Journal of Muscle Research and Cell Motility **19**, 877–886 (1998)

The case for a common ancestor: kinesin and myosin motor proteins and G proteins

F. JON KULL[1,2]*, RONALD D. VALE[1,2] and ROBERT J. FLETTERICK[2]

[1]*Howard Hughes Medical Institute and* [2]*Departments of Cellular/Molecular Pharmacology and Biochemistry/Biophysics, University of California, San Francisco, USA*

Received 26 March 1998; revised 18 May 1998; accepted 20 May 1998

Summary

Recent studies have shown surprising structural and functional similarities between the motor domains of kinesin and myosin. Common features have also been described for motor proteins and G proteins. Despite these similarities, the evolutionary relationships between these proteins, even among the motor proteins, has not been obvious, since the topological connectivities of the core overlapping structural elements in these transducing proteins are not identical to one another. Using secondary structure topology, comparison of functional domains and active site chemistry as criteria for relatedness, we propose a set of rules for determining potential evolutionary relationships between proteins showing little or no sequence identity. These rules were used to explore the evolutionary relationship between kinesin and myosin, as well as between motor proteins and other phosphate-loop (P-loop) containing nucleotide-binding proteins. We demonstrate that kinesin and myosin show significant chemical conservations within and outside of the active site, and present an evolutionary scheme that produces their respective topologies from a hypothetical ancestral protein. We also show that, when compared with various other P-loop-containing proteins, the cytoskeletal motors are most similar to G proteins with respect to topology and active site chemistry. We conclude that kinesin and myosin, and possibly G proteins, are probably directly related via divergent evolution from a common core nucleotide-binding motif, and describe the likely topology of this ancestor. These proteins use similar chemical and physical mechanisms to both sense the state of the nucleotide bound in the active site, and then transmit these changes to protein partners. The different topologies can be accounted for by unique genetic insertions that add to the edge of a progenitor protein structure and do not disrupt the hydrophobic core.

Introduction

When comparing proteins that share common structural and/or functional characteristics, an obvious question to ask is whether the proteins are related through a common ancestor or, alternatively, if they have evolved along separate pathways to end up resembling each other. While the former case describes divergent evolution, where the effects of amino acid mutations have changed a common, ancestral protein into two or more distinct proteins, the latter case describes convergent evolution, where random mutations over long periods of time result in proteins that share functionality but achieve it via different pathways. One can imagine that the evolutionary classification of proteins can be a formidable task, especially if the proteins do not share common amino-acid sequences.

Nevertheless, many examples exist for both evolutionary mechanisms. Divergent evolution is common in proteins, resulting in protein families and superfamilies (Orengo *et al*., 1994). Although many proteins related by divergent evolution share both structural and functional characteristics (Hwang & Fletterick, 1986), divergent evolution over tens of millions of years can produce proteins that share a common structure but have virtually unrelated functions. Examples of this type include the groups of actin, hsp70 and hexokinase (Flaherty *et al*., 1990, 1991; Bork *et al*., 1992; Gupta & Golding, 1993), biotin carboxylase and ADP-forming peptide synthetases (Fan *et al*., 1995; Artymiuk *et al*., 1996), E-cadherin and immunoglobulins (Wagner, 1995), N-terminal nucleophile

*To whom correspondence should be addressed at:
Department of Biophysics,
Max-Planck Institute for Medical Research,
Jahnstrasse 29,
69120 Heidelberg,
Germany.
E-mail: kull@mpimf-heidelberg.mpg.de

0142–4319/98

hydrolases (Brannigan *et al*., 1995), and glycogen phosphorylase and β-glucosyltransferase (Holm & Sander, 1995; Artymiuk *et al*., 1995). Although it is more difficult to demonstrate convergent evolution convincingly, clear examples are found in proteases and hydrolases, where various enzymes (sulfhydryl, serine, metallo, aspartyl, etc.) have developed similar catalytic sites but are otherwise different in sequence, shape and architecture (Doolittle, 1994). Examples of other proteins in which similar functionality has evolved independently include superoxide dismutases, aminoacyl tRNA synthetases, and topoisomerases (Doolittle, 1994).

Although clear cases of divergent evolution exist, many proteins with similar chemistry or structure are in an evolutionary 'grey zone' in which it is difficult to predict reliably a direct evolutionary relationship. For such proteins it is necessary to establish a set of criteria for evaluating whether two proteins are related. We propose four criteria to evaluate evolutionary relationships in the 'grey zone':

(1) Spatial conservation of secondary structural elements.
(2) Spatial conservation of domains or regions that share a common function. This could include regions such as the active site, as well as domains that interact with other proteins or that change conformation.
(3) Conservation of active site residues that use the same chemistry as part of their catalytic mechanism. These residues should share a similar spatial location and be positioned via a conserved protein backbone scaffold (as opposed to the catalytic triads of subtilisin and chymotrypsin families of proteases).
(4) Conservation of topological connectivity. However, topological connectivity is not always conserved in related proteins (Orengo *et al*., 1995) and in such cases, there must be a sound evolutionary explanation for connectivity differences.

Kinesin, myosin and G proteins belong in this 'grey zone', since they have no amino-acid sequence conservation, and since secondary structural elements in their 'common core' have similar but not identical connectivities. Kinesin and myosin represent two superfamilies of molecular motors that play distinct cellular roles. The kinesin motor proteins move along microtubule polymers and are involved in scores of different processes, including vesicle transport and chromosomal segregation (Bloom & Endow, 1995). In contrast, myosins are associated with actin filaments and are involved in muscle contraction, cytokinesis, vesicle transport and cell motility. Myosin's motor domain is much larger than kinesin's (860 v. 340 amino acids), and the two proteins share no amino-acid identity, as determined from computer alignment programs. In addition, their hydrolytic cycles differ, since ADP release is the rate-limiting step for kinesin, while phosphate release is rate-limiting for myosin. Because of these and other differences, it was widely believed two years ago that kinesin and myosin represent two unrelated families of molecular motors.

