Evolution of protein structures and functions

Lisa N Kinch* and Nick V Grishin†

Within the ever-expanding repertoire of known protein sequences and structures, many examples of evolving three-dimensional structures are emerging that illustrate the plasticity and robustness of protein folds. The mechanisms by which protein folds change often include the fusion of duplicated domains, followed by divergence through mutation. Such changes reflect both the stability of protein folds and the requirements of protein function.

Addresses
Howard Hughes Medical Institute and Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9050, USA
*e-mail: lkinch@chop.swmed.edu
†e-mail: grishin@chop.swmed.edu

Current Opinion in Structural Biology 2002, 12:400–408

0959-440X/02/$ – see front matter

Abbreviations

CPS	carbamoyl phosphate synthetase
FVIIa	coagulation factor VIIa
GAT	type I glutamine amidotransferase
MPP	mitochondrial processing peptidase
PDB	Protein Data Bank
PSI-BLAST	position-specific iterated basic local alignment search tool
rmsd	root mean square deviation
SH	Src homology
TF	tissue factor
Trbp	tRNA-binding protein
TTR	transthyretin
UQ-N	ubiquitin N-terminal fragment
VatN-N	VatN N-terminal domain

Introduction

The recent accumulation of thousands of protein three-dimensional structures and the development of sensitive tools for sequence similarity searches are beginning to shed light on the evolution of protein structures and functions. Several studies have revealed many impressive and convincing examples of the evolution of three-dimensional structure, which results in homologs possessing different structural folds. On the one hand, the emergence of a new paradigm that protein structures are evolutionarily plastic and changeable has important applications for protein design and opens new frontiers in the engineering of proteins that possess desired functional properties. On the other hand, the existence of proteins with similar sequences but different structures hinders homology modeling methods and our ability to detect such cases from the sequences is crucial.

Homology, or descent from a common ancestor, is often inferred from similarities in protein sequences or structures. Many strategies have been developed to detect such similarities. Of these methods, sequence profile based algorithms such as PSI-BLAST (position-specific iterated basic local alignment search tool) and hidden Markov model based algorithms such as HMMER provide the most straightforward evidence of homology. Substantial similarities in three-dimensional structure may exist in the absence of significant sequence identity. However, because similar folds can arise independently in evolution, structural similarity alone does not provide sufficient evidence of common ancestry. In this case, evaluation of evolutionary relatedness must include additional considerations, such as similarity of molecular function, retention of unusual structural features, a common domain organization or a combination of these features. Finally, methods have been employed to detect small, localized regions of sequence and/or structural similarity. The evolutionary meaning of these types of similarities remains ambiguous and arguments can support either convergent or divergent evolution. In this review, we attempt to provide recent structural examples illustrating the concepts behind the structural and functional evolution of proteins.

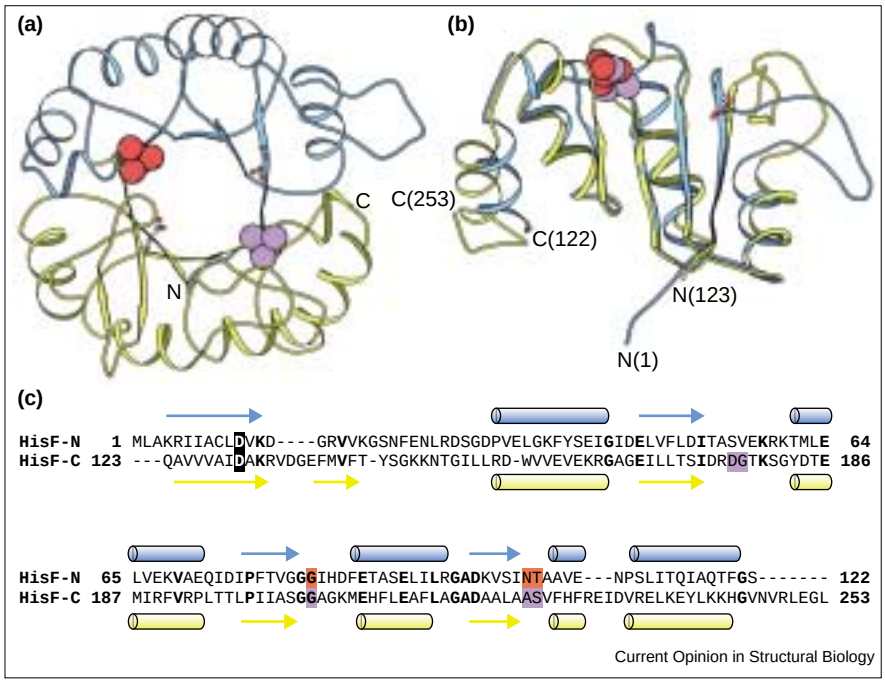
Evolution of protein structures

Protein structures evolve through a combination of mechanisms, which often include gene duplication followed by mutation and selection. Many examples of protein structural domains containing pseudo-twofold symmetry exist. Such symmetry often implies domain evolution from the fusion of two primitive half-domains. Many of these symmetric molecules retain a significant portion of sequence identity between the two halves, whereas some have mutated past the point of similarity detection at the sequence level and perhaps even at the structural level. Such mutation at the structural level can arise from a combination of mechanisms. Four of the most common mechanisms for protein fold change include insertion/deletion/substitution of secondary structural elements, circular permutation, β-strand invasion/withdrawal and β-hairpin flip/swap. Clear structural examples of these fold changes have been outlined [1]. Here, we illustrate several recent examples of fold evolution through gene duplication and discuss the concept of domain swapping as a mechanism driving oligomerization. We provide additional examples of clear sequence homologs that form alternative structures in order to illustrate the plasticity of protein folds.

