curve, but the variance of displacements for NK11 showed a broadened distribution that was best fit with a double gaussian curve (Fig. 4b). Ncd showed an average displacement of 6.6 ± 0.8 nm to the minus end (mean ± s.e.m., total = 98), whereas NK11 showed average displacements of 6.3 ± 1.2 nm to the plus end and 7.9 ± 1.1 nm to the minus end (total = 86). The rigidity of the linkage of the motors to the bead was calculated from the amplitude of the brownian noise and found to be stiff, 0.05–0.10 pN nm^{-1}, even at zero displacement (Fig. 4a, NK11). With a stiff linkage of the motor to the bead, any change in angle or movement of the motor will be amplified by the displacement of the bead, which acts as a lever arm. We therefore believe that we are measuring a conformational change within the motor, rather than a step of the motor along the microtubule.

x–*y* plots of the Ncd displacements showed that the movement was directional and biased towards the minus end and to the right (*x* displacement = 2.5 ± 0.6 nm at $y\rangle>0$ nm, mean ± s.e.m, n = 69), whereas NK11 movement was towards the plus or minus end and frequently to the left or right (Fig. 4c, d). Thus, Ncd attaches and/or makes a conformational change that is biased in one direction, whereas the mutant does so in either direction. Movement of Ncd towards the minus end and a protofilament to the right is consistent with the microtubule rotation caused by Ncd in gliding assays and its handedness[9]. In contrast, NK11 moves towards the plus or minus end and a protofilament to the left or right.

The residues mutated in NK11 and Sup11 are highly conserved among the C-terminal motor kinesins, which may all be minus-end motors[10]. The wild-type velocities of the mutant motor indicate that the neck is functional, but directionality is defective. We have shown here that single mutant motors can move on microtubules with either polarity of movement. This finding excludes differences in the motor–tubulin interface as the basis of Ncd 'reversed' directionality. Instead, a working stroke of the motor has a directional bias that displaces the motor towards the minus end. The working stroke of the NK11 neck mutant occurs in either direction, displacing the motor towards either the plus or minus end. The mutants indicate that the directional bias is dependent on neck/motor-core interactions. The directional bias of the wild-type motor could result from a levered movement that is dependent on interactions of the neck residue with the motor-core residue. Interactions that involve residues of the kinesin neck linker[11] or structural elements of myosin analogous to the Ncd neck (for example, the converter[12]) could explain the directional movement of kinesin to microtubule plus ends or the reversed movement of myosin VI to actin minus ends[13]. □

Methods

Plasmids

Plasmids for mutant motors were constructed from *pGEX/MC1* (ref. 14) using PCR to introduce mutations and were confirmed by DNA sequencing. Mutants contained glutathione *S*-transferase (GST) fused to the N terminus of Ncd residues 209–700. *pBioNcd* and *pBioNK11* were made by ligating a fragment from the *Propionibacterium shermanii* biotin-containing transcarboxylase gene to Ncd or NK11 DNA. The *pMW172*-based plasmids[15] encode a 120-residue peptide that is biotinylated in bacteria fused to the N terminus of Ncd or NK11 residues 293–700.

Protein expression

We prepared lysates of GST fusion proteins for microtubule gliding assays[16] and analysed them by immunoblotting with an anti-Ncd C-terminus antibody[17]. A band of the predicted relative molecular mass for the motor was detected in lysates of all three mutants and bands attributable to breakdown products were absent. BioNcd and BioNK11 proteins for laser-trap experiments were purified by S-Sepharose chromatography[16] followed by microtubule affinity purification. Reaction with Ncd antibody and streptavidin-alkaline phosphatase showed that the proteins were positive for both Ncd and the biotinylated tag. A control Ncd motor protein showed no staining with streptavidin-alkaline phosphatase.

