

JMBAvailable online at www.sciencedirect.com ScienceDirect

COMMUNICATION

A Periodic Table of Coiled-Coil Protein Structures

Efrosini Moutevelis¹ and Derek N. Woolfson^{1,2*}¹*School of Chemistry,
University of Bristol,
Bristol BS8 1TS, UK*²*Department of Biochemistry,
University of Bristol,
Bristol BS8 1TD, UK**Received 19 August 2008;
received in revised form
15 November 2008;
accepted 18 November 2008
Available online
25 November 2008*

Coiled coils are protein structure domains with two or more α -helices packed together via interlacing of side chains known as knob-into-hole packing. We analysed and classified a large set of coiled-coil structures using a combination of automated and manual methods. This led to a systematic classification that we termed a “periodic table of coiled coils,” which we have made available at http://coiledcoils.chm.bris.ac.uk/ccplus/search/periodic_table. In this table, coiled-coil assemblies are arranged in columns with increasing numbers of α -helices and in rows of increased complexity. The table provides a framework for understanding possibilities in and limits on coiled-coil structures and a basis for future prediction, engineering and design studies.

© 2008 Elsevier Ltd. All rights reserved.

*Edited by M. Sternberg**Keywords:* α -helix; coiled coil; knob-into-hole interaction; protein structure classification; periodic table

Protein structure is hierarchical: polypeptide chains fold locally to form α -helices and β -strands; these secondary structures combine to form tertiary structures, and independently folded tertiary structures may associate to form complexes or quaternary structures. These apparent simplicities of building blocks (the secondary structures) and order in assembly (secondary–tertiary–quaternary) mask the potential complexity of protein structures; in other words, there are a large number (in the order of thousands) of possible ways to fold protein chains into stable tertiary and quaternary structures.^{1–3} Over the past few decades, researchers have attempted to capture and relate the observed protein structures from nature in various classification schemes. The first of these was the straightforward $\alpha\alpha$, $\alpha+\beta$, α/β and $\beta\beta$ system of Chothia.⁴ Developing on this, expert-defined and semi-automated databases such as SCOP and CATH, respectively, provide the industry standard.^{5,6} Through such studies, the number of protein domains “used” by nature is currently estimated at 5000.⁷

In a related question, others have attempted to explore the theoretical potential for and limits on protein structures. Regarding all- α -helical struc-

tures, Murzin and Finkelstein presented an elegant hierarchy for helical globules;⁸ similar analyses have been performed for all- β structures,^{9–11} and, most recently, Taylor attempted to capture all combinations of α and β structures in a periodic table of protein structures.¹² Here, we present a classification system for a common protein-folding and association motif called the α -helical coiled coil, a motif that is either not formally included or not captured well by the existing classification schemes. We organised coiled-coil structures in a periodic table that sheds light on the likely limits on the coiled-coil architectures and topologies, as well as providing potential frameworks for improved protein structure prediction, engineering and design.

The α -helical coiled coil is a ubiquitous protein-folding motif¹³ with an estimated occurrence of 5%–10% in translated protein sequences.^{14,15} It is found in a variety of structural forms and in a wide range of proteins,^{16,17} including, for example, small units such as leucine zippers that drive the dimerisation of many transcription factors, longer segments that underpin dimeric and trimeric fibrous structural proteins of the cytoskeleton and extracellular matrix, and more-complex structures such as the family of viral proteins responsible for virus–host membrane fusion.

In structural terms, coiled coils are bundles of α -helices in which the helical axes are aligned slightly offset from one another (Fig. 1a). These assemblies are encoded at the protein sequence level by motifs in which hydrophobic (**H**) residues are usually alter-

*Corresponding author. School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. E-mail address: D.N.Woolfson@bristol.ac.uk.

Abbreviations used: KIH, knob-into-hole.

