Deriving folds of macromolecular complexes through electron cryomicroscopy and bioinformatics approaches

Wah Chiu*†§, Matthew L Baker*†, Wen Jiang*† and Z Hong Zhou*‡

Intermediate-resolution (7–9 Å) structures of large macromolecular complexes can be obtained by electron cryomicroscopy. This structural information, combined with bioinformatics data for the individual protein components or domains, can lead to a fold model for the entire complex. Such approaches have been demonstrated with the 6.8 Å structure of the rice dwarf virus to derive models for the major capsid shell proteins.

Addresses
*Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA
†Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA
‡Department of Pathology and Laboratory Medicine, University of Texas-Houston Medical School, Houston, Texas 77030, USA
§e-mail: wah@bcm.tmc.edu

Current Opinion in Structural Biology 2002, 12:263–269

0959-440X/02/$ – see front matter

Abbreviations

PDB	Protein Data Bank
RDV	rice dwarf virus
rmsd	root mean square deviation

Introduction

Structural biology using X-ray crystallography and NMR spectroscopy has reached a mature stage at which molecular biologists can use these techniques as part of their investigations to pursue structural and functional relationships of molecules of interest. Through structural genomics initiatives, structural biologists aim to represent all molecules with unique structures and/or functions in order to better understand the structure/function relationship, thus filling up 'fold space' [1,2•,3–5]. In this regard, fold space refers to all the possible configurations that a polypeptide might adopt. The fold of a protein might be described as the overall topology, which includes the arrangement and connectivity of the secondary structure elements of the polypeptide chain [6,7]. In a large protein, a single chain may form one or several distinct domains, each with a particular fold.

To date, the majority of structures solved to high resolution, at which the fold of a protein can be discerned, have molecular masses of less than 250 kDa. The current Protein Data Bank (PDB) contains more than 16 000 structures [8], of which only 93 have masses greater than 250 kDa, according to the biological macromolecular crystallization database [9]. With the rapid growth of cloning, sequencing and purification of new gene products, a by-product of the genome sequencing initiatives [10,11], the availability of large proteins and protein complexes will also increase. Therefore, structural biologists face the challenges of determining the structures of macromolecular complexes of increasing complexity and size [12••].

Strategies to study large complexes

In general, there are three approaches to deriving the fold models of large macromolecular complexes. The first approach is to pursue the structural study solely by X-ray crystallography. The success in determining the structures of large viruses (e.g. bluetongue virus [13]) and machines (e.g. RNA polymerase [14]) argues strongly for the merit of this approach. However, the limiting step in this approach is often the inability to obtain a suitable crystal. Another approach is the 'divide and conquer' approach, which is to determine the atomic structures of individual components and to map them into a low-resolution structure of the entire complex obtained by electron cryomicroscopy [15]. There have been many successful studies using this approach, which will continue to be valuable in many instances. The third approach is to use electron cryomicroscopy as a primary tool for structural studies, presently at intermediate resolution (7–9 Å), supplemented with bioinformatics methods to deduce the fold model of the complex. This method allows one to study a complex without a crystal. The information yielded is particularly reliable if a high-resolution structure of a similar or related complex, or of its domains is already known. This review will focus on the promise of this third approach to describing a moderately high resolution structure of a large complex.