HisF and HisA β/α barrels: evolution by gene duplication

The crystal structures of two histidine biosynthetic enzyme homologs (HisA and HisF) [2••] reveal $(\beta/\alpha)_8$ barrels, which display eightfold pseudo-symmetry (Figure 1a). Such a pattern suggests structural similarity between the N-terminal (N) and the C-terminal (C) halves of the $(\beta/\alpha)_8$ barrel. In fact, the half-barrel sequences detect each other with PSI-BLAST, and HisF-N and HisF-C superimpose with a rmsd of 1.58 Å (Figure 1b). Structure-based sequence

Figure 1



Gene duplication in the evolution of $(\beta/\alpha)_8$ barrels. (a) Ribbon diagram of the complete HisF $(\beta/\alpha)_8$ barrel. The N-terminal half (blue) and the C-terminal half (yellow) are related by pseudo-twofold structural symmetry. The active site phosphates are indicated in CPK representation (red and purple). The active site aspartates are indicated by ball-and-stick models. (b) Superposition of the two half-barrels, colored as (a). (c) Structural alignment of the N-terminal and the C-terminal halves of the barrel, indicating the conserved active site aspartate (black highlight) and additional invariant residues (bold). Phosphate-binding residues are highlighted in purple (C-terminal half) and red (N-terminal half). Structural elements of each half-barrel are indicated above and below the alignment (strands, arrows; helices, cylinders) and colored according to the ribbon diagrams.

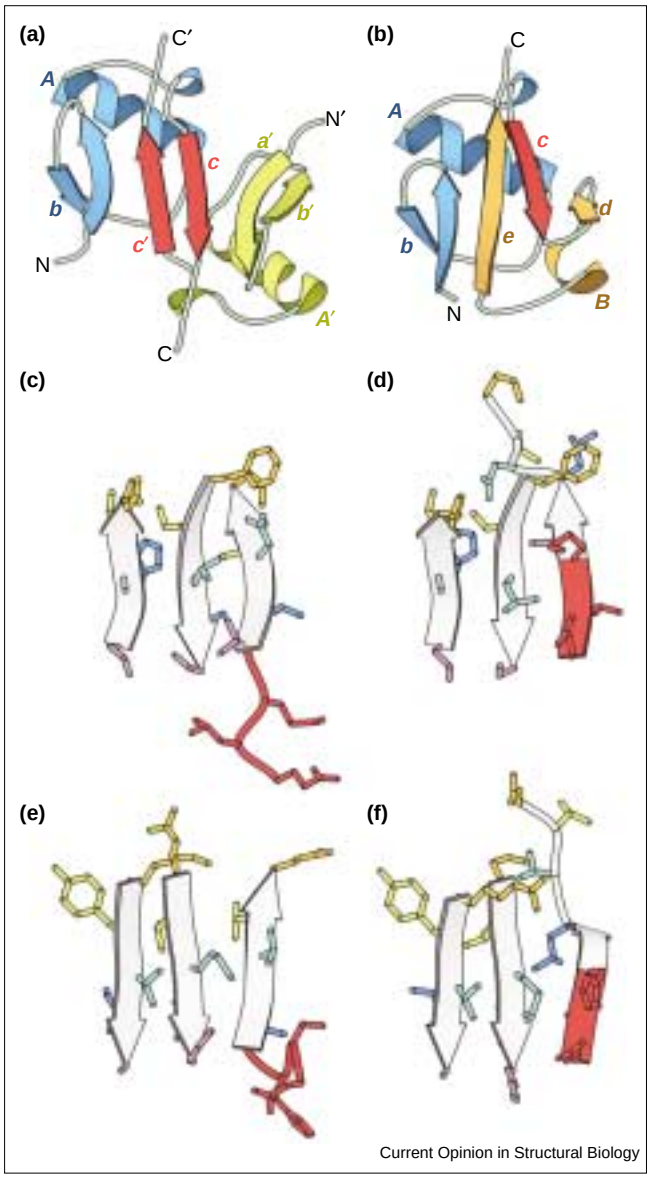
alignments of the four half-barrels indicate several invariant residues (Figure 1c), including a completely conserved aspartate in the active site. These structures support the notion that the $(\beta/\alpha)_8$ barrels arose from the duplication and fusion of an ancestral half-barrel [2••]. In support of this, coexpression *in vivo* and joint refolding *in vitro* of the HisF half-barrels produce a catalytically active enzyme [3•]. The HisF-N and HisF-C subunits form isolated, stable structures that tend toward oligomers, which are characteristics most probably mimicked by the ancestral proteins.

There are additional examples of putative ancestral homodimers giving rise to symmetric present-day structures through gene duplication and fusion events. βγ-Crystallins contain two similar domains retaining 35% sequence identity. This identity suggests that the proteins evolved from a single-domain ancestor that subsequently duplicated, fused and diverged into the present-day molecule. Consistent with this hypothesis, an engineered N-terminal domain of βB2-crystallin forms a homodimer that superimposes on the two domains of the native crystalline structure [4]. The adenylyl cyclase RV1625c gene of *Mycobacterium tuberculosis* encodes a membrane-anchored protein corresponding to one-half of the mammalian enzyme. The molecular architecture of the mycobacterial gene product includes six transmembrane spans and a single C-terminal catalytic domain. Indeed, the gene product forms a homodimeric cyclase with two active centers [5], suggesting that it may be a progenitor of mammalian adenylyl cyclases. In one last example, EMAPII, a component of the aminoacyl-tRNA synthetase complex, contains an N-terminal tRNA-binding domain homologous to bacterial tRNA-binding proteins (Trbps), followed by a domain lacking homology to any known sequence. Trbps form homodimers and interact with one tRNA molecule per dimer. The structure of EMAPII reveals the second domain to resemble the tRNA-binding domain, with a significant deletion. Despite this deletion, the interdomain interactions found in EMAPII mimic those of the homodimeric Trbps [6]. On the basis of the significant sequence divergence of the EMAPII second domain, the authors propose that the dimer mimicry seen in this structure is a result of convergent evolution. However, considering its functional and structural similarities to the homodimeric bacterial enzyme, a duplication event followed by fusion, divergence and deletion remains a plausible evolutionary scenario for the emergence of this protein.

