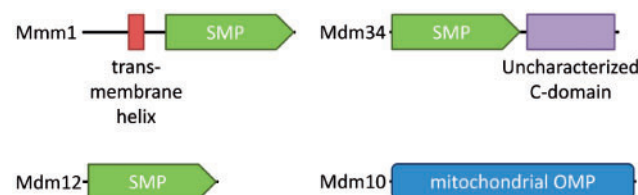
Mitochondria must uptake some phospholipids from the endoplasmic reticulum (ER) for the biogenesis of their membranes. They convert one of these lipids, phosphatidylserine, to phosphatidylethanolamine, which can be re-exported via the ER to all other cellular membranes. The mechanisms underlying these exchanges between ER and mitochondria are poorly understood. Recently, a complex termed ER–mitochondria encounter structure (ERMES) was shown to be necessary for phospholipid exchange in budding yeast. However, it is unclear whether this complex is merely an inter-organelle tether or also the transporter. ERMES consists of four proteins: Mdm10, Mdm34 (Mmm2), Mdm12 and Mmm1, three of which contain the uncharacterized SMP domain common to a number of eukaryotic membrane-associated proteins. Here, we show that the SMP domain belongs to the TULIP superfamily of lipid/hydrophobic ligand-binding domains comprising members of known structure. This relationship suggests that the SMP domains of the ERMES complex mediate lipid exchange between ER and mitochondria.

**Contact:** andrei.lupas@tuebingen.mpg.de**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on February 15, 2010; revised on June 5, 2010; accepted on June 7, 2010

## 1 INTRODUCTION

Mitochondria are organelles of endosymbiotic origin, found in virtually all eukaryotic organisms. They are the main generators of adenosine triphosphate (ATP)—the energy currency of the cell. Additionally, they participate in a series of other important processes, including apoptosis, amino acid and lipid metabolism, iron–sulphur cluster assembly and the regulation of calcium levels within the cell (Lill and Kispal, 2000; McBride *et al.*, 2006). However, only a small fraction of the biopolymers required to carry out these functions is synthesized in the mitochondria, the rest must be imported from the outside. For example, they only produce some of the phospholipids that make up their membranes, whereas the remainder originates from the endoplasmic reticulum (ER). Interestingly, mitochondria not only import phospholipids but also export a particular one, phosphatidylethanolamine (PtdEtn), to the ER, where



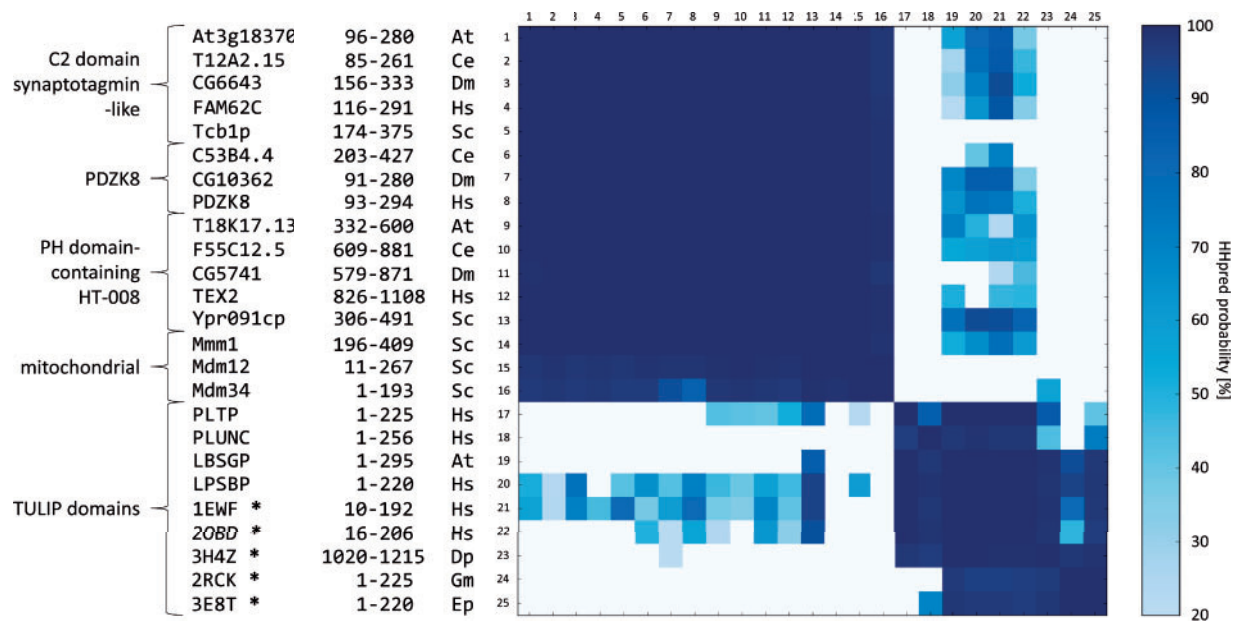
**Fig. 1.** Domain organization of the four ERMES proteins. All ERMES proteins, except the mitochondrial outer membrane protein (OMP) Mdm10, contain an SMP domain. The SMP domain in Mdm34 was discovered in this study.