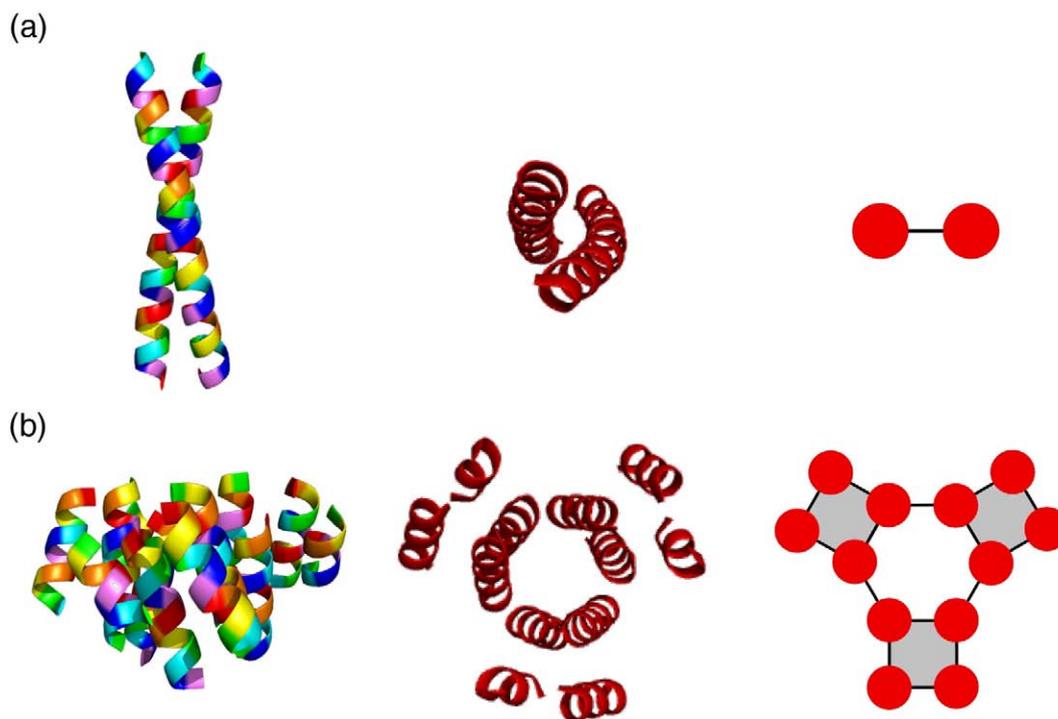


Fig. 1. Graphical representations of coiled-coil structures. (a) A classical coiled coil: the parallel and dimeric leucine zipper 1fos.¹⁸ (b) A complex coiled-coil assembly: a trimer of tetramers is found in the structure 1nig,¹⁹ which is a protein of unknown function from *T. acidophilum*. For both cases, and from left to right, the structures are first rendered as ribbons with rainbow colours of the heptad repeat using PyMOL;²⁰ second, down the coiled-coil axis; and, third, in a schematic where helices and KIH interactions are presented by red circles and by lines and gray shading, respectively.

nately spaced three and four residues apart and separated by polar (*P*) residues. This results in the pattern (*HPPHPPP*)_{*n* ≥ 3}, which is the classical heptad repeat and often denoted as *abcdefg*. Such motifs direct the folding of amphipathic helices, which assemble into bundles to bury their hydrophobic faces from solvent. Other patterns are compatible with the coiled coil—notably, various other short combinations of three- and four-residue spacing and hydrophobic amino acids^{21–23}—but these are less common.

These apparent simplicities in sequence and the relationship between sequence and structure are somewhat misleading, however. Coiled coils are structurally diverse: they can have different numbers of helices (different *architectures*, Fig. 1) that can be assembled in parallel, antiparallel and mixed orientations (i.e., different *topologies*). Thus, contemporary questions in coiled-coil research include those on what the limits on coiled-coil structures are and which sequence rules or other structural constraints distinguish the various structural alternatives and lead to faithful folding and assembly of the polypeptide chains in the cell. Here, we present a framework to address these issues by determining the current occupied coiled-coil structural space and analysing and classifying the structures that comprise it.