Recent X-ray crystallographic findings, however, have shown structural similarities between myosins and kinesins. Most strikingly, a spatial computer alignment search using the program DALI (Holm & Sander, 1993) showed a strong overlap between kinesin and the core domain of myosin, but not with any other protein whose atomic structure is known. Virtually the entire motor domain of kinesin, including seven of the eight β-strands in its core β-sheet and all six α-helices, aligned with the catalytic core domain of myosin (Kull *et al*., 1996; Sablin *et al*., 1996; Vale, 1996). Further examination of these two motors illuminated other similarities, such as a conservation of active site chemistry (Kull *et al*., 1996) and spatial conservation of their distinct polymer-binding site relative to their shared core domain (Woehlke *et al*., 1997). In addition to the similarities shared by these two molecular motors, recent papers have highlighted similarities in active site chemistry and conformational change strategy between motors and G proteins, a superfamily of GTP hydrolysing enzymes that act as molecular switches (Sablin *et al*., 1996; Smith & Rayment, 1996; Vale, 1996).

Although previous work has described structural and functional similarities between kinesin, myosin and G proteins, the evolutionary relationship between these proteins has not been rigorously explored. Conservation of structural elements and functional domains (the first two criteria described above) has already been reported for kinesin, myosin and G proteins (Kull *et al*., 1996; Sablin *et al*., 1996; Smith & Rayment, 1996; Vale, 1996). However, these criteria alone do not provide sufficient evidence for an evolutionary relationship between these proteins, since their topological connectivities are not identical. In this study, we explore the evolutionary relationships among the motor proteins kinesin and myosin, as well as among motor proteins and other nucleotide-binding proteins, with an emphasis on the third and fourth criteria. We conclude that kinesin and myosin are related directly by divergent evolution, and present a simple evolutionary scheme whereby their individual topologies could have arisen from a common ancestor. We also demonstrate that motor proteins are more closely related to the G protein family than to any other class of nucleotide-binding proteins, and that these two distinct protein families are also likely to share a common ancestor.

Results and discussion

Comparison of kinesin and myosin

Previous studies have shown that kinesin and myosin share common structural elements, including their core β-strands, associated loops and functionally essential amino-acid side-chains (Kull *et al*., 1996). In the nucleotide-binding site, kinesin and myosin share four motifs conserved both in sequence and in positional space. Figure 1 shows the nucleotide-binding regions of kinesin and myosin, highlighting three of these four conserved motifs. In Fig. 1, the green loop following the green β-strand is the phosphate-binding P-loop, or N-1 motif. The P-loop motif is characteristic of many nucleotide hydrolysing enzymes and contains a consensus sequence GXXXXGKS/T. The purple loop and β-strand is the N-2 motif (also called Switch I), which contains the conserved NXXSSR sequence in both kinesin and myosin. The first serine residue in this motif (orange in Fig. 1) is thought to interact directly with the nucleotide and may be involved in sensing the presence or absence of γ-phosphate. The third motif, N-3 (also called Switch II), is shown in blue in Fig. 1 and contains the conserved sequence DXXGXE in both motor proteins. The conserved aspartic acid in this motif, shown in red, is thought to interact with the nucleotide-bound magnesium via a bridging water molecule. The fourth conserved motif (N-4), which is involved in interactions with the adenine ring in kinesin and myosin, is not shown in Fig. 1. Finally, the yellow β-strand in this figure, although not directly involved in nucleotide binding, is conserved spatially in kinesin and myosin, and is part of a common core structure which will be discussed in more detail below.

To examine whether there are additional conserved residues between kinesin and myosin, we aligned the kinesin and myosin structures and looked for residues that are conserved both in sequence as well as positional space. As can be seen in Fig. 2, of the 165 residues that superpose in the two proteins, there are only 20 identities (shown in blue) and seven significant similarities (grey). The majority (15 out of 27 of these residues) lie within four conserved nucleotide-binding motifs (N-1 to N-4 in Fig. 2), which, although spread throughout the primary sequence, are positionally clustered in a conserved core that surrounds the bound nucleotide.

The 12 conserved residues located outside of the nucleotide binding pocket roughly fall into three regions that may play significant roles in kinesin and myosin motor function (Fig. 2). One cluster of conserved residues (INXSTXXXL) is found at the N-terminal end of helix α4. These residues are located in a helix that follows the N-3 (Switch II) motif, which is involved in sensing of the γ-phosphate. In myosin, this helix is found in significantly different positions in crystal structures solved in different nucleotide states (Fisher *et al*., 1995). The spatial conservation of residues in the N-3 loop as well as the α4 helix suggests that a conformational switch mechanism may be conserved in this region between the kinesin and myosin motors. A second region containing conserved residues is found at the C-terminal end of helix α5 (LXXS/T) and a proline at the end of β8. This region is of interest, since the secondary structural elements and loops from α5 to α6 are the ones that most closely superimpose between kinesins and myosins. This α5/β8/α6 module follows the polymer-binding site and may provide the link to a force-generating element in kinesin and myosin. The three conserved residues may function in this process for both kinesin and myosin. The third region containing conserved residues is found in helix α1 and strand β3. This region has not been implicated in a specific functional role; it leads directly into the highly conserved N-1 region (P-loop). Since N-1 is the most highly conserved region in both sequence and conformation, the preceding structural elements may have been conserved during evolution in order to retain the essential P-loop conformation. In summary, the data presented in Fig. 2 show a conservation of residues in the active site, as well as in other structurally and functionally important areas outside of the nucleotide binding pocket which may be involved in motor function.

The topology and connectivity of kinesin and myosin differ. As described above, previous studies have illuminated a high degree of structural overlap between these two motors. Comparison of topological connectivity shows that the core regions of the two motors are essentially identical, but differences are observed at the edge of the central β-sheet (see Fig. 3, areas boxed in grey). At the left edge of the β-sheet (directions throughout this paper are relative to the orientation in Fig. 1, in which the nucleotide is at the top-front of the β-sheet), the order of strands β1 and β2 is reversed in kinesin and myosin, and the direction of the outermost strand points down in myosin and up in kinesin. At the right edge, kinesin and myosin have more significant differences. Kinesin has an extra β-strand (β5), and the topological position of the edge β-strands in the two proteins is also changed. In kinesin, strands β4 and β5 are topologically located between helices α2 and α3, whereas in myosin a direct connection links the analogous helices, and the edge strand, β5, is located following strand β4. The topology is identical in the remainder of the structural elements.