it is methylated to form the phospholipid phosphatidylcholine (Voelker, 2003). Mitochondria synthesize PtdEtn by decarboxylating phosphatidylserine, a phospholipid imported from the ER. The mechanisms responsible for the influx and efflux of phospholipids are unclear.

Unlike most organelles, mitochondria do not exchange phospholipids via vesicular transport. Previous studies have suggested that this exchange takes place via ER–mitochondria associations (Achleitner *et al.*, 1999; Voelker, 2003). More recently, Kornmann *et al.* (2009) identified a complex, the ER–mitochondria encounter structure (ERMES), that acts as a molecular tether between ER and mitochondria in *Saccharomyces cerevisiae* and is required for efficient inter-organelle phospholipid exchange. However, it remained unclear whether ERMES merely tethers these organelles together, thereby aligning proteins that carry out the actual transport, or also recruits and transfers phospholipids itself.

ERMES comprises four proteins (Fig. 1): the mitochondrial outer membrane protein Mdm10, the putative outer membrane protein Mdm34 (Mmm2), the ER-resident Mmm1 and the cytosolic Mdm12 (Kornmann *et al.*, 2009). These proteins have also been implicated in other mitochondrial functions, including morphology maintenance (Okamoto and Shaw, 2005) and protein import (Meisinger *et al.*, 2004). Two of the ERMES components, Mmm1 and Mdm12, were reported to contain the uncharacterized SMP domain (synaptotagmin-like, mitochondrial and lipid-binding proteins), which is also present in a number of other eukaryotic membrane-associated proteins (Lee and Hong, 2006). SMP domain-containing proteins have been classified into four broad groups: C2 domain synaptotagmin-like, PH domain-containing HT-008, PDZK8 and mitochondrial protein families (Lee and Hong, 2006). The functions of these proteins are poorly understood.

\*To whom correspondence should be addressed.



**Fig. 2.** Pairwise HMM comparison of SMP and TULIP domains. Representatives of the four SMP domain-containing groups and of TULIP domains were chosen from *Arabidopsis thaliana* (At), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Dermatophagoides pteronyssinus* (Dp), *Epiphyas postvittana* (Ep), *Galleria mellonella* (Gm), *Homo sapiens* (Hs) and *Saccharomyces cerevisiae* (Sc). Group and protein names of these representatives, along with domain boundaries and names of source species are indicated (from left to right). HHpred was used to perform pairwise HMM comparisons between them. Cell color indicates the HHpred probability of the match as depicted in the scale on the right; probabilities <20% are shown as white cells. Proteins with known structures are marked with an asterisk. The abbreviations of the hydrophobic ligand binders are explained in the text.

In this study, we show that the SMP domain belongs to a superfamily of lipid/hydrophobic ligand-binding domains of known structure, which we call TULIP for tubular lipid-binding proteins, and propose a role for it in cellular phospholipid traffic.