Motility assays

Gliding assays of GST fusion proteins bound to coverslips were in 10 mM $NaPO_4$ pH 7.4, 1 mM EGTA, 1 mM $MgCl_2$ and 5 mM Mg-ATP using axoneme–microtubule complexes[18]. Laser-trap assays were carried out by diluting and attaching BioNcd or BioNK11 to beads[6] in buffer plus 200 mM NaCl, and by monitoring their binding and movement on fluorescent microtubules or asymmetrically labelled fluorescent axoneme–microtubules. Beads were trapped at a trap stiffness of 0.042 pN nm^{-1} and held next to a microtubule bound to the coverslip surface. Attenuated brownian motion of the bead indicated attachment of a motor on the bead to the microtubule. The standard deviation of brownian noise of the unbound beads was ± 9.9 nm.

Received 31 May; accepted 20 June 2000.

1. Endow, S. A. & Waligora, K. W. Determinants of kinesin motor polarity. *Science* **281,** 1200–1202 (1998).
2. Sablin, E. P. *et al.* Direction determination in the minus-end-directed kinesin motor ncd. *Nature* **395,** 813–816 (1998).
3. Kozielski, F., De Bonis, S., Burmeister, W. P., Cohen-Addad, C. & Wade, R. H. The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations. *Structure* **7,** 1407–1416 (1999).
4. Hoyt, M. A., He, L., Totis, L. & Saunders, W. S. Loss of function of *Saccharomyces cerevisiae* kinesin-related *CIN8* and *KIP1* is suppressed by *KAR3* motor domain mutations. *Genetics* **135,** 35–44 (1993).
5. Endow, S. A. Determinants of molecular motor directionality. *Nature Cell Biol.* **1,** 163–167 (1999). NB: See online version at http://www.nature.com/ncb/ for correct sequence alignment.
6. Iwatani, S., Iwane, A. H., Higuchi, H., Ishii, Y. & Yanagida, T. Mechanical and chemical properties of cysteine-modified kinesin molecules. *Biochemistry* **38,** 10318–10323 (1999).
7. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77,** 773–784 (1994).
8. Crevel, I. M. -T. C., Lockhart, A. & Cross, R. A. Weak and strong states of kinesin and ncd. *J. Mol. Biol.* **257,** 66–76 (1996).
9. Walker, R. A., Salmon, E. D. & Endow, S. A. The *Drosophila claret* segregation protein is a minus-end directed motor molecule. *Nature* **347,** 780–782 (1990).
10. Endow, S. A. in *Guidebook to the Cytoskeletal and Motor Proteins* (eds Kreis, T. & Vale, R.) 403–408 (Oxford Univ. Press, Oxford, 1999).
11. Kozielski, F. *et al.* The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell* **91,** 985–994 (1997).
12. Houdusse, A., Kalabokis, V. N., Himmel, D., Szent-Györgyi, A. G. & Cohen, C. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell* **14,** 459–470 (1999).
13. Wells, A. L. *et al.* Myosin VI is an actin-based motor that moves backwards. *Nature* **400,** 505–508 (1999).
14. Chandra, R., Salmon, E. D., Erickson, H. P., Lockhart, A. & Endow, S. A. Structural and functional domains of the *Drosophila* ncd microtubule motor protein. *J. Biol. Chem.* **268,** 9005–9013 (1993).
15. Way, M., Pope, P., Gooch, J., Hawkins, M. & Weeds, A. G. Identification of a region in segment 1 of gelsolin critical for actin binding. *EMBO J.* **9,** 4103–4109 (1990).
16. Chandra, R. & Endow, S. A. Expression of microtubule motor proteins in bacteria for characterization in *in vitro* motility assays. *Methods Cell Biol.* **39,** 115–127 (1993).
17. Hatsumi, M. & Endow, S. A. The *Drosophila* ncd microtubule motor protein is spindle-associated in meiotic and mitotic cells. *J. Cell Sci.* **103,** 1013–1020 (1992).
18. Song, H., Golovkin, M., Reddy, A. S. N. & Endow, S. A. *In vitro* motility of AtKCBP, a calmodulin-binding kinesin protein of *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **94,** 322–327 (1997).

Acknowledgements

We thank L. Yamaguchi and Y. Toyoshima for fluorescently labelled axoneme–microtubules, and J. Howard and W. Schief for valuable discussions. This work was supported by the NIH (S.A.E.), Japan Ministry of Education, Science, Sport & Culture (H.H.) and Human Frontier Science Program (S.A.E. and H.H.).

