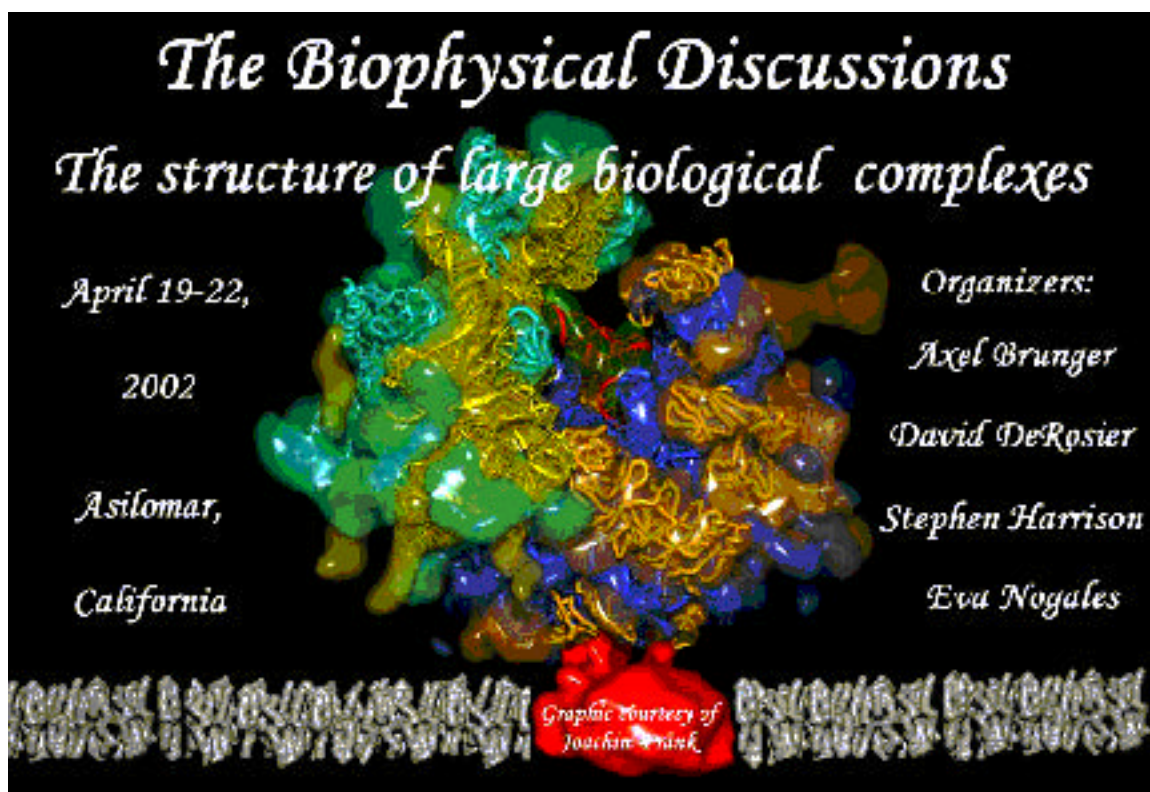# Biophysical Society Discussions

Frontiers in Structural Cell Biology:
How Can We Determine the Structures of Large Subcellular Machines
at Atomic Resolution?

April 19-22, 2002
Asilomar, California

Study Book

2002 Biophysical Discussions
April 19–22, 2002

DAILY SCHEDULE

	FRIDAY, APRIL 19	SATURDAY, APRIL 20	SUNDAY, APRIL 21	MONDAY, APRIL 22
7:30 AM		Breakfast *Crocker Dining Room*	Breakfast *Crocker Dining Room*	Breakfast *Crocker Dining Room*
9:00 AM		Session I: The State of Structural Biology of Large Structures *Chapel*	Session III: What Does the Future Hold for Electron Cryomicroscopy? *Chapel*	Session V: Computational Models *Chapel*
10:00 AM		Coffee Break *Chapel*	Coffee Break *Chapel*	Coffee Break *Chapel*
12:00 NOON		Lunch *Crocker Dining Room*	Lunch *Crocker Dining Room*	Checkout/Lunch *Crocker Dining Room*
3:00 PM	Check-in			
4:30 PM	Wine & Cheese Reception *Chapel*	Poster Session/ Cash Bar *Fred Farr Forum* & *Kiln Room*	Poster Session/ Cash Bar *Fred Farr Forum* & *Kiln Room*	
6:00 PM	Dinner *Crocker Dining Room*	Dinner *Crocker Dining Room*	Dinner *Crocker Dining Room*	
7:30 PM	Poster Set-up/Cash Bar *Fred Farr Forum* & *Kiln Room*	Session II: Extending X-Ray Crystallography to Ever Larger Structures *Chapel*	Session IV: Can Hybrid Methods Provide Credible Atomic Models? *Chapel*	

Biophysical Society Discussions

Frontiers in Structural Cell Biology:
How Can We Determine the Structures of Large Subcellular
Machines at Atomic Resolution?

Program

SUNDAY 9 AM–12 NOON

Session III: What Does the Future Hold for Electron Cryomicroscopy?

Moderator *Bob Glaeser*, University of California, Berkeley

SUNDAY 4:30 PM–6:00 PM

SUNDAY 7:30 PM–10:30 PM

Session IV: Can Hybrid Methods Provide Credible Atomic Models?

Moderator *Eva Nogales*, University of California, Berkeley

MONDAY 9:00 AM–12:00 NOON
Session V: Computational Methods
Moderator *Axel Brunger*, Stanford University

Speaker 1
Richard Henderson

No Speaker Paper Available

THE RIBOSOME – A MOLECULAR MACHINE IN MOTION

Joachim Frank; HHMI, Health Research, Inc. at Wadsworth Center, Empire State Plaza, Albany, New York 12201-0509.

Following the recent publication of the X-ray structures of ribosomal subunits (Wimberly et al., 2000; Ban et al., 2000; Yusupov et al., 2000; Schlünzen et al., 2000) and the 70S ribosome (Yusupov et al., 2001) – in Harry Noller's words, "the whole enchilada" --, the main focus has been on attempts to understand the workings of the functional centers: decoding center and peptidyltransferase center, the former being the site of codon-anticodon interaction, the latter, the site where a new peptide bond is catalyzed linking the polypeptide chain to the new-coming amino acid. In each case, the focus is quite limited both in space and time: in space, molecular interactions are being investigated in a very small region; in time, attention is confined to a narrow time period. However, what we wish to do is following the motion of the molecular machine in its entirety through the whole elongation cycle, the cycle during which a new amino acid is added to the nascent polypeptide chain.

There are intrinsic limitations of X-ray crystallography (need for crystallization, and spatial constraints produced by crystallization), which let us expect that cryo-electron microscopy will make significant contributions toward an understanding of the dynamics of the ribosome during elongation, as well as the other major processes of translation: initiation and termination. To date, in fact, the visualization of factor-ribosome interaction (Stark et al., 1997; Agrawal et al., 1998; 1999; Gomez-Lorenzo et al., 2001) and associated conformational dynamics of the ribosome (Agrawal et al., 1999; Frank and Agrawal, 2000; Valle et al., in prep.) have been done exclusively by cryo-EM. To be sure, the interpretations of the density maps in terms of molecular interactions and conformational changes all draw on structures obtained by X-ray crystallography and NMR, if necessary re-modeled by changing domain orientations around known "flex points."

Dynamic information is obtained by collecting data for 3D reconstruction in a series of "snapshots" at points that are closely spaced. All experiments that have been done to date (such as the works cited in above paragraph) rely on antibiotics or non-cleavable GTP analogs to stall the ribosome. The advantage is that the corresponding states of the system are well defined and highly populated, but the disadvantage is that the time points are preordained and cannot be chosen at will. We are currently making efforts toward physical trapping of intermediates by using a spray-freezing apparatus (White et al., 1998).

In the following I will make three contributions to the Discussions: (i) a consideration of obstacles to achieving higher resolution; (ii) experimental results obtained with the Philips/FEI F30 Tecnai with liquid He and N cooling; (iii) the most recent results on the structure of the ribosome, and under what conditions they were obtained.

1. What keeps us from achieving higher resolution?

In aiming at high resolution, we face a series of bottlenecks which we can categorize in terms of instrumental limitations, radiation damage, specimen heterogeneity, conformational variability, demanding data collection, and computational limitations. The bottleneck presented by ***Instrumentation*** has been shifted dramatically toward higher resolution with the addition of computer-controlled electron microscopes equipped with a field emission gun and high stability. First results with our new 300-kV instrument are reported below.

Radiation damage is a physical reality that cannot be controlled, except that there is hope for a reduction of beam sensitivity at temperatures close to liquid helium (Chiu et al., 1986). Radiation damage obviously directly interferes with the attempt to produce a three-dimensional image of the undamaged structure, but its effect is also indirect, since any compensatory dose reduction leads to more severe statistical limitations in all numerical operations aimed at comparing projection data, e.g., alignment and classification, which ultimately translate into lower resolution. Conversely, any reduction of radiation damage achieved by lowering the specimen temperature could be utilized by increasing the dose, making the statistical support of alignment operations more robust, thereby increasing the resolution of the end product.

Specimen heterogeneity is often associated with ligand binding experiments in which less than 100% occupancy is achieved. Consequently, density associated with the ligand is "diluted" in proportion to the occupancy ratio, leading to less certainty in the subsequent demarcation of molecular boundaries and fitting of X-ray structures. What is worse, particularly in the case of the ribosome, the ligand binding induces conformational changes in the host structure, which means that the entire density map is invalid since it is the result of a conformational mix. Low binding occupancy is often the result of the low sample concentration required for cryo-EM, which drives the equilibrium of the system toward a state where the ligand is dissociated. The problem might be overcome by a technique of specimen preparation whereby the sample is rapidly diluted at low temperature (~4° C) immediately prior to freeze-plunging. Another way to deal with specimen heterogeneity is "after the fact," by image classification. In Figs. 1 and 2, there is an example for the successful separation of a ribosome population

into subsets that are either bound with the aatRNA.EF-Tu.GTP ternary complex or unoccupied; in this case, the separation was achieved by using two 3D references, one bound with EF-G in the GDP form and one empty, taking advantage of the high similarity between the structures of EF-G and the ternary complex and mode of binding to the ribosome. Smaller ligands pose larger problems, however, making the separation much more fuzzy and ambiguous.
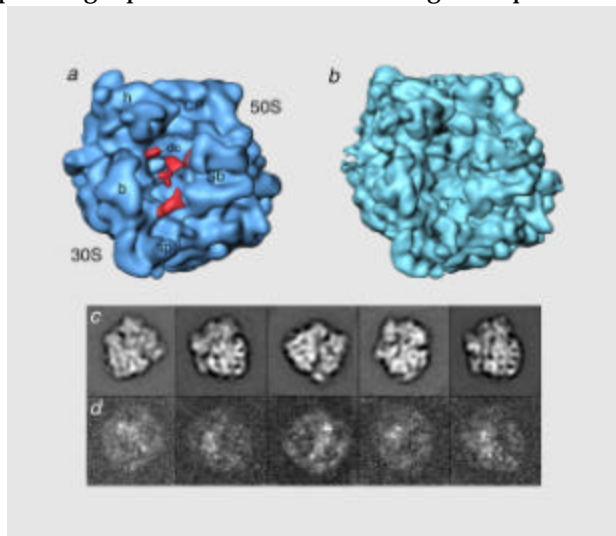
Fig. 1: (a) Cryo-EM map of ribosome bound with ternary complex (density in red) and (b) control. Mass due to ternary complex is scattered because of limited occupancy. (c,d) Averages and corresponding variance maps of ribosomes in selected orientations. Variance maps indicate high variability due to inconsistent occupancy in the intersubunit space. (from Valle et al., submitted).
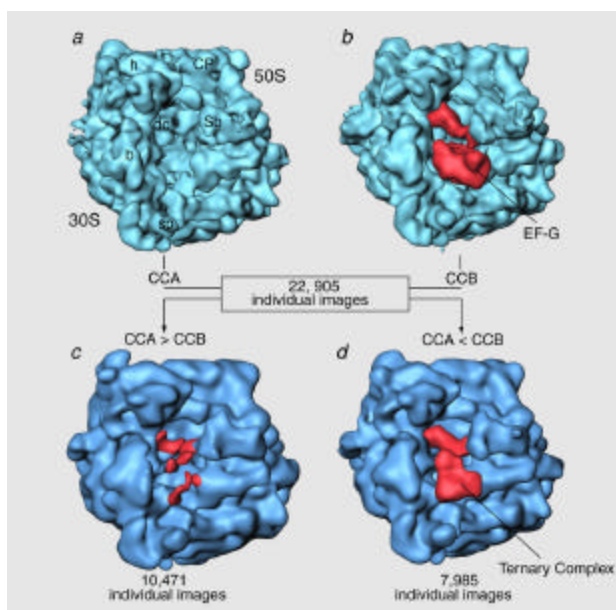
Fig. 2: Separation of particles based on similarity with two references [(a) empty and (b) bound with EF-G] as measured by cross-correlation (CCA and CCB). This separation leads to two volumes: (c) reconstruction from particles most similar to empty ribosome: only small pieces of mass remain, (d) reconstruction from particles most similar to EF-G.ribosome complex: ternary complex is recovered with high definition. (from Valle et al., submitted).

Conformational variability and the ensuing heterogeneity of the particle population may also occur without being induced by partial ligand binding, simply because flexible components of the molecule stick out into the solvent without mechanical support. In the ribosome, the most obvious example is presented by the two L7/L12 dimers, one of which forms the "stalk" of the 50S subunit, while the other is probably folded over. The presence of a long linker region in these proteins means that the distal C-terminal domains can "visit" different remote sites on the ribosome, as found by cross-linking studies and recently by Nanogold labeling and cryo-EM (Montesano-Roditis, 2001). The chances of obtaining high-resolution structural information by cryo-EM on the shape and position of the stalk are therefore remote. Other examples of structures "sticking out" and prone to variability induced during the preparation of the grid are proteins L9 and L1. Since evidence of

such variability is more likely with increasing resolution, as smaller and smaller structural features come into view, we can speak of a "conformational hierarchy," represented by a tree that sprouts multiple branches as we proceed toward atomic resolution. This has implications for the amount of data required to be collected: at a given level of resolution, we might be able, in principle, to distinguish 5 conformers occurring with equal frequency. This would mean that to reconstruct any of the 5 with that resolution, the amount of data collection would have to be five times the amount necessary for a completely homogeneous specimen.

Data collection must be expanded by a large factor in order to realize higher resolution. It is clear that current methods of data acquisition (via recording on film and microdensitometry) are inadequate. While software for automated data collection is becoming generally available, the field size covered by 2k x 2k CCD cameras is still marginal for single particle work, so the ten-fold or so increase in data volume over the current ~100,000 particles (see Gabashvili et al. (2000) and below) likely to be needed for significant progress toward atomic resolution must await availability of larger arrays.

Computational limitations to such explosive increase in data can be overcome with massively parallel computing schemes or use of distributed processing via the "publish and subscribe" mode. Currently we are working on modifications and expansions of the SPIDER software that will take full advantage of these new opportunities.

2. Experiments with the Philips/FEI F30 Tecnai electron microscope

The F30 now installed in Albany reaches 11.5° K as measured in close vicinity to the specimen. The most striking feature of the instrument is the extraordinary stability of the specimen stage, which measures in Angstroms per minute even at high tilt. While it is too early to say anything definite on the advantages of He cooling for single particle work, the data produces by the instrument with ribosome specimens are clearly superior for *both* liquid He and N cooling compared with those obtained from the F20 (see power spectrum reproduced in Fig. 3). The fact that the instrument can be operated at both
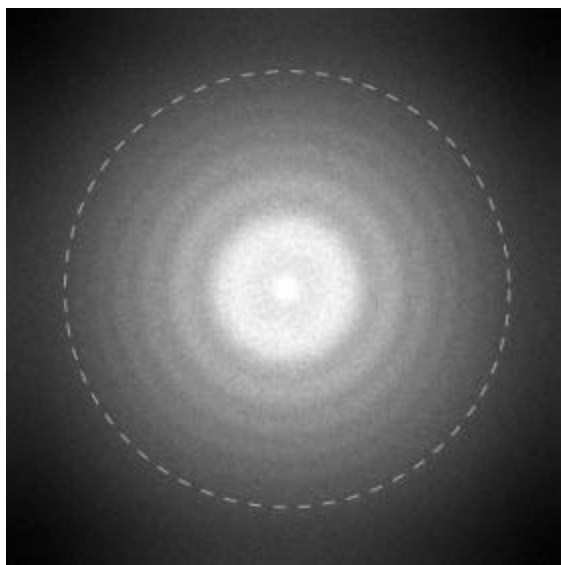
Fig. 3: Power spectrum of a micrograph of a frozen-hydrated ribosome specimen obtained with the F30. Thon rings are visible out to 6 Å.

temperatures, and converted from one mode to the other easily, makes it possible to perform experiments that test temperature dependency of a phenomenon while keeping all other parameters exactly constant. In a first such experiment, we have looked at ribosome specimens in order to see if the contrast in ice is affected by the temperature change. There exist some concern that the contrast might decrease since a significant increase of ice scattering density has been reported to take place around 30° K (Heide and Zeitler, 1985; Jenniskens and Blake, 1994). Micrographs were recorded for a range of defocus settings, from 1 to 4.5 •m). Particles were processed and classified into 83 standard orientation groups. Averages of particles falling into one of the orientations were analyzed, by averaging the optical density within the particle boundary and comparing the result with the average of the ice background. Within experimental error, no contrast change was found over the whole defocus range (Grassucci et al., 2002).

Data collected with the new instrument were used in a most recent reconstruction of the *E. coli* ribosome from ~86,000 particles (C.M.T. Spahn, R.A. Grassucci, K.H. Nierhaus, and J. Frank, work in progress) that will be outlined below.

3. Cryo-EM of ribosomes: most recent progress

The ribosome complex ("pre-translocational complex", obtained from Knud Nierhaus, Berlin) was programmed with poly(U) and contained Phe-tRNA[phe] in the A site and AcPhe-N-tRNA[phe] in the P site, a "trick" that prohibits peptidyl transfer as well as any following spontaneous translocation, thus leading to high occupancy of the A site with tRNA, and a high conformational purity of the complex. The complex poses no problems in terms of occupancy, and is therefore exceptionally well suited for high-resolution studies. 70% of the 86,000 particles were collected on the F30 (M= 39,000 and 59,000; U= 300kV), the rest on the F20 Tecnai. Defocus groups ranged from 0.5 to 6 •m (high/low defocus micrograph pairing was used to identify particles at defocus settings <1.5 •m). The resulting 10.5-Å density map was amplitude-corrected using low-angle X-ray solution scattering data, as was done for the previously published 11.5-Å reconstruction (Gabashvili et al., 2000).