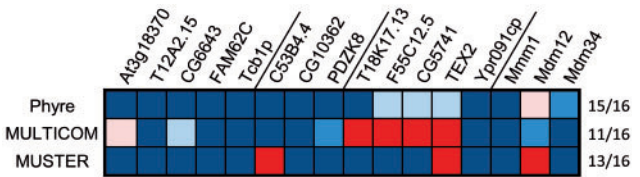
2 METHODS

All sequence similarity searches were carried out in the MPI bioinformatics toolkit (<http://toolkit.tuebingen.mpg.de>; Biegert *et al.*, 2006) using HHpred (Söding *et al.*, 2005) and HHSenser (Söding *et al.*, 2006) with default settings. HHpred searches were performed against a database comprising PDB70 (protein databank structures, as available on the 15th of April 2010, clustered at 70% sequence identity) and genomes of phylogenetically diverse organisms (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens* and *S.cerevisiae*). Representatives of the four SMP domain-containing groups from the aforementioned organisms were chosen as seeds for the searches in Figures 2 and 3, based on their presence in the core of their respective clusters in the sequence cluster map (Fig. 5).

To identify sequences for cluster analysis, we searched the non-redundant protein sequence database (nr) at NCBI for homologs of the SMP domain from the yeast protein Mmm1 (residues 196–409), the N-terminal domain of human cholesteryl ester transfer protein (2OBD, residues 16–206), the Takeout 1 protein from *Epiphyas postvittana* (3E8T), and the dust mite allergen Der p 7 (3H4Z) using HHSenser. We pooled the permissive sets returned by HHSenser to obtain 2033 sequences, which we clustered by their pairwise BLAST *P*-values (Altschul *et al.*, 1990) in CLANS (Frickey and Lupas, 2004). Clustering was done to equilibrium in 2D at a *P*-value cutoff of *e*-4 using default settings.

3 RESULTS

The number of structural solutions available to a polypeptide chain is limited, making protein structures multiply convergent (Cheng



**Fig. 3.** Fold predictions for SMP domains. The highest scoring PDB matches for the 16 representative SMP sequences in Figure 2 were collected with three of the top-scoring prediction servers in CASP8. Top matches to TULIP domains are shown in blue and to any other structure in red. The color saturation is scaled linearly between the maximum and minimum scores returned by the respective method. The value ranges corresponding to pale (low confidence), medium and dark (high confidence) saturation are: Phyre-estimated precision 0–33, 34–66 and 67–100, MULTICOM *e*-value 7.4–5, 5.1–2.6, 2.5–0, MUSTER Z-score 0–1.8, 1.9–3.5, 3.6–5.3. The number of matches to TULIP domains against the total is shown in the right-hand column.

*et al.*, 2008; Krishna and Grishin, 2004; Salem *et al.*, 1999), while the combinatorial possibilities in sequence space are nearly endless. For this reason, sequence similarity is considered the primary marker of homology. We thus used sensitive sequence comparisons, as implemented in HHpred, to find homologs of the SMP domain in a database concatenating several complete genomes with PDB70 (see ‘Methods’ section). The search was seeded with the SMP domain from Mmm1. The best hits were to other proteins that have previously been described to contain this domain (Fig. 2). In addition, we detected a hitherto unknown SMP domain in the ERMES protein Mdm34. This protein has been reported as an

integral outer membrane protein (Youngman *et al.*, 2004), but we were unable to identify any sequence motifs in it that would indicate membrane insertion. HHpred searches with other representatives and reciprocal searches with Mdm34 confirmed the presence of an SMP domain, raising the number of SMP domains in ERMES to three (Fig. 1).

We also found statistically significant matches to many eukaryotic proteins from the bactericidal/permeability-increasing protein-like (BPI-like) family (Fig. 2), including two with known structures: BPI (1EWF) and cholesteryl ester transfer protein (CETP; 2OBD). Other members of this family are lipopolysaccharide-binding protein (LPSBP), lipid-binding serum glycoprotein (LBSGP), phospholipid transfer protein (PLTP) and long and short paralogs of palate, lung and nasal epithelium carcinoma-associated protein (PLUNC). Some of these proteins have been shown to bind lipids, e.g. CETP facilitates lipid transport between different lipoproteins (Qiu *et al.*, 2007).