The differences in the topological connectivities between kinesin and myosin must be explained before the two proteins can be claimed to have evolved from a common progenitor. Figure 4 shows an evolutionary scheme in which the respective topologies

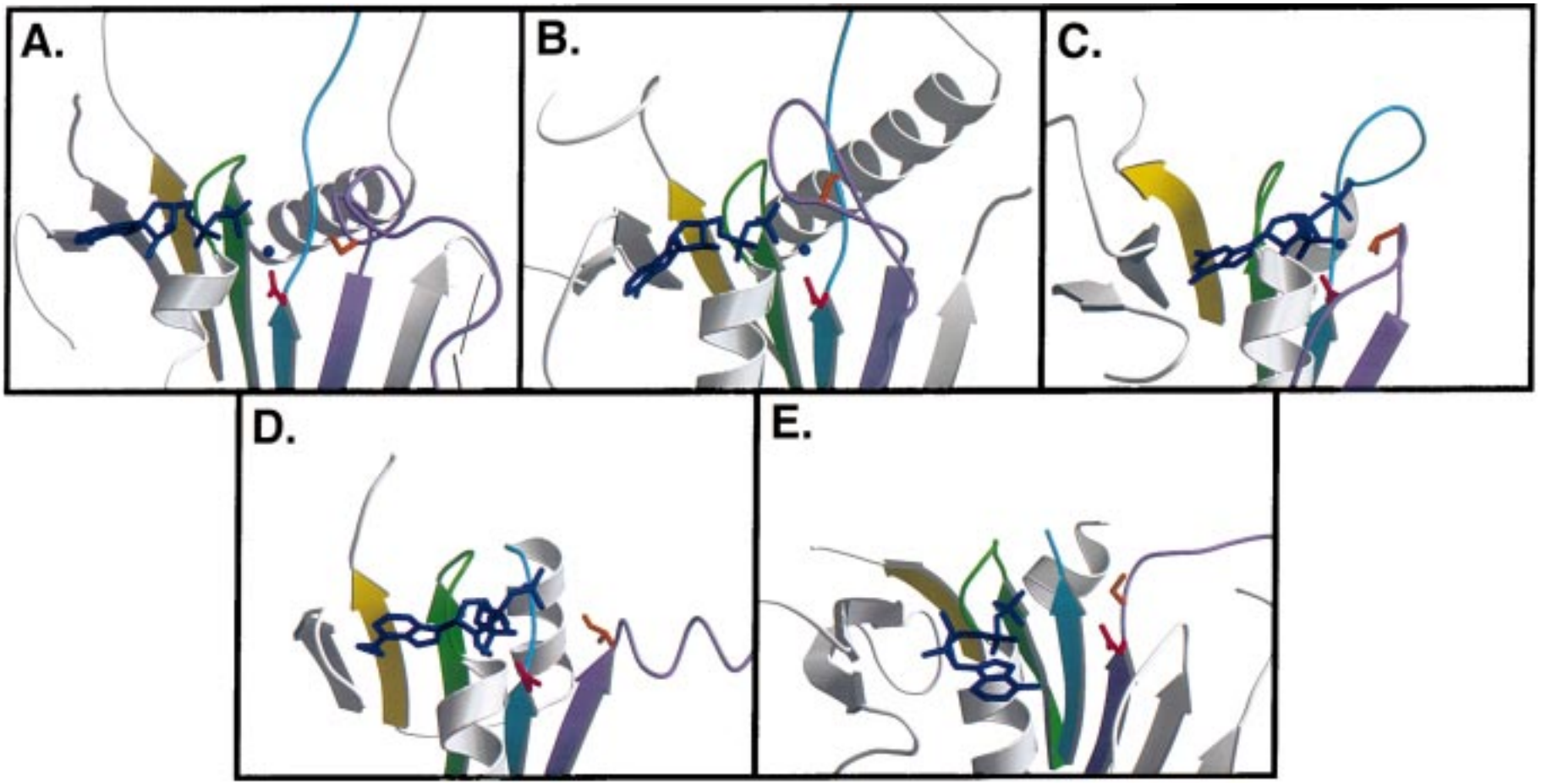

Fig. 1. Similarities in the nucleotide binding/γ-phosphate sensing region of various P-loop protein families. The core nucleotide-binding structural elements are shown in this view of: (A) kinesin; (B) myosin (Brookhaven Protein Data Bank accession number 1VOM); (C) $p21^{ras}$ (6Q21); (D) adenylate kinase (1ANK); and (E) recA (nucleotide coordinates from 1REA, sidechains from 2REB). The conserved nucleotide-binding elements (N-1 green, N-2 purple, N-3 blue) as well as the spatially conserved β-strand (yellow) are coloured consistently throughout all figures. The bound nucleotides (and Mg^{2+} ion, where available) are shown in blue. The conserved aspartic acid residue that is involved in metal coordination is shown in red, and a conserved serine/threonine which may interact with the nucleotide γ-phosphate is shown in orange. Note that the orientation of the rightmost conserved β-strand (purple) is reversed in adenylate kinase and recA. Also, an equivalent aspartic acid in recA, although in a similar spatial position, is located on a different β-strand than in the other proteins, being at the end of the purple rather than the blue strand.