The periodic table of coiled-coil structures

The coiled coil is almost unique amongst protein-folding motifs because, as first proposed by Crick,²⁴

there is a clear link between protein sequence and structure. The aforementioned amphipathic helices of coiled coils are cemented by a run of so-called knob-into-hole (KIH) interactions (Fig. 2). In canonical, heptad-based coiled coils, residues at the *a* and *d* positions form *knobs* that interact with diamond-shaped *holes* formed by four residues on a partnering helix (Fig. 2). For example, in parallel coiled coils, an *a* knob interacts with a hole formed by side chains at *d*₋₁, *g*₋₁, *a*, *d* on the partner helix. Similar direct relationships can be written for *d* knobs and for antiparallel helix arrangements. Previously, we introduced SOCKET, a program that identifies KIH interactions in protein structures and hence classical coiled-coil arrangements. In SOCKET, a side chain is identified as a knob if it contacts four or more side-chain centres of mass in the partnering helix within a specified packing cutoff. The four nearest side chains define the corresponding hole.²⁵

SOCKET was used to identify coiled-coil positive structures in the November 2006 release of the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB).²⁶ [Search criteria are given in the legend to Fig. 3.] Interestingly, through visual inspection of all 766 coiled coil-containing protein structures, it became clear that the structures could be arrayed based on the number of helices in the coiled-coil motifs and simple relationships between them (Fig. 3). The resulting table resembles the periodic table of elements and Taylor's periodic table of protein structures;¹² hence, we refer to it as the “periodic table of coiled coils.” The table highlights a

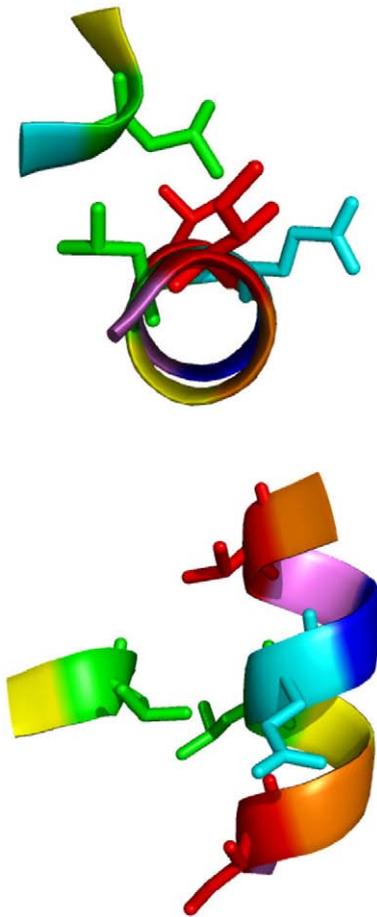


Fig. 2. Orthogonal views of a KIH interaction. These are the interactions that SOCKET identifies in classical coiled coils.²⁵ Rainbow colours from red to violet correspond to positions of the heptad repeat *abcdefg*, which are also assigned by SOCKET.

number of interesting features in these motifs and how the classes (equivalent to elements) are related at least at a structural level.

The table was populated using the following system of definitions:

1. A structure with one coiled coil and consequently one hydrophobic core is classified as *classical*. Classical coiled coils consisting of more than two helices are defined by having of a contiguous ring (or daisy chain) of KIH interactions between successive helices²⁵ and a single, consolidated central hydrophobic core. Classical coiled coils with two, three, four or five helices form the first row of the table and head each of the columns.
2. Protein structures that contain two or more classical coiled coils that are connected in some way are termed *complex*. Complex coiled-coil assemblies are separated into classes according to the highest-order participating classical coiled coil upon which they are based. These classes are arranged vertically according to the

total number of the highest-order classical coiled coil and the total number of participating helices they contain.

Any new coiled coil-based structure can be classified following the abovementioned rules.

Coiled-coil structure in detail

At this point, a walk-through of the table and a description of some of its members are useful.