Correspondence and requests for materials should be addressed to S.A.E. (e-mail: endow@ duke.edu).

A deeply knotted protein structure and how it might fold

William R. Taylor

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The search for knots in protein has uncovered little that would cause Alexander the Great to reach for his sword. Excluding knots formed by post-translational crosslinking, the few proteins considered to be knotted form simple trefoil knots with one end of the chain extending through a loop by only a few residues[1,2], ten in the 'best' example[3]. A knot in an open chain (as distinct from a closed circle) is not rigorously defined and many weak protein knots disappear if the structure is viewed from a different angle. Here I describe a computer algorithm to detect knots in open chains that is not sensitive to viewpoint and that can define the region of the

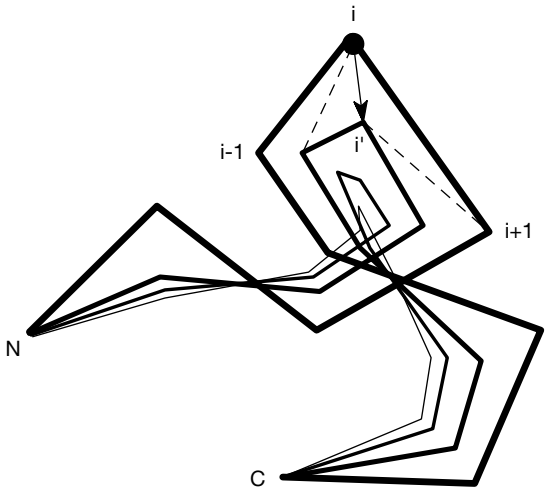

Figure 1 The basic chain smoothing algorithm. Protein chains are drawn as lines connecting the central carbon atom in the backbone of each residue unit running from the amino (N) terminus to the carboxy (C) terminus. Beginning at the second residue, for each residue point (i) in the starting conformation, the average coordinate of i, $i-1$ and $i+1$ was taken as the new position (i') for the residue. This procedure was then repeated, and the results of this are progressively smoother chains, shown as a series of fainter lines. Note that the termini do not move. With each move, it was checked that the chains did not pass through each other. This was implemented by checking that the triangles $\{i'-1, i, i'\}$ and $\{i, i', i+1\}$ (dashed lines in the Figure) did not intersect any line segment $\{j'-1, j'\}(j<i)$ before the move point or any line $\{j, j+1\}(j>i)$ following.

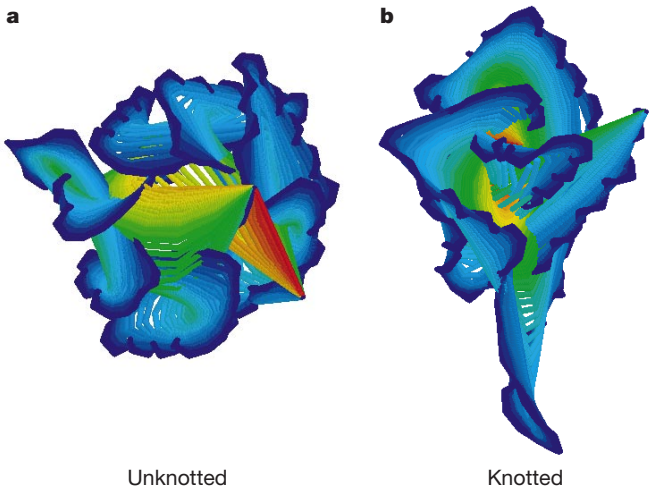

Figure 2 Smoothed protein structures. Applying the smoothing algorithm described in the text (and Fig. 1) to protein structures produces a series of increasingly smoothed chains, coloured from blue to red. (For clarity, the native starting structure is not shown). **a**, Applied to a protein that has no knots (triosephosphate isomerase, (1tph1)) results in a straight line (red) joining the termini. To reach this stage took 52 smoothing iterations. **b**, Applied to the knotted protein (the carboxy-terminal domain of acetohydroxy acid isomeroreductase, (1yvel)), a straight line is never attained and a small knot remains deep in the (red) core part of the protein. This is shown in Fig. 3b. Figure prepared using RasMol version 2.6 (ref. 9).

chain giving rise to the knot. It characterizes knots in proteins by the number of residues that must be removed from each end to abolish the knot. I applied this algorithm to the protein structure database and discovered a deep, figure-of-eight knot in the plant protein acetohydroxy acid isomeroreductase[4]. I propose a protein folding pathway that may explain how such a knot is formed.