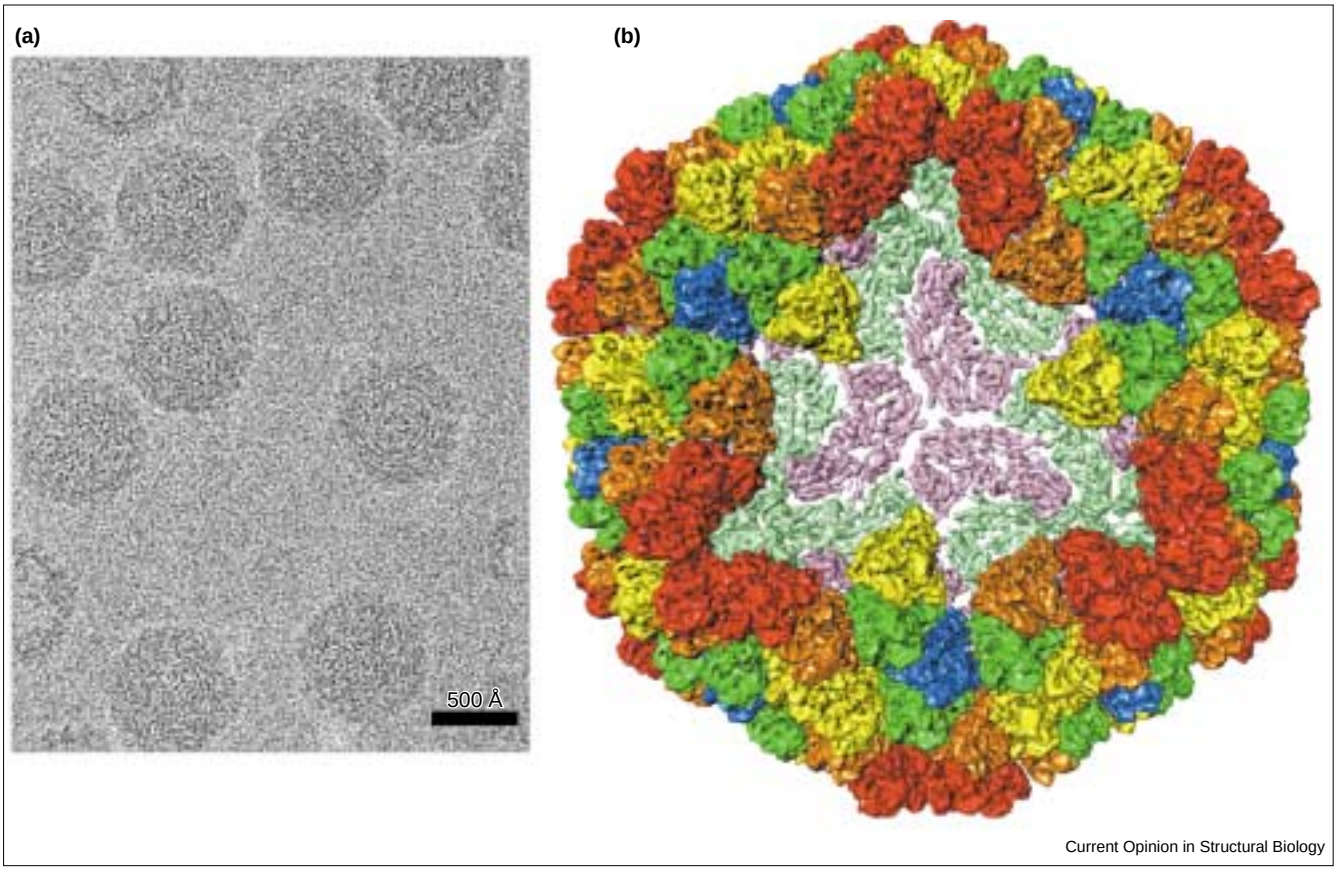
Electron cryomicroscopy at intermediate resolution (7–9 Å)

Advances in electron cryomicroscopy and computational methods of analysis [16•] have made it possible to determine 7–9 Å structures of macromolecular complexes with high internal symmetry without using a crystal [17–19,20••]. As an example, Figure 1 shows a raw image of ice-embedded individual rice dwarf virus (RDV) particles and the 6.8 Å structure reconstructed from ~4000 particle images. This virus contains two protein shells — the outer layer is made up of P8 (46 kDa) with a T = 13 icosahedral lattice and the inner shell is made up of P3 (114 kDa) with a T = 1 icosahedral lattice. There are other types of biological macromolecular complexes, notably membrane proteins, which have been studied to similar resolution or beyond when prepared as crystalline or helical arrays [21–26]. For complexes with less or no symmetry, structural determination has also been pursued towards and beyond 10 Å (e.g. [27–30]).

Localization of secondary structure elements

The mining of three-dimensional density maps of a large macromolecular complex determined at intermediate

Figure 1



(a) 400 kV electron cryomicroscopy image of RDV particles of ~700 Å in diameter (courtesy of Joanita Jakana at Baylor College of Medicine). (b) 6.8 Å structure of RDV reconstructed from ~4000 particle images. The virus is made up of two protein shells. Each asymmetric unit of the outer shell contains four one-third trimers, which are colored red, green, blue, yellow and orange. One icosahedral face of the outer shell is removed to reveal the inner shell. Each asymmetric unit of the inner shell consists of a dimer of P3 isoforms, which is colored in mauve and aquamarine.

resolution for secondary structure features requires both advanced graphics and computational feature-extraction techniques. Below, we will describe the visual and computational methods that have been employed.