At present, classical coiled coils are restricted to dimers, trimers, tetramers and pentamers and account for 92.8% of the structures identified from the PDB. There is one example of a six-stranded coiled coil in cobalamin adenosyltransferase, 2nt8;³⁰ however, this has peripheral coiled-coil interactions that make it a complex coiled-coil assembly, placing it farther down the hierarchy. The presence of this assembly implies the potential existence of a classical six-helix coiled coil, however. The members of each class can be grouped further according to their occurrence in the same or different protein chains and their topology. For example, the class of “Comprising 2 helices” (751 assignments) consists of “two-stranded” (541 assignments) and “dimeric” (210 assignments) structures depending on whether the helices belong to the same and different chains, respectively. These structures can also be divided further into parallel (184 assignments) and antiparallel (567 assignments in total) examples based on the orientation of the helices and into canonical and non-canonical categories based on their sequences being purely heptad based and not, respectively. For example, there are 69 two-stranded, canonical and parallel assignments and 6 dimeric, non-canonical and antiparallel assignments.

In the first column of the table, all the structures are based on classical two-helix structures. For instance, on the second row are what we term three-helix α -sheets in which the helices are connected via 2 two-helix interfaces; in effect, the central helix is shared between 2 two-helix coiled coils (Fig. 3). This is the most populated complex coiled-coil class. It contains 35 validated assemblies in 32 protein structures. Most of these are transferases, lyases or structures derived from structural genomics projects. The lengths of the shared helix in the middle of the α -sheet vary from 13 to 42 amino acids. However, in general, the more complex the assembly, the shorter the average helix length is (10 amino acids in the “Conjoined 6-stranded and 2-stranded assembly”). In the more-populated classical coiled coils, such as the two- and three-helix structures of the first row (Fig. 3), there are a variety of protein families and functions and helices are usually longer than those in the complex assemblies; in these cases, the lengths of the participating helices range from 11 to 122 and from 8 to 56 residues, respectively.

Sheets can of course close to form rings. This occurs in two examples in the table. The first is a