Pulling the ends of a given piece of string will usually decide whether it is knotted or not. Because we hold the ends, the string and our body form a closed circle and there is no danger of untying the knot as it is pulled. The definition of knots in circular strings is mathematically well defined[5], and one way to approach the problem in open strings is simply to join the ends (as we do when we pick up a string). This is fine for clear knots (where the ends of the string are remote from the knot site), but if the ends are tangled-up together with the knot then any algorithm devised to 'pick-up' the ends creates the risk that the external connections might either untie an existing knot, or create a new one. Fortunately, the ends of protein chains (being charged) tend to lie on the surface of the structure and so can often be joined unambiguously by a wide loop. However, not all termini lie on the surface and if we want to locate the knotted region, any deletion from the ends could drive the termini deeper.

An alternative approach is to invert the problem: rather than extending the termini outwards, these can be left fixed and the rest of the protein made to shrink around them. This was done by contracting the protein chain as if it were a rubber band: specifically, with residues represented by the coordinates of their α-carbon, each point was repeatedly replaced by the average of itself and its two neighbours (Fig. 1). This procedure, which has been used previously to visualize protein chains, quickly reduces the protein to a smooth curve but leaves the termini untouched (as they do not have two neighbours). If continued indefinitely, all the points will lie on the line connecting the termini. When investigating topological features of the chain, however, it is most important that two parts of the chain cannot pass through each other and this was checked with each move of a residue point. Any move that violated this check was not implemented, leaving the current residue in its original position. With this condition, unknotted strings will still be reduced to a straight line, but those containing knots will become blocked (Fig. 2).

Although this simple algorithm is sufficient to detect knots, some pragmatic refinements were made. Because of a limit in the

Table 1 Knots found in proteins

Protein	PDB code	Length	Knot	Core	Depth
Acetohydroxy acid iso/red	1yvel	513	LRRL	245–444	17,220
Acetohydroxy acid iso/red C-domain	1yvel	234	LRRL	16–215	170
Carbonic anhydrase IV	1zncA	262	RRR	31–261	64
Ubiquitin YUH1-UBAL	1cmxA	214	LLL	5–210	30
VP3 core protein (bluetongue virus)	2btvB	885	LRRL	203–879	1,632
S-adenoylsmethionine synthetase	1fugA	383	RRR	10–276	1,070
Carbonic anhydrase	1kopA	223	RRR	39–223	40
Carbonic anhydrase I	1hcb	258	RRR	28–256	87
Carbonic anhydrase V	1dmxA	237	RRR	6–234	28

A selection of 3,440 protein structures was extracted from the current protein structure databank (PDB) in which no two structures shared more than 50% sequence identity. (see ref. 11 for selection protocol). A separate entry is included for the core of the C-terminal domain of 1yvel (excluding the N-terminal domain and the last 50 residues). Residue numbers are the sequential numbering of the PDB file and the size of the protein is tabulated as length. The knots found were characterized by the string formed by the handedness of their successive crossovers (L, left; R, right). Note that while the two trefoils are distinct even as circular knots, the others can be created by introducing different break-points in a circular knot. The core of the knot was determined by a series of deletions from each terminus to find the smallest region that remained knotted under application of the method, and the product of the number of residues that must be deleted from the ends to free the knot is tabulated under depth. Note that the knot type in the minimal core of 1yvel is not the same as in the full C-terminal domain (or full protein), as the deletion of most of the first α-helix in the domain results in the chain being 'pulled' tight from a different direction. The knot in these more complete structures is the figure-of-eight type (LLRR) depicted in Fig. 3b.

numerical accuracy of the computer calculations, each line between residues was represented by an impenetrable tube (1 Å diameter). This allowed residues to be progressively removed: specifically, the middle residue of three that were almost collinear or where the separation of the outer two fell within the tube diameter. This not only improved execution time but led to an even simpler test for knots as any chain that can be reduced to just its two termini is not knotted. Those with more than two residues remaining are either knots or tangles in which a group of moves have become 'grid-locked' (like rush-hour traffic at junction). This latter condition was eased (but not completely eliminated) by making a slight reduction in the tube diameter any time the chain became stuck.

