Eur Biophys J (2000) 29: 457–462

BIOPHYSICS LETTER

Susana G. Kalko · Mónica Chagoyen
Natalia Jiménez-Lozano · Nuria Verdaguer
Ignacio Fita · José M. Carazo

The need for a shared database infrastructure: combining X-ray crystallography and electron microscopy

Received: 26 January 2000 / Revised version: 15 May 2000 / Accepted: 15 May 2000

Abstract Advances in structural biology are opening greater opportunities for understanding biological structures from the cellular to the atomic level. Particularly promising are the links that can be established between the information provided by electron microscopy and the atomic structures derived from X-ray crystallography and nuclear magnetic resonance spectroscopy. Combining such different kinds of structural data can result in novel biological information on the interaction of biomolecules in large supramolecular assemblies. As a consequence, the need to develop new databases in the field of structural biology that allow for an integrated access to data from all the experimental techniques is becoming critical. Pilot studies performed in recent years have already established a solid background as far as the basic information that an integrated macromolecular structure database should contain, as well as the basic principles for integration. These efforts started in the context of the BioImage project, and resulted in a first complete database prototype that provided a versatile platform for the linking of atomic models or X-ray diffraction data with electron microscopy information. Analysis of the requirements needed to combine data at different levels of resolution have resulted in sets of specifications that make possible the integration of all these different types in the context of a web environment. The case of a structural study linking electron microscopy and X-ray data, which is already contained within the BioImage data base and in the Protein Data Bank, is used here to illustrate the current approach, while a general discussion highlights the urgent need for integrated databases.

S.G. Kalko · N. Verdaguer · I. Fita (✉)
IBMB (C.S.I.C.), Jordi-Girona 18-26, 08034 Barcelona, Spain
Fax: +34-93-2045904

M. Chagoyen · N. Jiménez-Lozano · J. M. Carazo
Centro Nacional Biotecnología (C.S.I.C.),
Universidad Autónoma de Madrid,
28049 Madrid, Spain

Key words X-ray crystallography · Electron microscopy · Biological databases

Introduction

Structural biology is undergoing an explosive expansion at all levels of resolution. At the atomic scale there is rapid growth in the number and complexity of models from biomolecules determined by nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography and also, in a few cases, electron crystallography. Equally important is the increasing improvement in the number and quality of 3D reconstructed images of molecular aggregates and cellular complexes obtained by electron microscopy. The information afforded by these two types of data brings the promise of visualizing complex biological structures at a continuum of scales from the cellular to the atomic level of organization. In the last few years the combination of X-ray crystallography and three-dimensional electron microscopy (3D-EM) has yielded substantial progress in studies of large viruses (Butcher et al. 1997; Dokland et al. 1999; Grimes et al. 1998; Prasad et al. 1999; Wikoff et al. 1997), viral complexes (Bella et al. 1998; Hewat et al. 1997) and important molecular aggregates such as muscle-related filaments (Steinmetz et al. 1998), amyloid fibrils (Jimenez et al. 1999), microtubules (Hirose et al. 1999; Kozielski et al. 1998), chaperones (Nitsch et al. 1998; White et al. 1997), proteasomes (Wang et al. 1997), membrane-bound molecules (Celia et al. 1999; Luo et al. 1999) and the ribosome (Ban et al. 1998; Cate et al. 1999). A survey of the current extent of data generation in combination 3D-EM/X-ray methods is shown in Fig. 1.

Despite the complementary nature of atomic models and related 3D images, the availability of these two types of structural data is very different, mainly because they are stored in different ways. Atomic models are routinely deposited and can be retrieved efficiently from

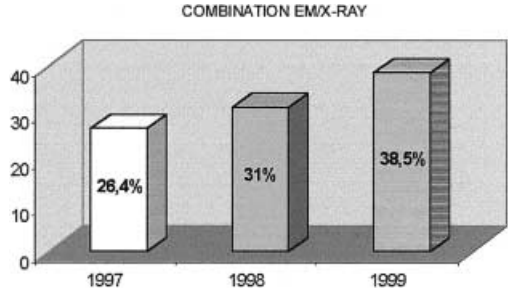

Fig. 1 The chart shows the percentage of 3D electron microscopy (3D-EM) studies where there is a combination with X-ray diffraction data over the total number of 3D-EM articles published. Note the increasing rate of combined studies during the past three years. A thorough bibliographic follow-up has been performed by personal inspection on a weekly basis in order to obtain the 3D-EM studies reported on 16 of the most popular journals in the field (around 200 articles containing close to 350 individual 3D structures have been considered)

a number of databases, including the Protein Data Bank (PDB) (Abola et al. 1997; Bernstein et al. 1977) by the entire scientific community. However, 3D reconstructed images are not systematically stored and have a limited access.