Domain-swapped dimers: protein evolution toward oligomerization

Many structures have evolved from monomeric to oligomeric folds. How are these monomeric folds, whose surfaces are optimized for solvent contacts, driven to oligomerize? One mechanism for monomers to acquire dimeric properties involves domain swapping. Domain swapping is based on the mutual exchange of domains or structural elements between each of the molecules of a dimer. Many structures exist naturally as domain-swapped dimers, which are often in a dynamic equilibrium with their monomeric counterparts. An SH3 domain of epidermal growth factor substrate 8 [7], some members of the ribonuclease superfamily [8] and some cell-cycle regulatory proteins [9,10] exist as monomers that exchange with domain-swapped dimers. The equilibrium between such

Figure 2



Evolution of protein folds through β-strand rearrangements. **(a)** Ribbon diagram depicting the structure (PDB code 1gjz) of the ubiquitin N-terminal dimer (UQ-N). Structural elements belonging to each monomer are colored blue and yellow, with the β strands involved in the domain swap colored red. **(b)** Structure of native ubiquitin (PDB code 1ubi) depicted as a ribbon diagram. Structural elements belonging to the N-terminal core fold are colored according to those found in the UQ-N dimer (a). The shifted β strand, as compared to the UQ-N structure, is colored red. The remaining structural elements are colored orange. Comparison of the β sheets found in **(c)** the zymogen form (PDB code 1jbu) and **(d)** the active form (PDB code 1dva) of FVIIa. Comparison of the β sheets found in **(e)** the amyloidogenic mutant (PDB code 1g1o) and **(f)** the native (PDB code 1f41) TTR structures. Residue sidechains are colored according to their corresponding positions in the sheets. Residues and structural elements involved in the β-strand shifts are colored red.

dimer and monomer states can be shifted through minor mutations in the domain-swapped loops [11•,12,13] and such changes can even influence protein function [11•]. Thus, domain swapping provides an evolutionary mechanism for changing protein quaternary structures via small changes in the protein sequence.

Ubiquitin: stabilization of a core protein fold through dimerization

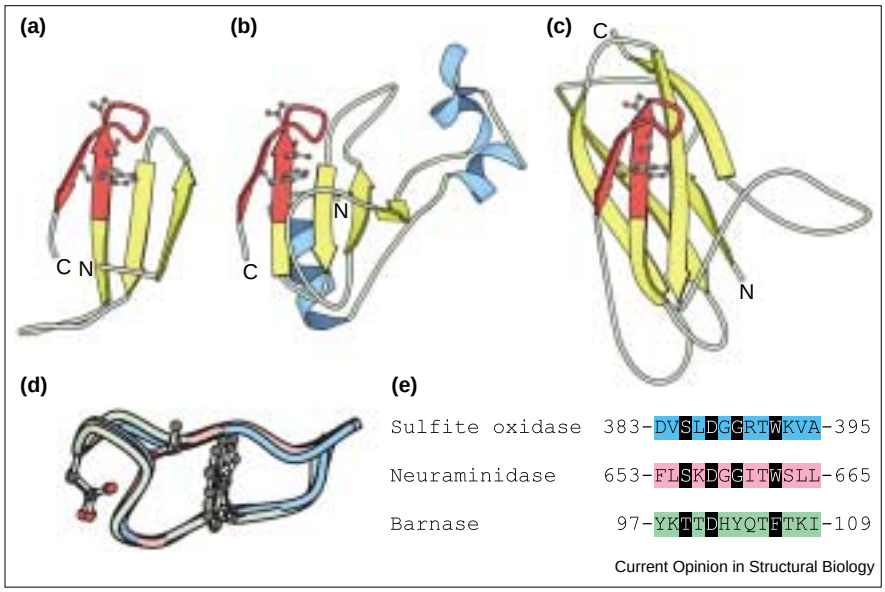
The structure of an engineered form of ubiquitin comprising the N-terminal half of the protein (UQ-N) provides an excellent example of the robustness of core protein folds [14••]. The fold of this truncated protein is stabilized by the formation of a strand-swapped dimer, which mimics the native ubiquitin fold [15] (Figure 2a,b). Both structures adopt an overall mixed orientation β sheet supported on one side by α helices. However, the truncated protein completes its core fold by replacing strand βe of native ubiquitin with strand βc′ of another monomer. The packing of the two helices of each structure is also remarkably similar, with αA′ of the dimer taking the place of 3_{10} helix B of ubiquitin. Each subunit of the UQ-N dimer adopts a hairpin-helix-strand fold, highlighting the previously revealed structural self-similarity of ubiquitin [15]. Such structural self-similarity is reminiscent of the pseudo-twofold symmetry seen in previously discussed examples and, by analogy, could reflect the evolution of the present-day ubiquitin fold by gene duplication followed by significant sequence and structural divergence.

β-Strand slip: two structures in one sequence

Comparison of the UQ-N structure with that of ubiquitin reveals an interesting difference in the arrangement of the four central strands. Although interactions between βa and βc′ of the dimer, and βa and βe of ubiquitin are parallel in each case, the strand registry has slipped by two residues, resulting in different relative orientations of the elements (Figure 2a,b). This conformational change is reminiscent of an incomplete strand invasion/withdrawal previously described for the serine protease inhibitors [1,16]. Interestingly, such a strand slip had been observed in the structure of the zymogen form of coagulation factor VIIa (FVIIa) and in an amyloidogenic mutant of transthyretin (TTR). In each of these cases, identical, or nearly identical, sequences exhibit rearranged structural elements, adopting different packing within the protein fold.

FVIIa contains a trypsin-like serine protease domain responsible for initiating a proteolytic cascade leading to blood clot formation. Like other trypsin-like proteases, FVIIa is produced as an inactive zymogen, requiring a proteolytic cleavage event for activation. The cleaved form of the enzyme, however, is devoid of proteolytic activity and requires an interaction with tissue factor (TF) for complete conversion to its active form. The recently solved structure of the FVIIa zymogen (PDB code 1jbu; [17•]) reveals the structural basis for this conversion. When compared to the structure of the activated enzyme [18,19], the zymogen displays a different registration between two β strands (Figure 2c,d), resulting in the withdrawal of residues that are part of the TF interaction site. This remarkable

Figure 3



The Asp-box motif. Ribbon diagrams of (a) a single blade of the *Vibrio cholerae* neuraminidase (PDB code 1kit: residues 617–684) β-propeller structure, (b) the microbial ribonuclease barnase (PDB code 1a2p: chain A) and (c) the C-terminal domain of sulfite oxidase (PDB code 1sox: chain A, residues 346–466). The α helices and β sheets are colored blue and yellow, respectively. Structural elements corresponding to the Asp-box motif are colored red, with conserved residues depicted as ball-and-stick models. (d) Structural alignment of the Asp-box motifs of neuraminidase (pink), barnase (green) and sulfite oxidase (blue), with conserved residues depicted as ball-and-stick models. (e) Alignment of the sequences of Asp-box motifs, with conserved residues highlighted in black.

conformational change includes the destruction of 15 hydrogen bonds between β strands and the reincorporation of 12 different hydrogen bonds in a three-residue β-strand shift. The authors suggest that the FVIIa sequence can support this shift because of the presence of a Leu-X-Val-Leu-X-Val tripeptide repeat [17•].

