

A tangled problem

William R. Taylor and Kuang Lin

From our experiences with shoelaces, we tend to think that tying a knot requires both dexterity and intent. As proteins lack both of these qualities, it was assumed for many years that there would be few, if any, knots found in protein chains. But perhaps shoelaces are not a good model for proteins. Before the Christmas decorations are packed away, it is possible to find a slightly better model in the strings of beads that are sometimes draped over the festive branches.

To conduct an unbiased experiment, take a metre length of these (around 150 beads, which represents a small protein), cover the end beads in Blu-Tack (adhesive putty) and 'pour' the beads from hand to hand. When the two end beads eventually stick together, check for knots. In an unscientifically small sample of 20 tries, one-quarter of our 'proteins' were knotted. From this result, one might now wonder why more proteins do not contain knots; but, until relatively recently, there were almost no protein chains that contained a knot that would not be laughed at by any qualified boy scout. Most of these had one end poking through a loop by only a few 'beads' (amino-acid residues).

Given that we now know the structures of around 2,000 different proteins, why are there so few knots? A likely explanation is that proteins are not free-flowing strings of beads, but rather are 'sticky'. (Dedicated experimentalists should now repeat the knot experiment, using beads coated in honey.) Interactions within the chain are therefore predominantly local, and few open loops will be formed for the protein's termini to pass through. However, a few years ago a knot in quite a different league was noticed in a protein; this knot had over 200 amino-acid residues on one side and 70 on the other. More recently, another respectable knot has appeared with 30 residues on its shorter side.

Unlike beads, real proteins do not (normally) have their termini joined. This presents a technical problem, as knots are only properly (mathematically) defined in circular strings. However, one common definition of a knot is "a loop in a string that tightens when pulled". This can be applied to a protein by repeatedly smoothing (averaging consecutive triples of points along the chain) while keeping the two termini fixed in place and seeing if a straight line is obtained. Using this method, the ends can also be progressively trimmed to find the exact location of the knotted core. Protein knots can then be quantified by the number of residues on either side of them, and a useful distinction can be made between 'deep' knots and 'shallow' ones, with the latter having less than 20 residues on their shorter side.

The 'true' or 'topological' knots considered above are defined by the path of the backbone chain alone. But there is another, more common, knot found in proteins that is created by crosslinks between parts of the chain. As the bonds involved are covalent (usually disulphide links), these knots are best referred to as 'covalent knots'. Unlike topological knots, there is no mystery about how covalent ones might form: they simply require amino acids with sulphide groups (cysteine residues) to lie close enough to become crosslinked either during or after folding. Like disulphide bonds in general, their function may be to give extra stability to the fold. Might topological knots have a similar function, or are they just 'harmless' tangles that have arisen accidentally? Intriguingly, both of the known examples of deep knots occur in the catalytic domains of their proteins, with one even running through the active site. It is difficult to imagine anything in the structure of a knot that could not just as easily be constructed by an unknotted piece of protein chain. Any advantage from their presence must therefore derive from an

Protein knots

Do knots give extra stability to protein folds or are they just 'harmless' tangles that have arisen accidentally?

indirect or entropic (ordering) effect, such as reducing thermal motion in a knotted active site or allowing large motions only in a restricted segment of chain.

Another twist to the protein-knot story is provided by the structure of a SET-domain protein. The 'knotted' region in this protein fold is neither a topological nor a covalent knot. Instead, it consists of a loop held by hydrogen bonds that 'traps' the carboxy-terminal part of the chain. Perhaps our concept of a protein knot should be broadened to include these 'pseudo-knots' that are formed by hydrogen bonding as well as covalent links. The definition of a protein knot then becomes a matter of energetics: how many hydrogen bonds are needed? Or perhaps we need a continuum of crosslinks from covalent bonds, through hydrogen bonds, to van der Waals forces?

With the deluge of data anticipated from the various structural-genomics programmes currently being undertaken, we may soon have a large enough collection of folds to get a better idea of how frequent and important all forms of knots are to protein structure and function. If some of the ideas discussed above are correct, most will be of the shallow kind, and the few deep knots may require unusual folding mechanisms to account for them. The analysis of these should provide useful insight into how proteins fold. If knots are selected for some advantageous reason, then it might be expected that they will provide greater advantages in thermophilic bacteria. As the structures of proteins from both mesophilic and thermophilic organisms are being determined (the latest knot is from a thermophile), their comparison should provide a natural test-bed for this idea. ■

William R. Taylor and Kuang Lin are in the Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

FURTHER READING

Protein Data Bank <http://www.rcsb.org/pdb/> (PDB accession numbers for knotted proteins: 1QMG, 1YVE, 1IPA, 1GOZ, 1K3R, 1MT6, 1MVH, 1H3I, 1ML9, 1MLY).
Taylor, W. R. *Nature* **406**, 916–919 (2000).
Taylor, W. R., May, A. C. W., Brown, N. P. & Aszodi, A. *Rep. Prog. Phys.* **64**, 517–590 (2001).
Taylor, W. R., Xiao, B., Gamblin, S. J. & Lin, K. *Comp. Biol. Chem.* (in the press).
Mansfield, M. L. *Nature Struct. Biol.* **1**, 213–214 (1994).



M. GEORGE