This undesirable situation motivated the proposal developed within the BioImage project (HIPERVÍNCULO, http://www.bioimage.org), aimed at providing a means of organization and access to microscopic data, as well as developing new capabilities to integrate different data sources – in the context of this work, 3D-EM and X-ray data – (Carazo and Steltzer 1999; Carazo et al. 1999). The central development was a modern database design that structured the data and the metadata in a unified manner (Lindek et al. 1999). In this way the BioImage project developed the prototype of a complete service, which included the submission of datasets, querying against certain key words as well as the retrieval of metadata and data. Also, special attention was given to the provision of appropriate visualization as well as data conversion tools to address these varied data sources (Pittet et al. 1999).

The present work emphasizes some of the possibilities that the BioImage database infrastructure offers within the context of high-resolution structural biology. Fitting of atomic models into 3D reconstructed images represents the most abundant application among the high-resolution combined studies, with a large and diverse number of important achievements. As a representative example of fitting studies, the interaction between foot and mouth disease virus with the Fab fragment of the neutralizing monoclonal antibody SD6 will be presented in some detail. Other advantages in combining X-ray diffraction data with 3D EM images will also be illustrated by briefly summarizing some relevant recent studies. The work initiated by BioImage will be further developed within the field of structural biology in close cooperation with the central data repositories, and more specifically with the EBI in Europe.

Steps towards an integrated resource for the structural biology of macromolecules

In the context of high-resolution structural biology, the first attempts to organize microscopic data and relate them to other types of data information such as X-ray studies were performed within the BioImage project. Specifically, BioImage provided a first implementation of a database that enabled structural searches and analyses of 3D volumes of molecules or molecular assemblies, in a similar way to how this is now done with atomic models in the PDB. The database opened the possibility of organizing a diversity of combined studies where 3D volumes were correlated with other types of structural information and, in particular, with molecular models. In recent years, research in high-resolution structural biology has increasingly required the combination of different experimental methodologies (Baker and Johnson 1996; Grimes et al. 1999). Many of these combined studies use atomic models determined by X-ray crystallography and 3D volumes obtained by microscopy, particularly cryo-electron microscopy (cryo-EM) together with image reconstruction techniques.

Studies that combine X-ray data and 3D-EM reconstructed images can be classified into three different categories according to the type of data combination used (Table 1):

1. Comparison/validation: this refers to studies where molecular structures or electron density maps are represented at lower than atomic resolution for a qualitative or a semi-quantitative comparison or validation of a 3D-EM volume.
2. Fitting: this category that, as indicated before, includes the largest number of high-resolution combined studies, consists in the docking of molecular structures into EM volumes, providing a pseudo-atomic resolution model according to the "divide and conquer" strategy. The information obtained in fitting studies ranges from the atomic modeling of a complex molecular assembly to the characterization of local movements within large macromolecules.
3. Phasing: this includes studies where 3D images are used as search models for the crystallographic phase determination by some variation of the molecular replacement method used extensively in crystallography (Rossmann 1990).

A feature common to all the studies referred to above, and that clearly shows in the context of database design, is that, besides the 3D-EM volume data in themselves produced by 3D-EM, some complementary information had also to be provided in order to be able to address the diversity of applications and analysis envisioned. The complementary information varies according to the experimental method used to obtain the volume data, in particular, according to the level of resolution attained.