TTR functions normally as a transport protein for thyroid proteins. However, TTR can form deposits of insoluble protein fibrils and is associated with two clinical forms of amyloidosis. The crystal structure of a triple mutant form of TTR (G53S/E54D/L55S [20•]) prone to amyloid characteristics displays the same type of conformational change seen in FVIIa. Comparison of the native TTR structure [21] with the amyloidogenic mutant reveals a three-residue β-strand shift, in which Leu58 of the mutant resides in the site previously occupied by Leu55 (Figure 2e,f).

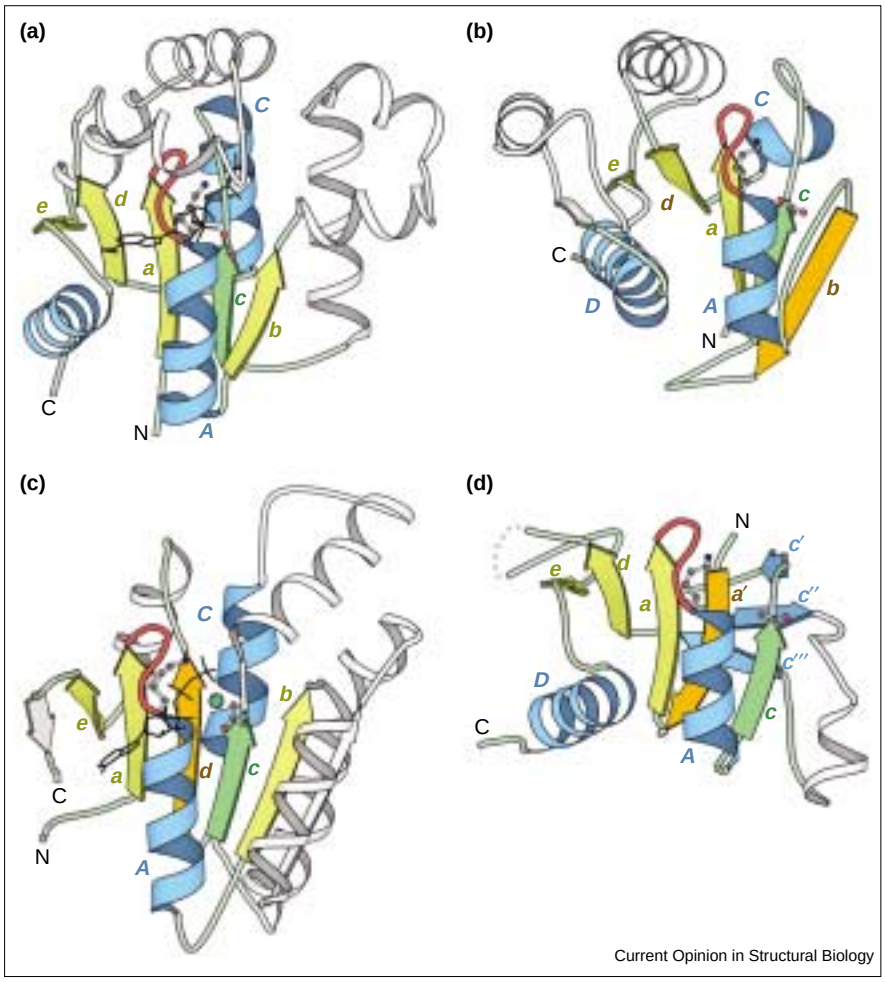
Protein structure plasticity: small mutations lead to big changes

In the previous examples of β-strand shifts, the stability of the protein fold is demonstrated by the ability of different sequences to generate identical structural elements. Alternatively, quite similar sequences can form completely different structural elements, highlighting the plasticity of these protein folds. Recent structures of several engineered proteins expand on this concept of protein plasticity and illustrate an ability to generate different structures with similar sequences. The structure of a double mutant of the Arc repressor protein (N11L/L12N) adopts a novel fold in which two 3_{10} helices replace a two-stranded, antiparallel β sheet found in the native structure [22]. A single mutant bearing the N11L substitution can adopt either fold and exists in a dynamic equilibrium between the two states, representing an evolutionary bridge through protein sequence/structure space [23••]. An additional example of protein plasticity is seen in the structure of an RNA-binding protein (ROP) that forms a canonical left-handed, all-antiparallel four-helix bundle. A single amino acid substitution in the turn region (A31P) results in the complete structural rearrangement of the protein. The ROP mutant forms a right-handed, mixed parallel and antiparallel bundle [24].

VatN: evolution of complex structures from simple elements

The structure of VatN (a substrate recognition domain in the AAA family of ATPases) provides an excellent example of the evolutionary concepts illustrated in this review. The VatN N-terminal domain (VatN-N) displays an internal sequence duplication (38% identity over 42 residues) that divides the domain into two halves and translates into pseudo-twofold structural symmetry (the two halves form identical ββαβ motifs). The two ββαβ motifs are completely interleaved, forming a double-psi β barrel with two short helices located in mirrored loops. Sequence and structural similarities between VatN-N and prokaryotic transcription factors, metabolic enzymes, protease cofactors and aspartic proteinases suggest a logical evolutionary path from simple, homodimeric transcription factors containing a single copy of the VatN-N ββαβ motif to complex enzymes containing as many as four copies of the motif [25]. A group of archaeal and eukaryotic transcription factors represent the putative ancestral element and relate to VatN-N by a circular permutation that omits the first β strand and includes the sequence-related fourth β strand (βαββ). The permuted transcription factor dimerizes to form a minimally interleaved barrel and lacks the psi loops of the VatN-N domain. All other VatN-N homologs can evolve from this simple βαββ structure through duplication and permutation events.