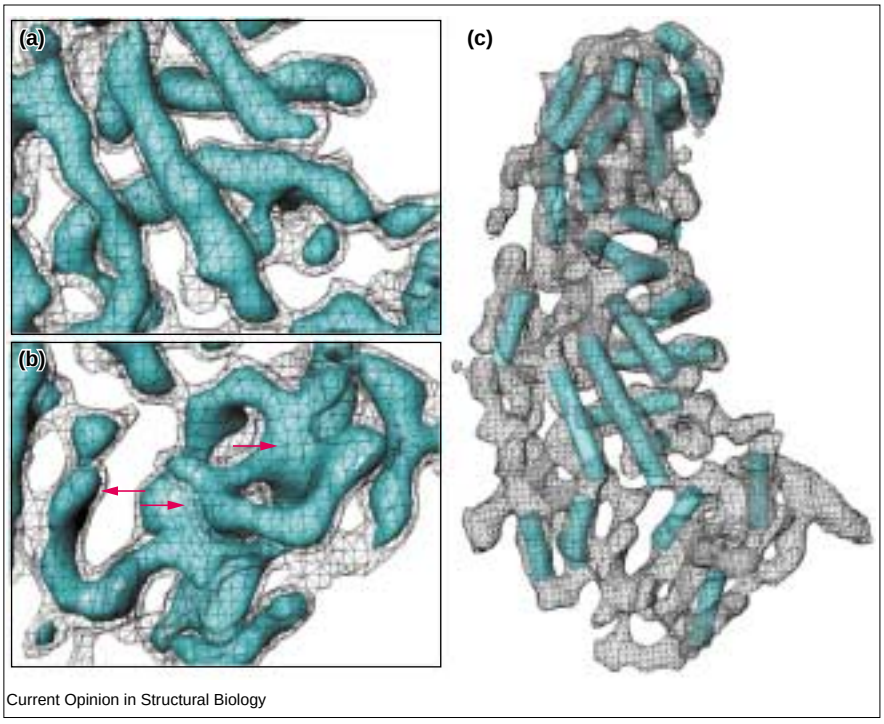
Direct visual detection

In general, because of the large volume of density maps, it is cumbersome to visualize the detailed features of an entire macromolecular complex using interactive graphics tools, even with high-end computer graphics. A practical approach to visualization is to computationally segment an asymmetric unit from the large complex for further examination. At intermediate resolutions (7–9 Å), it is often possible to recognize not only individual protein boundaries and domains, but also internal features, such as individual α helices and β sheets. This resolution is not yet sufficient to resolve individual strands, which are ~4.8 Å apart. With proper graphical contouring, helices appear as cylindrical mass densities with a diameter of 5–6 Å, whereas β sheets appear as continuous and flat densities with varying sizes (see Figure 2a,b, extracted from the 6.8 Å structure of RDV [20••]). However, visual recognition can be subjective, making it difficult to make a definitive decision in some of the regions. Therefore, a more quantitative method of analysis is preferred to yield higher accuracy in secondary structure element identification.

Computational recognition of helices

Software such as *helixhunter* has been developed to help objectively and automatically identify the locations of helices consisting of three or more turns within a density map [31••]. *Helixhunter* incorporates a multistep process including cross-correlation, density segmentation, segment quantification, helix identification and an explicit description of the identified helices. The final helices are represented as cylinders, each specified by six parameters (three for center positions, two for orientation and one for length). This abstract helix representation makes the visualization of helices easy, as well as allowing subsequent spatial fold recognition. Figure 2c shows an example of the helices localized in the density map of the inner shell P3 protein of RDV.

Figure 2



Segments of the 6.8 Å density map of the inner shell P3 protein of RDV, showing (a) bundle densities and (b) flat and continuous densities (indicated by red arrows), which are interpreted as α helices and β sheets, respectively. (c) The helices identified by *helixhunter* are shown as colored cylinders of 5 Å diameter, which are superimposed on the density map, shown as gray wire frames.

