Supporting Text

Sequence Threading

To search for homologs of the different domains in the motor unit of the cytosolic dynein of slime mold *Dictyostelium discodeum* (gb: P34036; refs. 1 and 2), we submitted the domain sequences for threading to 3DJury (http://bioinfo.pl/meta/)(3, 4), a consensus prediction server. We observed that the threading resulted in sequences of AAA domains containing an extra α-helical globular subunit in addition to the classical AAA domains. Furthermore, secondary structure predictions of the regions between dynein AAAs indicate that they primarily consist of α-helices. Thus, we submitted for threading also the sequences of AAA plus adjacent linker. Threading statistics for the final templates is summarized in Table 1.

Construction of Domains

Using the initial alignments from consensus threading as input (SI Fig. 9), we built the models of the domains and interdomain chains in the Homology suite of INSIGHTII (Accelrys Inc., San Diego, CA). We evaluated the models for correct residue packing (buried hydrophobic residues and exposed charged side chains) using Verify3D (http://nihserver.mbi.ucla.edu/Verify_3D/) (5). The Verify3D profiles were used to identify structurally erroneous sites due to misalignment of target sequence to template. We then made minor changes in the initial alignment to improve the pairing of residues with similar properties and improve the connectivity between residues.

Since the AAA1-AAA4 are highly conserved thus expected to resemble symmetric oligomerization of AAAs, we constructed a tetramer composed of AAA1-AAA4 by superposing the AAA1-AAA4 models to four contiguous AAA domains of sigma-54 RNA polymerase transcriptional activator (NtrC1) (PDB entry 1NY6; ref. 6), which is a ring-shaped homoheptamer. The root-mean-square deviation (RMSD) between an AAA model and the NtrC1 AAAs ranged from 2 Å to 6.2 Å.

Construction of Heptamer

To form a heptamer from the AAA1-AAA4 tetramer, AAA5, and AAA6, and C-domain, we extracted the densities corresponding to domains 1-4, 5, 6, and 7 using FLOODFILL in SITUS (7). Then using vector-quantization implemented in QRANGE (8), the tetramer AAA1-AAA4, AAA5, AAA6, and C-domain were separately fitted to the extracted densities 1-4, 5, 6, and 7, respectively. The correct fit of AAA1-AAA4 and C-domain was identified *ab initio*. We manually reoriented AAA5 such that its C terminus connects to the N-terminus of the putative coiled coil. We also manually reoriented AAA6 such that its N terminus connects to the C terminus of the coiled coil. This constraint on the peptide continuity fixed the orientation of AAA5 and AAA6. IDR4 was manually fitted to the overarching density. To add the linkers in AAA-AAA4, we docked the models IDR1, IDR2, and IDR3 to the heptamer using PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/) (9). Finally, the domains and linkers were connected together by searching through plausible loops derived from known structures in the Protein Data Bank.

To validate the overall fit, we fitted the model using COLORES of SITUS which employs two types of fitting methodologies: (i) standard cross-correlation fitting takes the whole density as a fitting criterion and (ii) cross-correlation with Laplacian filtering, which includes the contour of the density as fitting criterion (10). With the standard filter, the highest correlation (0.74) assigns the domains of the atomic model to the correct densities in the EM map. Using the Laplacian filtering, the second highest correlation (0.267) gives the correct domain assignment. (The highest correlation in the second method is 0.268.) The correlation values using the Laplacian filtering are lower due to masking (10), and the expected correlation values for non-spurious fit is 0.2 to 0.5.

Refinement of Model

To eliminate clashes, we performed discrete molecular dynamics (DMD) (11) on the model structure using two-bead representation of the for each residue. When the clashes are eliminated, we reconstructed the all-atom model of the motor. The detailed protocol described elsewhere (12). The RMS deviation between the model before simulation and reconstructed model is 2.5 Å.

Normal Mode Analysis

We used the ELNEMO server (http://igs-server.cnrs-mrs.fr/elnemo/index.html) (13) to compute the normal models of the model structure. Interactions between atoms are approximated with a harmonic potential with a cut-off of 8 Å. The equations of motion are calculated by diagonalization of the Hessian matrix (mass-weighted second derivatives of the potential energy matrix). Eigenvalues of the matrix correspond to the mode frequencies and the associated eigenvectors are the normal modes.