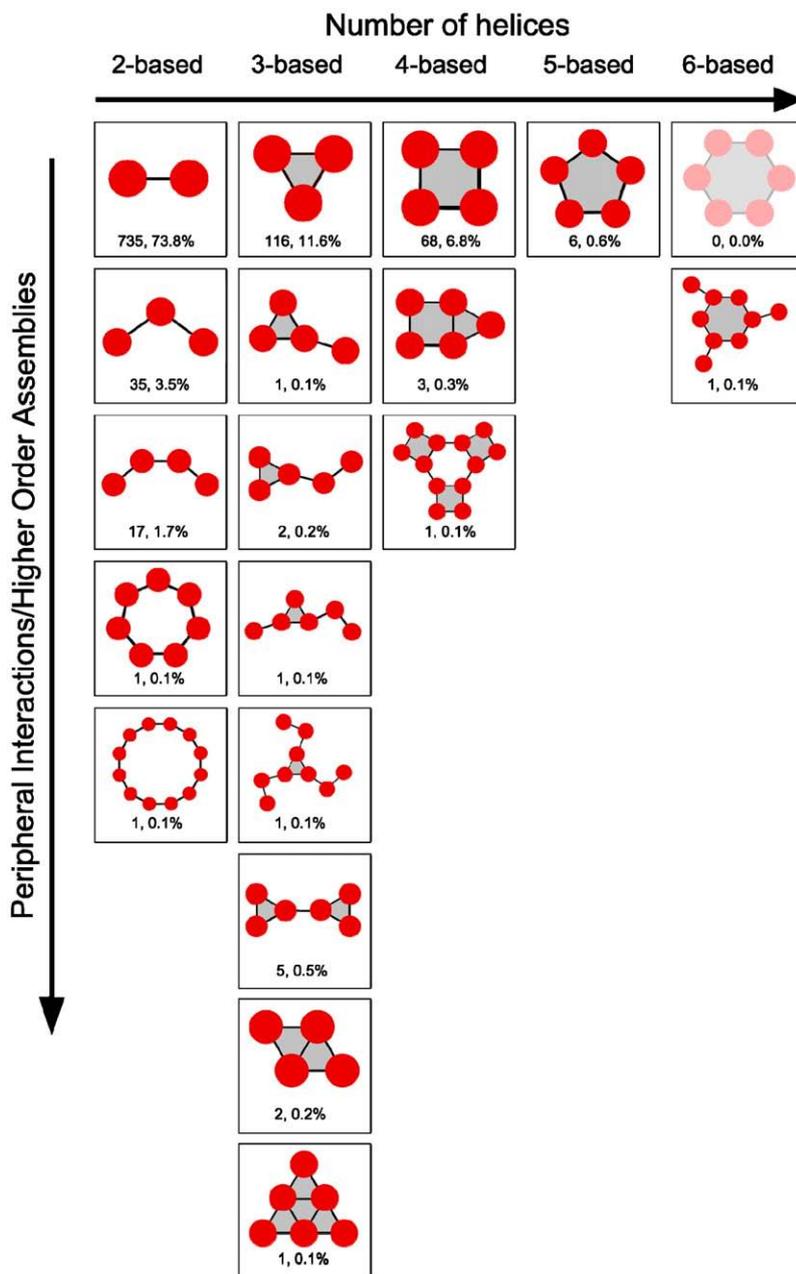


Fig. 3. Periodic table of coiled coils. The table is presented using cartoon diagrams of the architectures of the representative classes. Circles represent helices; lines, KIH interfaces; and gray regions, consolidated hydrophobic cores of classical coiled coils. The population and percentage of occupancy by each class are given. Structural validation of coiled coil-containing structures was implemented using SOCKET,²⁵ with increasing packing cutoffs from the default 7 Å to 7.5 Å, retrieving information from an in-house coiled-coil MySQL database²⁹ and visualising the structures using PyMOL²⁰ and Swiss-PDBViewer.²⁷ This process was repeated manually for each coiled-coil entry in our set. Where the biological unit was not specified by the PDB coordinate file,²⁶ the corresponding file from the PQS²⁸ server was used preferentially. The final data set included only coiled-coil segments that have less than 50% sequence identity. This set was generated by implementing an in-house sequence comparison algorithm described elsewhere.²⁹ Where two identical coiled coils were found within different PDB structures, the structure with an alphanumerically preceding PDB identifier was chosen. Perl scripts were used for parsing information and automatic generation of output directories, files and images.

heptameric engineered form of the GCN4 leucine zipper (2hy6),³¹ which in its wild-type form is a classical dimer (Fig. 1a). SOCKET reveals that the helices of the heptamer are linked by a series of back-to-back dimer interfaces rather than a daisy chain of KIH interactions that would constitute a classical coiled-coil heptamer. The second closed α -sheet is in the central domain of the periplasm-spanning pore of TolC (1tqq).³² In this case, 2-helix coiled coils assemble in a back-to-back fashion and close to form a ring of 12 helices. For this particular case, however, the packing cutoff used in SOCKET had to be increased to 7.7 Å to capture the assembly, which placed it outside the boundaries we would normally consider for a coiled-coil interaction. Analysis of the structure 1ek9,³³ a closed state of TolC, showed that

the packing of the helices in this region was not continuous.

We identified some coiled-coil architectures that, to our knowledge and from the available literature, have not been assigned as such before. Invariably, these were new classes of complex coiled-coil assemblies (Fig. 4). Although sparsely populated, they further highlight the structural space currently occupied by coiled coils. In the second row of the table, with increasing complexity, there is the “Conjoined 2-stranded and 3-stranded coiled-coil assembly” class with an example in 2chn,³⁴ a β -glucosaminidase, followed by the “Joined 2-stranded and 3-stranded coiled-coil assembly” class with an assignment in 1qdb,³⁵ a cytochrome *c* nitrite reductase. Lower in the same column, there is the