BPI and CETP have similar structures, each containing two tandem domains that adopt the same fold, comprising a long  $\alpha$ -helix wrapped in a highly curved anti-parallel  $\beta$ -sheet. All BPI-like proteins contain these two domains, the only exception being short PLUNC, which has only one. The domains show little sequence identity (<15%) and sequence comparisons do not yield significant matches between them. Instead, the C-terminal domain only shows matches to the Aha1 protein, a co-chaperone of Hsp90 in eukaryotes which shares the same fold (1USU; d.83.2). Nevertheless, the N- and C-terminal domains of BPI-like proteins are thought to have a common ancestry based on their structural similarity (Kleiger *et al.*, 2000). The Structural Classification of Proteins database (SCOP; Murzin *et al.*, 1995) also considers them to be homologous and classifies them into the same family (d.83.1.1).

HHpred searches with SMP domains yielded many statistically significant matches to the N-terminal domain of BPI-like proteins (Fig. 2), but not to the C-terminal domain. We confirmed these findings with reciprocal searches using both domains of BPI-like proteins. From the statistical significance of these matches, we conclude that SMP domains are homologous to BPI-like proteins and therefore predict that they share the same tubular fold and lipid-binding properties.

Further searches with the N-terminal domain of BPI-like proteins retrieved three more proteins of known structure: dust mite allergen Der p 7 (3H4Z), a juvenile hormone-binding protein from *Galleria mellonella* (JHBP, 2RCK), and a Takeout 1 protein from *Epiphyas postvittana* (3E8T). These proteins are exclusively found in arthropods and are involved in binding hydrophobic ligands. They are composed of a single domain homologous to the N-terminal domain of BPI-like proteins (Supplementary Fig. S1), a relationship that has been described previously (Hamiaux *et al.*, 2009; Kolodziejczyk *et al.*, 2008; Mueller *et al.*, 2010). In view of their similarities in sequence and structure, we propose to group the arthropod proteins together with the BPI-like family into the TULIP superfamily.

To confirm the membership of SMP domains in the TULIP superfamily, we generated fold predictions for 16 representative sequences using the servers Phyre (Kelley and Sternberg, 2009), MULTICOM (Wang *et al.*, 2010) and MUSTER (Wu and Zhang, 2008), all of which performed very well in the most recent Critical Assessment of Structure Prediction Experiment, CASP8 (Kryshtafovych *et al.*, 2009). All three methods yielded

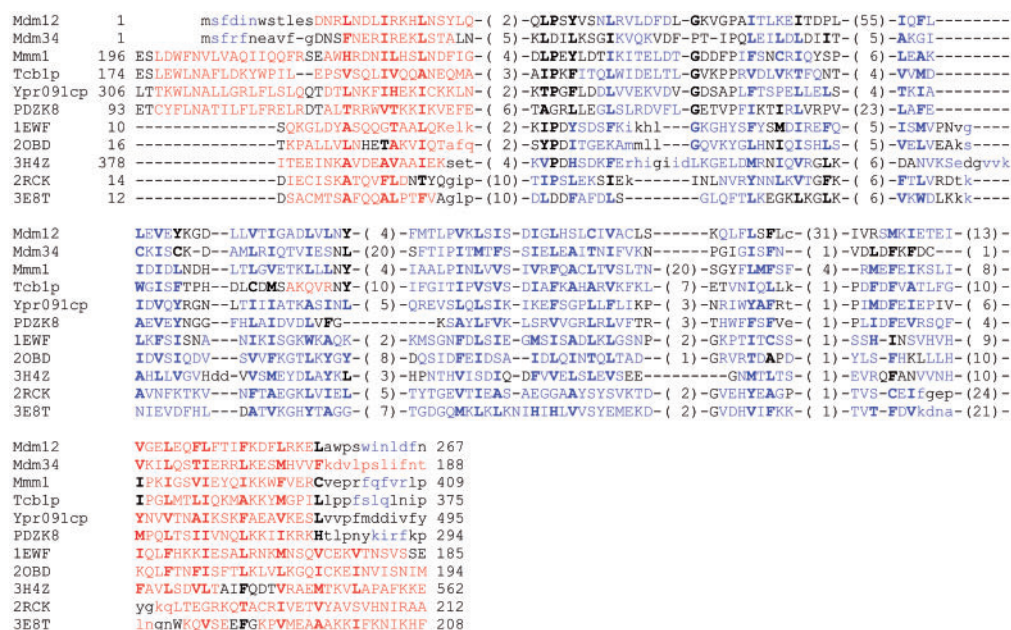
many highest-scoring matches to TULIP domains (Fig. 3 and Supplementary Fig. S2). Additionally, we queried the fold prediction metasever I-TASSER (Roy *et al.*, 2010), which was the top performing server in CASP 7 & 8 (Zhang, 2007, 2009). This server returned a TULIP domain as the top match for 10 of 16 queries and as one of the top three matches for all but one query (Supplementary Fig. S3). These matches included both BPI-like and Takeout-like proteins. A structure-assisted multiple sequence alignment of SMP domains to TULIP domains of known structure highlights the basis for these matches (Fig. 4). All sequences have similar length, distribution of (predicted) secondary structure, and pattern of hydrophobic residues. However, there are no conserved sequence motifs, unsurprisingly as such motifs are not even detectable within individual families (Beamer *et al.*, 1997; Kolodziejczyk *et al.*, 2008).

