This classification of structure has not explicitly considered topology but has concentrated on the prerequisite of defining the secondary and tertiary links. From this basis, the previously difficult issue of topology becomes almost trivial, because two proteins matching the same ideal Form will either have the same connection of their secondary structures or will not. (See Supplementary Information A.2 for nomenclature details). Topology descriptions were written for each protein in each grid-cell and sorted for uniqueness, thus giving the number of folds in each cell. This value was then plotted along with the number of unique secondary structure strings for the three-layer Forms (Fig. 5). This simple analysis shows that for the smallest Forms, all linear arrangements of secondary structure have been observed and there are typically twice as many folds as secondary structure arrangements. With over ten secondary structures (around 35 in the ranked Forms in Fig. 5), this balance changes, and almost every secondary structure arrangement corresponds to a unique fold. This suggests that with these larger $\beta\alpha$ proteins, sequence-structure matching (threading) methods¹⁹ need only concentrate on the correct prediction of secondary structure to find an unknown fold and not on their three-dimensional interaction.

If a structure matches, say, a 2-5-3 (α - β - α) Form then it will also match any substructure at least as well (2-5-2, 1-5-3, 2-4-2...). and two proteins which have a different topology when matched against the 2-5-3 Form might have the same core 2-5-2 topology. The question of whether two proteins have the same fold now becomes relative and should be posed thus: what is the largest ideal Form under which two proteins have the same fold? Although the largest Form with common topology will be of greatest interest, it is also informative to view all the subsidiary matches in the format of the grids used in Fig. 4. Such a representation allows not only pairs, but also whole groups of proteins to be analysed and it is simple to determine both visually and automatically if they share a common core. Each step through the grid from one protein to another (via subsidiary matches) represents a deletion or addition of a secondary structure element. This allows a path to be defined between any two structures in the same grid (no matter how dissimilar) and a minimum path length to be computed.

Besides the all- α class of protein, the only structures not 'captured' by the set of three basic Forms used above were β proteins that contain internal repetition. These included not only the series of propeller proteins but also those with a triangular arrangement of structure, including β -trefoils, β -prisms and β -helices²⁰. Clearly triangles do not map well onto layers (or cylinders) and the only solution for these may be to generate a 'triangular' version of Fig. 2, or else to treat them as exceptions because of their detectable internal sequence repeats. For the moment, the latter route will be followed but a series of propeller Forms will be added to the basis set, along with the neglected C3₁ Form from Fig. 2 (which includes the β -grasp structures).

Thus, by breaking down structures into their basic Forms I have opened up a new approach to the classification and analysis of protein structure that is both flexible and automatic. Unlike clustering methods, it does not require the comparison of all structures to each other and can therefore be incrementally updated as new structures are determined. The major result from this approach is that it allows the difficult problem of protein topology to be rigorously addressed. The fold of any individual protein can be specified as that of the largest matching Form, and for pairs or groups of proteins, the fold is that of the largest common Form. This implies that the topology of a protein can only be defined under the specification of a given ideal Form which, in turn, means that proteins do not have a unique intrinsic topology or a unique position in any classification that is based on topology. □

Received 12 October 2001; accepted 28 January 2002.

1. Chothia, C. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).

- Orengo, C. A., Jones, D. T. & Thornton, J. Protein superfamilies and domain superfolds. *Nature* **372**, 631–634 (1994).
- Eidhammer, L., Jonassen, I. & Taylor, W. R. Structure comparison and structure patterns. *J. Comput. Biol.* **7**, 658–716 (2000).
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
- Orengo, C. A. *et al.* CATH—a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1108 (1997).
- Mizuguchi, K., Deane, C. M., Blundell, T. L. & Overington, J. P. HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci.* **7**, 2469–2471 (1998).
- Holm, L. & Sander, C. DALI/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* **25**, 231–234 (1997).
- Hadley, C. & Jones, D. T. A systematic comparison of protein structure classifications SCOP, CATH and FSSP. *Structure* **7**, 1099–1112 (1995).
- Chothia, C. & Finkelstein, A. V. The classification and origins of protein folding patterns. *Ann. Rev. Biochem.* **59**, 1007–1039 (1990).
- Finkelstein, A. V. & Ptitsyn, O. B. Why do globular proteins fit the limited set of folding patterns? *Prog. Biophys. Mol. Biol.* **50**, 171–190 (1987).
- Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. Analysis and prediction of protein β -sheet structures by a combinatorial approach. *Nature* **285**, 378–382 (1980).
- Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. Analysis and prediction of the packing of α -helices against a β -sheet in the tertiary structure of globular proteins. *J. Mol. Biol.* **156**, 821–862 (1982).
- Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of β -sheet barrels in proteins: I. A theoretical analysis. *J. Mol. Biol.* **236**, 1396–1381 (1994).
- Taylor, W. R. Defining linear segments in protein structure. *J. Mol. Biol.* **310**, 1135–1150 (2001).
- Taylor, W. R. Searching for the ideal forms of proteins. *Biochem. Soc. Trans.* **28**, 264–269 (2000).
- Murzin, A. G. & Finkelstein, A. V. General architecture of the α -helical globule. *J. Mol. Biol.* **204**, 749–769 (1988).
- Taylor, W. R., Jones, D. T. & Green, N. M. A method for α -helical integral membrane protein fold prediction. *Protein Struct. Funct. Genet.* **18**, 281–294 (1994).
- Taylor, W. R. Protein structure domain identification. *Protein Eng.* **12**, 203–216 (1999).
- Jones, D. T., Taylor, W. R. & Thornton, J. M. A new approach to protein fold recognition. *Nature* **358**, 86–89 (1992).
- Chothia, C. & Murzin, A. G. New folds for all- β proteins. *Structure* **1**, 217–222 (1993).
- Taylor, W. R. Protein structure comparison using bipartite graph matching and its application to protein structure classification. *Mol. Cell. Proteom.* advance online publication 4 March 2002 (DOI 10.1074/mcp.T200001-MCP200).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: wtaylor@nimr.mrc.ac.uk).

correction

Transition of Mount Etna lavas from a mantle-plume to an island-arc magmatic source

Pierre Schiano, Roberto Clocchiatti, Luisa Ottolini & Tiziana Busà

Nature **412**, 900–904 (2001).

We acknowledge that the source of the data for major-element composition of sample 27.2 in the Supplementary Information should have been cited as the work of Kamenetsky and Clocchiatti (ref. 1). We also acknowledge that the discussion of our model for the origin of Mount Etna magmatism would have benefited from referencing the model of multi-stage melting contemporaneous with metasomatic enrichment presented in ref. 1. □

- Kamenetsky, V. & Clocchiatti, R. Primitive magmatism of Mt. Etna: insights from mineralogy and melt inclusions. *Earth Planet. Sci. Lett.* **142**, 553–572 (1996).