In the comparison (Fig. 4) with the previous map it is striking to see how much better protein shapes and RNA helices are defined in the new map, even though the nominal resolution is only improved by one Ångstrom. The clue to this paradox is provided by the comparison of the FSC curves (Fig. 5), which show that the 11.5-Å data fall off much earlier than the new data, indicating a high definition (and high SNR) up to the cutoff point. This demonstrates the importance of data quality even below the cutoff point, and also drives the point home that a single value is insufficient to characterize the quality of a reconstruction (see Frank et al., 2002).
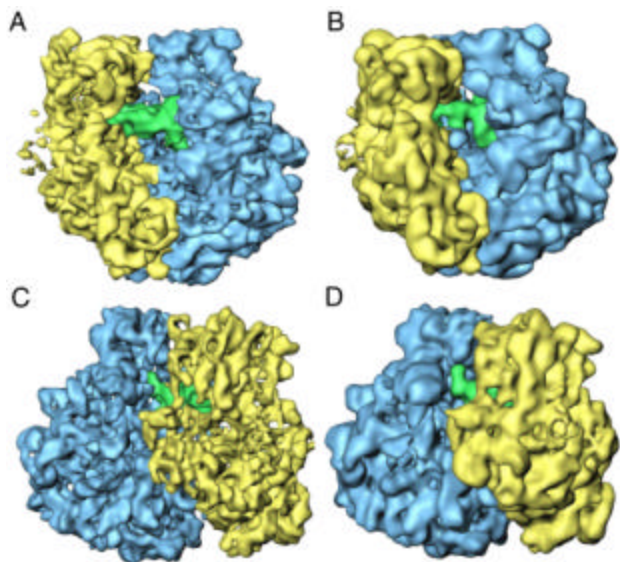
Fig. 4: Cryo-EM reconstructions of (A,C) pre-translocational ribosome complex at 10.5A resolution (Spahn et al., unpublished results); and (B.D) fMet-tRNA.ribosome complex at 11.5Å resolution (Gabashvili et al., 2000). 30S subunit in yellow, 50S subunit blue, tRNAs in green. In (A,C) both A and P site are solidly occupied with tRNA.
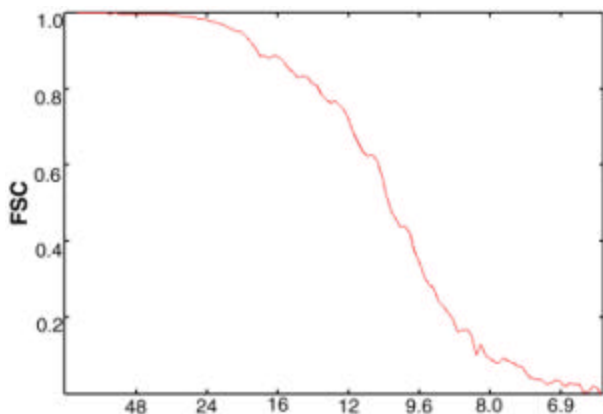
Fig. 5a: FSC for new 10.5-A reconstruction obtained by Spahn et al. (unpublished).
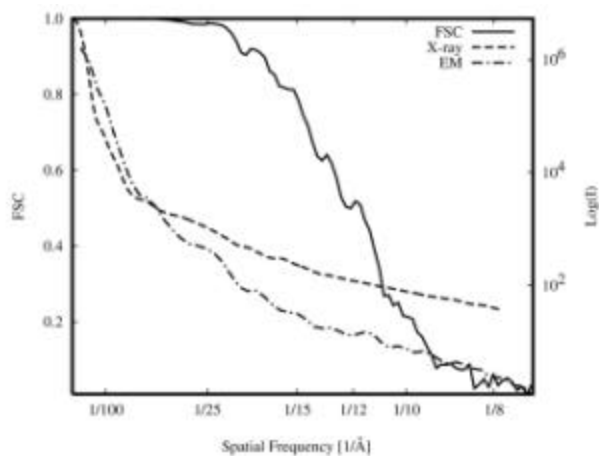
Fig. 5b: FSC for 11.5-Å reconstruction obtained by Gabashvili et al. (2000).

References

1. Agrawal, R.K. et al. (1998) PNAS 95:6134-6138.

2. Agrawal, R.K. et al. (1999) Nature Struct. Biol. 6:643-647.

3. Ban, N. et al. (2000) Science 289:905-920.

4. Chiu et al. (1986) J. Microsc. 141:385-391.

5. Frank, J. (2002) Ann. Rev. Biophys. Biomol. Str. Vol 31, in press.

6. Frank, J. and Agrawal. R.K. (2000) Nature 406:318-322.

7. Gabashvili, I.S. et al. (2000) Cell 100:537-549.

8. Gomez-Lorenzo et al. (2001) EMBO J. 19:2710-2718.

9. Grassucci, R.A. et al. (2002) abstract subm. to MSA Meeting in Quebec City, in press.

10. Heide, H.G. and Zeitler, Z. (1985) Ultramicrosc. 16: 151-160.

11. Jenniskens, P. and Blake, D.F. (1994) Science 265: 753-756.

12. Montesano-Roditis et al. (2001) J. Biol. Chem. 276:14117-23.

13. Schlünzen et al. (2000) Cell 102:615-

14. Stark, H. et al. (1997) Nature 389:403-

15. Stark, H. et al. (2000) Cell 100:301-

16. White et al. (1998) J. Struct. Biol. 121:306-313.

17. Wimberly et al. (2000) Nature 407:327-339.

18. Yusupov et al. (2001) Science 292:883-896.

BIOCHEMICAL BASIS FOR X-RAY CRYSTALLOGRAPHY OF THE RIBOSOME.

Jamie Cate, University of California, Berkeley

Protein synthesis in all organisms is carried out by the ribosome, a large ribonucleoprotein complex that is over 20 nm in each dimension. The mass of the bacterial ribosome (70S sedimentation) is approximately 2.5 MDa, and consists of a small (30S) and a large (50S) subunit composed of about two-thirds RNA (Green and Noller, 1997). Protein synthesis, or translation, involves reading the genetic code in messenger RNA (mRNA) by means of adaptor molecules, transfer RNAs (tRNAs) that couple each amino acid to its corresponding base triplet in the genetic code. Many cofactors assist the ribosome in each stage of translation, including initiation factors, elongation factors, and termination factors. The ribosome itself undergoes many conformational changes, still poorly understood, in each stage of translation (Frank, 2001b; Ramakrishnan, 2002).

Recent advances in our understanding of the structural basis for protein synthesis by the bacterial ribosome have come primarily from two highly complementary approaches, cryo-electron microscopy and x-ray crystallography (Frank, 2001a; Frank, 2001b; Ramakrishnan, 2002). The 30S and 50S ribosomal subunits have been imaged at atomic resolution, whereas the intact 70S ribosome has only been imaged at a resolution of 5.5 Å. Many states of the translation cycle have been seen at lower resolution by cryo-electron microscopy, but have eluded x-ray crystallographic analysis. The factors that aided the solution of the ribosome structure by x-ray crystallography to 5.5 Å resolution will be described (Cate, 2001; Yusupov et al., 2001). More importantly for the present meeting, factors holding back solution of the 70S ribosome structure to atomic resolution, and in different functional states, will be discussed.

First, the good news is that it is possible to solve the structures of supramolecular complexes the size of the ribosome by x-ray crystallography. In the case of the 70S ribosome, this meant generating a high-quality electron density map that could be interpreted based on a combination of previously determined structures and a wealth of biochemical information. The key to generating high-quality electron density maps depended on the determination of accurate structure factor phases. This was achieved by using molecular replacement with cryo-EM derived electron density maps, followed by Multiwavelength Anomalous Dispersion (MAD) phasing and density modification algorithms. The required steps were recently reviewed (Cate, 2001), thus I will expand the discussion here in a more editorial fashion.

The 7.8 Å electron density map of the ribosome is the largest generated by means of MAD phasing. It clearly demonstrates that there is essentially no upper limit to the size of a structure that can be solved by MAD phasing. This strategy worked for heavy-atom cluster compounds and structure factors to a resolution of 12 Å, and for more common metal ions, such as lanthanides (unpublished results). Due to their strong LIII anomalous dispersion, the hexa-hydrated magnesium analogs iridium (III) hexammine and osmium (III) hexammine provided the best experimental structure factor phases for ribosome complexes, which contain a large quantity of RNA (Ban et al., 2000; Cate et al., 1999; Wimberly et al., 2000). It is unclear which compound would be as generally useful for solving the structures of supramolecular complexes composed mainly of protein. Selenomethionine may be the best choice, although it has a relatively weak anomalous signal compared to lanthanides and other heavier elements.

Interestingly, the solution of the 70S ribosome structure required few modifications to existing crystallographic software, apart from array sizes, in order to generate the electron density maps. However, there were three situations that involved unconventional use of the standard software. First, since the structure factors used to calculate the final electron density maps extended to a resolution only 7.5 Å or 5.0 Å, and to lower resolution at earlier stages of the solution process, we relied heavily on the use of "pseudo-atom" x-ray scatterers (Cate, 2001). These artificial atoms were used as place-holders in the electron density in a way analogous to the use of "globbic" scatterers for protein structures at low resolution (Guo et al., 1999). In practice, pseudo-atoms allowed us to convert the cryo-EM electron density map into a form that could be used in any crystallographic program, namely a PDB file. It remains an open question as to whether pseudo-atoms could be used in any way to refine low-resolution structures of the ribosome (at 8 Å or 5.5 Å), given atomic-resolution structures of ribosomal components.

Second, refinement of heavy-atom parameters required careful monitoring, due to the large number of heavy atom sites (up to 200 for the 5.5 Å structure determination). Although the number of heavy-atoms was quite large, the corresponding number of structure factor amplitudes scaled appropriately to restrain refinement of the heavy-atom parameters. The program MLPHARE was used to refine the heavy-atom parameters (Otwinowski, 1991), due to its flexibility in recasting MAD phasing as a variant of multiple isomorphous replacement with anomalous scattering (MIRAS). However, refinement of the large number of heavy-atom sites was computationally time consuming. Fast Fourier Transform algorithms used to compute structure factors from the heavy-atom model have been implemented in CNS, which should

alleviate this problem, and did so for the archaeal 50S subunit structure determination (Ban et al., 2000; Brünger et al., 1998).

Third, methods of density modification typically used to improve structure factor phases derived from molecular replacement models or MAD phasing, such as solvent flipping (Abrahams and De Graaff, 1998), were modified in order to improve the low-resolution electron density maps. For example, the level of contrast between macromolecule and solvent at very low resolution (20 Å or lower) skewed the default settings for protein and solvent density in the program CNS (Brünger et al., 1998) to unexpected values, even when the structure factors extended to a resolution of 5 Å.

Modeling of the 70S ribosome structure at a resolution of 5.5 Å relied heavily on the known atomic-resolution structures of the individual subunits, of tRNA, and of individual RNA and protein components of the ribosome (Yusupov et al., 2001). The methods used for modeling were entirely manual, and thus incredibly time-consuming. The program O was used extensively for displaying the electron density maps and for moving parts of the protein and RNA models into place (Jones and Kjeldgard, 1997; Kleywegt and Jones, 1997). Even given the 30S and 50S subunit structures from bacterial sources (Ramakrishnan, 2002), it is not yet clear how these could be used in crystallographic refinement. At 5.5 Å, α-helices in proteins were clearly seen in the electron density, as were β-hairpins, but no other parts of the proteins were clearly visible as protein structure. RNA components, on the other hand, were more easily modeled at nucleotide resolution, since the length scales of the nucleotide chemical constituents match the resolution of the structure factors. In fact, it was often possible to recognize non-canonical RNA structures in the electron density, even at a resolution of 7.8 Å (Cate et al., 1999; Culver et al., 1999).

The 5.5 Å structure of the 70S ribosome in a complex with mRNA and two tRNAs represents just one snapshot of the elongation cycle of translation. What are the prospects for obtaining the structures of other states in the elongation cycle by x-ray crystallography, and for extending the structures to atomic resolution? Both questions are essentially biochemical in scope. The atomic-resolution structures of the 30S and 50S subunits, and MAD phasing of the 70S ribosome structure demonstrate that the crystallographic issues can be solved. Given structure factors to atomic resolution, it may be possible to use the known subunit structures for molecular replacement and subsequent crystallographic refinement. At lower resolution, molecular replacement phases are likely to be too poor to generate high-quality electron density maps, even with density modification. A large number of conformational changes in the ribosome are suspected to occur during translation, and many have been observed (Frank, 2001b; Ramakrishnan, 2002). MAD phasing will still be the best approach to imaging the ribosome to low resolution by x-ray crystallography.

Many research groups are now trying to trap the 70S ribosome in different steps of the elongation cycle of protein synthesis in order to probe translation by x-ray crystallography. We have taken the strategy of using the bacterium *E. coli*, due to the wealth of biochemical and genetic information about translation available from this model organism. In my view, there is no inherent advantage in using thermophiles in preference to mesophiles such as *E. coli* for structural studies (Batey et al., 2000; Harms et al., 2001). In fact, the dearth of biochemical results available in thermophilic organisms could hamper structural studies of other steps in translation. The prospect of obtaining an atomic resolution structure for any of the 70S ribosomal states in translation remains unclear. For DNA polymerases and RNA polymerase holoenzymes (Doublie et al., 1999; Landick, 2001), crystallographic success followed the development of sophisticated biochemical analyses, and took well over a decade to reach fruition.

References

1. Abrahams, J. P., and De Graaff, R. A. (1998). New developments in phase refinement. Curr Opin Struct Biol *8*, 601-605.

2. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution [see comments]. Science *289*, 905-920.

3. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B., and Doudna, J. A. (2000). Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science *287*, 1232-1239.

4. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

5. Cate, J. H. (2001). Construction of low-resolution x-ray crystallographic electron density maps of the ribosome. Methods *25*, 303-308.

6. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing [see comments]. Science *273*, 1678-1685.

7. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. N. (1999). X-ray crystal structures of 70S ribosome functional complexes. Science *285*, 2095-2104.

8. Culver, G. M., Cate, J. H., Yusupova, G. Z., Yusupov, M. M., and Noller, H. F. (1999). Identification of an RNA-protein bridge spanning the ribosomal subunit interface [see comments]. Science *285*, 2133-2136.

9. Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999). An open and closed case for all polymerases. Structure Fold Des *7*, R31-35.

10. Frank, J. (2001a). Cryo-electron microscopy as an investigative tool: the ribosome as an example. Bioessays *23*, 725-732.

11. Frank, J. (2001b). Ribosomal dynamics explored by cryo-electron microscopy. Methods *25*, 309-315.

12. Green, R., and Noller, H. F. (1997). Ribosomes and translation. Annu Rev Biochem *66*, 679-716.

13. Guo, D. Y., Blessing, R. H., Langs, D. A., and Smith, G. D. (1999). On 'globbicity' of low-resolution protein structures. Acta Crystallogr D Biol Crystallogr *55*, 230-237.

14. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell *107*, 679-688.

15. Jones, T. A., and Kjeldgard, M. (1997). Electron-Density Map Interpretation. Methods in Enzymology *277*, 173-208.

16. Kleywegt, G., and Jones, T. (1997). Model-building and refinement practice. Methods in Enzymology *277*, 208-230.

17. Landick, R. (2001). RNA polymerase clamps down. Cell *105*, 567-570.

18. Otwinowski, Z. (1991). In Isomorphous replacement and anomalous scattering, W. Wolf, P. R. Evans, and A. G. W. Leslie, eds. (Warrington, UK, Daresbury Laboratory), pp. 80-86.

19. Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell *108*, 557-572.

20. Wimberly, B. T., Brodersen, D. E., Jr, W. M. C., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327-339.

21. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 A resolution. Science *292*, 883-896.

DATA COLLECTION FROM MICRO-CRYSTALS. IS THIS A VIABLE (RATHER THAN POSSIBLE) TECHNIQUE?

F. Cipriani and A. Thompson; EMBL Grenoble Outstation, 6 Rue Jules Horowitz, 38042 Grenoble, France

Abstract
A specially designed goniostat for micro-crystal diffraction (Perrakis 1999) has been developed in Grenoble as a collaboration between EMBL and ESRF, and extensively used on the micro-focus beamline ID13 for data collection. A brief description of the design of the instrument is followed by a discussion of experience in micro-crystal data collection both on ID13 and other beamlines at the ESRF, and some comments made about the future of data collection on micro crystals.

Introduction
The problem of manipulating (or even seeing) very tiny protein crystals (10 – 20 micron in all dimensions) has been addressed by developing a diffractometer that is custom designed for manipulating such small samples (and which has also proved very useful in handling larger crystals).
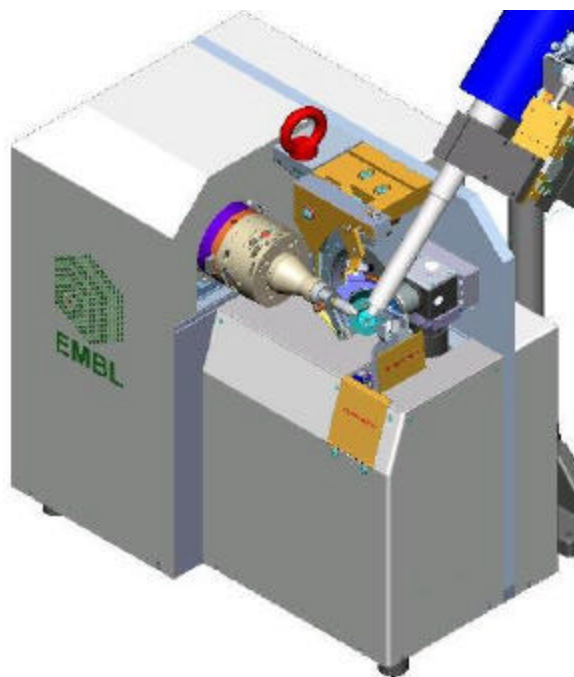
Fig 1 – Engineering drawing of the MK II microdiffractometer.

Sample Visualisation.
In order to be able to see such small crystals, an optical system where the X-ray beam passes through the lens prior to hitting the crystal sample has been developed. This type of approach offers 3 main advantages –

a)   The lens can be very close to the sample position, and occupies "dead" space around the experiment. Hence one can match the resolution of the digitized image to the resolution requirements of a "micro" sample size to obtain optimal image quality.

b)   The lens sees the crystal as the X-rays see it. So the problem of sample centring becomes one of determining where the X-ray beam is (with a high resolution phosphor and the same lens) and putting the sample in the correct place. This has a huge advantage when dealing with micro-beams, since it is extremely difficult (even with 3rd generation storage rings) to get an X-ray beam that is stable to more than a few microns which is significant when dealing with a 10 micron crystal.

c)   The lens is part of the defining geometry of the goniostat, and hence (to a first order) undergoes the same movements / vibrations as the crystal itself. The image is then a true (rather than a relative) measure of what is happening at the sample position.

A fluorescent single crystal can be automatically inserted into the direct beam. The crystal is displaced 1 mm from the X-ray beam position but rests within the cold gas stream. The only real disadvantage of this system is the price (the lenses are hard to make and thus expensive).
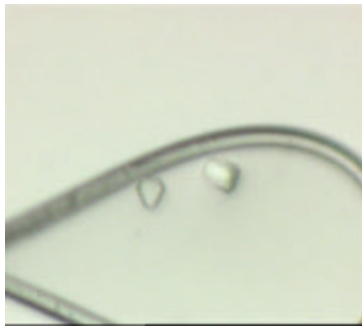
Fig 2 – Image of two 10 micron crystals in a loop taken using the MK I microdiffractometer visualization system. Such crystals are often not visible at all using typical beamline sample viewing systems!