To explore the relative positions in sequence space of proteins of the TULIP superfamily, we searched for homologs of SMP domains, N-terminal domains of BPI-like proteins, as well as allergens and Takeout proteins in the nr database using HHsenser, and clustered the obtained sequences in CLANS (see 'Methods' section). The resulting cluster map (Fig. 5) shows three distinct but connected regions corresponding to SMP, BPI and Takeout-like domain families, confirming the proposed homology between them. In addition to the SMP groups described by Lee and Hong (2006) and the group of Mdm34 proteins described in this article, the clustering revealed a further SMP group, the uncharacterized transmembrane 24 proteins. It also yielded a number of additional groups of BPI-like proteins, including the expression site-associated gene 5 proteins (ESAG5) from trypanosomes, whose homology to BPI-like proteins has been reported previously (Barker *et al.*, 2008). BPI and Takeout-like domains are connected by the arthropod allergens, one form of which is unique in containing tandem domains with clear sequence similarity, indicating a domain duplication that occurred in insects (yellow and orange clusters in Fig. 5).

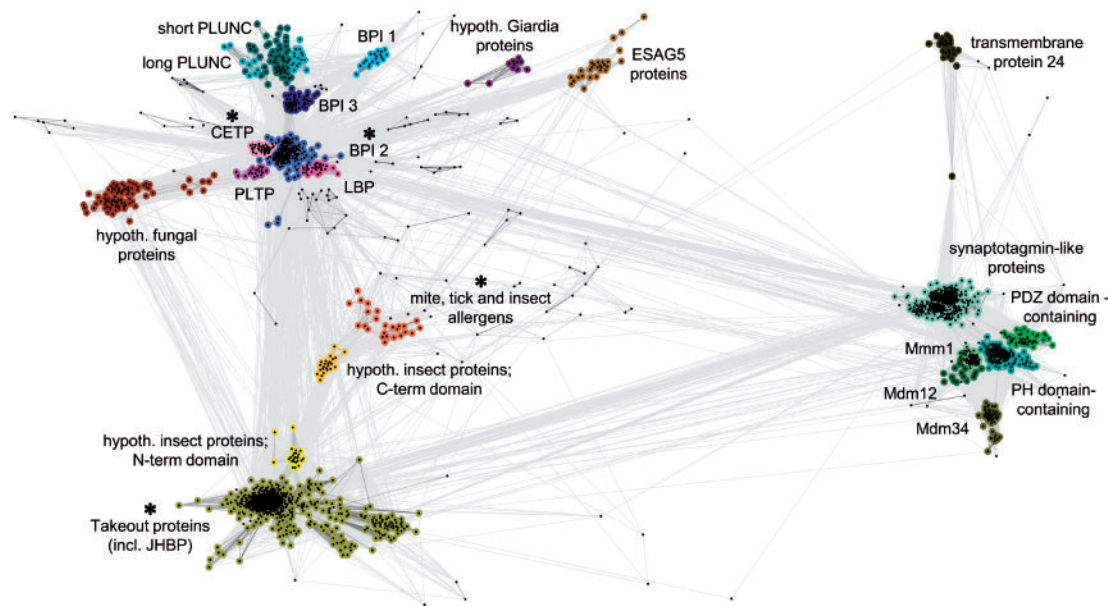
## 4 CONCLUSIONS

In this study, we have shown that the SMP domain belongs to the TULIP domain superfamily, a large group of proteins that bind lipids and other hydrophobic ligands within a central, tubular cavity (Fig. 6). In several cases (CETP, PLTP), members of this superfamily are known to exploit this binding activity in order to mediate lipid trafficking. Given the extensive lipid exchange between the ER and the mitochondrial outer membrane and the location of the ERMES complex as a connector between them, it is attractive to consider that this exchange is mediated by the SMP domains of the ERMES subunits. As the ERMES complex does not include a nucleotidase that could energize this process, we propose that it proceeds along an affinity gradient, amounting to facilitated diffusion. Although this could be envisaged as resulting from many short, structurally unspecific contacts between the SMP domains ('kiss-and-run' mechanism), we prefer to consider that the domains assemble into structurally well-defined complexes, which establish a lipophilic, tubular path between the two membranes. Since the stoichiometry of subunits within the ERMES complex is currently unknown, it is however not possible at this time to judge on whether 1:1:1 or some other ratio would most appropriately describe the composition of such complexes.