In practice, most chains of a few hundred residues are reduced to their termini in around 50 iterations. If by 500 iterations a chain was still not reduced two points, then the resulting configuration was analysed in more detail. As the termini are now well separated from the knot-site, they can be unambiguously joined and analysed as a 'proper' circular knot. This can be done using one of the knot-invariant polynomials, such as the Alexander or Jones polynomials[1,5]. However, the few knots encountered in proteins are so simple that they do not require any sophisticated analysis. Furthermore, from a theoretical perspective, not only are protein knots directional but also they have a unique break-point (between the termini) which is not taken into consideration by any of the polynomial forms. As a working tool, a simpler method was adopted to characterize these open knots based on the Dowker knot notation[5] (see legend to Table 1).

Applying this method to a non-redundant selection of protein structures (see Table 1 for selection details) revealed a surprising number of knots. A few of these proved to be unresolved tangles (often forming slip-knots), and some others were caused by breaks in the chain creating an unnatural short-cut. The former were all eliminated by running the program with a smaller 'tube' diameter but the latter could only be removed through visual inspection. Of the seven remaining structures (Table 1), five were right-handed trefoils including related carbonic anhydrase structures (1zncA, 1kopA, 1hcb, 1dmxA) and the protein *S*-adenoylmethionine synthetase (1fugA) both of which had been identified previously. In addition, three new knots were found: a left-handed trefoil in ubiquitin (1cmxA), and two figure-of-eight knots in a viral core protein (2btvB) and acetohydroxy acid isomeroreductase (1yveI; Fig. 3b). These last two are of particular interest as they include an additional crossover above the trefoil and are therefore less likely to be formed by a wandering chain during folding. This was confirmed by simulation of random and semi-random compact protein-like chains in which the trefoil was by far the most common knot type (data not shown). The location of the two figure-of-eight knots was determined by a series of deletions from both termini of the protein chain. This revealed that the knot in 2btvB required just the last eight residues, which is similar to the deepest trefoil knot. By contrast, the knot in 1yveI, which is contained in the carboxy-terminal domain of the protein, remained until 70 residues were deleted from the carboxy terminus and 245 residues (including a complete domain) were removed from the amino terminus (Fig. 3a).

It is interesting to speculate how a protein with such a deep and complex knot might fold—as it is difficult to imagine over 50 residues being 'fed' through a loop in a reproducible way during folding. Clues to the folding of this protein can be found in a clear internal duplication within the domain, comprising 88 residue pairs with 2.0 r.m.s. deviation (as measured by the program SAP (ref. 6) over the α-carbon positions). If it is assumed that the two most deeply buried, symmetrically equivalent helices initially pack together (A1 and B1, Fig. 4), then the remaining parts of each repeat (A2, A3 and B2, B3) can wrap around this core requiring only that the C-terminal segment can pass through the large loop between the repeats before this contacts and finally packs onto the core (Fig. 4). The symmetry in this arrangement suggests that the protein might have evolved from an exchange of structure or 'swap'[7] between the two duplicated domains in which the first