Assignment of observed helices to sequence segments

The putative helices identified in the density map do not contain any information on their sequence correspondence. In order to map amino acid segments to the observed helices, one can turn to sequence-based secondary structure predictions of the protein. There are numerous algorithms for predicting the secondary structure of proteins. Table 1 shows the outcomes of three algorithms applied to the outer shell P8 protein of RDV; these predict the nine sequence segments to be helices in the N- and C-terminal regions. These predictive algorithms yield similar but not identical predictions of the helical segments. In this case, one can match the lengths of the structurally observed helices to those predicted from a consensus secondary structure prediction, as shown in the case of P8 of RDV [20••]. This suggests that the N- and C-terminal regions occupy the lower domain of P8. Furthermore, in the lower domain of this protein, helical densities and their connecting densities are adequately resolved in the intermediate-resolution map (Figure 3a). Therefore, the connectivities of the two sets of helices in the N- and C-terminal regions can be established (Figure 3b). However, it should be pointed out that it is often not possible to establish the connectivity between contiguous helices, except in favorable circumstances, as seen in RDV and the hepatitis B virus core [18]. Obviously, it is not yet feasible to detect short helices that are less than three turns.

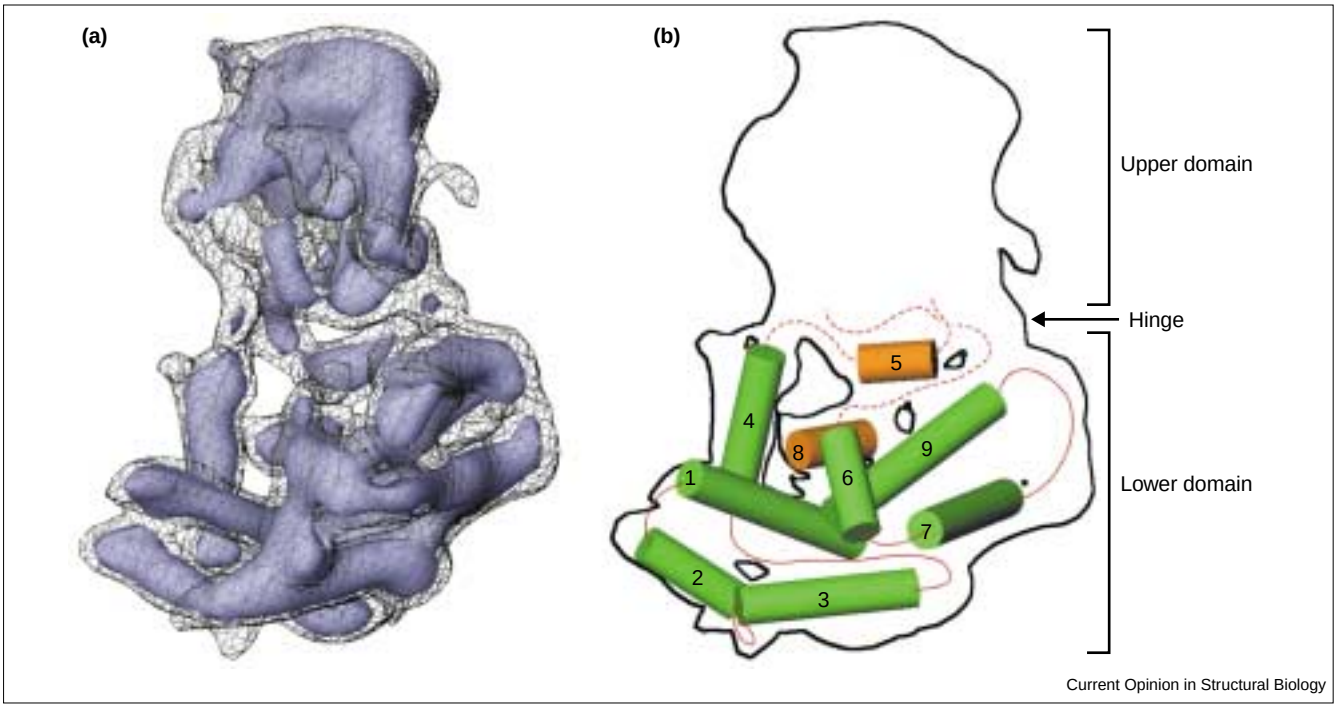
Membrane proteins represent a special class of proteins because of the predominant presence of transmembrane helices connected by loops. In the case of aquaporin-1, prediction of the sequence segments making up the transmembrane helices and functional loops was made through analyses of a large number of sequences from the aquaporin family. Although the connecting loops are invisible, the transmembrane helices can be clearly identified in an intermediate-resolution map [22–24,32]. Sequence analyses and correlated mutations, together with helix directionalities suggested by the cross-correlation of model helices with helices identified in the map, provided the

Table 1

Lengths of helices of RDV P8 determined with *helixhunter* from the 6.8 Å structure and from three different predictive algorithms.