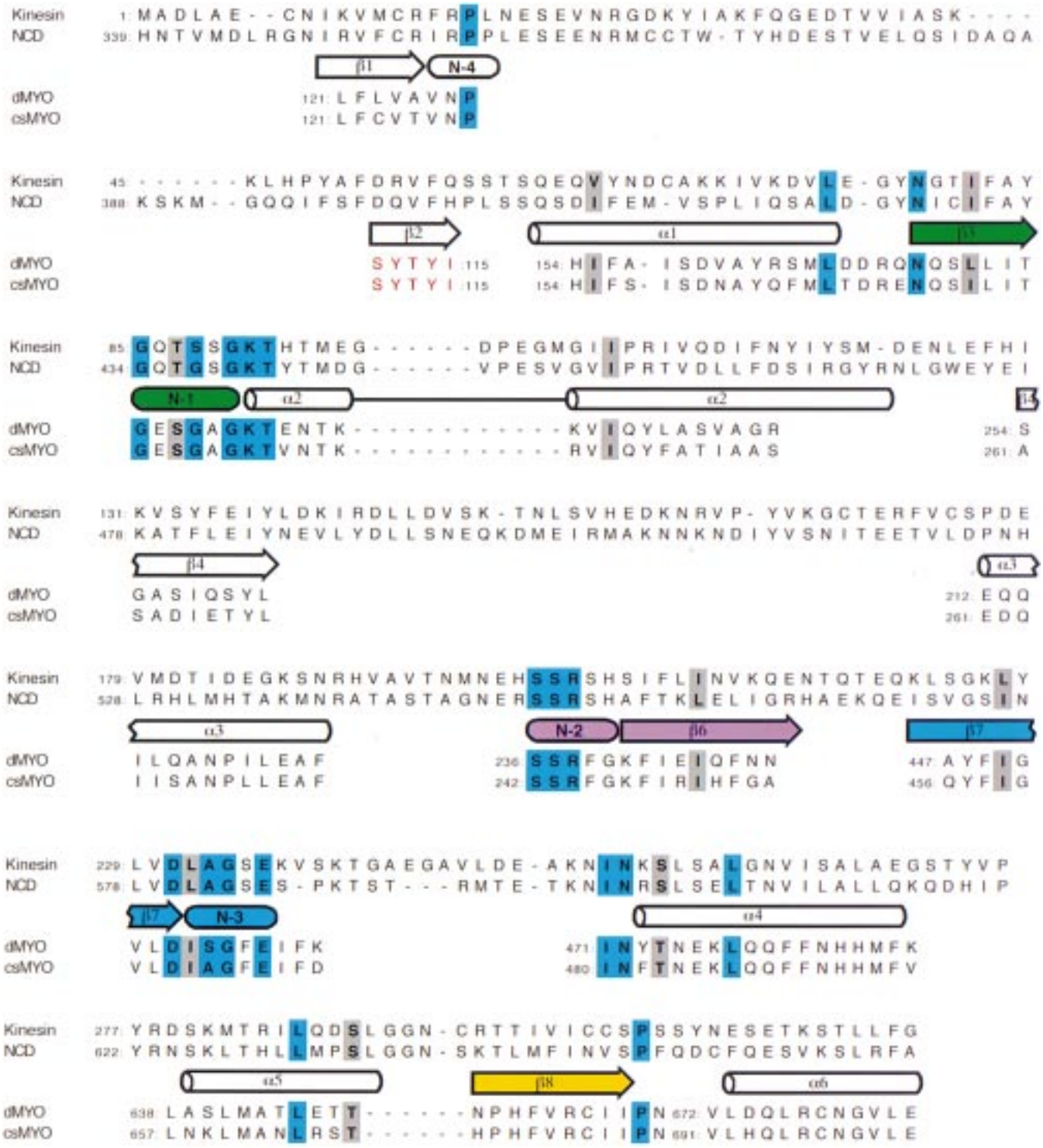

Fig. 2. Sequence alignment of common structural elements in kinesins and myosins. The motor domains of members of two distinct subfamilies of kinesin motors (kinesin–N-terminal motors; NCD–C-terminal motors) are aligned with the sequences of overlapping core structural elements from the myosin motor family (csMYO–chicken skeletal myosin; dMYO–*Dictyostelium* myosin). The sequence is continuous for the kinesin motors, but not continuous for the myosins (numbers indicate sequence position). The structural alignment was performed by superposing the α-carbon coordinates in the P-loop regions of the proteins. Overlapping structural elements are indicated and numbered according to the convention for the kinesin motors (Sablin *et al*., 1996). Conserved structural elements are coloured as described in Fig. 1. Aligned amino-acid residues overlap closely in positional space with an overall RMS deviation of 3.5 Å. Identical and highly conserved residues are shaded blue. Residues shaded in grey indicate conserved leucine/isoleucine or serine/threonine substitutions. Note the overlapping residues in β2 are coloured red for the myosins, as this β-strand runs in the opposite direction in kinesins and myosins.

of kinesin and myosin are derived from a common ancestor. The scheme follows two simple evolutionary propositions. The first states that new protein domains do not appear at once as completed units, but form via the addition of single strands or helices to an existing, stable core. Although whole-domain insertions and major rearrangements are certainly possible (e.g. as a result of gene transfer (Smith *et al*., 1992), domain swaps (Doolittle & Bork, 1993), or gene duplication (Doolittle, 1995)), they are excluded from this scheme, in which the formation of new domains is a linear process of small evolutionary steps. The second proposition is that any new secondary structure must appear at the edge of the progenitor structure, so that the existing hydrophobic core is not disrupted by the addition of new structure.

An evolutionary scheme obeying these propositions begins with an ancestral protein that may or may not have been a motor protein, but probably had all four nucleotide-binding elements N-1, N-2, N-3 and N-4 (Fig. 4), and represents a common core, probably found in all kinesins and myosins. Thus, this protein could conceivably bind to and hydrolyse nucleotides, as well as respond via conformational changes to the presence or absence of γ-phosphate in the nucleotide. The 'proto-motor' shown is somewhat smaller than modern motors, being composed of a central, five-stranded β-sheet (order 15243) flanked on either side by two helices. In the first hypothetical evolutionary step, an edge β-strand is added adjacent to the β3 strand. Although the spatial position of this strand is the same for both the kinesin and the myosin lineages, its position in the primary sequence is different. In the case of kinesin, this strand is added between two helices, whereas in myosin it is added between the rightmost β-strands. This relatively minor step (the formation of a new edge β-strand) has a significant consequence, since it causes the kinesin and myosin pathways to diverge. Subsequent additions of strands and helices can then build up the topology and connectivity of the modern motor families (Fig. 4). Once the correct topology is established, domain insertions (e.g. polymer-binding sites and force amplification domains) can be added via a similar insertional mechanism to build the functionally unique proteins. By our proposition, the conversion of a myosin topology to a kinesin topology (or the reverse) would be very unlikely, as it would require disruption of a stable protein core. Thus, although one