Table 1 Representative combined studies of 3D-EM reconstructed images with X-ray data (1997–1999)

Combination	Ref.	Remarks
Comparison	Wikoff et al. (1997)	Cucumber mosaic virus at 23 Å by EM agreed with X-ray data at 8 Å. Similarities with cowpea chlorotic mottle virus were demonstrated, and alignment of their sequences gave a probable distribution of residues in 3D structure of CMV
	Butcher et al. (1997)	*Pseudomonas* dsRNA phage 6 polymerase complex solved at 20 Å compared with blue tongue virus (BTV) X-ray data core shows striking similarities, supporting $T = 2$ arrangement. Functional correspondence was suggested
	Dokland et al. (1999)	X-ray structure of the closed procapsid of DNA virus ϕX174 solved at 3.5 Å was compared with the previously determined cryo-EM image of the open procapsid at 25 Å. The major structural rearrangements that must occur during assembly were derived
Fitting	Grimes et al. (1997)	The structure of the $T = 13$ VP7 layer of BTV was generated by fitting the X-ray crystal structure of individual VP7 trimers into the core particles determined by EM at 23 Å
	Hewat et al. (1997)	The structure of the FMDV-SD6 Fab complex at about 30 Å was determined by fitting the X-ray crystal structures of its components (see details in the main text and in Fig. 2)
	Nitsch et al. (1998)	The α-only thermosome map at 28 Å provides a first view on the open state of a group II chaperonin. Fitting of the atomic structures into the electron microscopy density allowed derivation of the domain rearrangement
	Malhotra et al. (1998)	The high quality 3D EM map of *Escherichia coli* 70 S ribosome at 15 Å provided important clues, in particular allowing the docking of the crystal structures from a bound tRNA molecule and from the L1 protein
	Kozielski et al. (1998)	The kinesin motor domain dimer-microtubule complex was determined at about 35 Å. Using the X-ray coordinates of the monomeric human kinesin motor domain, an atomic resolution of the motor complex was built
	Luo et al. (1999)	The 3D EM image of the insulin-insulin receptor complex at 20 Å allowed the fitting of the available atomic domain substructures which provided a detailed model of this transmembrane receptor
Phasing	Wang et al. (1997)	The 2.3 Å X-ray structure of the proteolytic component of the caseinolytic protease from *E. coli* was solved using as initial phases a cylindrical model determined by EM and solution small-angle X-ray scattering
	Ban et al. (1998)	The large ribosomal subunit from *Haloarcula marsmortui* was solved at 9 Å by a combination of data from multiple isomorphous replacement and anomalous scattering together with a molecular replacement solution using as a searching model a 20 Å EM map
	Grimes et al. (1998)	The 3.5 Å X-ray structure of the core particle of BTV (700 Å diameter) was solved by molecular replacement and phase extension using as a starting model the VP7 shell constructed using EM and X-ray data in a fitting experiment (see Grimes et al. 1997)
	Cate et al. (1999)	Structures of 70S ribosome functional complexes solved at 7.8 Å. Molecular replacement on a 25 Å cryo-EM single particle reconstruction of vacant *E. coli* ribosomes gave initial phases in the model
	Prasad et al. (1999)	The crystal structure of a calicivirus was determined at 2.3 Å by molecular replacement and phase extension using as a starting model the EM volume determined at 3 Å resolution

The case of a combined study currently in BioImage

To illustrate the diverse requirements that fitting studies demand in order to be efficiently stored in BioImage, we present the case of the complex between foot and mouth disease virus serotype C (FMDV-C) and a strongly neutralizing monoclonal antibody (SD6). This case is of interest because all the required data are now deposited in BioImage and in the PDB. The atomic model of this complex was generated using the 3D map determined by cryo-EM to 30 Å resolution (Hewat et al. 1997). The X-ray atomic structures were available for both the intact particle of FMDV-C and the SD6 Fab co-crystallized with a synthetic peptide corresponding to the dominant antigenic loop, the GH loop, from the viral capside protein 1 (VP1). An atomic model of the entire complex was obtained by docking the two crystallographic structures into the reconstructed EM map. The 3D reconstructed image of the FMDV-Fab complex was deposited in the BioImage database in the MRC map format, with ID 12. Also BioImage contains the links to the available coordinates in the PDB and the transformation matrices necessary to generate the new complex. Thus, when retrieving the 3D volume information from BioImage, the corresponding atomic model can be derived (Fig. 2).