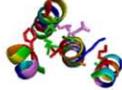
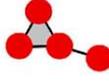
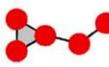
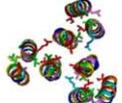
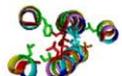
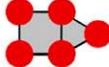
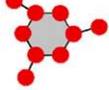
7-helix coiled coil			
1:2hy6	↑↑↑↑↑↑↑↑		
Conjoined 2-stranded and 3-stranded coiled coil			
1:2chn	↑↓↓-↑		
Joined 2-stranded and 3-stranded coiled coil			
2:1qdb, 1zb1	↑↑↓-↓↓, ↑↓↓-↓↑		
3-stranded partitioned alpha sheet			
1:2hr2	↑-↓↑↓-↑↓		
Trimer of 3-helix alpha sheets			
1:2e2a	↑↑↑(-↑↓) ₃		
Conjoined 3-stranded and 4-stranded coiled coil			
2:1hy1, 1tju	↑↑-(↑↓)-↓		
Trimer of 4-stranded coiled coils			
1:1nig	(↑↓↑↓) ₃		
Conjoined 6-stranded and 2-stranded coiled coil			
1:2nt8	↓↑↓↑↑↑(-↓) ₃		

Fig. 4. Newly observed coiled-coil structures. For each new class, the number of entries, the PDB code, a simplified topology diagram, a picture with orthogonal view of the assembly and a PT entry diagram are given.

“3-stranded partitioned α -sheet” class with an example in 2hr2, which is a hypothetical tetratricopeptide repeat-like protein. The last newly assigned assembly of the 2-helix-based column is the “Trimer of 3-helix α -sheets,” which occurs in 2e2a, a lactose-specific enzyme. In the third column of the table, two newly assigned coiled-coil assemblies are the “Conjoined 3-stranded and 4-stranded coiled coils” in 1hy1³⁶ and 1tju³⁷ and the “Trimer of 4-stranded coiled coils,” referred by the original authors as a helical super-bundle, in 1nig,¹⁹ a hypothetical protein from *Thermoplasma acidophilum*. The last newly assigned complex is on the sixth row of the table and is a “Conjoined 6-stranded and 2-stranded coiled coil” in 2nt8.³⁰ These and the other complex

coiled-coil assemblies contribute 7.2% of all the entries of the periodic table. Structural, geometric and energetic constraints in forming high-order complex structures are potential reasons for the underrepresentation of these structural assemblies in coiled-coil space. This is consistent with considering the small number of existing protein structural domains in comparison with the total number of possible arrangements of structural elements.³⁸⁻⁴²

Potentially, the number of protein structures in each element of the table could be expanded by searching the PDB and the non-redundant sequence database for proteins with sequence similarity using BLAST.⁴³ On this basis, the complex assembly, headed by the structure 2nt8,³⁰ also includes struc-

tures 3ci1,⁴⁴ 2r6t and 2r6x⁴⁵ and 2ah6, 1rty and 2idx⁴⁶ from the PDB with E values $<3e^{-19}$, as well as ~60 entries from the non-redundant sequence database with the same E value threshold. However, caution is needed for homologues with slightly different coordinate files, and we cannot check the reliability of entries without solved structures.

To our knowledge, the classification described herein is the first systematic analysis of different coiled coil-based architectures. The validation procedure enabled us to find new high-order assemblies and to push the limits of known coiled-coil architectures, providing insight on the potential structural space occupied by these structures. We have derived rules classifying coiled coils into a periodic table providing a simple method for processing the new architectures that will be observed as the PDB grows. We hope that the periodic table of coiled coils will be a valuable resource in the fields of protein design, engineering and structural prediction.

The Web interface of the periodic table of coiled-coil structures

We have collated the validated coiled-coil structures and presented the logical hierarchical groups of the periodic table in a series of pages on the Web† with supporting images, annotations, visualisation scripts, information on potential ligands in the structures, information on the sequence and the register and sequence profiles for certain sufficiently populated classical coiled-coil classes. These provide a navigable graphical interface to the periodic table. This sits alongside a dynamic interface for searching coiled-coil assignments‡. The dynamic interface is the Web interface of a relational database of all coiled-coil assignments automatically produced and updated monthly through SOCKET²⁵ searches of the PDB.²⁶ The dynamic interface allows queries of the coiled-coil assignments at different levels of complexity and in detail but only identifies classical coiled coils, although these can be components of complex coiled-coil assemblies of the type described in this article and organised in the periodic table. Thus, the periodic table is distinct from the dynamic interface in that it provides a manually validated and organised description of a subset of all the available coiled-coil structures. The manual validation and organisation procedures were very time-consuming. Therefore, we anticipate updating the periodic table less regularly than the dynamic interface. Together, the periodic table and dynamic interface form a comprehensive resource for visualising, relating and exploring the known coiled-coil structures, which we call CC+, and, to our knowledge, are unprecedented.