1. Koonce, M. P., Grissom, P. M. & McIntosh, J. R. (1992) *Journal of Cell Biology* **119**, 1597-1604.

2. Koonce, M. P., Grissom, P. M., Lyon, M., Pope, T. & McIntosh, J. R. (1994) *J. Eukaryot. Microbiol.* **41**, 645-651.

3. Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. (2003) *Bioinformatics* **19**, 1015-1018.

4. Ginalski, K. & Rychlewski, L. (2003) *Nucleic Acids Research* **31**, 3291-3292.

5. Bowie, J. U., Luthy, R. & Eisenberg, D. (1991) *Science* **253**, 164-170.

6. Lee, S. Y., De la Torre, A., Yan, D. L., Kustu, S., Nixon, B. T. & Wemmer, D. E. (2003) *Genes & Development* **17**, 2552-2563.

7. Wriggers, W., Milligan, R. A. & McCammon, J. A. (1998) *Journal of Molecular Graphics & Modelling* **16**, 283.

8. Wriggers, W. & Chacon, P. (2001) *Structure* **9**, 779-788.

9. Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R. *et al.* (2003) *Proteins-Structure Function and Genetics* **52**, 107-112.

10. Chacon, P. & Wriggers, W. (2002) *Journal of Molecular Biology* **317**, 375-384.

11. Ding, F. & Dokholyan, N. V. (2005) *Trends in Biotechnology* **23**, 450-455.

12. Ding, F., Prutzman, K. C., Campbell, S. L. & Dokholyan, N. V. (2006) *Structure* **14**, 5-14.

13. Suhre, K. & Sanejouand, Y. H. (2004) *Nucleic Acids Research* **32**, W610-W614.

14. Asai, D. J. & Koonce, M. F. (2001) *Trends in Cell Biology* **11**, 196-202.

15. Samso, M. & Koonce, M. P. (2004) *J. Mol. Biol.* **340**, 1059-1072.

16. Luthy, R., Bowie, J. U. & Eisenberg, D. (1992) *Nature* **356**, 83-85.

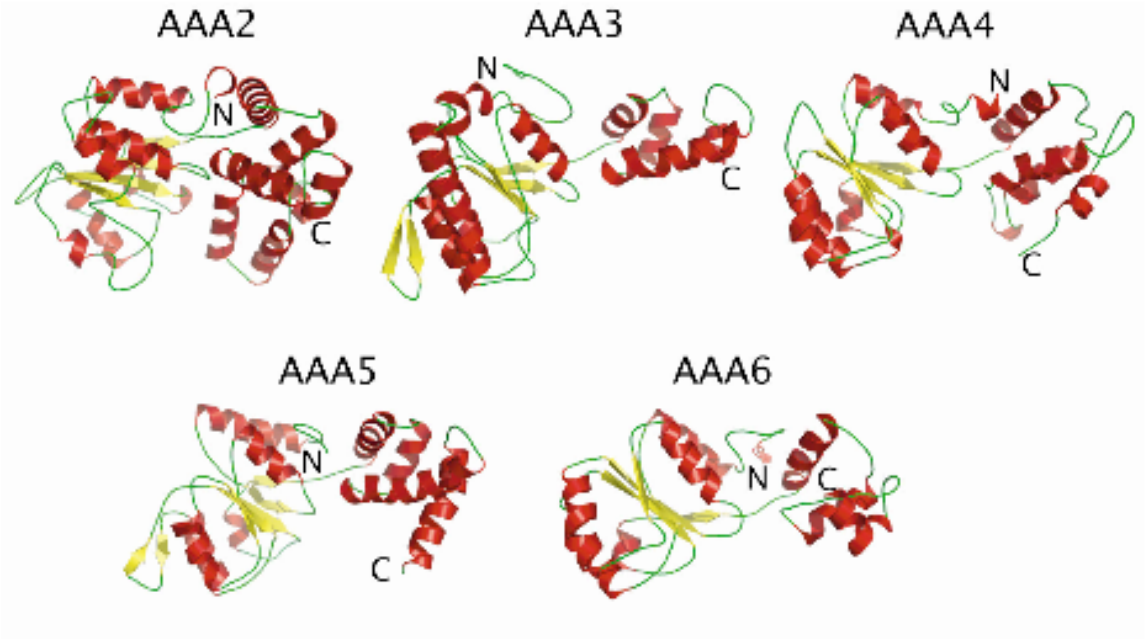

Fig. 5. Models of the AAA2 to AAA6 Domains of Dynein. The AAA domains exhibit the α/β- and all-α subunits of classical AAA enzymes. The individual domains also feature structures found only in some AAAs, such as the pre-Sensor1 β-hairpins of AAA3 and AAA5.