Assisted sample mounting and alignment
Suction holds a special "goniometer head" (which mounts most regularly used cryo-pin bases) which is placed into position in the cold stream on a 45 degree movable transfer arc (facilitating the transfer of crystals mounted in liquid N2 cryo-vials). The crystal is then automatically transferred to the beam position. The device is compatible with an automatic sample mounting system under development (Thompson 2001).(J-S. Aksoy 2002). Currently the sample position is identified by human intervention using a pair of cursors on the image display, and the sample moved to the beam position as identified by the phosphor. An algorithm to do this automatically is under development (Lavault 2002) The sequence of images (Fig 4) below show the process of manually mounting a crystal.

Fig 4 Sample Processing phases (manual mounting)

Sample Loading
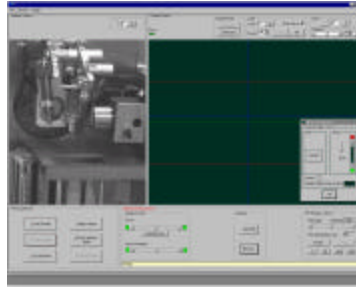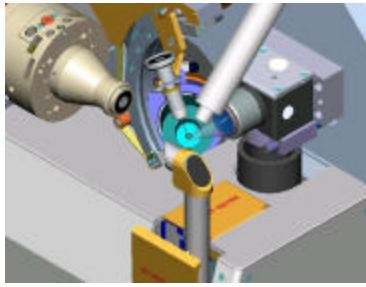
Sample Centring
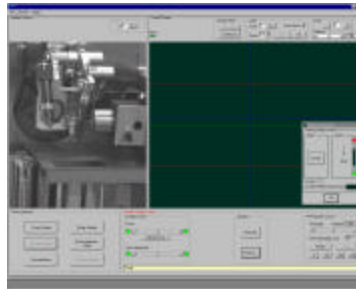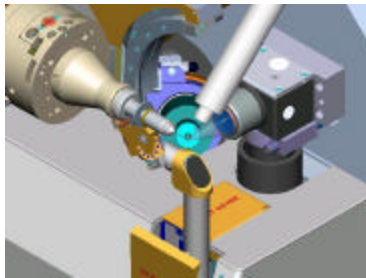
Beam position marking
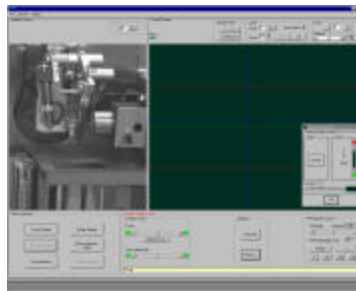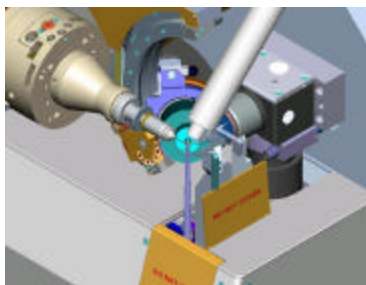
Sample alignment
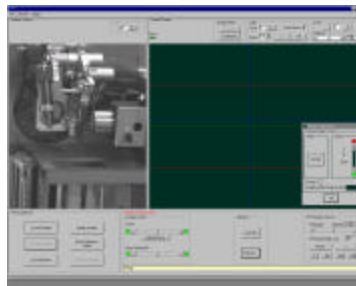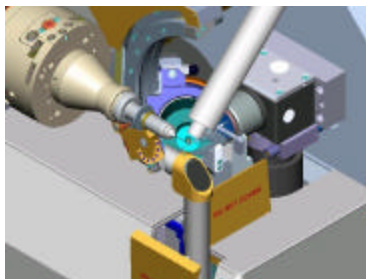
Data Collection
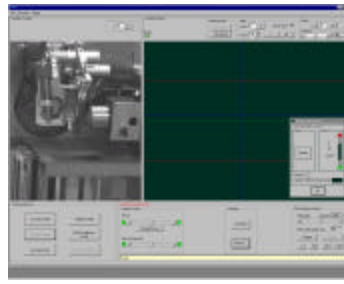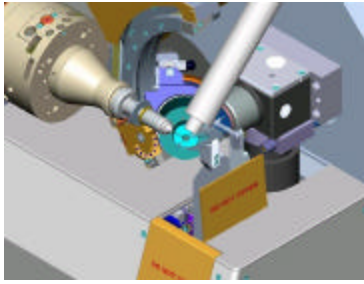
Control of image background
The device does not include a set of beam defining slits, though a number of precise pinhole apertures both for beam shaping (where diaphragms of between 5 – 100 microns are available) and scatter guard (50, 100 or 150 micron apertures) can be inserted into the X-ray beam. The beamstop is fabricated as a single piece along with this collimation tube, hence the two are aligned as a single unit. Both tube and beamstop are motorized, and can be driven to a position already identified using the visualization lens and phosphor screen. Alignment is then as simple as clicking a mouse button.

Data collection diagnostics
The device can currently operate with its own software or in SPEC "device server" mode, the device server having limited functionality. Diagnostic information is available via the diffractometer's own software, for example permitting a log of the beam intensity through the collimator tube to be measured as a function of rotation angle on a ms time scale.

Data collection from Micro-Crystals – some experiences on ID13 and elsewhere.

Radiation Damage.
The main problem of actually collecting data from micro-crystals is one of the onset of radiation damage. This issue has been extensively discussed (for example (Ravelli 2000), (Burmeister 2000), (Glaeser 2000), (Henderson 1990), (Gonzalez 1994)). A crystal of approximately 25 x 25 x 20 microns illuminated on beamline ID29 of the ESRF shows significant radiation damage after 100 degrees of data (540 s exposure) even in a highly attenuated X-ray beam at 12.7 KeV. Figure 3 shows that the change of B factor vs frame number for this crystal. The dose absorbed by the crystal in this measurement (calculated from the measured X-ray flux, the estimated sample size and absorption) is in rough agreement (250 s exposure corresponding to 83 frames in Fig 3 below) with the value predicted by Henderson (Henderson 1990) and experimentally confirmed by Gonzalez (Gonzalez 1992). Although numerous efforts have been made to overcome the effects of radiation damage by a number of means (free radical scavengers, lower temperature cooling) experience in protein crystallography to date shows no obvious solution to this problem.
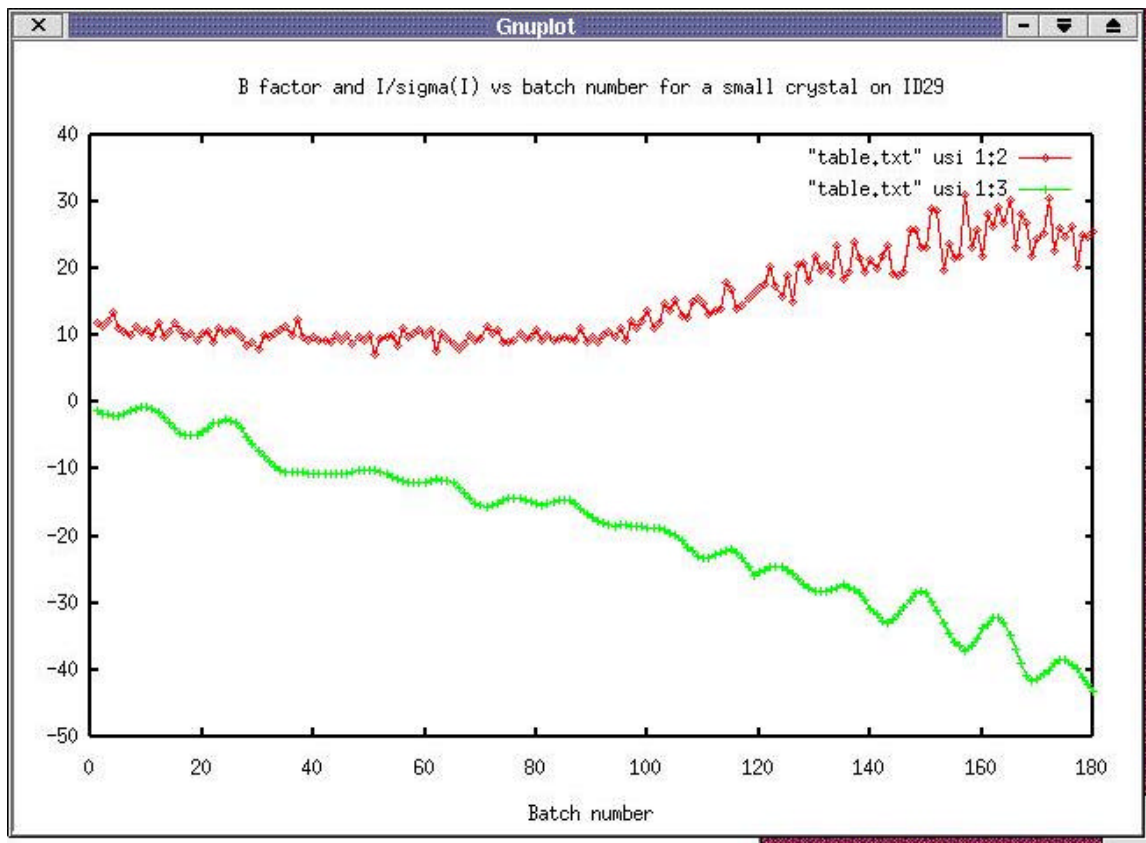
Figure 3. Evolution of B factor and I/sigma(I) vs batch number for data collected from the crystal referred to above. The decay of the crystal was verified by checking the diffraction pattern from a previously recorded image at the end of the data collection.

A similar sized crystal to the above was collected recently on ID29 from a "real" problem as opposed to a test data collection. The crystal belonged to the space group P21, and data collected from a single wavelength optimized to give maximum anomalous signal at the Se absorption edge was insufficiently complete or redundant in order to solve the Se partial structure and hence solve the crystal structure. In fact this is likely to be the case for the almost all micro-crystals of lower symmetry than tetragonal! Data must then of necessity be collected from several samples with the concomitant problems of scaling or lack of isomorphism.

Signal to noise ratio.
In order to optimise the signal to noise ratio for a micro-crystal data collection, sample slits need to be carefully aligned so that the crystal is just continually bathed in the X-ray beam. The improvement in "diffraction limit" detected from such a micro-crystal has been noted by Perrakis et al (Perrakis 1999) and Glaesser et al (Glaeser 2000). It is therefore a fundamental requirement of the technique to have an excellent level of beam position diagnostics (either "off line" like the visible insertion of a phosphor as described above, or preferably "on-line" in the form of an X-ray beam position monitor which determines the centroid of the X-ray beam to a few microns precision, as well as beam position stability and knowledge of the sample shape and size. Such beam position monitors are under development at the ESRF.

Limitations of Data Quality.

The optics required of beamline delivering a micro-focus are constrained by current synchrotron sources to strongly demagnify the X-ray source (generally between 10 and 20:1 demagnification). Therefore beams are strongly divergent after passing through the crystal (usually of the order of 0.5 – 2 mrads for current sources). For example considering the proposed micro-focus beamline for the DIAMOND synchrotron, convergence angles of 10 – 20 in the vertical and horizontal are necessary to attain a focal spot size of roughly 10 x 10 microns, giving an overall crossfire of 1.6 mrad in the worst (horizontal) direction. This has obvious implications for the maximum size of unit cell that can be studied and the diffraction resolution limit (reinforcing the need for very low background measurements). At the other extreme, if the unit cell size demands that crossfire be reduced (i.e. a focal spot of a hundred microns with a sample of say 10 microns), then the diffraction peaks would diverge 10 times less and may become wholly encompassed by a single pixel when measured on a low point spread function detector, hence causing a quantisation of their position in the diffraction pattern and leading to the requirement for new methods of predicting and integrating spots.

Beam intensity fluctuations.

A Kirkpatrick-Baez focusing system installed on beamline ID19 at the ESRF has produced a beam focus size of 0.61 x 0.2 microns FWHM independent of beam input aperture size, whereas raytacing calculations based on the real quality of the mirror optics suggest a spot size of 0.18 x 0.25 microns. This result strongly suggests spot broadening due to vibrations of the optical system. (Hignette 2001). This broadening corresponds to a sweeping of the "fine" focal spot against a broader aperture. Table 1 shows a measurement of the vibration levels sampled at various measurement points on the ID29 beamline at the ESRF, and Table 2 the major frequency components (LeSourd 2001). Several vibration in the low frequency domain match quite well the speed with which diffraction peaks pass the Ewald condition in a typical ESRF experiment (a fraction of a second to a few seconds exposure time).

Table 1

Sensor location	Vertical (Z)	Normal to X-ray (Y)	// to X-ray (X)
Floor	1.75 / *0.32*	0.77 / *0.15*	0.69 / *0.12*
Monochromator	1.85 / *0.33*	0.95 / *0.17*	0.88 / *0.14*
Floor	2.28 / *0.36*	0.75 / *0.13*	0.73 / *0.13*
Mirrors	2.30 / *0.36*	1.20 / *0.22*	0.88 / *0.14*
Floor	1.71 / *0.24*	0.73 / *0.11*	0.70 / *0.32*
Table	1.83 / *0.25*	1.33 / *0.20*	1.19 / *0.18*

Table 2

Freq.	9.8	16.9	17.9	26.3	26.4	29.3	29.5	34.5	35.2	36.5
Axis	Y	X/Y	Y	X	X	X	Y	Z/Y	X	X
Struct	Table	Table	Mirr.	Mono	Mirr.	Mono	Mirr.	Mono	Mirr.	Mono

Experiment and optics may then need to be mounted on the same stabilized or feedback controlled surface. For the spot broadening reported above for sub-micron focusing and assuming a naïve "Gaussian" model for the broadening due to beam movement, this would correspond to a 0.6 micron FWHM "sweep" of the beam over the sample due to vibration. Further crystal instability in the beam can be introduced by movements of the crystal in the cold N2 flow from the crystal cooling device. Depending on the frequency of these vibrations, data quality for "micron" sized crystals would be strongly limited by this (although of course multiple passes of exposure could be made. But it is not clear that reducing the beam intensity is the best way of getting higher resolution data from micro-crystals). There are clearly some major technical challenges to be addressed before "micro-crystallography" can attain the same level of development that "standard" crystallography.

Conclusion.
Using techniques developed in Grenoble and independently elsewhere, data can now be routinely collected from crystals as small as a few microns. However the quality and quantity of this data is such that it may not in all cases be useful without further developments in critical areas such as crystal cooling techniques, reduction in the speed of onset of radiation damage, integration of weak or broad diffraction spots, integration of very fine diffraction spots and improved relative stability of the beam and crystal.

Acknowledgements.

References.
1. Burmeister, W. P. (2000). "Structural changes in a cryo-cooled protein crystal owing to radiation damage." Acta Crystallographica D 56: 328-341.

2. Glaeser, R., Facciotti, M, Walain, P, Rouhani, S, Holton, J, MacDowell, A, Celestre, R, Cambie, D and Padmore, H. (2000). "Characterisation of conditions required for X-Ray diffraction experiments with protein microcrystals." Biophysical Journal 78: 3178-3185.

3. Gonzalez, A., Thompson, A. W and Nave, C. (1992). "Cryo-protection of protein crystals in intense X-ray beams." Rev. Sci. Intruments. 63(1): 1177-1180.

4. Gonzalez, A. a. N., C (1994). "Radiation damage in protein crystals at low temperature." Acta Crystallographica D50: 874-877.

5. Henderson, R. (1990). "Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction." Proc. R. Soc. London B241: 6-8.

6. Hignette, O., Rostaing,G, Cloetens,P, Rommevaux,A, Ludwig, W and Freund,A. (2001). "Submicron focussing of hard X-rays with reflecting surfaces at the ESRF." SPIE Conference Proceedings 4499: in press.

7. J-S. Aksoy, A. B., J-M. Bois,H. Casseroto, J-M. Chaize,F. Cipriani,S. Delangeniere, F. Felisaz, E. Homan, S. Laboure, B. Lavault, M. Lentini, T. Mairs, R. Steinman, O. Svensson, P. Theveneau and A. Thompson (2002). Towards the development of an automatic beamline for protein crystallography at the ESRF. The CCP4 Study Weekend, York.

8. Lavault, B., Cipriani, F, Thompson, A and Maurin, Y. (2002). Progress towards automated alignment of protein crystals in an X-ray beam. European Structural Biology Forum Workshop for Structural Genomics, Jena.

9. LeSourd, M. (2001). X-ray Intensity and Vibration Measurements on ID29, ESRF: 1-12.

10. Perrakis, A., Ci[priani, F, Castagna, J-C, Claustre, L, Burghammer, M, Riekel, C and Cusack, S. (1999). "Protein microcrystals and the design of a micro-diffractometer : current experience and plans at EMBL and ESRF ID13." Acta Cryst D 55: 1765 - 1770.

11. Ravelli, R. B. G., and McSweeney, S, M. (2000). "The fingerprint that X-rays can leave on structures." Structure 8(3): 315 - 328.

12. Thompson, A. W. (2001). The design of X-ray beamlines for measuring small anomalous signals. <u>Methods in Macromolecular Crystallography</u>. D. T. a. L. Johnson, IOS Press. Series 1:Life and behavioural sciences - Vol 325: 26-42.

THE CHALLENGE OF X-RAY FREE-ELECTRON LASERS

Janos Hajdu, Department of Biochemistry, Uppsala University, Biomedical Center, Box 576, S-751 23 Uppsala, Sweden.

The history of the living world can be summarised as an elaboration of ever more perfect eyes within a Cosmos in which there is always something new to be seen (Teilhard du Chardin, 1955). Today, the bottleneck in the atomic resolution imaging of biological systems is a fundamental need for crystals. This limits the scope of detailed structural analysis to macromolecules and macromolecular assemblies which can be crystallised. Many biologically important target complexes are difficult or impossible to crystallise. As a consequence, there are large, and systematically blank areas in structural biology. Only a handful of membrane protein structures are known, structural studies on large assemblies are problematic, and there is no hope of reaching high resolution with currently available methods on non-repetitive and non-reproducible structures (e.g cells). Recently, a new dynamic limit in radiation tolerance has been identified at extreme dose rates and ultra short exposure times (Neutze et al., 2000). This is the domain of X-ray free-electron lasers (XFELs). The predicted radiation tolerance in this regime is several orders of magnitude higher than theoretical limits in conventional X-ray experiments (Henderson, 1995). At the outer extremes (not reached on the first day of operation), scattering may be recorded from large single macromolecules, viruses, nanocrystals and nanoclusters of proteins without the need to amplify scattered radiation through Bragg reflections. Resolution under these conditions would not depend on sample quality as in conventional crystallography, but would be a function of radiation intensity, pulse duration, wavelength, detector parameters, and the extent of ionisation and sample movement during the exposure. Averaging procedures can be applied to extend resolution when a reproducible sample scatters a sufficiently large number of photons for its orientation to be determined. We believe that the potential benefits from studies in this field are substantial.