Figure 4



P-loop NTPases. Ribbon diagrams of the P-loop NTPases (a) adenylate kinase (PDB code 1aky), (b) ras (PDB code 1q21), (c) F_1-ATPase (PDB code 1bmf) and (d) HPr kinase (PDB code 1jb1). Core structural elements are colored blue for common α helices (and the β strands replacing an α helix in HPr kinase), yellow for common β strands and white for insertions. The P-loops are colored red, with conserved lysines represented as ball-and-stick models. The strands containing the Walker B motifs are colored green, with conserved aspartic acids represented as ball-and-stick models. When compared to the structure of adenylate kinase, β strands located in different positions (βd of F_1-ATPase), in different orientations (βb of ras) or with different connectivity (βa′ of HPr kinase replaces βb) are colored orange. Structural elements are labeled according to connectivity. The black stick structures in (a,c) represent nucleotide analogs.

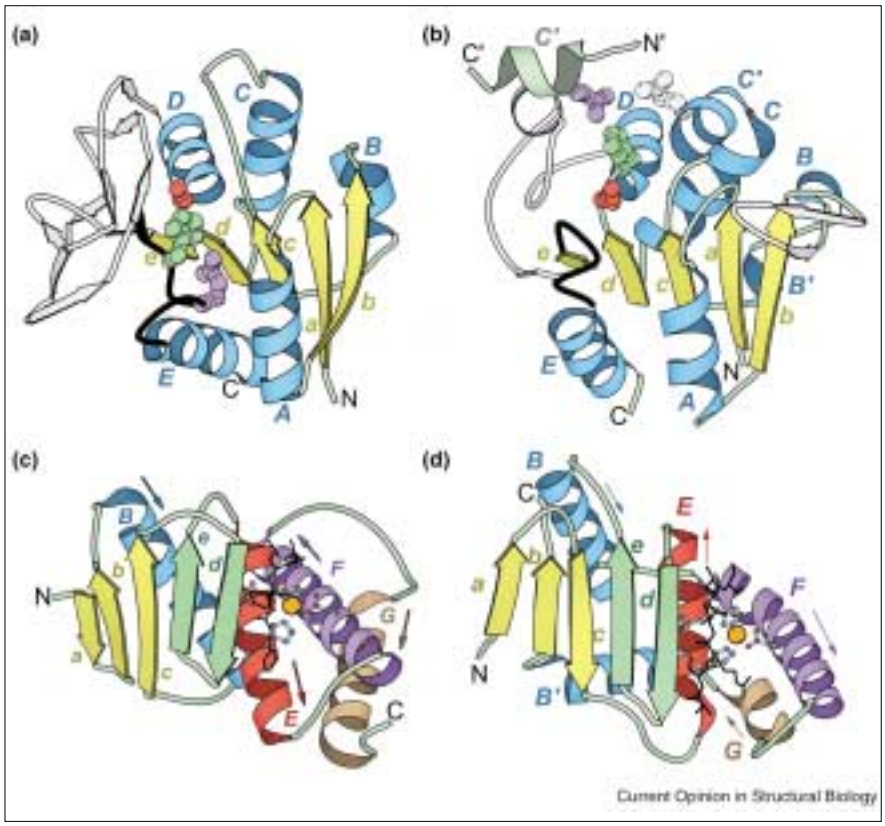
Localized structural motifs and fold evolution

Motifs exhibiting similar sequences and similar localized structures can exist within apparently nonhomologous protein folds. Examples of such motifs have been detected using sequence- and structure-based searching methods and include heme attachment motifs, P-loops (Walker A), FAD/NAD-binding motifs, Zn fingers, Fe-S-binding motifs, RNA-binding motifs and the Asp-box [26••,27,28•]. The differences in the overall folds of proteins containing these motifs bring into question their evolutionary pathways. Arguments can be made for convergent evolution based on functional selection. Alternatively, arguments can be made for divergent evolution of the overall folds from a common ancestor. Such divergence may arise from a common motif-containing ancestor through different mechanisms of fold change [1] or the motifs may represent the structural remains of a predomain world, with their present-day fold evolving from the assembly of primitive peptide proteins [26••,27,28•].

The Asp-box

The Asp-box is a β hairpin originally identified as a recurring motif in bacterial sialidases. By searching structural databases using the conformation of a typical Asp-box, Copley *et al.* [28•] have detected the motif in more than nine protein families. The motifs exhibit significant sequence and structural similarity, and localize within protein structures that have otherwise different sequences and folds (Figure 3). Sialidase [29] contains multiple copies of the Asp-box motif in a six-blade β-propeller structure. Such an arrangement provides a good argument for gene duplication followed by divergence of a common ancestral blade (Figure 3a). Other Asp-box proteins do not have clear evolutionary origins, as illustrated by the structures of barnase [30] and sulfite oxidase [31]. Barnase contains an Asp-box in an antiparallel β sheet that packs against a single α helix (Figure 3b) and sulfite oxidase contains an Asp-box in its C-terminal Greek key domain, which is characterized by seven antiparallel β strands that form two β sheets (Figure 3c). Although each of these structures has different overall folds, the localized structures encompassing their Asp-box motifs superimpose nicely (Figure 3d) and the sequences are quite similar (Figure 3e). A conserved water molecule is present in all but one of the detected Asp-box structures [28•]. However, a common functional significance for this motif remains

Figure 5



Evolution of protein function. Ribbon diagrams of the GAT superfamily homologs. (a) The CPS small subunit (PDB code 1c30: chain b) and (b) the intracellular protease from *P. horikoshii* (PDB code 1g2i: chain a; chain c residues 470–479) display identical active site residues located in different structural positions. The α helices and β strands of the GAT domain Rossmann-fold-like core are colored in blue and yellow, respectively. Inserts are depicted in white. The catalytic triad cysteine (red), histidine (green) and glutamate (purple) are represented as large ball-and-stick models. The loop containing catalytic residues common to most GAT domains is colored black. The helix from the second subunit contributing the catalytic residue (glutamate) across the dimer interface is colored in pale green, whereas the equivalent residue in the first subunit is depicted in white. The major structural elements of the GAT domain fold are labeled according to connectivity. Ribbon diagrams of Zn-dependent proteases. (c) Thermolysin from *Bacillus thermoproteolyticus* (PDB code 4fmn) and (d) MPP (PDB code 1hr8) show similar active sites formed in nonhomologous structures with different fold topologies. Structural elements present in both folds are colored and labeled according to the connectivity of thermolysin. Opposite orientations of helices are depicted with arrows. Active site residues are depicted as ball-and-stick molecules, Zn ligands are colored orange and peptide inhibitors/substrates are colored black.

unclear, making its substantial conservation enigmatic. In terms of evolution, the lack of functional conservation of this motif argues against functionally driven convergence.