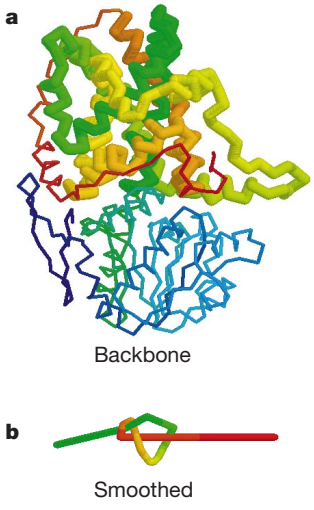

Figure 3 The knot in 1yvel. **a**, Backbone representation of the complete native protein structure (coloured from blue to red in the direction of the chain) with the core of the knotted domain drawn thickened. This region is preceded by a complete nucleotide binding domain (blue-green) and followed by a long loop (red) that wraps around the domain. **b**, The knotted core in the smoothed representation of 1yvel (Fig. 2b), coloured as in **a**. The figure-of-eight knot can be seen clearly. This form was attained after 50 cycles, and if continued an irreducible core consisting of eight points was attained. Figure prepared using RasMol version 2.6 (ref. 9).

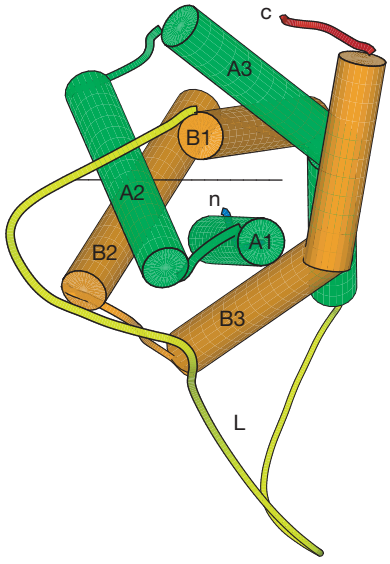

Figure 4 The core of the C-terminal domain in 1yvel. The duplicated portions of the chain (A1–A3 and B2–B3) are viewed down the packing axis of the two central α-helices. (The fine horizontal line between these is a least-squares fit to the pseudo-two-fold axis. The following parts of the duplicated regions (A2, A3 and B2, B3) wrap around this central pair creating the core of the knot. The final tying of the knot requires that the C-terminal part of the chain (including 45 residues following the point labelled c) passes through the long loop connecting the duplications (labelled L). The origin of this structure can be imagined as a gene duplication giving rise to a linked dimer in which the two central helices then exchanged positions. Figure prepared using MOLSCRIPT version 2.1.2 (ref. 10).

helix in the repeat has been transposed across the twofold axis of symmetry so creating the knot (see ref. 8 for a wider review of related processes). Intriguingly, the best example of a trefoil knot (in 1fugA) appears to have arisen in a similar manner, in which a β-strand on the edge of a sheet has been transferred from one duplicated domain to another. Although it cannot be stated that significant knots in proteins will not arise by other means, it appears that the swapping of elements of secondary structure between duplicated domains can provide a source of knotted proteins. □

Received 3 April; accepted 15 June 2000.

1. Mansfield, M. L. Are there knots in proteins? *Nature Struct. Biol.* **1,** 213–214 (1994).
2. Mansfield, M. L. Fit to be tied. *Nature Struct. Biol.* **4,** 116–"7 (1997).
3. Takusagawa, F. & Kamitori, K. A real knot in protein. *J. Am. Chem. Soc.* **118,** 8945–8946 (1996).
4. Biou, V. *et al.* The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog at 1.65 Å resolution. *EMBO J.* **16,** 3405–3415 (1997).
5. Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots* (Freeman, New York, 1994).
6. Taylor, W. R. Protein structure alignment using iterated double dynamic programming. *Protein Sci.* **8,** 654–665 (1999).
7. Bennet, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4,** 2455–2468 (1995).
8. Heringa, J. & Taylor, W. R. Three-dimensional domain duplication, swapping and stealing. *Curr. Opin. Struct. Biol.* **7,** 416–421 (1995).
9. Sayle, R. & Milner-White, E. J. RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.* **20,** 374 (1995).
10. Kraulis, P. J. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
11. Taylor, W. R. Multiple sequence threading: an analysis of alignment quality and stability. *J. Mol. Biol.* **269,** 902–943 (1997).

Correspondence and requests for materials should be addressed to W.R.T. (e-mail: wtaylor@nimr.mrc.ac.uk). The program used in this work can be found at ftp://mathbio.nimr.mrc.ac.uk.