Synchrotron machines were developed by physicists and materials scientists, and have evolved to benefit biology. When the first dedicated synchrotron light sources came to life more than twenty years ago, they produced X-rays with 50-100-fold higher intensities than a rotating anode generator at home. Yet synchrotron radiation triggered a revolution in life sciences, leading to almost 15,000 X-ray structures deposited in the protein data bank today. X-ray lasers promise to offer much more. If one takes the speed of walking and multiplies it by 10^{10}, one would get a speed which is 100 times faster than the speed of light. The expected improvement in peak brilliance expected from XFELs over currently existing synchrotrons is of this magnitude. These machines will bring us into a world for which only predictions exist. A lack of hands-on experimental data on the behaviour of matter under these conditions makes very detailed forecasts difficult. One thing is sure, however: XFELs will generate new thinking, new science and a new scientific community.

STRUCTURAL STUDIES WITH X-RAY LASERS

The underpinning for the planned biology experiments at emerging X-ray lasers (the Linac Coherent Light Source, LCLS, at Stanford and the TESLA X-ray laser at DESY, Hamburg) is provided by the calculation of Neutze et al. (2000) that usable diffraction data might be obtained from femtosecond

XFEL pulses, and by the demonstration of Sayre and coworkers (Miao et al., 1999; 2001) that x-ray diffraction data of non-crystalline specimens can be reconstructed to yield a real-space image. Recent papers with relevance to the exploitation of X-ray lasers in biology are Szoke (2001), Tegze & Faigel. (2001), Sayre, D. (2002), Persson et al. (2001); Ziaja et al. (2001; 2002).

- **A summary of the treatment of damage formation**

Radiation damage is caused by energy deposited into the sample by incoming photons, electrons, neutrons, etc. Cooling can slow down sample deterioration, but it cannot eliminate damage-induced sample movement within the time needed to complete conventional measurements (Henderson, 1990;1995; Nave 1995). Analysis of the dynamics of damage formation (Solem, 1986; Hajdu, 1990; Neutze et al. 2000; Persson et al., 2001; Ziaja et al. 2001; Ziaja et al., 2002) suggests that the conventional damage barrier (about 200 X-ray photons/•^{2} at 12 keV X-ray energy or 1• wavelength) (Henderson, 1990) may be extended at very high dose rates and very short exposure times.

A key difference between X-rays and electrons in imaging applications is the amount of signal that can be obtained before the sample gets destroyed. That is given by the ratio: sigma(elastic)/{sigma(inelastic) X (deposited energy)}. The numbers are 1 / { 3 x 30 eV } for 300 keV electrons and 1 / { 10 x 10 keV } for 10 keV X-rays. So the signal to noise ratio favors electrons by a factor of 1000 in a large sample. For small samples where the photoelectrons can escape, the energy deposited by X-rays drops to the energy of the Auger electrons (around 250 eV) or below (Ziaja et al., 2001; 2002), so the factor of 1000 advantage decreases to a factor of 25.

Photons are electrically neutral bosons and an unlimited number of them can be packed simultaneously into a small volume of phase space, so that in principle, a very large number of photons can be fired at the sample simultaneously. The speed of imaging with such ultrashort X-ray pulses limits damage-induced sample movements during the exposure (Neutze et al. 2000). Electrons on the other hand are charged fermions and cannot be focused on the same spot at the same time. This factor together with Coulomb interactions between electrons in a pulse preclude the use of similar strategies to improve electron microscopy (the Borsch effect, 1954). We estimate that the net gain in imaging with X-rays over electrons can be more than 1000-fold on very small samples if very short X-ray pulses can be produced.

With X-rays of 1• wavelength (12 keV energy), the photoelectric cross section of carbon is about 10 times higher than its elastic scattering cross section making the **photoelectric effect** the primary source of damage. The photoelectric effect is a resonance phenomenon in which a photon is absorbed and an electron ejected (Dyson, 1973) usually from a low lying orbital of the atom (about 95% of the photoelectric events remove K-shell electrons from carbon, nitrogen, oxygen and sulphur), producing a hollow ion with an unstable electronic configuration. Relaxation is achieved through a higher shell electron falling into the vacant orbital. In heavy elements, this usually gives rise to X-ray fluorescence, while in light elements, the electron falling into the lower orbital is more likely to give up its energy to another electron, which is then ejected in the **Auger process**. Auger emission is predominant in carbon, nitrogen, oxygen and sulphur (99-95%, Krause and Oliver, 1979), thus, most photoelectric events ultimately remove two electrons from these elements. These two electrons have different energies (around 12 keV for photoelectrons and around 0.25 keV for Auger electrons), and are released at different times. Relevant K-hole life-times can be determined from Auger line-widths (Krause and Oliver, 1979), and are

SP5-B

11.1 fs (C), 9.3 fs (N), 6.6 fs (O) and 1.3 fs (S). **Shakeup excitations** (multiple ionisation following inner shell ionisation, see, e.g. Svensson et al., 1987; Persson et al., 2001), initial- and final-state configuration interaction and interference between different decay channels will modulate this picture.

An additional effect is **inelastic** (Compton/Raman) **scattering**, which represents a direct momentum transfer from an X-ray photon to an electron, so that the X-ray photon is scattered with a reduced energy. If the energy taken up by a bound electron is greater than its shell binding energy, the atom will be ionised. The inelastic cross section of carbon, nitrogen and oxygen is around 3% of the corresponding photoelectric cross sections (Veigele, 1973; Hubbell et al., 1975; Henke et al. 1993), whereas the inelastic cross section of hydrogen is much higher than its photoelectric cross section.

Other factors: All plasmas start cold and dense and become hot later. The average velocities of 10 keV photoelectrons (43 nm/fs) and 0.25 keV Auger electrons (7 nm/fs) enable these electrons to escape a protein molecule in less than a fs during *early phases* of an exposure. At these velocities the inelastic electron scattering cross sections for carbon lie between 0.1 and 2.0 $\bullet^2$ such that approximately one electron in five would deposit additional energy into the molecule from which it escapes, and may also remove outer shell electrons (Ditmire, 1998). In *late phases* of an exposure, a significant fraction of the emitted electrons will not be able to escape the increased positive potential of the sample. Trapped electrons will increase the kinetic energy of the sample through thermal equilibration, but they will also slow down the Coulomb explosion of the sample by partially neutralising the positively charged protein core as a plasma is formed. These opposing effects have not been considered here, but are factors which are likely to influence the dynamics of larger systems (Ziaja et al., 2001; 2002).

It has been argued that in addition to the effects listed above, extreme X-ray intensities could strip all outer-shell electrons from their parent atoms on a time scale of attoseconds (Doniach, 1996). However, a quantum mechanical analysis of the electric field induced tunnelling demonstrates that atoms actually become stabilized against ionisation via this mechanism at very high photon frequencies (Reiss, 1997) like those of X-rays. As no experimental data are available for this regime with X-rays, stabilisation at high photon intensities has not been included in the present model, and changes in cross sections were computed without such effects using existing models (Veigele, 1973; Hubbell et al., 1975; Henke et al. 1993). Further work and experiments will be required to clarify this issue and to obtain numerical values. Our present model probably overestimates the ionisation of the sample.

EXPERIMENTS

4th-generation X-ray sources will produce radiation pulses with pulse-lengths on the time scale of molecular vibrations at wavelengths corresponding to interatomic distances in condensed-matter. These features will ensure direct experimental access to a wide range of fundamental processes in chemistry and biology for the first time.

- **An overview**

There are three general classes of experiments which could be performed in various combinations with each other. These include (i) experiments that utilises the short time structure; (ii) experiments that make use of the coherence of the beam; (iii) and experiments that exploit the potential for very large intensities in the XFEL beam, leading up to the new regime outlined in Neutze et al. (2000).

Some of these studies will include
- Crystalline diffraction
- Surface diffraction
- Powder diffraction
- Small angle scattering
- NEXAFS and EXAFS studies of excited species
- Diffraction experiments on single particles and biomolecules combined with spraying techniques and mass spectrometry

The short time structure of the source will allow for a range of novel time-dependent experiments. An exciting area of research includes mode-selective chemistry and the analysis of charge separations and vibrational energy transfers following excitations.

The high peak brilliance of the source will bring very small crystals, nanoclusters and single particles within the scope of structural investigation with X-rays with femtosecond time resolution.

The full transverse coherence of the beam may allow holographic techniques to be applied in structural studies.

In the following sections we will outline some exciting biological problems that become doable this way. Note that the coherence and the saturation of the LCLS plays only a secondary role in these experiments, but will be important later.

It was emphasised in the introduction that damage by the incident radiation is the ultimate limit to the resolution of imaging techniques in biological molecules. The verification of theories for this regime is, therefore, a key goal with the first set of experiments. This should include studies on non linear X-ray effects, and provide the experimental foundations for a detailed understanding of damage in biological materials.

In order to utilise the advantages outlined in the previous paragraph, two additional developments are needed: focusing of the beam to about 0.5-0.1 micron, and the development of single particle injection techniques (Tito et al., 2000). We anticipate that very mildly ionised droplets will be injected into the focused beam just in time. When mild focusing of the beam becomes available, e.g. the focusing of the projected 10^{12} photons in one pulse into a focal diameter of 1/2 micron, giving a flux of 2 10^4 photons/$\bullet^2$ at the focus, single particle images can be obtained (e.g. from large viral particles) using averaging techniques similar to those used in electron microscopy (van Heel, 1987; van Heel et al., 1996; 1997). This should open the door to virus structures, including their genomes, to the study of difficult proteins, to very high time resolution studies of other, known structures.

In summary, we will argue below that new and exciting biological problems will be open for investigation from the first day when the beam of an X-ray laser becomes available. We foresee that our interim technical efforts in sample handling, attaching proteins to viruses and developing algorithms for single particle imaging will complement developments with X-ray lasers. The experiments outlined below represent major research lines far beyond the limits of currently available methodologies. We also hope that new research areas may emerge in a field as explosive as biology today by the time LCLS (the planned Linac Coherent Light Source in Stanford) becomes available. Shortening the LCLS pulse length

would be a key improvement in all planned applications, including those described in the biological proposal. With shorter pulses, tighter focusing, and development of accurate injection techniques, biomolecular imaging should reach the extreme regime outlined in the theoretical part of the proposal.

- **Nanocrystals**

Nanocrystals are open periodic structures with submicron dimensions. LCLS with its planned initial parameters may offer completely new avenues for structural studies on nanocrystalline samples. No such studies are currently possible at synchrotrons. When a crystal is small, the Bragg peaks are broadened and the intensity between the Bragg peaks is not negligible. Both the Bragg peaks and the intensity between the peaks carries structural information. The oversampled diffraction pattern visible between Bragg peaks can directly provide phase information. Applying positivity constraints to eliminate the conjugate ambiguity (cf. twin image problem in holography), phase information can be obtained from the oversampled diffraction pattern. Based on various formulations of the Gerchberg-Saxton algorithm (see e.g. Fienup, 1982), an iterative algorithm has been developed (Miao et al., 1998; Miao & Sayre, 2000) and is available for the planned experiments. Computer phasing experiments demonstrate the feasibility of this technique. By employing a finite support, a random phase set and the positivity constraints on the real part of the electron density, it was possible to reconstruct the electron density of the whole small crystal in an iterative process (Miao & Sayre, 2000). It could be shown that the reconstruction of the structure from the diffraction pattern was still possible with deteriorating signal to noise ratios of 20, 10, and 5.

- **Two-dimensional crystals of macromolecules**

Various estimates show that there are at least as many different membrane proteins as soluble proteins, yet there are only a handful of structures known today (52 entries in the PDB on 12 April 2002) for integral membrane proteins compared to almost 18,000 structures for soluble proteins. An understanding of the structure-function relationships in membrane proteins would make invaluable contributions to biochemistry, physiology and medicine, and would produce a substantial socio-economical impact (about 70% of all known drugs target membrane proteins). Two-dimensional crystals of membrane proteins have been used for structural studies in electron cryomicroscopy (see, e.g. Stowell et al., 1998). Such arrays may also be obtained from a number of other proteins by epitaxial crystal growth methods. The intensity of the scattered X-rays from a small two-dimensional array was simulated, using 300 x 300 copies of a simple test protein (lysozyme) in the calculations. These protein molecules were arranged on a primitive 2D lattice with 32 • lattice spacing in two perpendicular directions. The surface area of the array was about 1 mm x 1 mm, and the size of the X-ray beam was selected to match this area. Pulse intensity (4×10^5 photons/•^{2}) was chosen to ionise about 10% of all non-hydrogen atoms in the sample. The results show diffraction rods with 200-400 photons/rod at 3-4 • resolution. Integration of the diffraction rods requires images with different sample orientations to be recorded, and this could be achieved by merging data from several randomly oriented samples. This procedure could provide structural data on membrane proteins in single bilayers. We will use this technique to obtain structural data on a variety of key membrane proteins.

- **Nanoclusters for structural studies**

In contrast to nanocrystals (which are small but open periodic structures with translational symmetry),

nanoclusters are closed structures, which may be periodic (like oligomeric proteins or viral capsids) but have no translational symmetry. Such nanoclusters come in very well defined sizes (see, e.g. Lin et al., 1996). We are developing procedures for assembling proteins of choice into regular nanoclusters for subsequent structural studies. These methods will be applied to structural studies on soluble and membrane proteins at the LCLS. Assembling protein molecules into nanoclusters will increase the intensity of scattered radiation from otherwise small proteins.

Expected outcome: structures for "uncrystallisable" proteins attached to the surface of regular templates. These studies can also cover organic/inorganic nanostructures like nanotubes and nanomotors, etc.

- **Virus structures and the structure of viral genomes**

No three-dimensional structure is available for genomes today. Small spherical viruses are among the simplest replicating systems in biology, yet the packing of the nucleic acid inside the capsid, and the factors affecting viral assembly, stability and disassembly remain elusive. Only a very superficial picture is available today on the packing of the genetic material in intact viruses (ValegŒrd et al. 1990; Fisher and Johnson, 1993; Bottcher and Crowther, 1996; Gouet et al., 1999). This is due to the fact that the nucleic acid inside a viral capsid does not follow the symmetry of the outer capsid; when crystallised, the outer shell obeys the space group symmetry, while the inner material does not. As a consequence, the image of the otherwise tightly packed nucleic acid inside the virion is rotationally averaged over some angular range. Viruses are big (around 10-100 nm diameter) and as a consequence give relatively strong scattering patterns. From images collected, we propose holographic reconstruction to recover the structure of the inner part (the genome) of the virus. Other methods of phasing will also be explored (e.g., oversampling, Miao et al., 1999; Szske, 1999, or classical methods based on derivatives). Further experiments will focus on the assembly/disassembly of the virus. Of particular interest are viruses that cannot be crystallised (e.g. human immunodeficiency virus, herpes simplex virus, etc.).

- **X-ray diffraction tomography** - **tomoholography**

Electron tomography has been used to obtain ~10 nm 3D resolution reconstructions of small prokaryotic cells have been obtained at ~10 nm 3D resolution (Grimm et al., 1998; Wang et al., 2000; Weiss et al., 2000), and pattern matching is beginning to be applied to attempt to locate proteins within the volume (Bohm et al., 2000). However, multiple scattering limits the applicability of this approach to specimens that are no more than about 1 micrometer thick (insufficient for whole eukaryotic cells), even if energy filters are used in intermediate voltage microscopes (Jacobsen et al., 1998). Electron images must have molecule positions relative to the focal plane determined so that the contrast transfer function can be estimated and applied to the data, while far-field x-ray diffraction data needs no such estimation and correction to be applied to photon-statistics-limited data. As a result, diffraction tomography with X-rays continues to represent an approach for imaging of whole hydrated eukaryotic cells, which are typically 5-20 micrometers in size.

- **Time-resolved experiments**

Biological function is a four-dimensional property. Time-resolved studies on structure, function and dynamics can cover catalysis, protein folding, nucleic acid folding, the assembly/disassembly of biomolecular complexes, viral uncoating, viral infections and so on. Such studies will first become

SP5-F

possible on nanoclusters and nanocrystals as they require less stringent beam parameters from the laser than studies on single molecules. Later on, these experiments *may* be extended to individual biomolecules or cells.

Photochemical reactions. X-ray lasers will produce radiation pulses with pulse-lengths on the time scale of molecular vibrations at wavelengths corresponding to interatomic distances in condensed-matter. These features will ensure direct experimental access to a wide range of fundamental processes in chemistry and biology for the first time. Marriage of femtosecond laser spectroscopy with femtosecond time-resolved structural studies will be possible. An exciting area of research includes studies on coherent reaction dynamics, mode-selective chemistry, and the analysis of charge separations and vibrational energy transfers following excitations.

Diffusion triggering and stop-flow studies. While certain key reactions in life are photochemical, most enzymes participate in diffusion-dominated processes with their reactants and partners. Time-resolved structural studies on diffusive processes in crystalline enzymes are difficult due to problems with mixing enzyme and reactant. Structural studies are only possible on intermediates which accumulate transiently in the crystal during a reaction. This requires a relatively fast binding followed by a relatively slow reaction. Due to the generally lower activity of crystalline enzymes, uniform catalysis can often be triggered by diffusing reagents (e.g. substrates) into crystals. However, the speed of diffusion and ligand binding sets an upper limit to the speed of reactions which can be analysed this way. In an average-sized protein crystal (0.2-0.3 mm overall dimensions), half saturation binding with small ligands can be reached within about a minute. An obvious possibility for lowering diffusion barriers is to reduce the sample size, and to use nanoclusters or nanocrystalline assemblies of enzymes for stopped-flow type time-resolved experiments. With very small samples, the vast majority of solution techniques and methodologies will become available for time-resolved structural investigations. Kinetic studies based on diffusion triggering may become possible on micron-, submicron-sized assemblies, and they will diffract to high resolutions in the LCLS beam. A new nanocluster/nanocrystal would be needed for each exposure. We foresee that container-free sample handling methods based on spraying techniques, will open up new horizons here. A microdrop-based stop flow apparatus has recently been developed by one of us (Carol Robinson), linking stopped flow and spraying methods. The overall impact of these developments on structural and functional biology is immeasurable.

- **Towards structural studies on single protein molecules**

If and when sufficiently short pulses (5-20 fs) and a reasonable tight focus (around 100 nm diameter focal spot) will become available, studies on large single individual molecules may become routinely possible. In order to *prepare* for such experiments, we intend to start testing the behaviour of individual protein molecules in the LCLS beam. In the tests, proteins of different molecular masses and oligomeric states will be selected and injected into the X-ray beam. We intend to analyse their fragmentation patterns by mass spectrometry in order to verify theoretical predictions. Success in this area may bring about a new era in life sciences.

From molecular to modular and cellular biology

A comprehensive understanding of function on the supramolecular and subcellular level is still to come, and may shed light on questions related to the emergence of life from a collection of complex chemicals.