P-loop NTPases

P-loop NTPases represent an extremely diverse and abundant protein superfamily characterized by Walker A (P-loop) and Walker B motifs, which bind nucleotide and divalent cation, respectively. Detection of proteins with these motifs by sequence analysis tools is relatively straightforward and a monophyletic origin for P-loop NTPases has been proposed [32]. Representative structures for most of the distinct families of P-loop NTPases have been determined [33–36]. All these structures display an overall αβα sandwich architecture, with a mainly parallel central β sheet composed of βα units (Figure 4). However, the connectivity between βα units is not the same in different NTPase families and the relative location of the Walker A and Walker B motifs changes. The structure of adenylate kinase [34] represents a prototypical P-loop NTPase fold (Figure 4a). A Walker A motif with a conserved lysine residue (shown in red) follows the first β strand (βa). Walker B is structured as an adjacent β strand (βc), with a conserved acidic Mg^{2+}-binding residue (shown in green). Ras-like G proteins [36] maintain this same relative location of Walker A and Walker B motifs (Figure 4b), but their second β strand (βb, shown in orange, Figure 4c) orients itself in reverse due to a deletion of structural elements. In the structure of the central domain of the mitochondrial F_1-ATPase [35], β strands βa and βc are separated by β strand βd (Figure 4c), which can be rationalized in terms of strand invasion/withdrawal. Such a structural change may be functionally selected by a requirement for the correct positioning of the P-loop. Finally, the structure of HPr kinase (PDB code 1jb1; [33]) diverges the most from the other P-loop structures. An N-terminal β strand (βa′, colored orange, Figure 4d) has inserted between βa and βc. The βb strand is no longer present and an antiparallel β sheet (βc′, βc″, βc‴, colored blue, Figure 4d) has replaced α helix αC.

Evolution of protein functions

With the classification of vast amounts of emerging sequence and structural data into evolutionary families (SCOP or CATH databases), we can begin to understand the relationships between protein sequence and structure. In order to gain a global understanding of sequence and structural evolution, however, we must also consider the role of protein function in these processes. Clearly, protein three-dimensional structure is more conserved than either sequence or function. Different protein folds can converge on the same function, with identical catalytic residues built

on completely different scaffolds. Conversely, a single architectural fold can adopt diverse sets of functions, with differing catalytic residues found in similar sites, similar catalytic residues found in different sites or a combination of the two. Todd *et al.* [37••] assessed such variation by comparing the function, as defined by the Enzyme Commission (EC) scheme, of enzymes within homologous protein superfamilies. This work explores the structural context of functional diversity in terms of conservation and variation of catalytic residues, reaction mechanisms and substrate specificity. They find functional diversity to be significant below a threshold of 30% sequence identity and describe numerous examples of functional variation, whereby superfamilies retain conservation of reaction chemistry, as opposed to conservation of substrate specificity. A second dominant theme from this study includes the concept of 'migration of catalytic residues' within homologous protein folds. Members of the hexapeptide repeat superfamily, the α/β hydrolase superfamily and the $(\beta/\alpha)_8$-barrel superfamily contain similar catalytic residues in different structural elements of their homologous protein folds [37••,38]. Such examples complicate the prediction of functional residues based on homology alone.

GAT superfamily: divergent catalytic sites in homologous structures

GAT (type I glutamine amidotransferase) domains are generally found in a group of biosynthetic enzymes that catalyze the transfer of water to free glutamine to create ammonia. Classic GAT domains perform this chemistry using a conserved catalytic triad (Cys-His-Asp/Glu) similar to the triad found in serine and cysteine proteases. GAT superfamily members, which include carbamoyl phosphate synthetase (CPS), have been evolutionarily linked to a domain in *E. coli* HPII catalase (lacks the catalytic triad) and to the thiJ putative protease domain [39]. Several GAT-containing proteins have been structurally characterized and contain a variation of the Rossmann fold topology, with a central sheet formed by five parallel β strands (order edcab). The strands are all connected in a right-handed fashion by α helices that flank the sheet on either side (Figure 5a,b, yellow and blue). A large insertion of variable sequence is also common in GAT domains (Figure 5a,b, white).

The structure of the CPS small subunit [40] represents a classic GAT domain with a conserved catalytic Cys-His-Glu triad required for activity (Figure 5a). The loop in the last β/α unit connecting βe and αE (Figure 5a, black) houses His353 and Glu355 of the triad. The third member of the triad (Cys269) resides in a sharp turn at the N terminus of αD, referred to as the nucleophile elbow [40]. The structural location of these functionally important residues is conserved among classic GAT domains. However, this conservation breaks down in the thiJ members of the GAT superfamily and multiple sequence alignment does not help in localizing active site residues [39]. The crystal structure of one thiJ member, an intracellular protease from *Pyrococcus horikoshii* (PH1704), illustrates this 'migration of catalytic residues' within the GAT superfamily [41•]. When compared to classic GAT domains, PH1704 retains Cys100 in the same nucleophile elbow, but two catalytic triad members reside on completely different structural elements (Figure 5b). His101 neighbors the catalytic Cys101 and Glu474 is provided by an adjacent monomer [41•]. Thus, PH1704 has adapted portions of its triad to different structural elements than those found in its GAT domain homologs and has included oligomerization as a strategy for creating this diversity.

Zn-dependent proteases: similar catalytic sites converge in nonhomologous structures

In contrast to the previous example, the Zn-dependent proteases display evolution of function through convergence. These enzymes built very similar active sites into different structural templates (as in trypsin and chymotrypsin), suggesting that the same catalytic triad arose independently during evolution. Knowledge of the spatial arrangements of such functional sites may help in predicting protein function [42]. The structures of two nonhomologous Zn-dependent proteases, thermolysin [43] and mitochondrial processing peptidase (MPP) [44,45], display striking structural convergence of overall molecular architecture, in addition to functional convergence of active site residues [46]. Thermolysin resembles a Rossmann fold and belongs to the family of zincin-fold proteases. The mechanism of this enzyme is well characterized and requires a key catalytic glutamate and a single zinc ion for its reaction chemistry. Two out of three thermolysin Zn ligands and the catalytic glutamate are located on helix αE (Figure 5c) in a signature HEXXH Zn-binding motif. By contrast, MPP possesses a modified ferredoxin fold and the inverted signature pattern HXXEH (Figure 5d).