Helixhunter		SSPro		PSIPRED		PHD	
Helix	Length (Å)	Sequence	Length (Å)	Sequence	Length (Å)	Sequence	Length (Å)
1	28	4–19	24	3–21	27	5–18	21
2	16	36–48	19	36–49	21	36–45	15
3	23	61–73	19	60–72	19	61–71	16
4	17	132–141	15	131–141	16	135–142	12
5	9	154–161	12	153–160	12	155–159	7
6	17	326–335	15	321–337	25	324–337	21
7	24	345–359	22	341–359	28	341–359	28
8	11	376–391	24	374–381	12	380–385	9
9	30	393–414	33	393–413	31	396–416	31

Figure 3



(a) 6.8 Å density map of the outer shell P8 protein of RDV displayed in two contour levels: the lower contour level as wire frame and the higher contour level as shaded surfaces. (b) The lower domain has been interpreted to consist of two sets of helices, which are in the N- and C-terminal regions. Their suggested connectivities are substantiated by the match of the helix lengths between the observed structures and the predicted sequence segments (Table 1).

basis for building the fold of the six transmembrane helices in the aquaporin-1 molecule [33••,34•]. The predicted fold is in good agreement with those subsequently determined by electron crystallography at 3.8 Å [26] and X-ray crystallography at 2.2 Å [35].

Intermediate-resolution structure-based fold recognition

The resulting set of helices identified in the intermediate-resolution structure can also be used as elements for probing the PDB for similar structures. To this end, *DejaVu* and *COSEC* [36–38] are particularly useful for this type of 'spatial fold recognition'. A successful matching of the helix arrangement of the *helixhunter* results for an unknown protein to a structure in the PDB is suggestive that the intermediate-resolution structure under investigation has a possible homolog within fold space. In the RDV example, six of the nine helices within the lower domain of P8 were matched to bluetongue virus VP7 (rmsd of helix centroid <5 Å) [13,39]. Although these proteins appear to have only remote sequence similarity (<20%), they appear to have similar structures and perform the same function within the capsid. Conversely, if no match can be found through such a search, one may conclude that the fold of the protein is likely to be novel.

Fold recognition with homologous structures

In certain instances, an individual protein or portion of a protein within the macromolecular complex may contain a homologous structure detectable through sequence comparisons using a variety of algorithms. For instance, the sequence of interest may be queried against all sequences whose structures are known by using *PsiBlast* with the PDB [40]. Although this type of analysis may work well with related sequences, it is much more difficult to detect remote homologs and/or sequence unrelated but structurally similar proteins. For example, the middle sequence segment of RDV P8 has no significant sequence homologs. However, by using fold recognition methods such as the UCLA DOE fold recognition server and 3D-PSSM [41,42•], which uses a variety of statistical measures besides the raw sequence alignment, it was possible to identify that this region is structurally similar to the β-sheet domain of bluetongue virus VP7 [43]. It should be pointed out that this approach works only for proteins in which the domain fold is formed from consecutive sequence elements. Currently, it is not possible to accurately predict the fold of a domain that comprises multiple chains or nonconsecutive sequence segments.

Once a homologous fold has been identified, the task at hand becomes localizing this domain to the entire complex. To this end, several approaches can be taken [31••,44–46]. For instance, *foldhunter*, a template-based cross-correlation tool, automatically searches in six dimensions for the best fit of the putative fold of a domain to the density map of the complex. Figure 4 shows an example of the match of

the putative jelly-roll β-sandwich fold to the density in the upper domain of P8 of RDV. The localization of this β-sheet domain represents the best possible fit; however, some structural differences are evident, as supported by their relatively low sequence similarity [20••].