SP5-G

It is obviously not enough to simply mix such compounds. Function in biology is embedded into a network of interacting components. Biologists and others have thus begun to think about systems and networks, and we can anticipate research picking up on system analysis of labile, functional collections of interacting molecules, or modules. Cellular functions (e.g. signal transduction, respiration, protein synthesis, etc.) are carried out by modules made up of many species of interacting molecules. Certain modules may be localised and analysed *in situ*. An understanding of how modules work and interact with each other is a key aim to understand organisation within a cell, and ultimately, in life. Structural information is only available on a handful of stable modules (macromolecular complexes) which could be crystallised (e.g. the cytochrome bc1 complex or the ribosome). Most modules are loose assemblies of interacting molecules, and studies on them require new approaches in which LCLS may play a role. We believe, that studies on modules are likely to become a new norm of biological investigation, crucially aided by the other traditional disciplines and also by computational simulations.

At the extremes of the operational space of LCLS, novel diffraction and data processing methods may allow phase retrieval and image reconstruction from images of whole biological cells, molecular clusters and single biomolecules. These will not be reached on the first day of operations, but a pursuit in this direction is already in progress.

REFERENCES

1. Bohm, J., et al., Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. Proceedings of the National Academy of Sciences of the United States of America, 2000. 97(26): p. 14245-14250.
2. Boersch, H., Experimentelle Bestimmung der Energieverteilung in thermisch ausgelosten Elektronenstrahlen. Z. Phys. 139, 115 (1954).
3. Bottcher, B., Wynne, S. A., Crowther, R. A. Determination of the fold of the core protein of hepatitis b virus by electron cryomicroscopy. *Nature* **386**, 88-91 (1997).
4. Devaney, A.J., Structure determination from intensity measurements in scattering experiments. Physical Review Letters, 1989. 62(20): p. 2385--2388.
5. Ditmire, T. Simulation of exploding clusters ionized by high-intensity femtosecond laser pulses. *Phys. Rev.* **A57**, 4094-4097 (1998).
6. Doniach, S. Studies of structure of matter with photons from an X-ray free electron laser. *J. Synchr. Rad.* **3**, 260-267 (1996).
7. Dyson, N. A. X-rays in atomic and nuclear physics. Longman, (1973).
8. Fienup, J. R. (1982). Phase retrieval algorithm: a comparison. *Appl. Opt.* **21**, 2758-2769.
9. Fisher, A.J., Johnson, J.E. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361**, 176-179 (1993).
10. Gouet, P., Diprose, J.M., Grimes, J.M., Malby, R., Burroughs, J.N., Zientara, S., Stuart, D.I., Mertens, P.P.C. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. *Cell* **97**, 481-490 (1999).
11. Grimm, R., et al., Electron tomography of ice-embedded prokaryotic cells. Biophysical Journal, 1998. 74: p. 1031-1042.
12. Hajdu, J. Laue crystallography of macromolecules. In Frontiers in drug research, Alfred Benzon Symposium 28, Jensen, B., Jorgensen, F. S., Kofod, H., editors, pp. 375-395. Munksgaard, Copenhagen (1990).

13. Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and X-ray diffraction. *Proc. R. Soc.* **241**, 6-8 (1990).
14. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quart. Rev. Biophys.* **28**, 171-193 (1995).
15. Henke, B, L., Gullikson, E. M. & Davis, J. C. X-ray interactions: photoabsorption, scattering, transmission, and reflection at E = 50 - 30,000 eV, Z = 1- 92. *At. Data Nucl. Data Tables* **54**, 181-342 (1993).
16. Hubbell, J. H., Veigele, W. J., Briggs, E. A., Brown, R. T., Cromer, D. T., Howerton, R. J. Atomic form factors, incoherent scattering functions, and photon scattering cross sections. *J. Phys. Chem. Ref. Data* **4**, 471-494 (1975).
17. Krause, M. O. Oliver, J. H. Natural width of atomic K and L level Ka X-ray lines and several KLL Auger lines. *J. Phys. Chem. Ref. Data* **8**, 329-338 (1979).
18. Jacobsen, C., R. Medenwaldt, and S. Williams, A perspective on biological x-ray and electron microscopy, in X-ray Microscopy and Spectromicroscopy, J. Thieme, et al., Editors. 1998, Springer-Verlag: Berlin. p. II-93--102.
19. Lin, T. W., Porta, C., Lomonossoff, G., Johnson, J. E. Structure-based design of peptide presentation on a viral surface: The crystal structure of a plant/animal virus chimera at 2.8•ngstršm resolution. *Fold. & Des.* **1**, 179-187 (1996).
20. Miao, J., Hodgson, K.O., Sayre, D. (2001) An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. *Proc. Natl. Acad. Sci. USA* **98**, 6641Ð6645.
21. Miao, J. W., Charalambous, P., Kirz, J., Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**, 342-344 (1999).
22. Miao, J. Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transform of nonperiodic objects. *J. Opt. Soc. Am.* **A15**, 1662-1669 (1998).
23. Miao, J. & Sayre, D. On possible extensions of X-ray crystallography through diffraction pattern oversampling. *Acta Cryst.* **A**, to be published (2000).
24. Nave, C. Radiation-damage in Protein Crystallography. *Radiat. Phys. Chem.* **45**, 483-490 (1995)
25. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. Hajdu, J. Femtosecond imaging of biomolecules with X-rays. *Nature* **406**, 752-757 (2000).
26. Persson, P., Lunell, S., Szoeke, A., Ziaja, B. & Hajdu, J. (2001) Shake-up and shake-off excitations with associated electron losses in X-ray studies of proteins. *Protein Science* **10**, 2480-2484.
27. Peuker, M., High-efficiency nickel phase zone plates with 20 nm minimum outermost zone width. Applied Physics Letters, 2001. 78(15): p. 2208-2210.
28. Reiss, H. R. Physical basis for strong-field stabilization of atoms against ionization. *Laser Physics* **7**, 543-550 (1997).
29. Sayre, D. (2002) X-Ray Crystallography: The Past and Present of the Phase Problem. *Structural Chemistry* **13**, 81-96.
30. Sayre, D. and H.N. Chapman, X-Ray Microscopy. Acta Crystallographica A, 1995. 51: p. 237--252.
31. Sayre, D. (2002) X-Ray Crystallography: The Past and Present of the Phase Problem. *Structural Chemistry* **13**, 81-96.
32. Solem, J. C. Imaging biological specimens with high-intensity soft X-rays. *J. Opt. Soc. Am.* **B3**, 1551-1565 (1986).
33. Spector, S., C. Jacobsen, and D. Tennant, Process optimization for production of sub-20 nm soft x-ray zone plates. Journal of Vacuum Science and Technology, 1997. B 15(6): p. 2872--2876.
34. Stowell, M. H. B., Miyazawa, A., Unwin, N. Macromolecular structure determination by electron microscopy: new advances and recent results. *Curr. Opin. Struct. Biol.* **8**, 595-60 (1998).

35. Svensson, S., Martensson, N., Gelius, U. Observation of autoionizing resonances in core-electron shakeup spectra. *Phys. Rev. Letts.* **58**, 2639-2641 (1987).

36. Szoke, A. (2001) Diffraction of partially coherent X-rays and the crystallographic phase problem. *Acta Crystallogr* **A57**, 586-603.

37. Szske, A. Time-resolved holographic diffraction at atomic resolution. *Chem. Phys. Letts.* **313**, 777-788 (1999).

38. Tegze, M., Faigel, G. (2001) X-ray holography: theory and experiment. *J Phys-Condens Mat* **13**, 10613-10623.

39. Teilhard de Chardin, P. (1955) The Phenomenon of Man.

40. Tito, M. A., Tars, K., Valegard, K., Hajdu, J., Robinson, C. V. Electrospray time of flight mass spectrometry of the intact MS2 virus capsid. *J. Am. Chem. Soc.* **122**, 3550-3551 (2000).

41. van Heel, M. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction, *Ultramicroscopy* **21**, 111-124.

42. van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M. (1996) A new-generation of the IMAGIC image-processing system. *J. Str. Biol.*, **116**, 17-24.

43. van Heel, M. Orlova, E. V., Harauz, G. Stark, H. Dube, P. Zemlin, F, Schatz, M. Angular reconstitution in three-dimensional electron microscopy. Historical and theoretical aspects, *Scanning Micros.* **11**, 195-210 (1997).

44. ValegŒErd., K. Liljas, L., Fridborg, K., Unge, T. The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36-41 (1990).

45. Veigele, W. J. Photon cross sections from 0.1 keV to 1 MeV for elements with Z=1 to Z=94. *Atomic Data* **5**, 51-111 (1973).

46. Wang, Y., et al., Soft x-ray microscopy with a cryo STXM: II. Tomography. Journal of Microscopy, 2000. 197: p. 80-93.

47. Weiss, D., et al., Computed tomography of cryogenic biological specimens based on x-ray microscopic images. Ultramicroscopy, 2000. 84: p. 185-197.

48. Wolf, E., Three-dimensional structure determination of semi-transparent objects from holographic data. Optics Communications, 1969. 1: p. 153--156.

49. Ziaja, B., van der Spoel, D., Szoke, A. & Hajdu, J (2001) Auger-electron cascades in diamond and amorphous carbon. *Physical Review B* **64**, 214104 (8 pages).

50. Ziaja, B., van der Spoel, D., Szoke, A. & Hajdu, J (2002) The spatial evolution of low energy electron cascades in diamond. *Physical Review B* (submitted).

THE PHASE PROBLEM: DOES SIZE MATTER?

Randy J. Read; Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 2XY, U.K.

Defining "the phase problem"
In the X-ray crystallographic experiment, the interaction of X-rays with electrons in a crystal gives rise to a diffraction pattern. Because of the physics of diffraction, this diffraction pattern is the Fourier transform of the electron density distribution. Fourier transforms can, in principle, be reversed so it is possible to reconstruct the distribution of electrons (hence of atoms) in the crystal. Unfortunately, to do this we need not only the amplitudes of the diffracted waves, which are obtained from the intensities of the diffraction spots, but also their relative phase shifts, which cannot be measured directly. This is the well-known phase problem.

Often the phrase "The Phase Problem" is used to mean the challenge of determining the phases from one set of measured amplitudes. In early days this was done by interpreting Patterson maps, which show peaks for vectors between pairs of atoms, but this has largely been superceded by probabilistic direct methods. For direct methods, size matters very much, as the sharpness of phase probability distributions is inversely proportional to the square root of the number of atoms. The most powerful direct methods are the recent dual-space methods (Weeks & Miller, 1999; Sheldrick, 1997), but even these are currently limited to relatively small protein structures (up to, say, 1000 non-hydrogen atoms) with atomic (better than 1.2Å) resolution data.

For structures that are bigger or diffract to lower resolution, the phase problem is solved case-by-case using experimental phasing techniques, which supplement the native amplitudes with other measurements from which phases can be deduced. It is more interesting to consider whether size matters for these methods. We know, from successes in determining the structures of very large viruses, that the relevant measure of size is the unique part of the structure. Internal symmetry provides strong constraints, making it easier to solve highly symmetric structures.

The question I propose to address is the following. Are current phasing methods sufficient to solve the structure of any molecular assembly that might be crystallized? Or are there assemblies so large that new breakthroughs will be needed?

The basis of experimental phasing methods
Most new structures are solved by experimental phasing methods based either on isomorphous replacement or anomalous dispersion. Both approaches give phase information indirectly by perturbing the diffraction pattern through an effect involving a subset of atoms. In isomorphous replacement, "heavy" atoms are added to the native crystal, changing the diffraction pattern, whereas in anomalous dispersion, certain atoms (those with absorption maxima near the energy of the X-rays) introduce a phase shift to their contribution to the diffraction pattern, which perturbs the equality that would otherwise be found between the intensities of "Friedel pairs" of diffraction spots. If we know the contribution of the subset of atoms to the diffraction pattern, then the change in amplitude gives information about the original phase angle. Instead of solving the entire structure to get phase information, it is then only necessary to solve the substructure of the special (heavy or anomalous) atoms, which is generally a much simpler problem.

How many atoms must there be in the substructure?
For the isomorphous replacement or anomalous dispersion experiment to give any phase information, it is obvious that the size of the predicted change in amplitude must be at least comparable to the experimental uncertainties in the amplitudes. The size of the change in amplitude will depend on the scattering power of the atoms in the substructure (hence the requirement for "heavy" atoms or for atoms with a large anomalous dispersion contribution), and also on the number of such atoms in the substructure. The expected fractional change in amplitude can be estimated with simple diffraction ratio formulas, which assume that the special atoms are distributed randomly through the crystal.

Crick and Magdoff (1956) first estimated the signal that would be found in an isomorphous replacement experiment, obtaining an equation equivalent to the following:

$$\frac{\text{rms}(\Delta F)}{\text{rms}(|F|)} \approx \sqrt{\frac{N_H}{2N_T} \frac{f_H}{Z_{eff}}}$$

In this equation, Z_{eff} is the effective atomic number of the atoms in the native structure (N_T in number), which can be taken as 6.7 for proteins (Hendrickson & Ogata, 1997), and N_H is the number of heavy atoms, with scattering power f_H. Hendrickson & Ogata (1997) give a similar equation for the anomalous dispersion experiment:

$$\frac{\text{rms}(\Delta F_{ano})}{\text{rms}(|F|)} \approx \sqrt{\frac{N_A}{2N_T} \frac{2f_A^{''}}{Z_{eff}}}$$

There is an extra factor of two in the signal for anomalous dispersion because the anomalous scattering component contributes in opposite directions for the two Friedel mates. For convenience in computing the expected signal, the approximate number of non-hydrogen atoms in a protein can be obtained by assuming that their average atomic weight is 14.7Da.

How many atoms can we find in the substructure?
Until a few years ago, the positions of atoms in the substructure were found by interpreting difference Patterson maps, either manually or automatically (Sheldrick, 1991). The number of peaks in a Patterson map goes up with the square of the number of atoms, so this procedure quickly becomes very difficult above about 10 atoms. The introduction of dual-space direct methods for substructure solution has pushed the limits tremendously, particularly for anomalous dispersion data. These direct methods work to solve substructures at much worse than atomic resolution, because the heavy or anomalous atoms are typically separated by distances large compared to the resolution of the data. Every year the record for the number of anomalous scatterers found by these methods increases dramatically. To the best of my knowledge, the current record is 160 Se sites in the crystal of ketopantoate hydroxymethyltransferase, which contains two decamers (with a monomer molecular weight of 26kDa) in the asymmetric unit (Von Delft & Blundell, personal communication).

Given the rate of improvement, it seems reasonable to postulate that the dual-space methods will eventually be able to solve substructures of similar size to the structures solved with atomic resolution data, *i.e.* up to about 1000 heavy or anomalous atoms.

If that is not sufficient, a bootstrap approach using heavy atom cluster compounds has been shown to work for large structures such as ribosomal subunits (Schluenzen *et al.*, 2000). Heavy-atom clusters diffract as superatoms, at resolutions below the dimensions of the cluster, giving a very clear signal in the isomorphous or anomalous differences. If the clusters are well-ordered and the orientation of the cluster can be determined, they provide high resolution phase information (Schluenzen *et al.*, 2000; Clemons *et al.*, 2001). But even low resolution phase information is sufficient to determine the positions of a large number of special atoms, which will typically be quite well separated (Clemons *et al.*, 2001), thus pushing the phasing to much higher resolution. Low resolution information from a complementary technique such as electron microscopy image reconstruction can also be used in such a bootstrap approach. Clemons *et al.* (2001), in discussing the solution of the small ribosomal subunit, provide an excellent review of the current state of the art in solving very large structures.

What are the limits?
Let us assume that it will be possible to grow crystals that diffract to a sufficient resolution, that it will be possible to collect data within a precision of 5-10%, that the crystals can be derivatized with 1000 heavy atoms or anomalous scatterers, and that we will be able to solve the substructure of these 1000 special atoms. Then we can use the diffraction ratio equations given above to work out the size of assembly for which useful phase information would be obtained, as shown in the following table.

Size of protein assembly phased with 1000 atom substructure

"Special" atom and type of experiment	f or f''	Size of assembly phased with 5% signal (MDa)	Size of assembly phased with 10% signal (MDa)
Se anomalous	6e	9.5	2.4
Os anomalous	20e	105.	26.

| Os isomorphous | 76e | 380. | 95. |
| U isomorphous | 92e | 560. | 140. |

From this table we see that the potential exists to phase enormous assemblies, without even using the bootstrapping approaches of low resolution phasing with cluster compounds or EM images.

Up to now, the biggest molecular complex that has been studied crystallographically (excluding viruses and other objects with high internal symmetry) is the whole prokaryotic ribosome. This has a molecular weight of about 2.5MDa. However, there are larger molecular machines with a defined structure, the largest of which is probably the vertebrate nuclear pore complex. With a molecular weight of about 125Mda (Reichelt *et al.*, 1990), it is fifty times the size of the prokaryotic ribosome, but still within the range of structures for which phase information could be obtained.

In fact, given good crystals, the structure of the nuclear pore complex would be easier to determine than the results of the table imply. The nuclear pore complex has eight-fold rotational symmetry, which reduces the size of the unique object to about 16MDa, and much of the core has sixteen-fold symmetry (Rout *et al.*, 2000).

What is the problem?
In a practical sense, the phase problem of macromolecular crystallography has been solved with methods that can work with assemblies as large as anything we can imagine crystallizing. Size does matter, in that the experiments will get harder with increasing size, but it is not an insuperable obstacle and no new breakthroughs are needed. If we can obtain crystals that diffract to sufficient resolution, collect data to reasonable precision, and arrange to have a suitable number of special (heavy or anomalous) atoms bound to the crystal, then we will be able to estimate phases to solve the structure. These are big ifs, but it moves the emphasis from phasing as the problem to all the other aspects of the experiment.

Crystallization is likely to be the biggest bottleneck for large assemblies, as it often is for smaller proteins. The larger the molecular machine, the more complicated its mechanism will be, and the more difficult it will become to ensure that every assembly in solution is in the same functional state, with the same collection of accessory molecules bound.

Data collection will also not be trivial for extremely large unit cells. Nonetheless, the determination of structures of very large viruses has shown that it is possible to collect data from such cells. The HK97 virus capsid (Wikoff *et al.*, 2000) crystallizes in a unit cell with an asymmetric unit volume of about 1.4×10^8 Å^3, which is larger than the volume that would be required to accommodate one-quarter of a nuclear pore complex (assuming this crystallized with its 8-fold axis along a crystallographic 4-fold axis). Data of reasonable quality were collected to 3.6Å resolution, although a total of 61 crystals was required.

In general, radiation damage is more of a problem for crystals of large assemblies, leading to the necessity to merge data from multiple crystals. All else being equal, crystals exposed to the same X-ray source should decay at the same rate, but when the cell is larger the diffracted radiation is divided into more diffraction spots, each of which must be observed longer to get sufficient signal-to-noise. Of course, there is a great deal of variation from sample to sample; for instance, complete data sets could be collected from one or two crystals of the large ribosomal subunit (Ban *et al.*, 2000), whereas only a few degrees of data could be collected from each crystal of the small subunit (Clemons *et al.*, 2001).