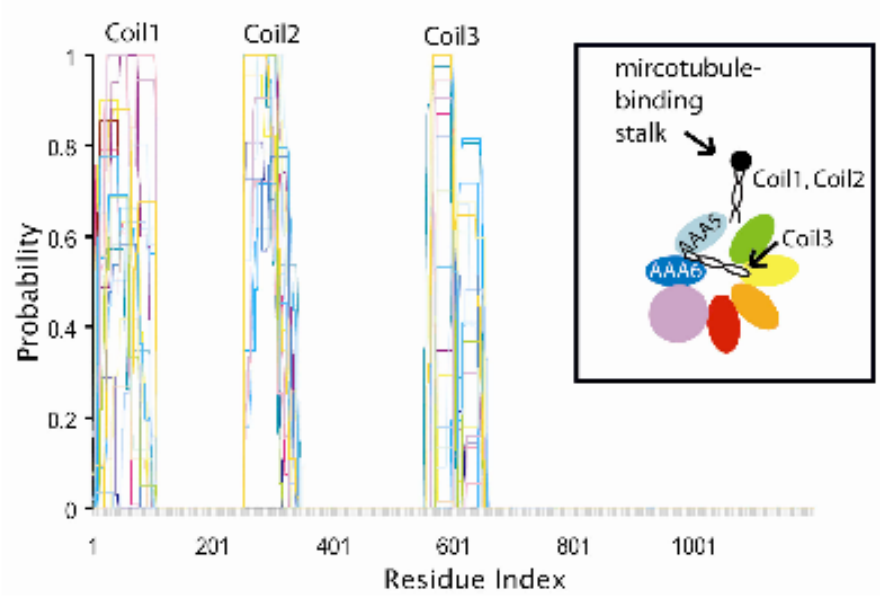

Fig. 6. Coil Probability from the Binding Stalk to AAA6. Plot of the coil probability in the region starting from the microtubule-binding stalk to AAA6 of dyneins from different organisms. It has been established that the first two peaks form the dynein stalk (14). The third peak putatively forms the long intermediate sequence between AAA5 and AAA6. Dynein sequences included in the coil probability calculation come from *Rattus norvegicus* (P38650), *Homo sapiens* (Q14204, Q9NYC9, Q8TE73, Q96DT5), *Mus musculus* (Q9JHU4, Q8VHE6), *Chlamydomonas reinhardtii* (Q39565, Q39610, Q39575, Q9SMH3), *Dictyostelium discoideum* (P34036), *Drosophila melanogaster* (P37276), *Caenorhabditis elegans* (Q19020), *Neurospora crassa* (P45443), *Aspergillus nidulans* (P45444), *Paramecium tetraurelia* (Q27171), *Saccharomyces cerevisiae* (P36022), *Kluyveromyces lactis* (Q9C1M7), *Tripneustes gratilla* (P23098), *Anthocidaris crassispina* (P39057), and *Schizosaccharomyces pombe* (O13290).