Conclusions

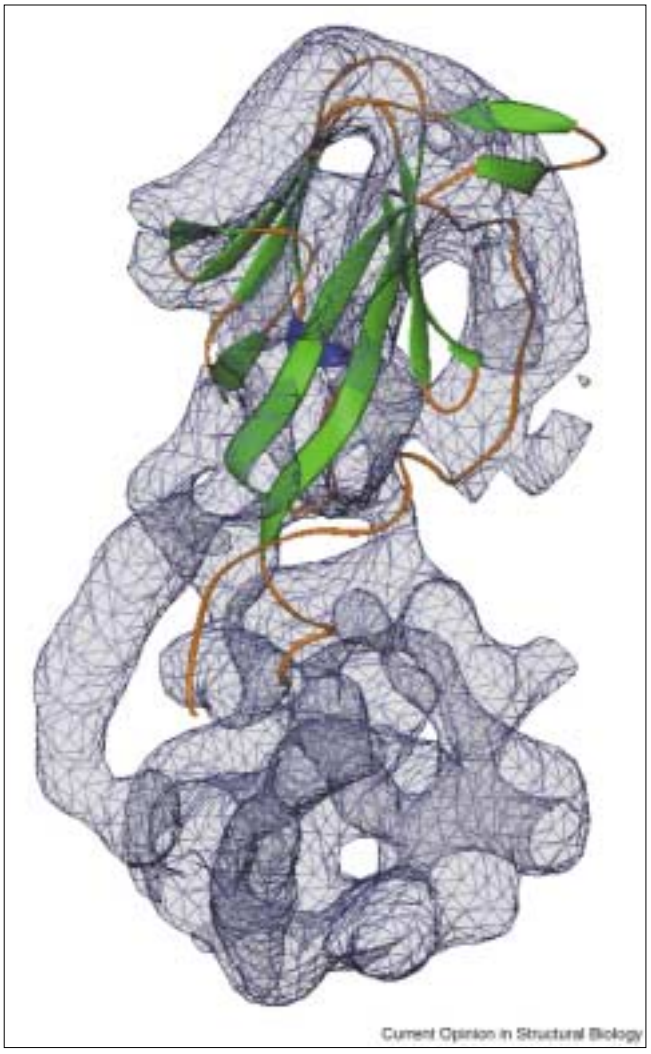
One of the trends in structural biology is to study large macromolecular complexes at increasing complexity and higher resolution. The opportunity now arises that one may consider using a combination of intermediate-resolution (7–9 Å) electron cryomicroscopy and bioinformatics approaches to obtain a model of a protein fold for part of or an entire complex. The examples cited here include icosahedral viruses and two-dimensional membrane crystals. Nevertheless, this imaging technology is ready to extend to macromolecular complexes with less or no symmetry. It is inevitable that the technology will be extended to tackle this class of structural problems. With the availability of sequence information, the increasing accuracy of some of the structure prediction algorithms and, in some instances, given other biochemical, biophysical and genetics information, we can now begin to use this information to support the structural analysis to obtain a quasi-atomic model for large complexes.

With the increasing knowledge of fold space due to the structural genomics efforts, a significant portion of the folds in a large complex may indeed be familiar ones. This, in turn, may lead to better and more accurate interpretation of the macromolecular complex at intermediate resolution. Conversely, the increasing number of intermediate-resolution structures, coupled with structural analysis, may facilitate structural genomics efforts in their pursuit of novel folds and coverage of fold space. Most importantly, however, understanding the structure of the entire complex is critical to understanding the function of cellular machinery, as it is only then that we understand the dynamics of the machine and intermolecular interactions. With the eventual incorporation of molecular dynamics into this multiresolution schema, it may be possible to further understand the structural and functional significance of the individual components within the complex that provide the chemical basis of the structural and functional diversity of proteins.

Update

Using a combination of sequence and structural analysis tools, together with the 8.5 Å resolution structure of the herpes simplex virus-1 capsid, it has been possible to derive a model of the entire major capsid protein, VP5 [47]. In VP5, extracted from the averaged hexon subunit, 33 helices were identified, of which 10 were shown to be structurally homologous to the family of annexins. Using sequence analysis, this particular domain could then be mapped back to two regions of the VP5 sequence. Additionally, nine of the remaining helices could be correlated to secondary structure predictions, ultimately producing a model for the structure of this large protein.

Figure 4



Docking of the putative β-sandwich fold into the intermediate-resolution density map of P8 of RDV.