Despite this structural nonhomology, all major functionally important structural elements are present in both thermolysin and MPP with different topological connections; the reverse orientation of αE explains the 'inversion' of the signature motifs. Each structure displays a loose bundle of α helices packed on one side of a five-stranded, mixed orientation β sheet. Makarova and Grishin [46] point out that superposition of the Cα atoms of the three Zn ligands and the catalytic glutamate results in an rmsd of 1.56 Å and leads to striking overlap of secondary structural elements, which superimpose with an rmsd of 3.4 Å (for 85 Cα atoms, ignoring connectivity and orientation). In fact, the striking resemblance of the overall structural elements led these authors to postulate a similar mode of substrate binding for MPP as is seen in thermolysin [46], in which the substrate peptide extends the β sheet, forming hydrogen bonds with βd in an antiparallel manner (Figure 5c). The recent crystal structure of MPP bound to substrate confirmed this hypothesis [45]. Such an example of convergence in structure and function is rare, and nicely illustrates the concept of functional selection. In this case, two different folds arose independently, converging on both similar overall structure and similar catalytic residue spatial orientation. This

convergence was presumably influenced by a requirement for catalysis, in addition to a limited number of alternatives to build a scaffold that can utilize hydrogen bonding between a peptide substrate and an enzyme β sheet.

Conclusions

Generally, the primary sequence of a protein dictates its fold and function, and cumulative changes in this sequence lead to the evolution of protein structures and functions. However, examples exist whereby very similar or even identical sequences fold into different structures. These examples signify an alternative to the concept that structures are more conserved than sequences and demonstrate that protein structures can evolve and change, possibly generating new folds and topologies. Just as in the structural evolution of proteins, the functional evolution of proteins can be divergent, with a 'migration of catalytic residues' within homologous folds, or it can be convergent, with identical functional residues forming similar spatial arrangements in completely different protein folds. How sequence dictates such structural and functional changes plays a vital role in the understanding of evolution.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Grishin NV: Fold change in evolution of protein structures. *J Struct Biol* 2001, 134:167-185.

2. •• Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M: Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 2000, 289:1546-1550.
The atomic structures of two $(\beta/\alpha)_8$ barrel histidine biosynthetic enzyme homologs suggest the evolution of protein structure from ancient gene duplications and fusions of two half-barrels.

3. • Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R: Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nat Struct Biol* 2001, 8:32-36.
The N- and C-terminal halves of the HisF $(\beta/\alpha)_8$ barrel were produced separately and fold into independent units. Upon coexpression or joint refolding, these half-barrels interact to form a catalytically active protein. The sequence similarity between the two half-barrels, combined with their ability to oligomerize from two independent domains, supports the notion that each half-barrel descended from a common ancestor, which may have existed as a homodimer of identical half-barrels.

4. Clout NJ, Basak A, Wieligmann K, Bateman OA, Jaenicke R, Slingsby C: The N-terminal domain of βB2-crystallin resembles the putative ancestral homodimer. *J Mol Biol* 2000, 304:253-257.

5. Guo YL, Seebacher T, Kurz U, Linder JU, Schultz JE: Adenylyl cyclase Rv1625c of ***Mycobacterium tuberculosis***: a progenitor of mammalian adenylyl cyclases. *EMBO J* 2001, 20:3667-3675.

6. Renault L, Kerjan P, Pasqualato S, Menetrey J, Robinson JC, Kawaguchi S, Vassylyev DG, Yokoyama S, Mirande M, Cherfils J: Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry. *EMBO J* 2001, 20:570-578.

7. Kishan KV, Newcomer ME, Rhodes TH, Guilliot SD: Effect of pH and salt bridges on structural assembly: molecular structures of the monomer and intertwined dimer of the Eps8 SH3 domain. *Protein Sci* 2001, 10:1046-1055.

8. Liu Y, Hart PJ, Schlunegger MP, Eisenberg D: The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution. *Proc Natl Acad Sci USA* 1998, 95:3437-3442.

9. Bourne Y, Arvai AS, Bernstein SL, Watson MH, Reed SI, Endicott JE, Noble ME, Johnson LN, Tainer JA: Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci USA* 1995, 92:10232-10236.

10. Bourne Y, Watson MH, Hickey MJ, Holmes W, Rocque W, Reed SI, Tainer JA: Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1. *Cell* 1996, 84:863-874.

11. • Schymkowitz JW, Rousseau F, Wilkinson HR, Friedler A, Itzhaki LS: Observation of signal transduction in three-dimensional domain swapping. *Nat Struct Biol* 2001, 8:888-892.
Single point mutations shift the equilibrium between a monomer and a domain-swapped dimer of suc1. The function of suc1, DNA binding, is influenced by the oligomeric state.

12. Canals A, Pous J, Guasch A, Benito A, Ribo M, Vilanova M, Coll M: The structure of an engineered domain-swapped ribonuclease dimer and its implications for the evolution of proteins toward oligomerization. *Structure* 2001, 9:967-976.

13. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci USA* 2001, 98:10687-10691.

14. •• Bolton D, Evans PA, Stott K, Broadhurst RW: Structure and properties of a dimeric N-terminal fragment of human ubiquitin. *J Mol Biol* 2001, 314:773-787.
The atomic structure of an engineered N-terminal fragment of ubiquitin folds into a domain-swapped dimer that mimics the symmetry-related fold of ubiquitin. The structure of the ubiquitin N-terminal fragment illustrates the robustness of a protein fold. The protein overcomes its N-terminal deletion by dimerizing to stabilize a native ubiquitin-like fold.

15. Vijay-Kumar S, Bugg CE, Cook WJ: Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* 1987, 194:531-544.

16. Loebermann H, Tokuoka R, Deisenhofer J, Huber R: Human α1 proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J Mol Biol* 1984, 177:531-557.

17. • Eigenbrot C, Kirchhofer D, Dennis MS, Santell L, Lazarus RA, Stamos J, Ultsch MH: The factor VII zymogen structure reveals reregistration of beta strands during activation. *Structure* 2001, 9:627-636.
The atomic structure of the FVIIa zymogen reveals its mechanism of activation to require an unusual reregistration a β strand. The reregistration of the β strand requires two different sequence portions to form the same structural element.