Conclusion
For even the largest molecular machines we can contemplate crystallizing, the phase problem is not the problem for determining their structures by crystallography.

Acknowledgements

References

1. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4A resolution", *Science* **289**: 905-920.

2. Clemons, W.M.Jr., Brodersen, D.E., McCutcheon, J.P., May, J.L.C., Carter, A.P., Morgan-Warren, R.J., Wimberley, B.T. & Ramakrishan, V. (2001). "Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination", *J. Mol. Biol.* **310**: 827-843.

3. Crick, F.H.C. & Magdoff, B.S. (1956). "The theory of the method of isomorphous replacement for protein crystals", *Acta Crystallog.* **9**: 901-908.

4. Hendrickson, W.A. & Ogata, C.M. (1997). "Phase determination from multiwavelength anomalous diffraction measurements", *Methods Enzymol.* **276**: 494-523.

5. Reichelt, R., Holzenburg, A., Buhle, E.L.Jr., Jarnik, M., Engel, A. & Aebi, U. (1990). "Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components", *J. Cell Biol.* **110**: 883-894.

6. Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y & Chait, B.T. (2000). "The yeast nuclear pore complex: composition, architecture and transport mechanism", *J. Cell. Biol.* **148**: 635-651.

7. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). "Structure of functionally activated small ribosomal subunit at 3.3 Ångstroms resolution", *Cell* **102**: 615-623.

8. Sheldrick, G.M. (1991). "Tutorial on automated Patterson interpretation to find heavy atoms", in *Crystallographic Computing 5*, D. Moras, A.D. Podjarny & J.C. Thierry (Eds.), Oxford University Press: Oxford, pp. 145-157.

9. Sheldrick, G.M. (1997). "Direct methods based on real/reciprocal space iteration", in *Recent Advances in Phasing: Proceedings of the CCP4 Study Weekend*, K.S. Wilson, G. Davies, A.S. Ashton, and S. Bailey (Eds.). CCLRC, Daresbury Laboratory: Warrington, UK, pp. 147-158.

10. Weeks, C.M. & Miller, R. (1999). "The design and implementation of *SnB* v2.0", *J. Appl. Cryst.* **32**: 120-124.

11. Wikoff, W.R., Schildkamp, W. & Johnson, J.E. (2000). "Increased resolution data from a large unit cell crystal collected at a third generation synchrotron X-ray source", *Acta Crystallogr.* **D56**: 890-893.

SINGLE PARTICLES ALWAYS FIT THE MOLD

Nikolaus Grigorieff, Howard Hughes Medical Institute, Brandeis University, MS029, 415 South Street, Waltham, MA 02454-9110

Introduction

An explanatory footnote to the Preface of the latest edition of the popular text "Three-Dimensional Electron Microscopy of Macromolecules and Their Assemblies" [1] explains: "X-ray crystallography was a common technique to determine the 3D structure of proteins around the turn of the century. It required extensive amounts of protein and cumbersome crystallization to obtain diffraction patterns, leading to the well-known phase problem. It is still used today to study small molecules at ultrahigh-resolution". The Preface notes that by far, the most common technique to determine 3D protein structures is electron microscopy and single particle averaging because of "its in-house accessibility, its universal scope, reaching from protein molecules to complex assemblies, its simplicity and ease of use and, most importantly, its guarantee for an atomic structure". It concludes that "this technique was instrumental in the completion of one of our most challenging quests, the visualization of the Human Proteome".

Though this book is not yet written, it is useful for this discussion to project into the future and evaluate what will remain science fiction. Using Single Particle Electron Microscopy (SPEM), structures have been obtained at about 7 Å resolution with highly symmetrical viruses [2-4], and at about 10 Å with the asymmetrical ribosome [5, 6]. SPEM has a unique niche in the visualization of large protein complexes, as exemplified over the last 30 years by the ribosome structure. It is gaining scope in image processing of helical structures [7] and 2D crystals [8-10], pointing to a single method for all possible sample geometries. However, with the potential of a universally applicable technique comes also the potential of strongly biasing a structure towards an initial (sometimes wrong) reference used for alignment. SPEM is also being introduced as a "new powerful technique" in many publications and grant applications. However, when compared to X-ray crystallography, the number of publications reveals that the latter is at least 50 times more powerful [1], and when comparing the amount of data contained in a structure on average in each case, X-ray crystallography is about 300 times as powerful. What are the reasons for this stark contrast?

Problems

<u>Automation of data collection and processing:</u> Rapid data collection and robust, semi-automatic data processing has made X-ray crystallography accessible to a large community. Automatic data collection on the EM is being developed [11, 12]. However, many EM labs are developing their own customized software that is often not easy to use by outsiders. In addition, programs are often not well documented and tend to have bugs, which led a prominent crystallographer to ask for the development of "programs and packages in a way that could be used by merely outstanding mortals, rather than demi-gods" [13].

<u>Conformational variability:</u> Conformational inhomogeneity in single particles cannot always be readily detected, especially when the conformational differences are small. Though insignificant at low resolution, conformationally variable parts might not be resolved at the atomic level. This is also a problem in crystallographic approaches though crystal contacts tend to promote a single conformation. Further discussed by Edward Egelman.

<u>Charging and beam-induced movement:</u> Both lead to blurring and degraded high-resolution contrast. They appear to be more severe for single particle work than 2D crystallography as single particles are often imaged while suspended in ice over a hole in the carbon support film [14].

<u>Quantitative understanding of image contrast:</u> In order to determine a 3D structure the experimental data is usually interpreted by a model. Its accuracy therefore depends on a complete and quantitative accounting of the image contrast. This should include not only single elastic scattering events, but also inelastic scattering and, if significant, multiple scattering.

---

[1] Someone may say there are many more X-ray crystallographers than electron microscopists, but one could reply that there are many more electron microscopes than syncrotrons.

Furthermore, the solvent and support film should be taken into account, including contrast originating from rough ice and carbon surfaces. The theory of image contrast is well developed in materials science, where the quantitative interpretation of electron micrographs and diffraction patterns is done routinely (see for example, [15]). Several studies of inelastic image contrast of frozen-hydrated biological specimens embedded in ice have been performed [16-18], of which the analysis by

Langmore & Smith [17] is the most complete. It has been suggested to use energy filtering to remove inelastically scattered electrons [17-19]. However, energy filters like the Omega filter built by Zeiss, or the Imaging Electron Energy Loss Spectrometer sold by Gatan, are presently not suitable for high resolution since scattering angles are restricted [16], the size of the CCD camera only records a small area of an image, and the filters introduce non-linear distortions in the images. Even more serious is the fact that most inelastically scattered electrons (with energy losses of about 20 eV) retain partial coherence and carry a significant amount of image contrast [20]. The image formed by the inelastically scattered electrons is a replica of the elastic image [15, 20]. To exclude those electrons from the analysis would increase the number of images necessary to obtain a given signal-to-noise ratio in the final structure. A better approach would be to include inelastically scattered electrons in the image analysis by modeling inelastic image contrast.

CTF correction: Although the CTF of an electron microscope is well-understood, its precise parameters have to be determined not only for each image, but for each particle in an image since the vertical particle positions are all different [21]. At high resolution, the curvature of the Ewald sphere also has to be taken into account. Apart from the defocus amount and astigmatism (3 parameters), the beam tilt also has to be determined for each image (2 parameters). This is routinely done in electron crystallography [22] but poses a problem for an image of single particles since the signal at high resolution is much weaker than that from a 2D crystal.

Accurate particle alignment: To determine a high-resolution structure, the 5 parameters determining position and orientation of each particle must be accurately known [23]. Inaccurate determination of particle parameters is likely to be one of the most severe resolution-limiting factors at the moment. Due to the weakening signal towards higher resolution, accurate alignment is difficult at best.

Quality measurement: A common way to evaluate a 3D reconstruction is its resolution. Unfortunately, agreement on measuring resolution is hard to reach [21, 24]. Usually, the resolution is based on the Fourier Shell Correlation (FSC) between reconstructions derived from two halves of the data set, and calculated in resolution zones. One proposal is to set the resolution at the crossover point between the FSC and three standard deviations (3• criterion) of the correlation of random data [21, 25]. Although the 3• criterion indicates a constant probability of presence of data at a particular resolution, it does not say much about the interpretability of the data. Furthermore, the 3• criterion depends on the size of the volume the reconstructed density sits in: Padding the volume will lower the 3• threshold. Another proposal is to set the resolution at the point where the FSC drops below 0.5 [3, 26]. This criterion does not depend on the size of the volume, and it indicates the point where the signal-to-noise ratio (SNR) in each reconstruction is 1 [27]. An additional complication is the bias of the FSC towards higher values as the two reconstructions are not independent but are usually calculated from images which were all aligned against a single reference [28].

Discussion

Developing the technique: When following the development of X-ray crystallography, it is clear that automation of data collection and processing is an important ingredient for a technique to become widely used and accepted. Given the progress already made [11, 12, 29-35] it is reasonable to expect that advances in instrumentation and software will eventually give us high throughput and ease of use: Samples will be applied to engineered grids, allowing for fully automated data collection overnight on a field-emission, liquid-helium electron microscope. As each image is recorded on a CCD, a preliminary evaluation of data quality will be carried out and particles will be selected, aligned and classified without user assistance. Programs will be fully documented, will not crash, and will provide the user with standardized, reliable diagnostic data. Data and a history of operations will be stored in a comprehensive data base. The programs will run on a Linux cluster with 1000 terraflop capacity, and soon after data collection finishes, the user will be presented with a 3D model of his/her protein, together with resolution curves, real-space reliability maps, and a projection of how much more data is needed to reach a specified resolution. The map will be stored in a standard map format, and tools will be provided to fit other models (atomic or otherwise) from a library of known structures into the map.

A fundamental question: A more fundamental problem may be the resolution one could hope to achieve with SPEM. Most of the problems listed above have a direct impact on resolution. Given a sufficiently high SNR, as with 2D crystals, one could obtain an atomic model using EM. Due to their size and symmetry, viruses are leading the way to atomic resolution in the realm of single particles, and this has prompted the use of viruses as scaffolds for other proteins [36]. But is there any

reason to believe that viruses, let alone asymmetric particles like the ribosome, could ever be aligned with sufficient accuracy to obtain an atomic model? It has been predicted that this could be done if the average SNR is sufficiently high to reliably detect the correct correlation peak in a cross-correlation map [37, 38]. This prediction is based on an average signal across the entire resolution range considered and does not account for the rapid signal decay normally observed towards higher resolution [39]. Furthermore, additional degrees of freedom due to parameters such as magnification, defocus/astigmatism, and beam tilt are not considered. *Can an accurate alignment at high resolution be achieved solely based on low-resolution signal?*

This will depend strongly on a detailed understanding of the low-resolution contrast and the noise distribution in the images. New approaches to particle alignment and 3D reconstruction, such as the use of maximum likelihood functions [40], or labeling with heavy-atom clusters as anchor points [41], may help. Furthermore, the signal at higher resolution can be enhanced by reducing charging effects, for example, by using higher electron beam energies, or by preparing the samples with a sandwich of two carbon layers which also reduces beam-induced movement [42, 43].

The question of resolution is intimately linked with the problem of resolution measurement. If one assumes for a moment that a reliable measurement exists, it is possible, in principle, to determine all necessary parameters to obtain any desirable resolution (assuming the images contain the necessary data). All one would have to do is to carry out an exhaustive search of all relevant parameters, and to evaluate the corresponding 3D reconstruction for every test parameter set. The problem of conformational flexibility could be handled in a similar manner by testing all possible subsets of a data set to find a homogeneous set of particles. The main problem with this type of approach is the prohibitively high computational cost but, as the past twenty years suggest, this problem will be solved by time.

Measuring resolution: The FSC would be an unbiased resolution measurement if the two reconstructions calculated from two half data sets were truly independent, i.e. if two references were used to refine parameters in each set. The disadvantage is the reduced signal-to-noise ratio of the two references as they would each contain only half the data. An alternative would be the exclusion of a certain percentage of data, say 5 or 10%, from the refinement-reconstruction cycle. A correlation coefficient or phase residual in resolution zones could then be calculated between the excluded images and the reconstruction to measure resolution. This idea, which is already implemented in the program FREALIGN [44], is analogous to the free R-factor used in X-ray crystallography [45].

References

1. Frank, J. and M. van Heel (2010). *Three-dimensional electron microscopy of macromolecules and their assemblies.* 10th ed. San Diego: Academic Press.
2. Conway, J.F., et al. (1997). *Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy.* Nature 386, 91-94.
3. Böttcher, B., S.A. Wynne, and R.A. Crowther (1997). *Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy.* Nature 386, 88-91.
4. Zhou, Z.H., et al. (2001). *Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus.* Nat Struct Biol 8, 868-73.
5. Gabashvili, I.S., et al. (2000). *Solution structure of the E. coli 70S ribosome at 11.5 A resolution.* Cell 100, 537-49.
6. Matadeen, R., et al. (1999). *The Escherichia coli large ribosomal subunit at 7.5 A resolution.* Structure Fold Des 7, 1575-83.
7. Egelman, E.H. (2000). *A robust algorithm for the reconstruction of helical filaments using single-particle methods.* Ultramicroscopy 85, 225-34.
8. Sherman, M.B., et al. (1998). *Multivariate analysis of single unit cells in electron crystallography.* Ultramicroscopy 74, 179-99.
9. Stahlberg, H., et al. (1998). *Are the light-harvesting I complexes from Rhodospirillum rubrum arranged around the reaction centre in a square geometry?* J Mol Biol 282, 819-31.
10. Walz, T. and R. Ghosh (1997). *Two-dimensional crystallization of the light-harvesting I-reaction centre photounit from Rhodospirillum rubrum.* J Mol Biol 265, 107-11.
11. Oostergetel, G.T., W. Keegstra, and A. Brisson (1998). *Automation of specimen selection and data acquisition for protein electron crystallography.* Ultramicroscopy 74, 47–59.
12. Rouiller, I., et al. (2001). *Automated image acquisition for single-particle reconstruction using p97 as the biological sample.* J Struct Biol 133, 102-7.
13. Harrison, S. (2002). Personal communication.
14. Böttcher, B. (1995). *Electron cryo-microscopy of graphite in amorphous ice.* Ultramicroscopy 58, 417-424.

15. Hirsch, P.B., et al. (1965). *Electron microscopy of thin crystals.* London: Butterworths.

16. Angert, I., et al. (1996). *Elastic and inelastic scattering cross-sections of amorphous layers of carbon and vitrified ice.* Ultramicroscopy 63, 181-192.

17. Langmore, J.P. and M.F. Smith (1992). *Quantitative energy-filtered electron-microscopy of biological molecules in ice.* Ultramicroscopy 46, 349-373.

18. Zhu, J., et al. (1997). *Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S Escherichia coli ribosome.* Journal of Structural Biology 118, 197-219.

19. Grimm, R., et al. (1997). *Energy filtered electron tomography of ice-embedded actin and vesicles.* Biophysical Journal 72, 482-489.

20. Rossouw, C.J. and M.J. Whelan (1981). *Diffraction contrast retained by plasmon and K-loss electrons.* Ultramicroscopy 6, 53-66.

21. van Heel, M., et al. (2000). *Single-particle electron cryo-microscopy: towards atomic resolution.* Q Rev Biophys 33, 307-69.

22. Henderson, R., et al. (1986). *Structure of purple membrane from Halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution.* Ultramicroscopy 19, 147-178.

23. Jensen, G.J. (2001). *Alignment error envelopes for single particle analysis.* J Struct Biol 133, 143-55.

24. Ruprecht, J. and J. Nield (2001). *Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction.* Prog Biophys Mol Biol 75, 121-64.

25. Saxton, W.O. and W. Baumeister (1982). *The correlation averaging of a regularly arranged bacterial cell envelope protein.* J Microsc 127, 127-38.

26. Beckmann, R., et al. (1997). *Alignment of conduits for the nascent polypeptide chain in the ribosome- Sec61 complex.* Science 278, 2123-6.

27. Frank, J. and L. Al-Ali (1975). *Signal-to-noise ratio of electron micrographs obtained by cross correlation.* Nature 256, 376-379.

28. Grigorieff, N. (2000). *Resolution measurement in structures derived from single particles.* Acta Crystallogr D Biol Crystallogr 56, 1270-7.

29. Downing, K.H. and F.M. Hendrickson (1999). *Performance of a 2k CCD camera designed for electron crystallography at 400 kV.* Ultramicroscopy 75, 215-33.

30. Frank, J., et al. (1996). *SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields.* Journal of Structural Biology 116, 190-199.

31. Lata, K.R., P. Penczek, and J. Frank (1995). *Automatic particle picking from electron micrographs.* Ultramicroscopy 58, 381–391.

32. Ludtke, S.J., P.R. Baldwin, and W. Chiu (1999). *EMAN: semiautomated software for high-resolution single-particle reconstructions.* J Struct Biol 128, 82-97.

33. Nicholson, W.V. and R.M. Glaeser (2001). *Review: automatic particle detection in electron microscopy.* J Struct Biol 133, 90-101.

34. Stewart, P.L., et al. (2000). *Digitally collected cryo-electron micrographs for single particle reconstruction.* Microsc Res Tech 49, 224-32.

35. van Heel, M., et al. (1996). *A new generation of IMAGIC image processing system.* Journal of Structural Biology 116, 17-24.

36. Kratz, P.A., B. Bottcher, and M. Nassal (1999). *Native display of complete foreign protein domains on the surface of hepatitis B virus capsids.* Proc Natl Acad Sci U S A 96, 1915-20.

37. Glaeser, R.M. (1999). *Review: electron crystallography: present excitement, a nod to the past, anticipating the future.* J Struct Biol 128, 3-14.

38. Henderson, R. (1995). *The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules.* Q Rev Biophys 28, 171-93.

39. Saad, A., et al. (2001). *Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination.* J Struct Biol 133, 32-42.

40. Sigworth, F.J. (1998). *A maximum-likelihood approach to single-particle image refinement.* Journal of Structural Biology 122, 328-339.

41. Jensen, G.J. and R.D. Kornberg (1998). *Single-particle selection and alignment with heavy atom cluster- antibody conjugates.* Proc Natl Acad Sci U S A 95, 9262-7.

42.   Brink, J., et al. (1998). *Reduction of charging in protein electron cryomicroscopy.* J Microsc 191, 67-73.

43.   Jakubowski, U., W. Baumeister, and R.M. Glaeser (1989). *Evaporated carbon stabilizes thin, frozen-hydrated specimens.* Ultramicroscopy 31, 351–358.

44.   Grigorieff, N. (1998). *Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice.* Journal of Molecular Biology 277, 1033-1046.

45.   Brünger, A.T. (1992). *Free R value: a novel statistical quantity for assessing the accuracy of the structure.* Nature 355, 472-475.