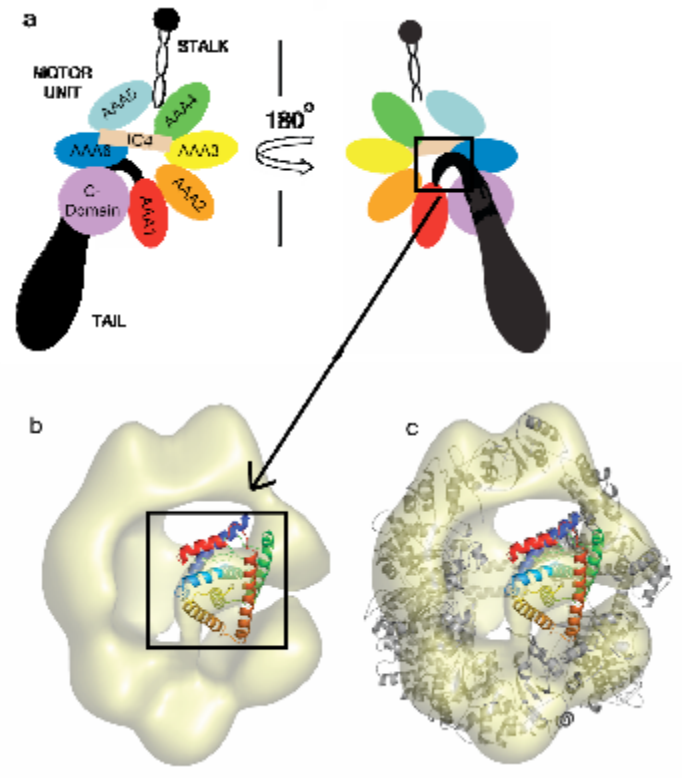

Fig. 7. Model of a Dynein Tail Fragment. Following the same protocol used in modeling the motor unit domains, we constructed a model of a tail fragment (1665E-1879H) using estrogen receptor beta (PDB: 1L2J) as template. **a,** The tail fragment precedes AAA1 and putatively docks to the center of the motor ring. **b,** Model of the tail fragment docked to the smaller facial lump of the EM map derived from negative staining by Samso *et al* (15). **c,** The tail fragment shown in the context of the motor unit model. However, we were not able to ascertain how much of the tail fragment accounts for the smaller facial lump.