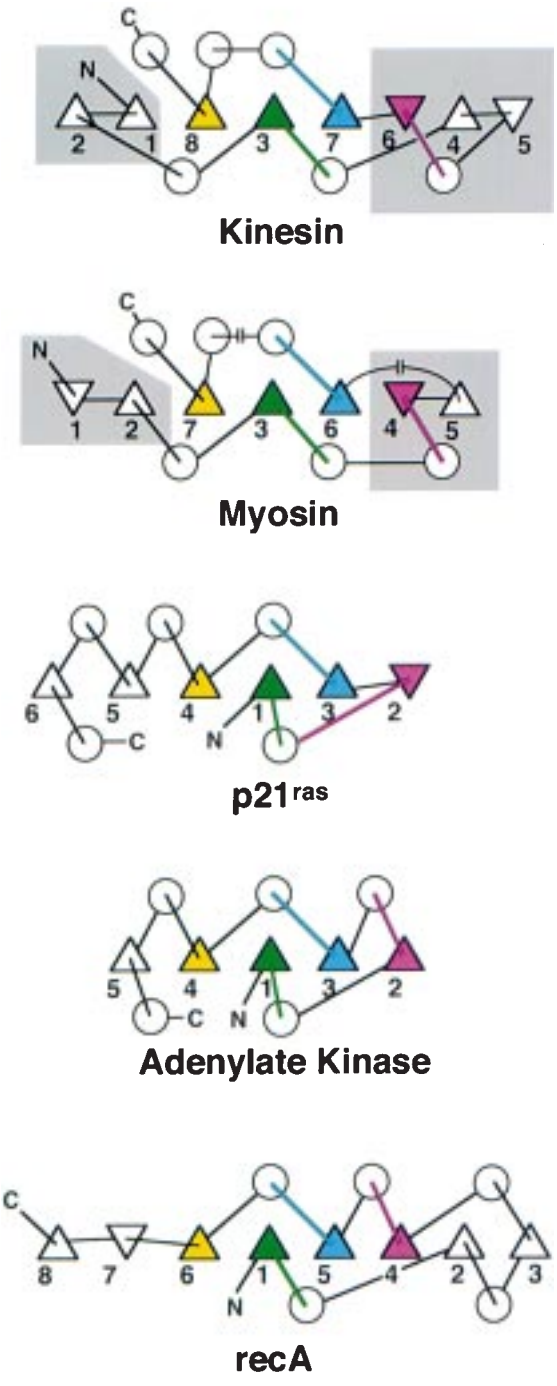

Fig. 3. P-loop protein topology. Topology diagrams (inspired by the program TOPS; Flores *et al*., 1994) for kinesin, myosin and three other representative members of P-loop-containing protein families ($p21^{ras}$ represents the G proteins). In this scheme, β-strands are shown as triangles (up or down indicating direction) and α-helices are shown as circles (up or down direction indicated by loop attachment). Numbers indicate the sequence order of the central β-strands. The core structures involved in nucleotide binding, which are conserved in all P-loop-containing proteins with the Walker fold, are indicated by colour-filled shapes (P-loop/N-1/G-1–green; N-3/G-3–blue; N-2/G-2-purple; although not involved in nucleotide binding, the yellow β-strand is always topologically conserved in these proteins). Note that in adenylate kinase and recA, the rightmost β-strand (purple) runs parallel to the other core β-strands, as opposed to anti-parallel in the case of the motors and the G proteins. The colour scheme in this figure is consistent with that in Fig. 1.