**Fig. 4.** Multiple sequence alignment of TULIP domains. An alignment comprising representatives of the SMP domain family and TULIP domains of known structure is shown. Sequences are labeled as in Figure 2. The alignment was generated by a three-step approach. First, a multiple alignment of SMP sequences was obtained using HHpred in local maximum accuracy (MAC) alignment mode. Second, a structure-based sequence alignment of TULIP domain structures was derived from a multiple structure superimposition calculated using MAMMOTH-mult (Lupyan *et al.*, 2005). In the final step, these two alignments were merged manually using as guide an alignment between 1EWF and Mmm1 obtained with HHpred. The  $\alpha$ -helices are shown in red and  $\beta$ -strands in blue. Secondary structure predictions for the SMP domains were performed with Ali2D (Biegert *et al.*, 2006). Numbers in parentheses represent length of omitted segments. Positions in the alignments that are highly conserved or strongly hydrophobic are shown in boldface. Residues that could not be aligned in structure or sequence are shown in lower case.



**Fig. 5.** Cluster map of the TULIP domain superfamily. We searched for relatives of SMP domains, Takeout proteins, dust mite allergens and N-terminal domains of BPI-like proteins in the non-redundant database using HHSenser and clustered them in CLANS based on their all-against-all pairwise similarities as measured by BLAST *P*-values. Dots represent sequences. Sequences within one group are indicated by the same color; sequences that could not be assigned to a group are not colored. Line coloring reflects BLAST *P*-values; the darker a line, the lower the *P*-value. Protein families containing members with known structure are indicated with an asterisk. The lipid-binding proteins cluster (LBP) also comprises LPSBP and LPSBP. Abbreviations are as in the text. Accession details for representatives of all clusters are provided in Supplementary Table 1.



**Fig. 6.** View along the ligand-binding tunnel of a Takeout protein (3E8T, residues 5–211). The ligand is shown as red sticks.