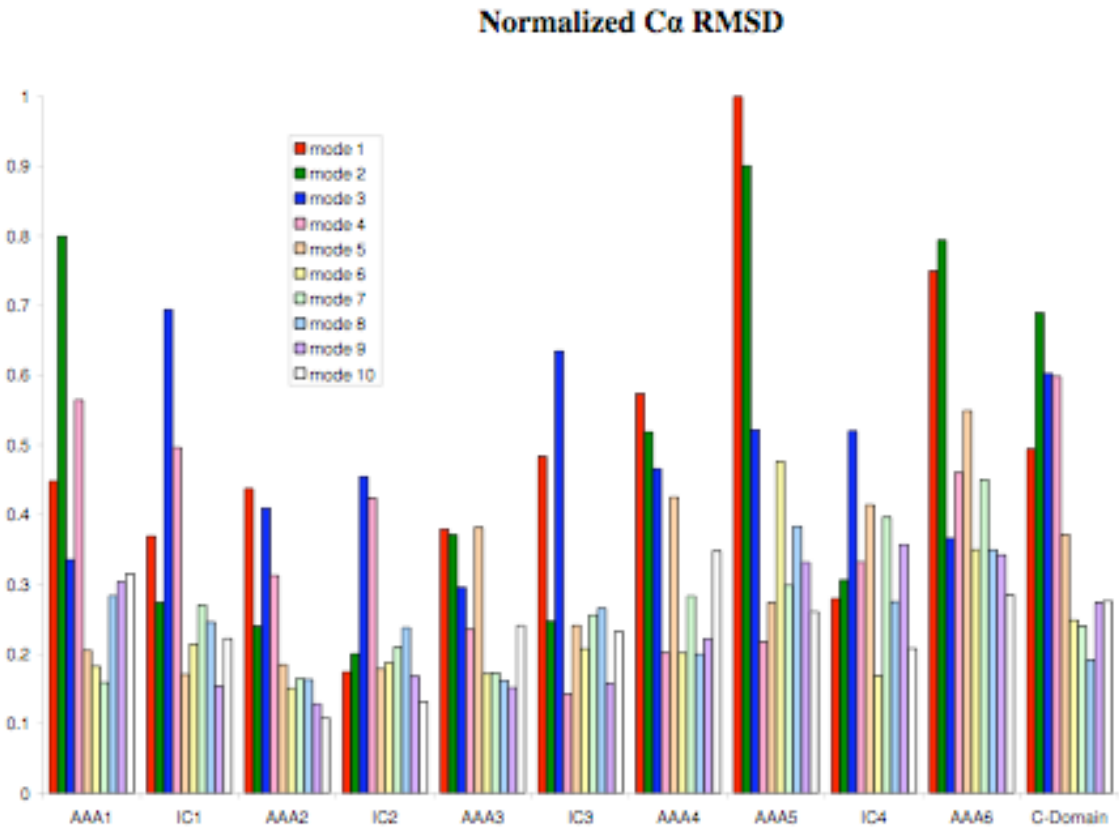

Fig. 8. Lowest Ten Normal Modes of the Dynein Motor Unit Model. Mean RMSD of Cα in a domain, normalized by the largest displacement and weighted by inverse frequency. The interdomain chain IDR4 between AAA5 and AAA6 exhibits smaller displacements relative to the rest of the motor. This result indicates that IDR4 is a rigid structure. See main text for discussion.

Fig. 9. Sequence alignment of *Dictyostelium discoideum* dynein and template proteins. Alignment is annotated using Chroma. Consensus markers and color represent residue groupings. The groupings in decreasing order of precedence are as follows: "-" indicates negative residues DE; "*" indicates ST; "|" indicates aliphatic residues ILV; "+" indicates positive residues HKR; "t" indicates tiny residues AGS; "a" indicates aromatic FNWY; "c" indicates charged residues DEHKR; "s" indicates small residues ACDGNPSTV; "p" indicates polar residues CDEHKNQRST; "b" indicates big residues EFHIKLMQRWY; and "h" indicates hydrophobic residues ACFGHILMTVWY.

Movie 1. Dynein structure projected along mode 1. Front view. Domains are colored as follows: AAA1 (red), AAA2 (orange), AAA3 (yellow), AAA4 (green), AAA5 (cyan), AAA6 (blue), and C-domain (magenta). (MPG; 920 KB)

Movie 2. Dynein structure projected along mode 1. Side view. (MPG; 720 KB)

Movie 3. Dynein structure projected along mode 2. Front view. (MPG; 945 KB)

Movie 4. Dynein structure projected along mode 2. Side view. (MPG; 706 KB)

Movie 5. Dynein structure projected along mode 3. Front view. (MPG; 1032 KB)

Movie 6. Dynein structure projected along mode 3. Side view. (MPG; 723 KB)

Table 1. Sequence Threading Statistics

Domains*	3DJury Score†	PDB Code	Template Protein Description
AAA1+IDR1, AAA1	109, 103‡	1HQC	Holliday Junction Motor Protein RuV
AAA2+IDR2, AAA2	194, 154‡	1JR3	DNA Polymerase III
AAA3+IDR3, AAA3	94, 85 ‡	1HQC	Holliday Junction Motor Protein RuV
AAA4	87	1SXJ_A	DNA Clamp Loader
AAA5	68	1HQC	Holliday Junction Motor Protein RuV
IDR4	35	1QU7	Cytoplasmic Domain of Serine Chemotaxis Receptor
AAA6	107	1SXJ_B	DNA Clamp Loader
C-Domain (alpha)	51	1GHQ	Complement Component C3d
C-Domain (beta)	38	1TXD	Dh/pH Domains of Leukemia-Associated Rhogef

*Domain boundaries are defined in the annotations to the *Dictyostelium discoideum* dynein (gene bank entry P34036) (1, 2).
†Statistically significant score is greater than 50, which implies that there is a 90% chance that the alignment is correct (http://bioinfo.pl/meta/queue.pl) (3, 4).
‡Threading scores are improved when the adjacent linker is included in the search for homologs.

Table 2. Evaluation of Model Quality

Domains	Model Score	Template Score
AAA1, IDR1	0.36, 0.05	0.50
AAA2, IDR2	0.18, 0.11	0.42
AAA3, IDR3	0.31, 0.11	0.50
AAA4	0.28	0.43
AAA5	0.22	0.50
IDR4	0.15	0.25
AAA6	0.25	0.45
C-domain	0.18	0.36*

*Combined score for 1GQH and 1TXD, templates for the first 290 residues and trailing 100 residues, respectively.

The above scores describe the accuracy of each model subunit using Verify3D (www.doe-mbi.ucla/Services/Verify_3D) (5, 16). For both the models and their templates, we compared the local environments of each residue to the population of averaged residue environments determined from known structures.