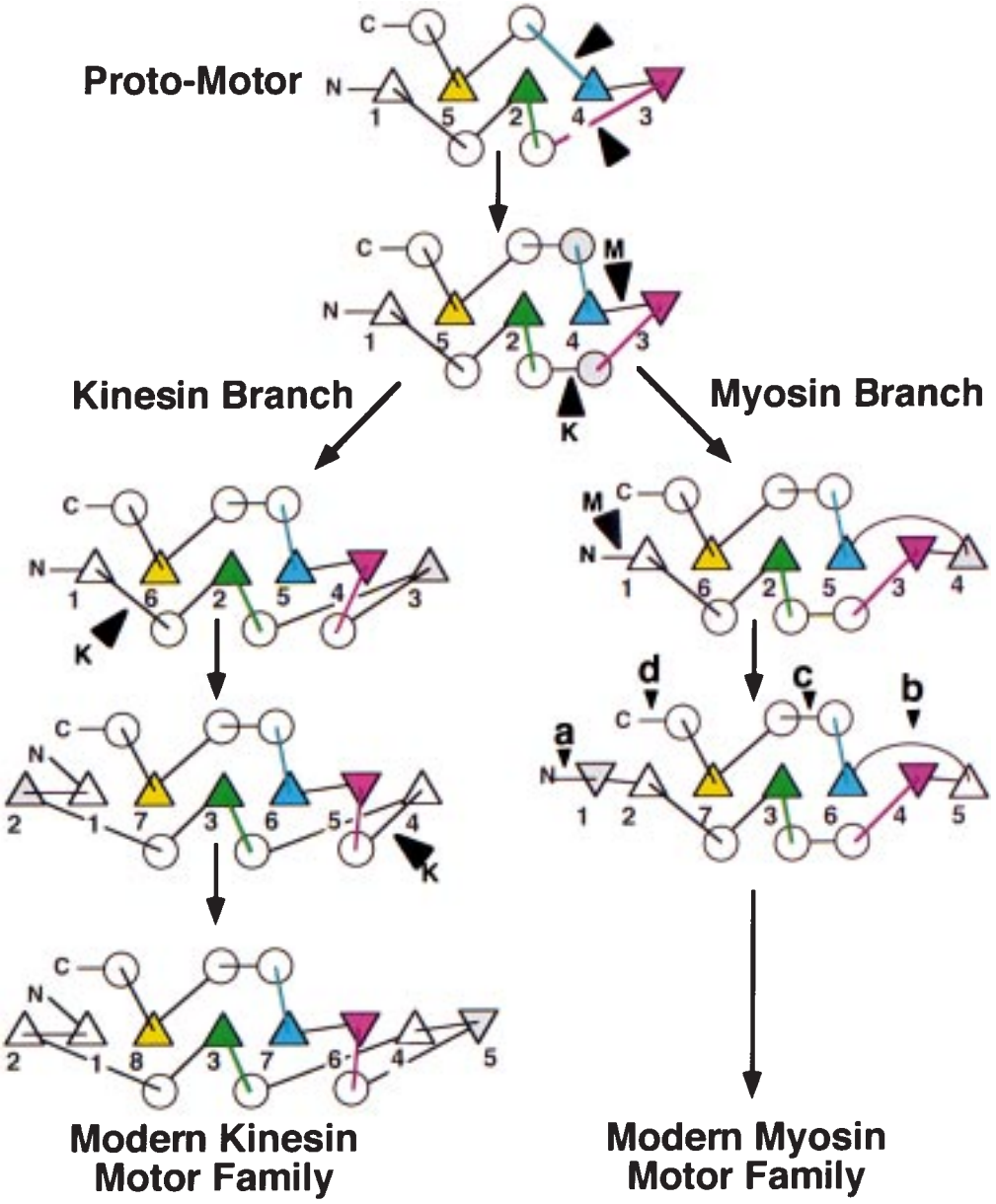

Fig. 4. Evolution of motor proteins. A potential pathway is shown for the evolution of the kinesin and myosin motor families from an ancestral proto-motor. Insertion sites are indicated by 'K'-wedges in the kinesin pathway and 'M'-wedges in the myosin pathway. New structural elements for each step are shown by grey-filled shapes. In the last step of the myosin pathway, letters indicate: a, N-terminal extension (aa 1–115); b, insertion 1 (aa 270–450); c, insertion 2 (aa 506–649); d, C-terminal extension (aa 710–843) (Kull *et al*., 1996). Numbers indicate the sequence order of the central β-strands. The major difference in topology is generated at the branch, where kinesin and myosin insert a β-strand in the same positional location but in different sequential order.

cannot rule out a scheme whereby kinesin and myosin have convergently evolved to resemble each other, the above argument presents a plausible genetic and structural foundation for divergent evolution.

Comparison of motors with other P-loop-containing proteins

We next compared motor proteins with other P-loop-containing proteins in order to determine which nucleotide-binding proteins may be related to the motors. P-loop-containing proteins are notoriously diverse, but can be grouped into distinct families based on the topological fold surrounding the nucleotide-binding region. The core of one family of P-loop proteins contains the 'Walker fold' (Walker *et al*., 1982) in which a central β-sheet of three or more β-strands (order 312) is flanked by two α-helices, one on each side of the sheet. In addition to the motor proteins, three other protein families containing the

Walker fold but having distinct overall topologies (Murzin *et al*., 1995) are typified by adenylate kinase (a metabolic enzyme), recA (an ATP-dependent recombination enzyme) and $p21^{ras}$ (a G protein) (see Fig. 3). In contrast, another family of nucleotide-binding proteins contain the classic or Rossmann fold (Rao & Rossmann, 1973; Rossmann *et al*., 1974; Schulz, 1992; Swindells, 1993), in which a central sheet of three parallel β-strands (order 123) has two α-helices packed against the front of the sheet. Given this large fundamental difference in the topology of the core nucleotide-binding region, it is unlikely that Rossmann fold-containing proteins (including tubulin (Nogales *et al*., 1998) and ftsZ (Lowe & Amos, 1998), two GTP-binding proteins) are directly related to motor proteins.

Of the enzymes containing the Walker fold, adenylate kinase has the simplest structure, composed of a central five-stranded β-sheet in the order 54132 (Fig. 3). Adenylate kinase differs from the motor proteins in topology, since one of the central β-strands (β2) runs in the opposite direction when compared with the spatially analogous strand in the motor proteins (β6 in kinesin, purple in Fig. 3). Another difference from the motor proteins is seen in the sequence of the N-1 region, in which adenylate kinase has a glycine inserted between the conserved lysine and threonine, which is missing in other P-loop proteins. Despite these differences, there are some similarities in the other conserved nucleotide-binding residues of adenylate kinase and motor proteins (Fig. 1). In the N-2 region, a serine and a threonine in adenylate kinase are located in an analogous position to the N-2 motif (SSR) in the motor proteins. Furthermore, adenylate kinase has an aspartic acid in a nearly identical position to the N-3 aspartic acid found in motors. In summary, although similarities exist in the active sites of adenylate kinase and motor proteins, their different topologies make a direct relationship with the motor proteins unlikely.

RecA is built around a core β-sheet composed of eight, mostly parallel β-strands with an order 87615423 (Fig. 3). Surprisingly, recent crystal structures of the F1-ATPase (Abrahams *et al*., 1994), a helicase from *Bacillus stearothermophilus* (Subramanya *et al*., 1996), and a helicase from the hepatitis C virus (Yao *et al*., 1997) reveal a common fold and similar active site chemistry with the recA protein. Like adenylate kinase, recA contains a β-strand (β4) that runs in the opposite direction to the motor protein counterpart (purple in Fig. 3). In the nucleotide-binding area, the purine ring is in a very different orientation compared with other P-loop proteins (Fig. 1), and there are no identifiable N-2 or N-3 motifs. However, conserved residues in other locations appear to fulfill similar roles (Story & Steitz, 1992). The conserved aspartic acid residue in the N-3 motif of both motors and adenylate kinase is absent in recA, but another aspartic acid is located at the top of the adjacent β-strand (in what would be the N-2 motif in the motors and G-proteins) and interacts with an active-site metal ion through a bridging water molecule (Fig. 1). The position of a functionally similar metal-binding aspartic acid on an entirely different β-strand compared with motor proteins would seem to be a result of convergent, rather than divergent, evolution. Interestingly, this metal-binding aspartic acid is followed by a conserved serine (Ser 145). Although this serine has not been implicated in nucleotide interactions, its location is similar to the serine/threonine seen in the motor proteins and adenylate kinase. In conclusion, the differences in topology and in the active site suggest that a direct evolutionary relationship of recA with the motor proteins is unlikely.