**Funding:** This work was supported by institutional funds from the Max-Planck-Society.

**Conflict of Interest:** none declared.

## REFERENCES

- Achleitner, G. *et al.* (1999) Association between the endoplasmic reticulum and mitochondria of yeast facilitates interorganelle transport of phospholipids through membrane contact. *Eur. J. Biochem.*, **264**, 545–553.
- Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Barker, A.R. *et al.* (2008) Bioinformatic insights to the ESAG5 and GRESAG5 gene families in kinetoplastid parasites. *Mol. Biochem. Parasitol.*, **162**, 112–122.
- Beamer, L.J. *et al.* (1997) Crystal structure of human BPI and two bound phospholipids at 2.4 angstrom resolution. *Science*, **276**, 1861–1864.
- Biegert, A. *et al.* (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.*, **34**, W335–W339.
- Cheng, H. *et al.* (2008) MALISAM: a database of structurally analogous motifs in proteins. *Nucleic Acids Res.*, **36**, D211–D217.
- Frickey, T. and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, **20**, 3702–3704.
- Hamiaux, C. *et al.* (2009) Crystal structure of *Epiphyas postvittana* takeout 1 with bound ubiquinone supports a role as ligand carriers for takeout proteins in insects. *J. Biol. Chem.*, **284**, 3496–3503.
- Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols*, **4**, 363–371.
- Kleiger, G. *et al.* (2000) The 1.7 Å crystal structure of BPI: a study of how two dissimilar amino acid sequences can adopt the same fold. *J. Mol. Biol.*, **299**, 1019–1034.
- Kolodziejczyk, R. *et al.* (2008) Insect juvenile hormone binding protein shows ancestral fold present in human lipid-binding proteins. *J. Mol. Biol.*, **377**, 870–881.
- Kornmann, B. *et al.* (2009) An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science*, **325**, 477–481.
- Krishna, S.S. and Grishin, N.V. (2004) Structurally analogous proteins do exist! *Structure*, **12**, 1125–1127.
- Kryshtafovych, A. *et al.* (2009) Protein structure prediction center in CASP8. *Proteins*, **77** (Suppl. 9), 5–9.
- Lee, I. and Hong, W. (2006) Diverse membrane-associated proteins contain a novel SMP domain. *FASEB J.*, **20**, 202–206.
- Lill, R. and Kispal, G. (2000) Maturation of cellular Fe-S proteins: an essential function of mitochondria. *Trends Biochem. Sci.*, **25**, 352–356.
- Lupyan, D. *et al.* (2005) A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics*, **21**, 3255–3263.
- McBride, H.M. *et al.* (2006) Mitochondria: more than just a powerhouse. *Curr. Biol.*, **16**, R551–R560.
- Meisinger, C. *et al.* (2004) The mitochondrial morphology protein Mdm10 functions in assembly of the preprotein translocase of the outer membrane. *Dev. Cell*, **7**, 61–71.
- Mueller, G.A. *et al.* (2010) The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. *J. Allergy Clin. Immunol.*, **125**, 909–917, e904.
- Murzin, A.G. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
- Okamoto, K. and Shaw, J.M. (2005) Mitochondrial morphology and dynamics in yeast and multicellular eukaryotes. *Annu. Rev. Genet.*, **39**, 503–536.
- Qiu, X. *et al.* (2007) Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. *Nat. Struct. Mol. Biol.*, **14**, 106–113.
- Roy, A. *et al.* (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protocols*, **5**, 725–738.
- Salem, G.M. *et al.* (1999) Correlation of observed fold frequency with the occurrence of local structural motifs. *J. Mol. Biol.*, **287**, 969–981.
- Söding, J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
- Söding, J. *et al.* (2006) HHsenser: exhaustive transitive profile search using HMM-HMM comparison. *Nucleic Acids Res.*, **34**, W374–W378.
- Voelker, D.R. (2003) New perspectives on the regulation of intermembrane glycerophospholipid traffic. *J. Lipid Res.*, **44**, 441–449.
- Wang, Z. *et al.* (2010) MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics*, **26**, 882–888.
- Wu, S. and Zhang, Y. (2008) MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*, **72**, 547–556.
- Youngman, M.J. *et al.* (2004) Mmm2p, a mitochondrial outer membrane protein required for yeast mitochondrial shape and maintenance of mtDNA nucleoids. *J. Cell Biol.*, **164**, 677–688.
- Zhang, Y. (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*, **69** (Suppl. 8), 108–117.
- Zhang, Y. (2009) I-TASSER: fully automated protein structure prediction in CASP8. *Proteins*, **77** (Suppl. 9), 100–113.