TUBULIN AND MICROTUBULE STRUCTURES – THE BEST OF ELECTRON CRYSTALLOGRAPHY AND SINGLE PARTICLE APPROACHES

Kenneth H. Downing. Life Sciences Division, Donner Laboratory, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Abstract
The solution of the structure of tubulin by electron crystallography provided great insight into some of the mechanical aspects of microtubule functions. Docking the crystal structure into a very high resolution 3-D reconstruction from cryo-EM images of intact microtubules now provides even fuller insight, suggesting differences in conformation between the two structures that may be functionally significant.

Since the discovery of microtubules (MTs) nearly forty years ago, much has been learned about their essential roles on the life of all eukaryotic cells. Their functions are intimately tied to their mechanical properties, as they are long, stiff elements that can both provide structural rigidity within the cell and serve as tracks on which various organelles are carried within the cell. A requirement for fulfilling several of their roles is the ability to rapidly polymerize and depolymerize, illustrated most dramatically by the restructuring of the disperse, interphase MT network into the metaphase spindle which is responsible for separating chromosomes to the daughter cells. Dynamic instability, referring to the switching between phases of growth and disassembly, is an especially important feature of MTs as they explore the cell to find and capture the chromosomes. The process of cell division is delayed until each chromosome is attached to the correct number of MTs from each centrosome – the organizing center from which the MTs radiate. These dynamic properties, along with the importance of MTs in the life of the cell and their consequent value as targets for chemotherapy, have captured the interest of a large number of scientists from a range of disciplines over the years.

The basic structure of the MT has been known for many years. Protofilaments, around thirteen in number, associate in parallel to form the wall of the hollow MT. The protofilaments, in turn, are composed of alternating alpha and beta tubulin monomers. However, tubulin itself resisted many years of attempts to determine its structure. Tubulin polymerizes into a wide variety of forms, all of which appear to based on the same protofilament structure. As these are all flexible assemblies and highly polymorphic, none have yet been found to be suitable for x-ray crystallography. Ordered arrays of MTs have given some promising x-ray fiber diffraction data, which could probably be significantly improved with modern sources, but this data has not been particularly productive. One of the polymers, the two-dimensional crystalline sheets that form in the presence of zinc ions, is highly ordered and very well suited to study by electron microscopy, and it was this form that finally yielded the tubulin structure [1].

Electron crystallography is a tool particularly well adapted for determining protein structures from small, monolayer crystals such as can be grown with tubulin. The crystals are typically two microns or less on edge, and yet electron diffraction patterns extending to around 2.5 Å have been recorded. Collecting a full 3-D data set is not possible, though, and in fact the crystals are flexible enough that our ability to collect data within the accessible region of the 3-D reciprocal space is often compromised. However, the ability to determine phases directly from high resolution images produces density maps that are interpretable with little or no refinement, even to rather good resolution.   The initial structure of tubulin (PDB 1TUB) was as built directly into the density, which in turn was computed at 3.7 Å resolution. More recently the structure has been refined to 3.5 Å (PDB ID 1JFF), in the course of which a number of substantial improvements were made to the model [2].   The structure now maps almost entirely to allowed regions of a Ramanchandran plot, except for some poorly defined loop regions. The final R-factor is 23 % (free R-factor 29%), which is very good for electron crystallographic structures. Thus this is overall a very well defined structure.

The structure is basically a beta sheet core surrounded by alpha helices. It consists of a conventional Rossmann-fold nucleotide binding domain at the N-terminal region, an intermediate domain containing a mixed, four-strand beta sheet and three helices, and a third domain containing mainly two antiparallel helices that cross the first two domains. There are some regions that were apparently too disordered to build the backbone, including the C-terminal 10-15 residues, which are presumed to be normally disordered, and the B2-H3 loop, which probably suffers from interaction with the support film when the crystals are prepared for electron microscopy. In the rest of the structure, almost all side chain

orientations are clear, along with the two nucleotides and a molecule of paclitaxel which was added to stabilize the crystals. In addition, one Mg ion was identified near the GTP, and one Zn ion at the inter-protofilament interface. The ability to identify these cations is a testament to the power of electron crystallography and the quality of the map even at 3.5 Å.

As good as the crystal structure is, it still does not tell us everything we need to know about MT structure, which includes all the interactions between subunits of the structure. In the zinc-sheet crystals, adjacent protofilaments are antiparallel, while in MTs they are all parallel. Thus the crystal structure shows directly the longitudinal interactions within the protofilaments, but the inter-protofilament interactions are different in the two complexes.

In order to identify the lateral interactions, the crystal structure was docked into a 3-D reconstruction obtained by cryo-EM and helical reconstruction from intact MTs [3]. Although the resolution of the MT was only around 20 Å, this exercise resolved a number of ambiguities. For example, it defined the polarity of the crystal structure and confirmed that beta tubulin sits at the plus end of the MT, in a position that exposes the exchangeable nucleotide and explains the available GTP binding and exchange data.

In docking the crystal structure, there is essentially just one free parameter, the rotation about the protofilament axis. Determination of this rotation angle was judged to be accurate to about +/- 5 degrees. Although a rotation in this range can make a significant difference concerning the details of the interface, it was possible to clearly define which elements of the structure were involved in the inter-protofilament contacts. The M-loop, which was also involved in contacts in the crystals, plays a major role on one protofilament. Part of helix H3 and the N-terminal loops are involved from the adjacent protofilament. There were, though, some significant limitations to this docking that we sought to overcome by extending the MT resolution. First was the precision of the rotation angle, and thus details of the interaction. Then there was always the possibility that the tubulin conformation itself might be different in the two polymers. Particularly the M-loop could be expected to adopt different conformations, given the differences in the interactions it makes.

We approached the MT reconstruction in several novel ways. First, images were recorded on a 400 kV microscope. The main advantage of the higher voltage in this context, aside from the usual issues of improved resolution, was the ability to work with thicker specimens. From the correlation between the apparent ice thickness and the quality of the images, we hypothesize that the thicker ice may provide sufficiently greater mechanical rigidity that the resolution is noticeably improved. Whereas typical "good" MT images from previous work show detectable diffraction on the 20-A layer line, the best of the 400 kV images show about as much signal on the 10-A layer line. Thus the starting images were substantially better than previously available. In order to be able to reach higher resolution, though, the images were taken somewhat closer to focus than before, which combined with the higher accelerating voltage (and possibly thicker ice) produced images with much lower contrast. We then used a much larger data set of images in order to raise the signal-to-noise ratio and extend the effective resolution. Altogether 85 MT images were selected, containing about 200,00 monomers.

Several methods for processing images of MTs have been used with success [4, 5], but we chose a single-particle approach which had not previously been used. Most MT reconstructions, including MTs decorated with motor molecules, have been done using helical methods. Aside from potential limitations of these methods, most MTs are not true helices, since the MT lattice often has a seam where the inter-protofilament interactions are different. Usually alpha tubulin monomers are apposed with alphas across the interface, and betas with betas, but at the seam alphas are apposed with betas. We chose to ignore the difference between alpha and beta, though, since to the expected resolution of around 8 Å the structures are nearly indistinguishable. Thus we could have considered the MTs to be true helices. However, there are some irregularities in the structure that seemed to be easier to deal with using single particle methods. With the helical methods the various distortions in the structure are determined, and the helix is then computationally straightened and flattened. The size scale of the distortions found in MTs, and the fact that all of the power in the Fourier transform of a MT is localized to a very small number of peaks, tend to favor dealing with the distortions in the single particle context. Perhaps more significantly, we also chose to use MTs containing 13 protofilaments, which are the majority in most preparations, and these have the problem of having overlapping Bessel functions on all layer lines. While this problem could presumably be overcome computationally, the single particle approach was intuitively more straightforward.

Microtubule images were boxed into segments that contained thirteen repeats along the axis. A typical segment is shown in fig. 1a. With this axial repeat and the thirteen-fold rotational symmetry, these might not be considered true single particles, and indeed some of the processing steps took advantage of the "particle" symmetry. In the first step, the in-plane rotation for each segment was determined using a Radon transform of the power spectrum, as shown in fig 1c. The Fourier transform of a helix comprises a set of layer lines, and projection precisely along these layer lines produces a distinctive set of peaks. Since the protofilaments in the 13-3 (13 protofilament, 3-start helix) MTs run parallel to the axis, without the supertwist found in other MT lattices, the layer lines are compact with no splitting. Thus the projection along

the equator is a very sensitive indicator of orientation in the Radon transform. With all segments from a given image rotationally aligned, they could be averaged to produce an image with remarkable detail (fig. 1d), with very well defined layer line amplitudes out to around 10 Å (fig. 1e). However, without taking into account other distortions in the images, the 3-D reconstruction obtained by combining data from over a dozen MTs did not reach significantly eyond 10 Å resolutiion.

Variations in the axial twist and out-of-plane tilt were determined by cross correlation of each segment with a set of reference images derived from a 3-D model. This model was built from the crystal structure, using a preliminary 3-D reconstruction to determine the placement of the atomic structure. Since the alignment parameters refer to the MT segment as a whole, rather than the monomers within the MT, using the crystal structure in constructing the reference did not introduce any unacceptable bias toward a particular structure. We checked this assertion experimentally by using a reference constructed with the monomers rotated 90 degrees about their axis, and obtained the same result as with the proper model orientation.

Figure 1: MT segments. (a) Boxed segment from one MT image. (b) Fourier transform of a. (c) Radon transform of b. (d) Average of aligned 28 segments from one micrograph. (e) Fourier trransform of d showing strong signal to 10 Å.
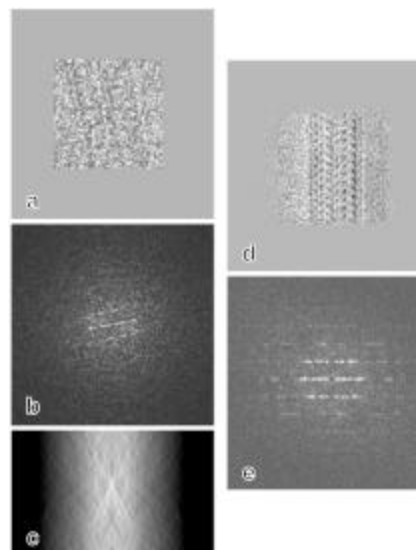
The final reconstruction, shown in fig. 2, represents a dramatic improvement in resolution over all previous MT reconstructions. Using criteria based on the Fourier shell correlation and differential phase residual, the resolution is approximately 8 Å. The map itself shows a wealth of detail consistent with this resolution estimate. Essentially all of the alpha helices are resolved as distinct densities, although not all at the same density. The beta sheet core is well separated from most of the rest of the structure, and even some of the loop regions show interpretable density.

This level of detail in the map allows an uncommonly high degree of precision in docking the atomic structure into the map, as shown in fig. 2. Docking can be done by eye with very little ambiguity. A quantitative measure of the precision, based on cross correlation between the map and a density computed from the crystal structure, indicates an accuracy of +/- 2 degrees in axial rotation and +/- 2 Å in translation. The docking is now seen to be slightly different from that originally determiined, with the monomers rotated about the axis in a way that brings the M-loop into more intimate contact with the N-terminal loops of the adjacent monomer. This contact presumably stabilizes the loops and thus accounts for the strong density that has been seen in previous MT reconstructions in this region, as opposed to the apparent disorder seen in the crystal structure. Two other notable differences between the new reconstruction and the crystal structure concern the M-loop itself and the helix H6 region. The M-loop appears to be shifted a few Angstroms along the axis, consistent with its presumed flexibility that allows a variety of MT sizes and polymer forms. H6 and the H6-H7 loop show only weak density in the map, suggesting that they are either more disordered in the MT than in the crystal or that they could be in different conformations in alpha and beta monomers.
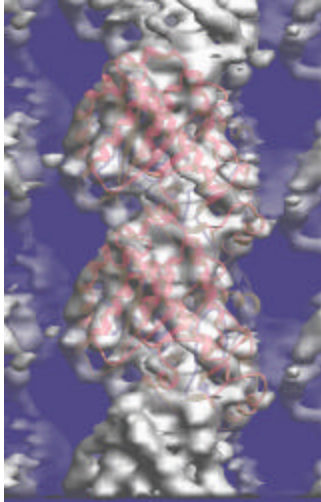
Figure 2: Surface view of the MT reconstruction, seen from the outside of the MT, with a ribbon diagram of the dimer docked in place.

Microtubules are such large, flexible structures that obtaining structural information at very high resolution from an ensemble, such as by x-ray diffraction, seems highly unlikely. The ability to use electron microscopy of individual MTs, though, can provide information at near atomic resolution. Combined with the crystal structure of the component subunits, EM provides a valuable structural model of the polymer. However, we have reached a level of resolution that suggests significant differences between the protein conformation in the crystal and polymer. This will always be a worrisome problem as one tries to dock structures together, until the resolution of the map into which the crystal structure is docked reaches sufficient resolution to resolve most of the secondary structure.

Literature Cited
1. Nogales, E., Wolf, S.G. & Downing, K.H. (1998). Structure of the ab tubulin dimer by electron crystallography. *Nature* **391**, pp. 199-203.
2. Löwe, J., Li, H., Downing, K.H. & Nogales, E. (2001). Refined structure of alpha beta-tubulin at 3.5 A resolution. *J. Mol. Biol.* **313**, pp. 1045-1057.
3. Nogales, E., Whittaker, M., Milligan, R.A. & Downing, K.H. (1999). High resolution model of the microtubule. *Cell* **96**, pp. 79-88.
4. Metoz, F., Arnal, I. & Wade, R. (1997). Tomography without tilt: three dimensional imaging of microtubule/motor complexes. *J. Struc. Biol.* **118**, pp. 159- 168.
5. Sosa, H., Hoenger, A. & Milligan, R.A. (1997). Three different approaches for calculating the three dimensional structure of microtubules decorated with kinesin motor proteins. *J. Struct. Biol.* **118**, pp. 149-158.

Speaker 9
Wolfgang Baumeister

No Speaker Paper Available

POLYMORPHISM, CAN WE DETECT IT? CAN WE USE IT? CAN WE CONTROL IT? EXAMPLES FROM ACTIN AND NUCLEOPROTEIN COMPLEXES
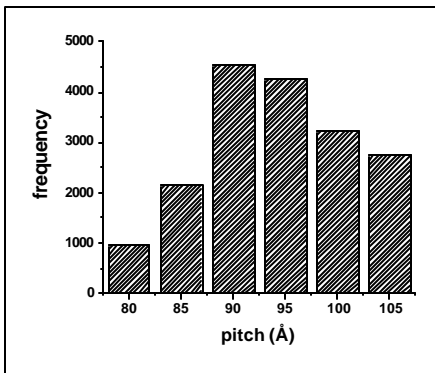
Vitold E. Galkin, Shixin Yang, Margaret S. VanLoock, Xiong Yu, Albina Orlova, and Edward H. Egelman, Dept. of Biochemistry and Molecular Genetics, University of Virginia Health Sciences Center, Charlottesville, VA 22908-0733

Disorder and Polymorphism in Helical Assemblies

Helical assemblies of proteins are ubiquitous in biology. These include, among many other complexes: F-actin, microtubules, myosin thick filaments, phage tails, and bacterial pili and flagella. The bacterial RecA protein, which can be the most abundant protein in the cell after DNA damage, forms a helical nucleoprotein filament with DNA. Many fields are dominated by our understanding of the structure of such helical macromolecular assemblies. For example, most current thought on the molecular basis for muscle contraction grows out of helical complexes of actin and myosin (Jontes et al., 1995; Rayment et al., 1993). The centrality of helical polymers to biology makes methods for structural studies of such polymers important.

It is thus not coincidental that the first application of electron microscopic three-dimensional reconstruction techniques was to a helical polymer (DeRosier and Klug, 1968). In the years since, there have been many refinements and advances, thus making it possible to use Fourier-Bessel methods to image at high resolution (better than 10 Å) specimens such as tobacco mosaic virus (Jeng et al., 1989), the acetylcholine receptor (Miyazawa et al., 1999), the sarcoplasmic reticulum calcium pump (Zhang et al., 1998) and bacterial flagella (Trachtenberg et al., 1998). To paraphrase Gertrude Stein's comment about Ezra Pound, these methods are excellent if the specimens are exactly helical, if not, not.

The application of Fourier-Bessel helical methods is not simple, particularly when the specimens are flexible and disordered. One initial approach to dealing with flexibility was to computationally straighten filaments (Egelman, 1986), but this has the potential for introducing artifacts due to the assumptions that need to be made (e.g., that there is no coupling between bending and twisting, that the filament undergoes a purely elastic normal mode of bending, etc.). We also know that these assumptions, to the extent that they can even apply at low resolution, must break down as the resolution improves. This approach of unbending, however, is now widely used. Another difficulty with helical analysis is that the indexing of a pattern can be ambiguous, and the wrong symmetry can be chosen (Egelman and Stasiak, 1988). Further complications exist when the filament does not have a precisely defined helical symmetry, such as F-actin that has a random, variable twist (Egelman et al., 1982). The magnitude of the problem can be seen (above, left) for *E. coli* RecA-DNA filaments. The RecA protein has been the most intensively studied enzyme in homologous genetic recombination, but structural studies of these filaments have been very difficult. The histogram above shows the pitch distribution found when 17,903 short segments of RecA-DNA filaments are analyzed. It thus becomes obvious why techniques which rely upon imposing an average pitch and twist will fail.