We conclude from the work described above that, additionally to the EM maps and the primary links to the PDB, the assembly matrices needed to construct a macromolecule from its basic parts as well as the matrices to dock the X-ray data on EM maps are required to reproduce the atomic model of the complex. Thus, these complementary data need to be submitted for such combined studies. The scheme shown in Fig. 3a represents a possible submission data flow. Fitting and assembly matrices are defined as compulsory and will allow the database user to know the precise geometric relations between the different structures involved. This additional information has to be entered during data submission, but it will be automatically generated in the near future by the fitting programs themselves. The actual submission step for this information is shown in Fig. 3b.

Additionally, a first step into the real integration between EM and X-ray data has been done: we have

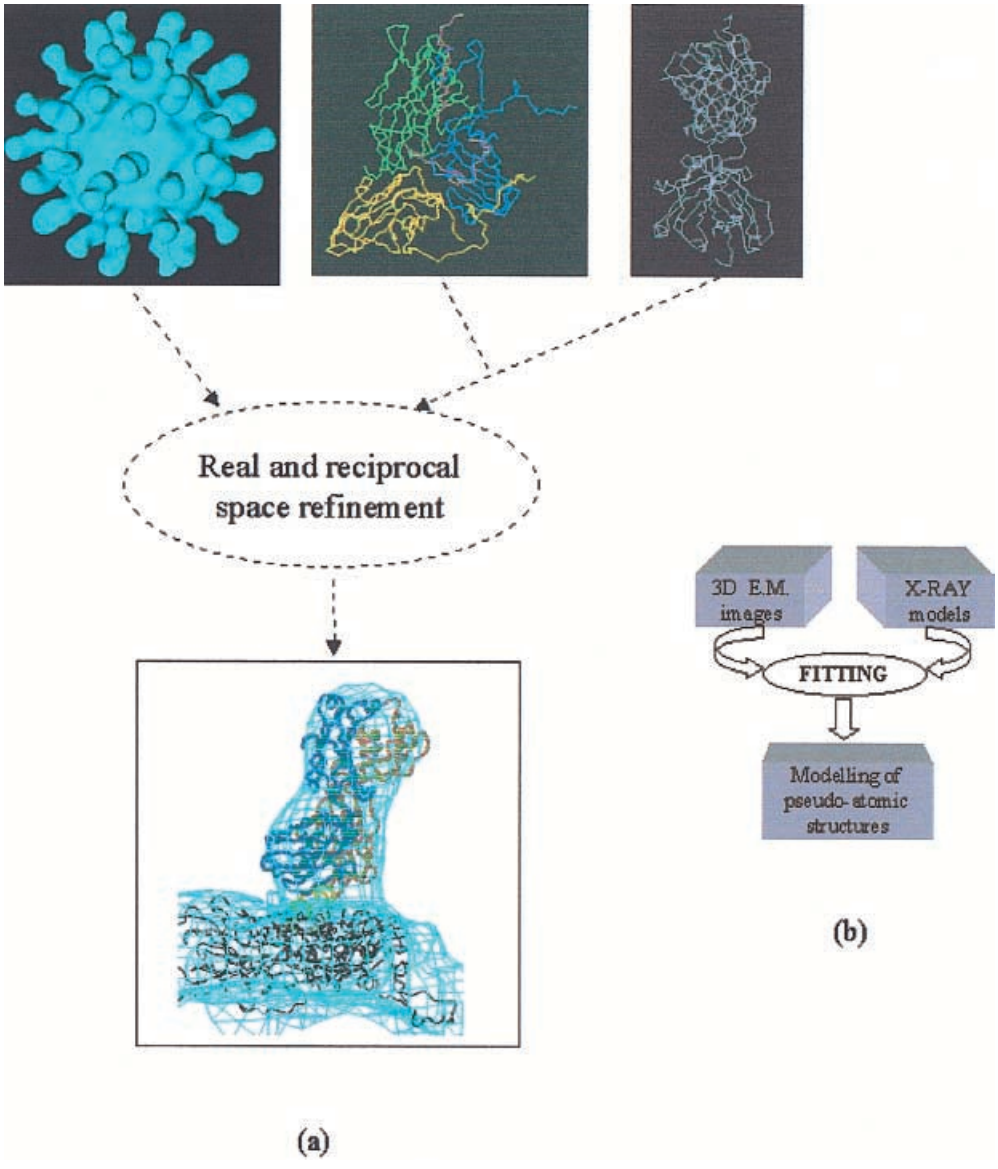

Fig. 2a, b Docking study of the FMDV-Fab complex. **a** The *top part* of the figure shows the 3D reconstructed image of the virus-antibody complex (*left*) and the atomic structures of both the FMDV capside protomer (*middle*) and the Fab fragment of the complex of the monoclonal antibody SD6 with the cognate peptide (*right*). The *bottom part* of the figure shows the result of the fitting procedure. The relative disposition of the Fab in the complex allowed the modeling of residues situated in the hinge of the epitopic loop (see reference for details). **b** Scheme of the general path followed in most docking studies. Fitting matrices that define the movements to be applied to the different PDB structures are stored in BioImage together with the corresponding links to the PDB (see text and Fig. 3). In turn, the new coordinates (like the ones corresponding to the hinge in the example presented) derived from this modeling are proposed to be deposited in the PDB, indicating their origin and peculiarities (see text)

submitted to PDB the coordinates of the FMDV-SD6 complex, with ID 1QGC. This submission represents the first structure in PDB pointing to a 3D volume in BioImage. We have stated the fundamentals for a new protocol in the frame of the PDB for this kind of study, according to the previously defined protocol in BioImage.

Closing the circle: the need for tighter links between structural databases

Considering the combined structural studies analyzed in the context of developing BioImage, two questions arise. These may have key strategic implications for the further relationships between low-resolution structural databases (such as BioImage) and atomic resolution databases (such as PDB).

Should the atomic resolution database reference low/medium resolution (3D-EM) databases? An increasing number of submissions in the PDB have used EM data as referred above and in Table 1. A link from the PDB to the appropriate low/medium resolution entries would therefore be essential to convey a complete picture of the actual experimental data used to derive that PDB entry.

Should the atomic resolution database contain fitted atomic models? Fitting studies, particularly those which allow flexibility in the fitted molecule, generate "pseudo-atomic" models. These data have no precise resolution value, but in the light of their increasing number and wider use, it seems convenient to have them available in the database, providing the appropriate links to different pieces of information actually used. This would increase the quantity of useful and manageable structural information and enhance the power of the links described here.