18. Dennis MS, Eigenbrot C, Skelton NJ, Ultsch MH, Santell L, Dwyer MA, O'Connell MP, Lazarus RA: Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* 2000, 404:465-470.

19. Banner DW, D'Arcy A, Chene C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D: The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* 1996, 380:41-46.

20. • Eneqvist T, Andersson K, Olofsson A, Lundgren E, Sauer-Eriksson AE: The beta-slip. A novel concept in transthyretin amyloidosis. *Mol Cell* 2000, 6:1207-1218.
The atomic structure of an amyloidogenic mutant of TTR reveals a β-strand reregistration that may account for the amyloidogenic folding properties of the protein.

21. Hornberg A, Eneqvist T, Olofsson A, Lundgren E, Sauer-Eriksson AE: A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J Mol Biol* 2000, 302:649-669.

22. Cordes MH, Walsh NP, McKnight CJ, Sauer RT: Evolution of a protein fold ***in vitro***. *Science* 1999, 284:325-328.

23. •• Cordes MH, Burton RE, Walsh NP, McKnight CJ, Sauer RT: An evolutionary bridge to a new protein fold. *Nat Struct Biol* 2000, 7:1129-1132.
A single mutation in the Arc repressor protein results in a dynamic equilibrium between two folds, one including a two-stranded β sheet and one with two 3_{10} helices. The conformational exchange between structural elements of a single protein sequence provides an example of structural plasticity and suggests a general mechanism of fold change through incremental small mutations.

24. Glykos NM, Cesareni G, Kokkinidis M: Protein plasticity to the extreme: changing the topology of a 4-alpha-helical bundle with a single amino acid substitution. *Structure* 1999, 7:597-603.

25. Coles M, Diercks T, Liermann J, Groger A, Rockel B, Baumeister W, Koretke KK, Lupas A, Peters J, Kessler H: The solution structure of VAT-N reveals a 'missing link' in the evolution of complex

enzymes from a simple βαββ element. *Curr Biol* 1999, 9:1158-1168.

26. •• Lupas AN, Ponting CP, Russell RB: On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J Struct Biol* 2001, 134:191-203.
A structure/sequence similarity detection scheme detects localized similar regions of different overall structural folds and defines seven classes of motifs. The evolutionary meaning of such motifs is discussed.

27. Fetrow JS, Godzik A: Function driven protein evolution. A possible proto-protein for the RNA-binding proteins. *Pac Symp Biocomput* 1998:485-496.

28. • Copley RR, Russell RB, Ponting CP: Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. *Protein Sci* 2001, 10:285-292.
A structural search against representatives of all protein families in the SCOP database reveals Asp-boxes in more than nine protein families. Possible functions and evolutionary scenarios of the motif are discussed.

29. Crennell S, Garman E, Laver G, Vimr E, Taylor G: Crystal structure of ***Vibrio cholerae*** neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* 1994, 2:535-544.

30. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A: Molecular structure of a new family of ribonucleases. *Nature* 1982, 297:162-164.

31. Kisker C, Schindelin H, Pacheco A, Wehbi WA, Garrett RM, Rajagopalan KV, Enemark JH, Rees DC: Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell* 1997, 91:973-983.

32. Neuwald AF, Aravind L, Spouge JL, Koonin EV: AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 1999, 9:27-43.

33. Fieulaine S, Morera S, Poncet S, Monedero V, Gueguen-Chaignon V, Galinier A, Janin J, Deutscher J, Nessler S: X-ray structure of HPr kinase: a bacterial protein kinase with a P-loop nucleotide-binding domain. *EMBO J* 2001, 20:3917-3927.

34. Egner U, Tomasselli AG, Schulz GE: Structure of the complex of yeast adenylate kinase with the inhibitor P1,P5-di(adenosine-5′-)pentaphosphate at 2.6 Å resolution. *J Mol Biol* 1987, 195:649-658.

35. Abrahams JP, Leslie AG, Lutter R, Walker JE: Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 1994, 370:621-628.

36. Tong LA, de Vos AM, Milburn MV, Kim SH: Crystal structures at 2.2 Å resolution of the catalytic domains of normal ras protein and an oncogenic mutant complexed with GDP. *J Mol Biol* 1991, 217:503-516.

37. •• Todd AE, Orengo CA, Thornton JM: Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 2001, 307:1113-1143.
The correlation between 31 homologous enzyme superfamilies and their function is assessed, and the modulation in sequence and structure that results in functional changes is discussed.

38. Heikinheimo P, Goldman A, Jeffries C, Ollis DL: Of barn owls and bankers: a lush variety of alpha/beta hydrolases. *Structure* 1999, 7:R141-R146.

39. Horvath MM, Grishin NV: The C-terminal domain of HPII catalase is a member of the type I glutamine amidotransferase superfamily. *Proteins* 2001, 42:230-236.

40. Thoden JB, Huang X, Raushel FM, Holden HM: The small subunit of carbamoyl phosphate synthetase: snapshots along the reaction pathway. *Biochemistry* 1999, 38:16158-16166.

41. • Du X, Choi IG, Kim R, Wang W, Jancarik J, Yokota H, Kim SH: Crystal structure of an intracellular protease from ***Pyrococcus horikoshii*** at 2-A resolution. *Proc Natl Acad Sci USA* 2000, 97:14079-14084.
The atomic structure of a GAT superfamily member is presented. The structure illustrates a 'migration of catalytic residues', with two catalytic triad members residing on different structural elements than those exhibited by classic GAT superfamily members.

42. Russell RB, Sasieni PD, Sternberg MJ: Supersites within superfolds. Binding site similarity in the absence of homology. *J Mol Biol* 1998, 282:903-918.

43. Holden HM, Tronrud DE, Monzingo AF, Weaver LH, Matthews BW: Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues. *Biochemistry* 1987, 26:8542-8553.

44. Braun HP, Schmitz UK: The bifunctional cytochrome c reductase/processing peptidase complex from plant mitochondria. *J Bioenerg Biomembr* 1995, 27:423-436.

45. Taylor AB, Smith BS, Kitada S, Kojima K, Miyaura H, Otwinowski Z, Ito A, Deisenhofer J: Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure* 2001, 9:615-625.

46. Makarova KS, Grishin NV: Thermolysin and mitochondrial processing peptidase: how far structure-functional convergence goes. *Protein Sci* 1999, 8:2537-2540.