† http://coiledcoils.chm.bris.ac.uk/ccplus/search/periodic_table

‡ http://coiledcoils.chm.bris.ac.uk/ccplus/search/dynamic_interface

Acknowledgements

This study was funded by the Biotechnology and Biological Sciences Research Council (BB/D003016/1). We thank Oliver Testa for assistance in the Web design of the periodic table and useful discussions. We also thank Andrei Lupas, Gail Bartlett and members of the Woolfson group for valuable discussions.

References

1. Chothia, C. (1992). One thousand families for the molecular biologist. *Nature*, **357**, 543–544.
2. Blundell, T. L. & Johnson, M. S. (1993). Catching a common fold. *Protein Sci.* **2**, 877–883.
3. Alexandrov, N. N. & Go, N. (1994). Biological meaning, statistical significance, and classification of local spatial similarities in nonhomologous proteins. *Protein Sci.* **3**, 866–875.
4. Chothia, C. (1984). Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **53**, 537–572.
5. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
6. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
7. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M. *et al.* (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40.
8. Murzin, A. G. & Finkelstein, A. V. (1988). General architecture of the [alpha]-helical globule. *J. Mol. Biol.* **204**, 749–769.
9. Chirgadze, Y. N. (1987). Deduction and systematic classification of spatial motifs of the antiparallel structure in globular proteins. *Acta Crystallogr., Sect. A*, **43**, 405–417.
10. Hutchinson, E. G. & Thornton, J. M. (1993). The Greek key motif—extraction, classification and analysis. *Protein Eng.* **6**, 233–245.
11. Woolfson, D. N., Evans, P. A., Hutchinson, E. G. & Thornton, J. M. (1993). Topological and stereochemical restrictions in beta-sandwich protein structures. *Protein Eng.* **6**, 461–470.
12. Taylor, W. R. (2002). A ‘periodic table’ for protein structures. *Nature*, **416**, 657–660.
13. Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.
14. Wolf, E., Kim, P. S. & Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6**, 1179–1189.
15. Mewes, H. W., Frishman, D., Gruber, C., Geier, B., Haase, D., Kaps, A. *et al.* (2000). MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40.
16. Walshaw, J. & Woolfson, D. N. (2003). Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *J. Struct. Biol.* **144**, 349–361.
17. Lupas, A. N. & Gruber, M. (2005). The structure of alpha-helical coiled coils. *Adv. Prot. Chem.* **70**, 37–78.
18. Glover, J. N. M. & Harrison, S. C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature*, **373**, 257–261.