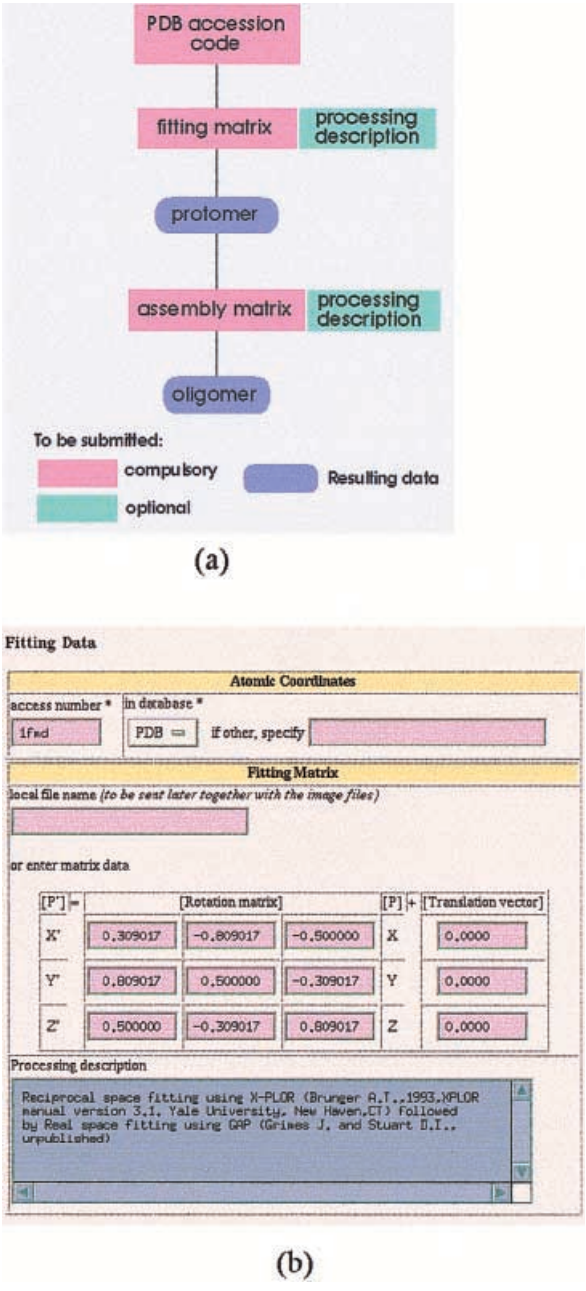

Fig. 3 **a** Complementary information to be added in a combined study: fitting and assembly matrices, proposed as compulsory data in the BioImage design. This information allows the actual oligomer merged in the 3D reconstruction image to be assembled from atomic models of the protomers deposited in the PDB. **b** Submission interface on BioImage where a fitting matrix of the docking study considered in Fig. 2 is shown. This matrix corresponds to the merging of the viral capside protomer onto the 3D reconstruction image of the virus-antibody complex

This forward and backward linking, from and into low/medium resolution databases and other structural databases (with PDB as the major one), will generate a close relation between all the deposited structures, from low to high resolution. In the light of the experience generated in the context of this first prototype work with BioImage, it is now clear that future efforts towards data integration from the structural databases point of view should address issues such as the submission and organization of (1) 3D images, (2) basic geometric relations, as matrices to merge such images with atomic models, and (3) links to the repositories of such atomic models.

Conclusion

In conclusion, the BioImage project has developed a first approach aimed at filling a "gap" as far as the existence of appropriate database infrastructures for structural biology is concerned. The 3D electron microscopy field clearly needs an infrastructure similar in concept to what PDB represents for data at atomic resolution, since in most cases atomic resolution will not be directly obtained by 3D-EM. Similarly, this first step towards the storage and organization of X-ray/EM combined studies has proven the validity and usefulness of an integrated approach.

Acknowledgements We are very grateful to A. Engel and B. Heymann for their helpful comments and suggestions. Many thanks are due to Elizabeth Hewat for facilitating the EM data of the FMDV-SD6 complex. Initial support from the European Molecular Biology Network, as well as the many discussions and proof-of-concept work in the field of macromolecular structure performed in collaboration with Dr. Joachim Frank (Albany, USA), is acknowledged by J.C. Work on this project has been supported by the European Union through grant PL 960472. Special actions or extensions are being made possible by core national grants (BIO95-0768, BIO97-1485-CE and BIO98-0761, to J.C., and PB95218 to I.F.).

References

Abola EE, Sussman JL, Prilusky J, Mannin NO (1997) Protein Data Bank archives of three-dimensional macromolecular structures. Methods Enzymol 277: 556–571

Baker TS, Johnson JE (1996) Low resolution meets high: towards a resolution continuum from cells to atoms. Curr Opin Struct Biol 6: 585–594

Ban N, Freeborn B, Nissen P, Penczek P, Grassucci RA, Sweet R, Frank J, Moore PB, Steitz TA (1998) A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. Cell 93: 1105–1115

Bella J, Kolatkar PR, Marlor CW, Greve JM, Rossmann MG (1998) The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. Proc Natl Acad Sci USA 95: 4140–4145

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouci T, Tasumi M (1977) The Protein Data Bank: a computer-based archival for macromolecular structures. J Mol Biol 112: 535–542

Butcher SJ, Dokland T, Ojala PM, Bamford DH, Fuller SD (1997) Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. EMBO J 16: 4477–4487

Carazo JM, Stelzer EHK (1999) The BioImage database project: organizing multidimensional biological images into a new database. J Struct Biol 125: 166–175

Carazo JM, Stelzer EHK, Engel A, Fita I, Henn C, Machtynger J, McNeil P, Shotton DM, Chagoyen M, de Alarcón PA, Fritsch R, Heymann JB, Kalko S, Pittet JJ, Rodriguez-Tomé P, Boudier T (1999) Organising multidimensional biological

image information: the BioImage database. Nucleic Acids Res 27: 280–283

Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF (1999) X-ray crystal structures of 70S ribosome functional complexes. Science 285: 2095–2104