Acknowledgements

We thank the support of grants from the National Institutes of Health (P41RR02250, R01AI38469, R01AI43656, R01AI46420), the Robert Welch Foundation and the Agouron Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sali A: 100,000 protein structures for the biologist. *Nat Struct Biol* 1998, 5:1029-1032.

2. • Vitkup D, Melamud E, Moult J, Sander C: Completeness in structural genomics. *Nat Struct Biol* 2001, 8:559-566.

This paper evaluates strategies taken in structural genomics for optimizing fold space coverage using the least number of structures. The authors state that ~16 000 structures, carefully selected, could cover the majority of known folds.

3. Mittl PR, Grutter MG: Structural genomics: opportunities and challenges. *Curr Opin Chem Biol* 2001, 5:402-408.

4. Thornton J: Structural genomics takes off. *Trends Biochem Sci* 2001, 26:88-89.

5. Brenner SE: A tour of structural genomics. *Nat Rev Genet* 2001, 2:801-809.

6. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM: CATH – a hierarchic classification of protein domain structures. *Structure* 1997, 5:1093-1108.

7. Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C: SCOP: a structural classification of proteins database. *Nucleic Acids Res* 2000, 28:257-259.

8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: The Protein Data Bank. *Nucleic Acids Res* 2000, 28:235-242.

9. Gilliland GL, Tung M, Ladner J: The biological macro molecule crystallization database and NASA protein crystal growth archive. *J Res Natl Inst Stand Technol* 1996, 101:309-320.

10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860-921.

11. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: The sequence of the human genome. *Science* 2001, 291:1304-1351.

12. •• Baumeister W, Steven AC: Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem Sci* 2000, 25:624-631.
This paper describes the promise that electron cryomicroscopy holds for the study of large macromolecules, including their interactions and dynamics. This paper also relates the importance of this type of work combined with X-ray crystallography of components from these complexes.

13. Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PP, Stuart DI: The atomic structure of the bluetongue virus core. *Nature* 1998, 395:470-478.

14. Cramer P, Bushnell DA, Kornberg RD: Structural basis of transcription: RNA polymerase II at 2.8 Å. *Science* 2001, 292:1863-1876.

15. DeRosier DJ, Harrison SC: Macromolecular assemblages. Sizing things up. *Curr Opin Struct Biol* 1997, 7:237-238.

16. • Unger VM: Electron cryomicroscopy methods. *Curr Opin Struct Biol* 2001, 11:548-554.
This paper reviews the current state-of-the-art electron cryomicroscopy techniques used in structural studies of macromolecular complexes.

17. Conway JF, Cheng N, Zlotnick A, Wingfield PT, Stahl SJ, Steven AC: Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* 1997, 386:91-94.

18. Böttcher B, Wynne SA, Crowther RA: Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 1997, 386:88-91.

19. Zhou ZH, Dougherty M, Jakana J, He J, Rixon FJ, Chiu W: Seeing the herpesvirus capsid at 8.5 Å. *Science* 2000, 288:877-880.

20. •• Zhou ZH, Baker ML, Jiang W, Dougherty M, Jakana J, Dong G, Lu G, Chiu W: Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nat Struct Biol* 2001, 8:868-873.
This paper represents the current state of the art, as the 6.8 Å structure of RDV is the highest resolution single-particle structure determined by electron cryomicroscopy. A combination of computational and bioinformatics tools was used to develop a rough backbone model for the inner and outer capsid proteins.

21. Miyazawa A, Fujiyoshi Y, Stowell M, Unwin N: Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel wall. *J Mol Biol* 1999, 288:765-786.

22. Li H, Lee S, Jap BK: Molecular design of aquaporin-1 water channel as revealed by electron crystallography. *Nat Struct Biol* 1997, 4:263-265.

23. Walz T, Hirai T, Murata K, Heymann JB, Mitsuoka K, Fujiyoshi Y, Smith BL, Agre P, Engel A: The three-dimensional structure of aquaporin-1. *Nature* 1997, 387:624-627.

24. Cheng A, van Hoek AN, Yeager M, Verkman AS, Mitra AK: Three-dimensional organization of a human water channel. *Nature* 1997, 387:627-630.

25. Unger VM, Kumar NM, Gilula NB, Yeager M: Three-dimensional structure of a recombinant gap junction membrane channel. *Science* 1999, 283:1176-1180.

26. Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y: Structural determinants of water permeation through aquaporin-1. *Nature* 2000, 407:599-605.

27. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M *et al.*: Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* 2000, 33:307-369.

28. Gabashvili IS, Agrawal RK, Spahn CM, Grassucci RA, Svergun DI, Frank J, Penczek P: Solution structure of the ***E. coli*** 70S ribosome at 11.5 Å resolution. *Cell* 2000, 100:537-549.

29. Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Horwich AI, Saibil HR: ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 2001, 107:869-879.

30. Ludtke SJ, Jakana J, Song J-L, Chuang D, Chiu W: A 11.5 Å single particle reconstruction of GroEL using EMAN. *J Mol Biol* 2001, 314:253-262.

31. •• Jiang W, Baker ML, Ludtke SJ, Chiu W: Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 2001, 308:1033-1044.
The tools described here show a practical and computational approach to identifying structural features within intermediate-resolution density maps. This approach is capable of identifying potential homologous folds or novel folds when used in conjunction with existing secondary structure matching techniques.

32. Mitsuoka K, Murata K, Walz T, Hirai T, Agre P, Heymann JB, Engel A, Fujiyoshi Y: The structure of aquaporin-1 at 4.5-Å resolution reveals short alpha-helices in the center of the monomer. *J Struct Biol* 1999, 128:34-43.

33. •• Heymann JB, Engel A: Structural clues in the sequences of the aquaporins. *J Mol Biol* 2000, 295:1039-1053.
The authors present a substantial analysis of hydrophobic and conservation periodicity, and of correlated mutations of the amino acid sequences of the aquaporin protein family. The structural information from this sequence analysis is integrated with a 4.5 Å resolution electron crystallography map of aquaporin-1 to propose an arrangement for the six transmembrane helices in each aquaporin monomer.

34. • de Groot BL, Heymann JB, Engel A, Mitsuoka K, Fujiyoshi Y, Grubmuller H: The fold of human aquaporin 1. *J Mol Biol* 2000, 300:987-994.
The authors present a template-based correlation method for determining the directionalities of transmembrane α helices in maps at ~4 Å resolution. The fold of human aquaporin-1 was proposed based on integration of its 4.5 Å map with other structural information derived from sequence analyses.

35. Sui H, Han B-G, Lee JK, Walian P, Jap BK: Structural basis of water-specific transport through the AQP1 water channel. *Nature* 2001, 414:872-878.

36. Kinoshita K, Kidera A, Go N: Diversity of functions of proteins with internal symmetry in spatial arrangement of secondary structural elements. *Protein Sci* 1999, 8:1210-1217.

37. Kleywegt GJ, Jones TA: Detecting folding motifs and similarities in protein structures. In *Methods in Enzymology*, vol 277. Edited by Charles W, Carter J, Sweet RM. London: Academic Press; 1997:525-545.

38. Mizuguchi K, Go N: Comparison of spatial arrangements of secondary structural elements in proteins. *Protein Eng* 1995, 8:353-362.

39. Grimes J, Basak AK, Roy P, Stuart D: The crystal structure of bluetongue virus VP7. *Nature* 1995, 373:167-170.

40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389-3402.

41. Fischer D, Eisenberg D: Protein fold recognition using sequence-derived predictions. *Protein Sci* 1996, 5:947-955.

42. • Kelley LA, MacCallum RM, Sternberg MJ: Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 2000, 299:499-520.
This paper describes the current best method for structure prediction in the category of fold recognition. In particular, this method is targeted at the identification of remote homologs using multiple sequence profiles and secondary structure prediction. Also of interest is that this program is available through a web interface (http://www.bmm.icnet.uk/~3dpssm/).

43. Lu G, Zhou ZH, Baker ML, Jakana J, Cai D, Wei X, Chen S, Gu X, Chiu W: Structure of double-shelled rice dwarf virus. *J Virol* 1998, 72:8541-8549.

44. Wriggers W, Chacon P: Modeling tricks and fitting techniques for multiresolution structures. *Structure* 2001, 9:779-788.

45. Wriggers W, Milligan RA, McCammon JA: Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 1999, 125:185-195.

46. Volkmann N, Hanein D: Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol* 1999, 125:176-184.

47. Baker ML, Jiang W, Zhou ZH, Dougherty M, Chiu W: Building a model for the HSV-1 major capsid protein using electron cryomicroscopy and bioinformatics. *Biophys J* 2002, 82:491a.