The $p21^{ras}$ protein contains a core β-sheet composed of six β-strands in the order 654132 (Fig. 3). This topology is typical of the G protein family. In contrast to the situation with adenylate kinase and recA, the orientation and order of the core β-strands in the G proteins is the same as in motor proteins. The nucleotide-binding regions of the G proteins and the motor proteins are also very similar (Fig. 1). As has been previously described, the conserved binding motifs N-1, N-2 and N-3 in motors have direct counterparts in the G protein family (Sablin *et al*., 1996). There is also likely to be conservation of general function, since both motor proteins and G proteins use N-2 and N-3 as 'switch' regions that undergo conformational changes in response to phosphate release following nucleotide hydrolysis (Vale, 1996). Thus the G proteins and the motor proteins display conservation of structure and active site chemistry.

Although the four core β-strands of the motors and G proteins share identical topology (coloured yellow, green, blue and purple in the figures), the areas outside of the core have different connectivities. An evolutionary scheme abiding by the previously described rules can be constructed that produces the two distinct folds from a common ancestor (Fig. 5). In this scheme, the ancestor is a simple protein containing three of the four nucleotide binding/interaction motifs seen in motor proteins and G proteins (N-1, N-2 and N-3). The variable fourth motif (N-4) need not have been present in this protein, as it is absent in some nucleotide-binding proteins that contain P-loops. The nucleotide-binding ancestor, which consists of a four-stranded β-sheet (order 4132) with at least two connecting helices, represents a core structural motif which is probably common to all members of the motor and G-protein families. This protein could evolve into motor proteins and G proteins in a straightforward manner by adding β-strands to the N-terminus for motors, and to the C-terminus for G-proteins.

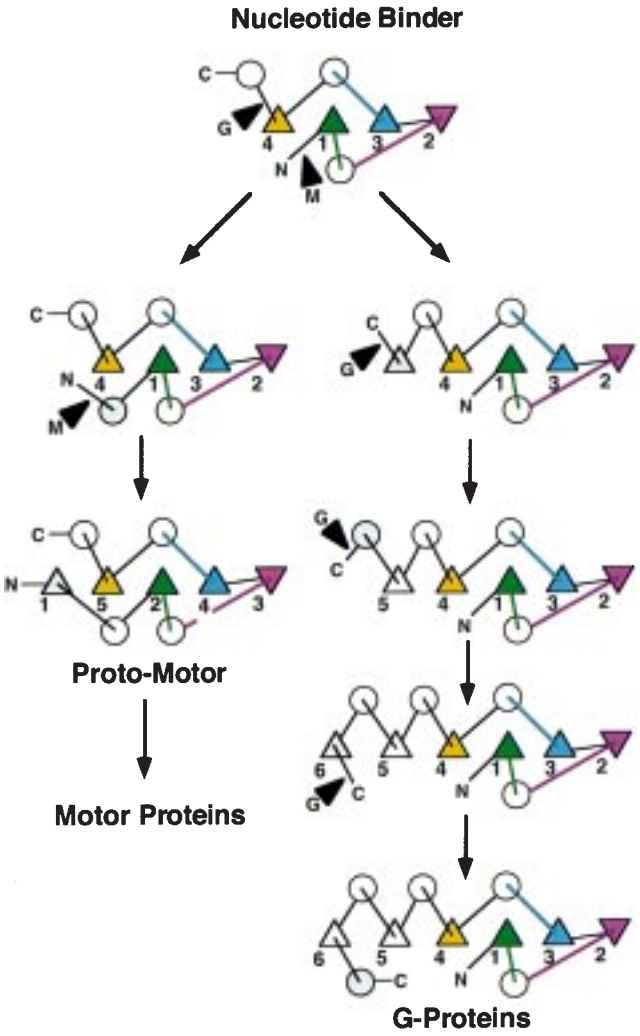

Fig. 5. Evolution of G-proteins. A potential pathway is shown for the divergent evolution of G proteins and motor proteins from a simple, nucleotide-binding protein. Insertion sites are indicated by 'G'-wedges in the G protein pathway and 'M'-wedges in the motor pathway. New structural elements for each step are shown by grey-filled shapes. Numbers indicate the sequence order of the central β-strands. The major difference in topology is generated by adding new structures at the N-terminus for motors and at the C-terminus for G proteins. The proto-motor structure is identical to the one at the start of the pathway in Fig. 4. Note that the ancestral nucleotide-binding protein contains all of the conserved core structural elements discussed throughout this paper.

Conclusion

We have proposed a set of criteria that help to define evolutionary relationships among proteins lacking significant amino-acid sequence identity. Proteins within this 'grey zone' may be considered to be related if they possess spatial conservation of core structural elements, conservation of functional and chemical elements, and conservation of topological connectivity. By applying these criteria to kinesin and myosin, we have identified a shared core motif, described its topology, and shown that the two motor proteins are likely to be related by a common ancestor. Furthermore, we have presented a scheme by which divergent evolution could have produced their somewhat distinct topological connectivities.

It will be interesting to ascertain whether a third family of cytoskeletal motors, the dyneins, also contains a similar conserved motor fold. Dynein is a P-loop-containing protein, but, unlike kinesin and myosin, it contains four P-loops, one of which is involved in ATP catalysis while the other three bind nucleotides and may serve a regulatory function (Mocz & Gibbons, 1996). Interestingly, the four P-loops are spaced at ~350 amino-acid intervals, suggesting that they may have arisen by gene duplication (Gibbons, 1995). Although this element is comparable in size to the kinesin motor, we have been unable to find clear N-2, N-3 and N-4 motifs in dynein, and therefore it is difficult to predict whether dynein is related to kinesin/myosin without crystal structure data.

Finally, we have compared the motor protein family with other proteins containing nucleotide-binding P-loops. Our analysis indicates that, although the motor family is not directly related to the majority of these other proteins, it shares a topologically identical nucleotide-binding core motif with the family of molecular switches known as G proteins. While analysis based only on sequence data leads to the conclusion that these proteins are unrelated, our analysis of the topologies of these proteins shows that it is quite likely that kinesin, myosin and G proteins are related by an ancestral nucleotide-hydrolysing enzyme which has divergently evolved to form the modern protein families.

References

ABRAHAMS, J. P., LESLIE, A. G., LUTTER, R. & WALKER, J. E. (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–8.