Celia H, Wilson-Kubalek E, Milligan RA, Teyton L (1999) Structure and function of a membrane-bound murine MHC class I molecule. Proc Natl Acad Sci USA 96: 5634–5639

Dokland T, Bernal RA, Burch A, Pletnev S, Fane BA, Rossmann MG (1999) The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus ϕX174. J Mol Biol 288: 595–608

Grimes JM, Jakana J, Ghosh M, Basak AK, Roy P, Chiuu W, Stuart D, Venkataram Prasad BV (1997) An atomic model of the outer layer of the blue tongue virus core derived from X-ray crystallography and electron microscopy. Structure 5: 885–893

Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PPC, Stuart DI (1998) The atomic structure of the blue tongue virus core. Nature 395: 470–478

Grimes JM, Fuller SD, Stuart DI (1999) Complementing crystallography: the role of cryo-electron microscopy in structural biology. Acta Crystallogr Sect D 55: 1742–1749

Hewat EA, Verdaguer N, Fita I, Blakemore W, Brooks S, King A, Newman J, Domingo E, Mateu MG, Stuart D (1997) Structure of the complex of a Fab fragment of a neutralizing antibody with foot-and-mouth disease virus: positioning of a highly mobile antigenic loop. EMBO J 16: 1492–1500

Hirose K, Löwe J, Alonso M, Cross RA, Amos L (1999) Congruent docking of dimeric kinesin and ncd into three-dimensional electron cryomicroscopy maps of microtubule-motor ADP complexes. Mol Biol Cell 10: 2063–2074

Jimenez JL, Guijarro JI, Orlova E, Zurdo J, Dobson CM, Sunde M, Saibil HR (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. EMBO J 18: 815–821

Kozielski F, Arnal I, Wade RH (1998) A model of the microtubule-kinesin complex based on electron cryomicroscopy and X-ray crystallography. Curr Biol 8: 191–198

Lindek S, Fritsch R, Machtynger J, de Alarcón PA, Chagoyen M (1999) Design and realization of an on-line database for multidimensional microscopic images of biological specimens. J Struct Biol 125: 103–111

Luo RZ, Beniac DR, Fernandes A, Yip CC, Ottensmeyer FP (1999) Quaternary structure of the insulin-insulin receptor complex. Science 285: 1077–1080

Malhotra A, Penczek P, Agrawal RK, Gabashvili IS, Grassucci RA, Jünemann R, Burkhardt N, Nierhaus KH, Frank J (1998) *Escherichia coli* 70 S ribosome at 15 Å resolution by cryo-electron microscopy: localization of fMet-tRNA$_f^{Met}$ and fitting of L1 protein. J Mol Biol 280: 103–116

Nitsch M, Walz J, Typke D, Klumpp M, Essen LO, Baumeister W (1998) Group II chaperonin in an open conformation examined by electron tomography. Nat Struct Biol 5: 855–857

Pittet J-J, Henn C, Engel A, Heymann JB (1999) Visualizing 3D data obtained from microscopy on the internet. J Struct Biol 125: 123–132

Prasad BVV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK (1999) X-ray crystallographic structure of the Norwalk virus capsid. Science 286: 287–290

Rossmann MG (1990) The molecular replacement method. Acta Crystallogr Sect A 46: 73–82

Steinmetz MO, Hoenger A, Tittmann P, Fuchs KH, Gross H, Aebi U (1998) An atomic model of crystalline actin tubes: combining electron microscopy and X-ray crystallography. J Mol Biol 278: 703–711

Wang J, Hartling JA, Flanagan JM (1997) The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. Cell 91: 447–456

White HE, Chen S, Roseman AM, Yifrach O, Horovitz A, Saibil HR (1997) Structural basis of allosteric changes in the GroEl mutant Arg197→Ala. Nat Struct Biol 4: 690–694

Wikoff WR, Tsai CJ, Wang G, Baker T, Johnson JE (1997) The structure of cucumber mosaic virus: cryoelectron microscopy, X-ray crystallography, and sequence analysis. Virology 232: 91–97