19. Sanishvili, R., Pennycooke, M., Gu, J., Xu, X., Joachimiak, A., Edwards, A. M. & Christendat, D. (2004). Crystal structure of the hypothetical protein TA1238 from *Thermoplasma acidophilum*: a new type of helical super-bundle. *J. Struct. Funct. Genomics*, **5**, 231–240.
20. DeLano, W. L. (2002). DeLano Scientific, Palo Alto, CA. <http://www.pymol.org>.
21. Brown, J. H., Cohen, C. & Parry, D. A. (1996). Heptad breaks in alpha-helical coiled coils: stutters and stammers. *Proteins*, **26**, 134–145.
22. Hicks, M. R., Holberton, D. V., Kowalczyk, C. & Woolfson, D. N. (1997). Coiled-coil assembly by peptides with non-heptad sequence motifs. *Folding Des.* **2**, 149–158.
23. Gruber, M. & Lupas, A. N. (2003). Historical review: another 50th anniversary—new periodicities in coiled coils. *Trends Biochem. Sci.* **28**, 679–685.
24. Crick, F. H. C. (1953). The packing of alpha-helices—simple coiled-coils. *Acta Crystallogr., Sect. A*, **6**, 689–697.
25. Walshaw, J. & Woolfson, D. N. (2001). SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.* **307**, 1427–1450.
26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
27. Nicolas Guex, M. C. P. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
28. Henrick, K. & Thornton, J. M. (1998). PQS: a protein quaternary structure file server. *Trends Biochem. Sci.* **23**, 358–361.
29. Testa, O. D., Moutevelis, E. & Woolfson, D. N. (2008). CC⁺: a relational database of coiled-coil structures. *Nucleic Acids Res.* doi:10.1093/nar/gkn675.
30. Maurice, M. S., Mera, P. E., Taranto, M. P., Sesma, F., Escalante-Semerena, J. C. & Rayment, I. (2007). Structural characterization of the active site of the PduO-type ATP:Co(I)rrinoid adenosyltransferase from *Lactobacillus reuteri*. *J. Biol. Chem.* **282**, 2596–2605.
31. Liu, J., Zheng, Q., Deng, Y., Cheng, C.-S., Kallenbach, N. R. & Lu, M. (2006). A seven-helix coiled coil. *Proc. Natl Acad. Sci. USA*, **103**, 15457–15462.
32. Higgins, M. K., Eswaran, J., Edwards, P., Schertler, G. F. X., Hughes, C. & Koronakis, V. (2004). Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein TolC. *J. Mol. Biol.* **342**, 697–702.
33. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*, **405**, 914–919.
34. Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J. *et al.* (2006). Structure and mechanism of a bacterial [beta]-glucosaminidase having O-GlcNAcase activity. *Nat. Struct. Mol. Biol.* **13**, 365–371.
35. Einsle, O., Messerschmidt, A., Stach, P., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Kroneck, P. M. H. (1999). Structure of cytochrome *c* nitrite reductase. *Nature*, **400**, 476–480.
36. Sampaleanu, L. M., Vallee, F., Slingsby, C. & Howell, P. L. (2001). Structural studies of duck delta 1 and delta 2 crystallin suggest conformational changes occur during catalysis. *Biochemistry*, **40**, 2732–2742.
37. Sampaleanu, L. M., Coddling, P. W., Lobsanov, Y. D., Tsai, M., Smith, G. D., Horvatin, C. & Howell, P. L. (2004). Structural studies of duck delta2 crystallin mutants provide insight into the role of Thr161 and the 280s loop in catalysis. *Biochem. J.* **384**, 437–447.
38. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.
39. Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations: II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.
40. Hoang, T. X., Trovato, A., Seno, F., Banavar, J. R. & Maritan, A. (2004). Geometry and symmetry presculpt the free-energy landscape of proteins. *Proc. Natl Acad. Sci. USA*, **101**, 7960–7964.
41. Söding, J. & Lupas, A. N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. *BioEssays*, **25**, 837–846.
42. Fitzkee, N. C. & Rose, G. D. (2004). Steric restrictions in protein folding: an α -helix cannot be followed by a contiguous β -strand. *Protein Sci.* **13**, 633–639.
43. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
44. St. Maurice, M., Mera, P., Park, K., Brunold, T. C., Escalante-Semerena, J. C. & Rayment, I. (2008). Structural characterization of a human-type corrinoid adenosyltransferase confirms that coenzyme B12 is synthesized through a four-coordinate intermediate. *Biochemistry*, **47**, 5755–5766.
45. Mera, P. E., St. Maurice, M., Rayment, I. & Escalante-Semerena, J. C. (2007). Structural and functional analyses of the human-type corrinoid adenosyltransferase (PduO) from *Lactobacillus reuteri*. *Biochemistry*, **46**, 13829–13836.
46. Schubert, H. L. & Hill, C. P. (2006). Structure of ATP-bound human ATP:cobalamin adenosyltransferase. *Biochemistry*, **45**, 15188–15196.