ARTYMIUK, P. J., POIRRETTE, A. R., RICE, D. W. & WILLETT, P. (1996) Biotin carboxylase comes into the fold. *Nature Struct. Biol.* **3**, 128–32.

ARTYMIUK, P. J., RICE, D. W., POIRRETTE, A. R. & WILLETT, P. (1995) beta-Glucosyltransferase and phosphorylase reveal their common theme. *Nature Struct. Biol.* **2**, 117–20.

BLOOM, G. & ENDOW, S. (1995) Motor proteins 1: kinesin. *Protein Profiles* **1**, 1112–38.

BORK, P., SANDER, C. & VALENCIA, A. (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl Acad. Sci. USA* **89**, 7290–94.

BRANNIGAN, J. A., DODSON, G., DUGGLEBY, H. J., MOODY, P. C., SMITH, J. L., TOMCHICK, D. R. & MURZIN, A. G. (1995) A protein catalytic framework

with an N-terminal nucleophile is capable of self-activation *Nature* **378**, 416–19.

DOOLITTLE, R. F. (1994) Convergent evolution: the need to be explicit. *Trends Biochem. Sci.* **19**, 15–18.

DOOLITTLE, R. F. (1995) The multiplicity of domains in proteins. *Ann. Rev. Biochem.* **64**, 287–314.

DOOLITTLE, R. F. & BORK, P. (1993) Evolutionarily mobile modules in proteins. *Sci. Amer.* **269**, 50–56.

FAN, C., MOEWS, P. C., SHI, Y., WALSH, C. T. & KNOX, J. R. (1995) A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine: d-alanine ligase of Escherichia coli. *Proc. Natl Acad. Sci. USA* **92**, 1172–6.

FISHER, A J., SMITH, C. A., THODEN, J. B., SMITH, R., SUTOH, K., HOLDEN, H. M. & RAYMENT, I. (1995) X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.AlF4-. *Biochemistry* **34**, 8960–72.

FLAHERTY, K. M., DELUCA-FLAHERTY, C. & MCKAY, D. B. (1990) Three-dimensional structure of the ATPase fragment of a 70 K heat-shock cognate protein. *Nature* **346**, 623–8.

FLAHERTY, K. M., MCKAY, D. B., KABSCH, W. & HOLMES, K. C. (1991) Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc. Natl Acad. Sci. USA* **88**, 5041–5.

FLORES, T. P., MOSS, D. S. & THORNTON, J. M. (1994) An algorithm for automatically generating protein topology cartoons. *Protein Engineering* **7**, 31–7.

GIBBONS, I. R. (1995) Dynein family of motor proteins: present status and future questions. *Cell Motil. Cytoskel.* **32**, 136–44.

GUPTA, R. S. & GOLDING, G. B. (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. *J. Mol. Evol.* **37**, 573–82.

HOLM, L. & SANDER, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–38.

HOLM, L. & SANDER, C. (1995) Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. *Embo J.* **14**, 1287–93.

HWANG, P. K. & FLETTERICK, R. J. (1986) Convergent and divergent evolution of regulatory sites in eukaryotic phosphorylases. *Nature* **324**, 80–84.

KULL, F. J., SABLIN, E. P., LAU, R., FLETTERICK, R. J. & VALE, R. D. (1996) Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* **380**, 550–55.

LOWE, J. & AMOS, L. A. (1998) Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203–6.

MOCZ, G. & GIBBONS, I. R. (1996) Phase partition analysis of nucleotide binding to axonemal dynein. *Biochemistry* **35**, 9204–11.

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**, 536–40.

NOGALES, E., WOLF, S. G. & DOWNING, K. H. (1998) Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, 199–203.

ORENGO, C. A., JONES, D. T. & THORNTON, J. M. (1994) Protein superfamilies and domain superfolds. *Nature* **372**, 631–4.

ORENGO, C. A., SWINDELLS, M. B. MICHIE, A. D., ZVELEBIL, M. J., DRISCOLL, P. C., WATERFIELD, M. D. & THORNTON, J. M. (1995) Structural similarity between the pleckstrin homology domain and verotoxin: the problem of measuring and evaluating structural similarity. *Protein Sci.* **4**, 1977–83.

RAO, S. T. & ROSSMANN, M. G. (1973) Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–56.

ROSSMANN, M. G., MORAS, D. & OLSEN, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**, 194–9.

SABLIN, E. P., KULL, F. J., COOKE, R., VALE, R. D. & FLETTERICK, R. J. (1996) Crystal structure of the motor domain of the kinesin-related motor ncd. *Nature* **380**, 555–9.

SCHULZ, G. E. (1992) Binding of nucleotides by proteins. *Curr. Opin. Struc. Biol.* **2**, 61–7.

SMITH, C. A. & RAYMENT, I. (1996) Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J.* **70**, 1590–1602.

SMITH, M. W., FENG, D. F. & DOOLITTLE, R. F. (1992) Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* **17**, 489–93.

STORY, R. M. & STEITZ, T. A. (1992) Structure of the recA protein-ADP complex. *Nature* **355**, 374–6.

SUBRAMANYA, H. S., BIRD, L. E., BRANNIGAN, J. A. & WIGLEY, D. B. (1996) Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379–83.

SWINDELLS, M. B. (1993) Classification of doubly wound nucleotide binding topologies using automated loop searches. *Protein Sci.* **2**, 2146–53.

VALE, R. D. (1996) Switches, latches, and amplifiers: common themes of G proteins and molecular motors. *J. Cell. Biol.* **135**, 291–302.

WAGNER, G. (1995) E-cadherin: a distant member of the immunoglobulin superfamily. *Science* **267**, 342.

WALKER, J. E., SARASTE, M., RUNSWICK, M. J. & GAY, N. J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J.* **1**, 945–51.

WOEHLKE, G., RUBY, A. K., HART, C. L., LY, B., HOM-BOOHER, N. & VALE, R. D. (1997) Microtubule interaction site of the kinesin motor. *Cell* **90**, 207–16.

YAO, N., HESSON, T., CABLE, M., HONG, Z., KWONG, A. D., LE, H. V. & WEBER, P. C. (1997) Structure of the hepatitis C virus RNA helicase domain. *Nature Struct. Biol.* **4**, 463–7.