A clear alternative to such methods is to use single-particle techniques (Penczek et al., 1992). A number of such applications to helical filaments have been reported. A real-space approach to dealing with the variable pitch of sickle-cell hemoglobin helical fibers has been used (Bluemke et al., 1988), and other groups have also employed single-particle methods to reconstruct helical filaments (Sosa et al., 1997; Burgess et al., 2000; Jimenez et al., 1999). Real-space methods offer many advantages over helical approaches, particularly in surmounting the problem of disorder and flexibility. For highly ordered specimens, such as tobacco mosaic virus (Jeng et al., 1989) or helical tubes of membrane proteins (Zhang et al., 1998; Miyazawa et al., 1999), the advantages may be less obvious. What we show is that it is possible to significantly extend the simple application of single-particle techniques by a specialized approach based heavily on the local helical geometry of the objects being studied.
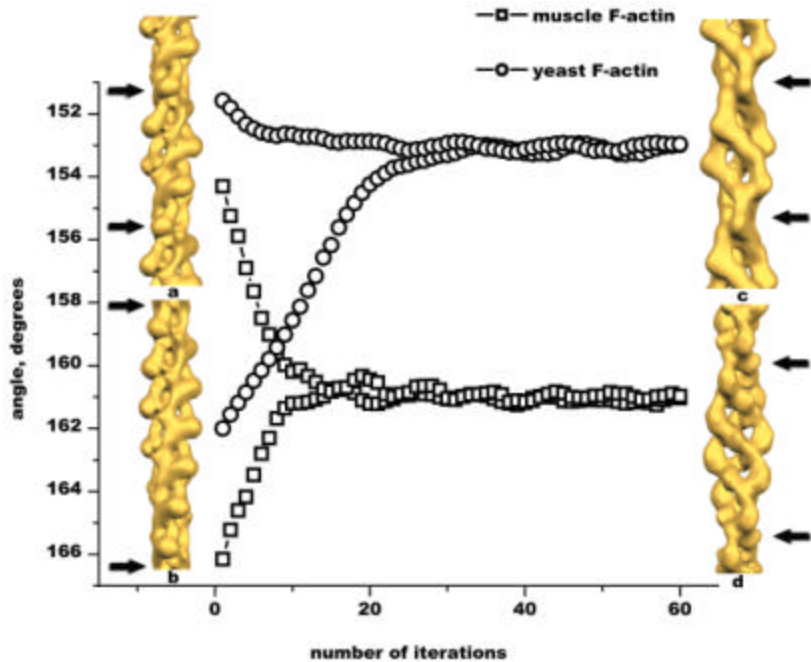

Pitch distribution for 17,903 segments of RecA-DNA filaments

We have developed a novel method to do this (Egelman, 2000). We call the method Iterative Helical Real Space Reconstruction (IHRSR), and it differs significantly from all previous approaches in that it is based upon a refinement and imposition of the helical geometry each cycle. We have found that this is key to the success of the procedure. The method has been extensively tested in our lab: (Yu et al., 2001; Yang et al., 2001a; Yang et al., 2001b; Lukoyanova et al., 2002; Orlova et al., 2001; Galkin et al., 2001). We show an example (above) of the application of this procedure to an unusual state of F-actin, found when ADF (Actin Depolymerizing Factor) or cofilin are bound to actin. Shown (left) are two starting reference volumes (a,b), which are low-resolution versions of the Holmes model for F-actin (Holmes et al., 1990), and resulting reconstructions for yeast actin (c) and muscle actin (d). The reference volume (a) has been generated with a symmetry of 152° per subunit, while the reference volume (b) has a symmetry of 166° per subunit. This difference in twist leads to a large difference in the distance between crossovers of the right-handed long pitch helical strands in F-actin, indicated by the arrows (a,b). The 503 segments of muscle actin, selected as being undecorated within filaments extensively decorated with hADF, were reconstructed with IHRSR, starting with the references having symmetries of either 166° or 154°. The open squares show the convergence of the symmetry to the value of 161°, and both paths resulted in the reconstruction shown in (d), even though the starting models have very different symmetries and conformation from the result found. For the yeast actin, a similar procedure is shown (open circles), where 1,240 segments have started IHRSR using the initial models with symmetries of 152° and 162°, but converge to a symmetry of 153° after 60 cycles, with the resulting reconstruction shown in (c).

Cycles of IHRSR for unusual state of F-actin

Polymorphism in Ring Helicases

We have previously shown that the *E. coli* DnaB helicase can exist in two states: a symmetrical hexamer, having six equivalent subunits, and a hexameric ring that is a trimer of asymmetric dimers, with two different subunit conformations (Yu et al., 1996). This work was done in two-dimensions using averages of projections. The averages above are reference-free (Penczek et al., 1992), with no symmetry imposed, and show how images can be sorted into the two classes. We have now extended this work to three-dimensions, motivated in large part by the fact that crystal structures exist for the the DnaB N-terminal domain (Fass et al., 1999) and for the homologous helicase domain from another protein (Singleton et al., 2000). The extension to three-dimensions in the presence of this polymorphism is not trivial, and we have been very surprised to learn how artifactual solutions can be obtained that appear to be completely consistent with the data.
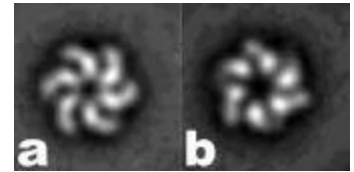
2D averages for 6-fold (a) and 3-fold (b) DnaB
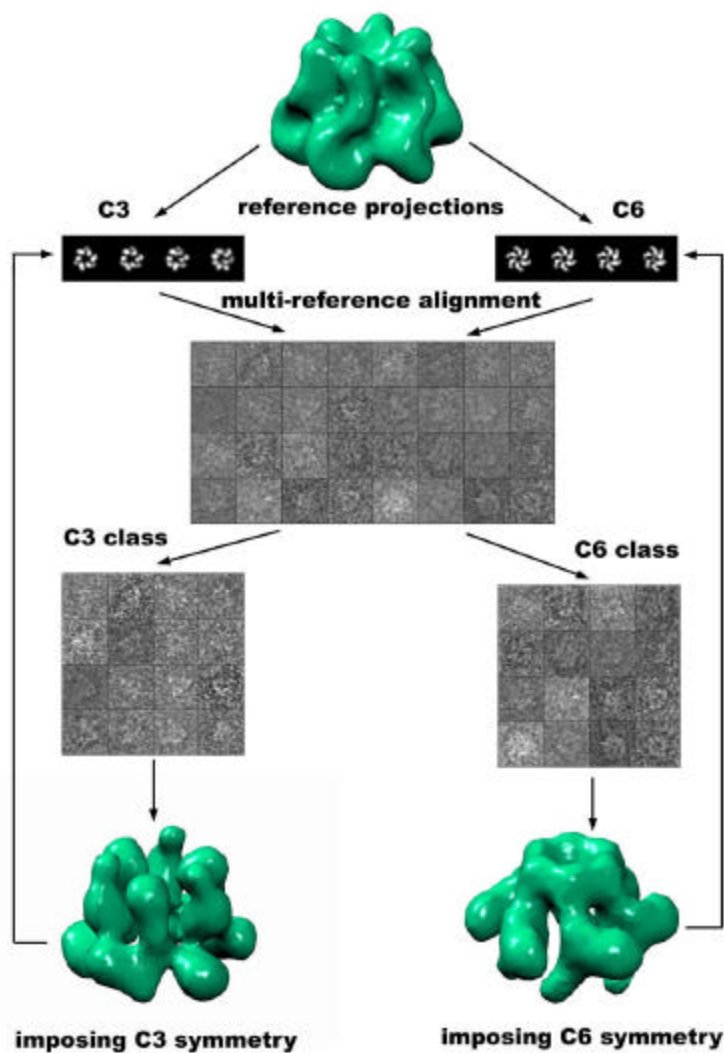
A frequent assumption in the field of electron microscopic image analysis is that if the projections of a three-dimensional reconstruction match all classes of characteristic averages generated from the raw images, the reconstruction is correct except for the possibilities of enantiomorphic ambiguity. We show that in the presence of structural heterogeneity this assumption can be completely wrong. Specifically, we used images of DnaB to generate two different reconstructions and in the process iteratively sort the images into two classes (schematic above) . The procedure was started by using a reconstruction of the hexameric T7 gp4 protein (VanLoock et al., 2001) (top of figure). The class averages of images, generated by reference-free approaches, were in excellent agreement with the projections of the final reconstructions (bottom of figure). However, the reconstructions were completely artifactual due to the fact that the sorting of images was wrong. That is, the images were never properly sorted by the correct symmetry (3-fold or 6-fold) and angular orientation (Euler angles). This was only learned when trying to interpret the reconstructions in terms of the existing crystal structures and the mechanism for conversion between the two forms. Using these crystal structures and model-building, very different three-dimensional reconstructions were eventually generated that appear to be physically plausible, can be used to sort the images, and match the image data.

Thus, one needs to determine whether structural heterogeneity is present when generating three-dimensional reconstruction, since heterogeneity can lead to highly artifactual solutions.

Scheme for sorting DnaB by symmetry

Reference List

1. Bluemke,D.A., Carragher,B., and Josephs,R. (1988). The reconstruction of helical particles with variable pitch. Ultramicroscopy *26*, 255-270.

2. Burgess, S. A., Knight, P., Walker, M., Schmitz, S., Sparrow, J. C., and Trinick, J. Real-space 3D reconstruction of frozen-hydrated arthrin and actin filaments at 2 nm resolution. Biophysical Journal 78[1], 8a. 2000. Ref Type: Abstract

3. DeRosier,D.J. and Klug,A. (1968). Reconstruction of three-dimensional structures from electron micrographs. Nature *217*, 130-134.

4. Egelman,E.H. (1986). An algorithm for straightening images of curved filamentous structures. Ultramicroscopy *19*, 367-373.

5. Egelman,E.H. (2000). A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy *85*, 225-234.

6. Egelman,E.H., Francis,N., and DeRosier,D.J. (1982). F-actin is a helix with a random variable twist. Nature *298*, 131-135.

7. Egelman,E.H. and Stasiak,A. (1988). Structure of helical RecA-DNA complexes. II. Local conformational changes visualized in bundles of RecA-ATP- -S filaments. J. Mol. Biol. *200*, 329-349.

8. Fass,D., Bogden,C.E., and Berger,J.M. (1999). Crystal structure of the N-terminal domain of the DnaB hexameric helicase. Structure. Fold. Des *7*, 691-698.

9. Galkin,V.E., Orlova,A., Lukoyanova,N., Wriggers,W., and Egelman,E.H. (2001). Actin Depolymerizing Factor Stabilizes an Existing State of F-Actin and Can Change the Tilt of F-Actin Subunits. J. Cell Biol. *153*, 75-86.

10. Holmes,K.C., Popp,D., Gebhard,W., and Kabsch,W. (1990). Atomic model of the actin filament. Nature *347*, 44-49.

11. Jeng,T.W., Crowther,R.A., Stubbs,G., and Chiu,W. (1989). Visualization of alpha-helices in tobacco mosaic virus by cryo-electron microscopy. J. Mol. Biol. *205*, 251-257.

12. Jimenez,J.L., Guijarro,J.I., Orlova,E., Zurdo,J., Dobson,C.M., Sunde,M., and Saibil,H.R. (1999). Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. EMBO J. *18*, 815-821.

13. Jontes,J.D., Wilson-Kubalek,E.M., and Milligan,R.A. (1995). A 32 degree tail swing in brush border myosin I on ADP release. Nature *378*, 751-753.

14. Lukoyanova,N., VanLoock,M.S., Orlova,A., Galkin,V.E., Wang,K., and Egelman,E.H. (2002). Each actin subunit has three nebulin-binding sites: Implications for steric blocking. Cur. Biol. *in press*.

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

16. Orlova,A., Galkin,V.E., VanLoock,M.S., Kim,E., Shvetsov,A., Reisler,E., and Egelman,E.H. (2001). Probing the structure of f-actin: cross-links constrain atomic models and modify actin dynamics. J. Mol. Biol. *312*, 95-106.

17. Penczek,P.A., Radermacher,M., and Frank,J. (1992). Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy *40*, 33-53.

18. Rayment,I., Holden,H.M., Whittaker,M., Yohn,C.B., Lorenz,M., Holmes,K.C., and Milligan,R.A. (1993). Structure of the Actin-Myosin Complex and Its Implications for Muscle Contraction. Science *261*, 58-65.

19. Singleton,M.R., Sawaya,M.R., Ellenberger,T., and Wigley,D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589-600.

20. Sosa,H., Hoenger,A., and Milligan,R.A. (1997). Three different approaches for calculating the three-dimensional structure of microtubules decorated with kinesin motor domains. J. Struct. Biol. *118*, 149-158.

21. Trachtenberg,S., DeRosier,D.J., Zemlin,F., and Beckmann,E. (1998). Non-helical perturbations of the flagellar filament: Salmonella typhimurium SJW117 at 9.6 A resolution. J. Mol. Biol. *276*, 759-773.

22. VanLoock,M.S., Chen,Y.J., Yu,X., Patel,S.S., and Egelman,E.H. (2001). The primase active site is on the outside of the hexameric bacteriophage t7 gene 4 helicase-primase ring. J. Mol. Biol. *311*, 951-956.

23. Yang,S., VanLoock,M.S., Yu,X., and Egelman,E.H. (2001a). Comparison of Bacteriophage T4 UvsX and Human Rad51 Filaments Suggests that RecA-like Polymers May Have Evolved Independently. J. Mol. Biol. *312*, 999-1009.

24. Yang,S., Yu,X., Seitz,E.M., Kowalczykowski,S.C., and Egelman,E.H. (2001b). Archaeal RadA Protein Binds DNA as Both Helical Filaments and Octameric Rings. J. Mol. Biol. *314*, 1077-1085.

25. Yu,X., Jacobs,S.A., West,S.C., Ogawa,T., and Egelman,E.H. (2001). Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. Proc. Natl. Acad. Sci. U. S. A *98*, 8419-8424.

26. Yu,X., Jezewska,M.J., Bujalowski,W., and Egelman,E.H. (1996). The hexameric *E. coli* DnaB helicase can exist in different quaternary hexameric states. J. Mol. Biol. *259*, 7-14.

27. Zhang,P., Toyoshima,C., Yonekura,K., Green,N.M., and Stokes,D.L. (1998). Structure of the calcium pump from sarcoplasmic reticulum at 8-A resolution. Nature *392*, 835-839.

ATOMIC MODEL OF THE CELL: DOCKING IN A TOMOGRAPHIC ENVIRONMENT

Niels Volkmann**,** The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, **niels@burnham.org,**
Phone: (858) 646-3187, Fax: (858) 646-3195

Introduction

In three-dimensional (3D) structure determination of macromolecules, X-ray crystallography covers the full range from small molecules to very large assemblies such as viruses with molecular masses of megadaltons. The limiting factors are expression, crystallization and the stability and homogeneity of the structure. In the case of NMR, atomic structures can be determined from molecules in solution, but the size limit is of the order of 100 kDa. Electron microscopy (EM) is able to tackle large isolated assemblies (Chiu et al., 1999; Frank, 2001; Kühlbrandt and Williams, 1999; Saibil, 2000; Stahlberg et al., 2001; van Heel et al., 2000) as well as cellular and subcellular structures (Baumeister et al., 1999; McEwen and Frank, 2001; McIntosh, 2001; Perkins et al., 1997). Cryo-methods enable the observation of molecules close to physiological conditions in their native aqueous environment (Dubochet et al., 1988). Thus, EM can potentially bridge the gap between detailed structural knowledge of isolated macromolecules and their cellular context.

For most biological specimen it has not yet been possible to determine their structure beyond 6-30Å resolution by using EM, thus precluding atomic modeling directly from the data. Recently, algorithms were developed that allow objective and automatic combination of atomic resolution structures with lower resolution EM reconstructions (Jiang et al., 2001; Roseman, 2000; Rossmann, 2000; Volkmann et al., 2000; Wriggers and Chacon, 2001). These algorithms allow building atomic models of large complexes imaged by EM using the atomic structures of their components. In this paper we explore the potential usefulness of such docking studies in the context of cellular electron tomograms. We present calculations on two test systems, actomyosin and the ribosome, that address the influence of the peculiarities of electron tomography on the docking process.

Electron Tomography

An electron micrograph consists of two-dimensional projections of a 3D object. To retrieve the structure of the underlying 3D object, sufficiently sampled angular views need to be aligned and combined. An object might possess crystalline, helical, icosahedral, or rotational symmetry; or no symmetry at all. The presence of symmetry or the availability of similar views of the object means that redundant motifs are provided in the specimen, thereby giving the opportunity to enhance the signal-to-noise ratio (SNR) of the image by averaging. This use of multiple copies at a sufficient number of different angular views and their proper alignment is the key to obtaining high-resolution reconstructions with low SNR.

The most general method for obtaining 3D information by EM is tomography. In this technique, a series of images is taken of a single object as the object is tilted over a wide range of angles. Tomography is the only method available for reconstruction of specimen with unique structure (no multiple copies). An entire cell for example would fall into this category, as it would be impossible to find two cells that are exactly identical, which prevents averaging to enhance the SNR. Today, with the use of computer-controlled microscopes and the availability of charge-coupled device (CCD) cameras, it has become possible to image large-scale structures at a resolution of about 50-100Å. Tomography is undergoing considerable growth at the present time due to the realization that molecular information can also be obtained from unstained, frozen-hydrated whole cells (Grimm et al., 1998) and isolated organelles (Nicastro et al., 2000).

The main disadvantage of the tomographic approach is that radiation damage builds up during the multiple exposures as the specimen is being tilted, thus limiting the possible electron dose that is available for the single exposures. As a consequence, the SNR of electron tomograms tends to be extremely low and the limited number of angular steps restricts the maximum achievable resolution. This limit is currently at about 50Å for frozen-hydrated samples, but it is anticipated that a resolution of 20-40Å is achievable in the future (Böhm et al., 2000). Specimen holders restrict the possible tilt angles of the specimen to about ±70° for cryogenic data collection, leading to a missing wedge of data. As a consequence, features perpendicular to the electron beam are better resolved than features parallel to the beam, resulting in anisotropic resolution and distortions of the structure. Here, we examine the influence of limited resolution, low SNR and missing wedges on docking of atomic models into calculated electron tomograms of actomyosin and the ribosome.

Tomograhic Docking

It has been demonstrated, that density correlation can be used efficiently to distinguish between macromolecules such as GroEL and the 20S proteasome in electron tomograms of resolution as low as 80Å (Böhm et al., 2000). A refinement procedure based on a density correlation-like measure was used successfully to improve the fit of pre-built actomyosin models with averaged 3D motifs from insect flight muscle electron tomograms at a resolution of about 70Å (Chen et al., 2001). These studies indicate that density correlation is a good choice for docking studies in a tomographic environment. Here, we use the program system COAN (Volkmann and Hanein, 1999) to evaluate docking of atomic models into calculated electron tomograms of actomyosin and ribosomes. COAN contains a module for very fast, accurate density correlation-based docking. For example, a fully automatic fit of myosin into its corresponding density at 20Å resolution, using a global search protocol, completes in less than 3 minutes on an SGI R12000 processor as well as on an Athlon XP1900 processor running Linux. The accuracy of the resulting fit is better than 0.5Å.

In order to run the test calculations, densities at resolution between 25-100Å were calculated from the respective atomic models of the 70S ribosome of *thermus thermophilus* (Yusupov et al., 2001) and of smooth muscle actomyosin in the rigor state (Volkmann et al., 2000). Random noise was added to reflect SNR levels expected for tomographic reconstructions. Missing wedges were generated by deleting a corresponding region in the 3D Fourier transform of the respective density map (Figure 1).

In case of the ribosome, we employed two tests. First, the accuracy of docking the atomic coordinates of the 50S subunit into the complete density of the 70S ribosome was evaluated. As a second test, the 70S density was segmented into the 50S and 30S components using a 3D watershed-based segmentation procedure (Volkmann, in press), then the 50S atomic model was fitted into the 50S segment of the density. All segmentations were done at 25Å. A similar accuracy of segmentation at lower resolution can be achieved by incorporating the high-resolution shape information of the fragment into the segmentation process (Volkmann, unpublished). Before docking or segmentation was used on densities corrupted with noise, the map was convoluted with a 3D Gaussian kernel in order to reduce the noise level (see Figure 2).

For the actomyosin densities, a single myosin molecule was segmented from the calculated actomyosin map and the myosin atomic model was fitted into the corresponding segment. Again, before docking or segmentation of densities corrupted with noise, the maps were convoluted with a 3D Gaussian kernel (see Figure 3).

Results

The main parameter influencing modular docking of the 50S atomic model into the complete 70S density appears to be the resolution of the tomogram. In the noise-free case the accuracy of docking at 25Å is indicated by a 2.41Å root-mean-square deviation (rms) from the correct solution. At 40Å resolution the accuracy is almost the same (2.64Å) and at 50Å the rms is
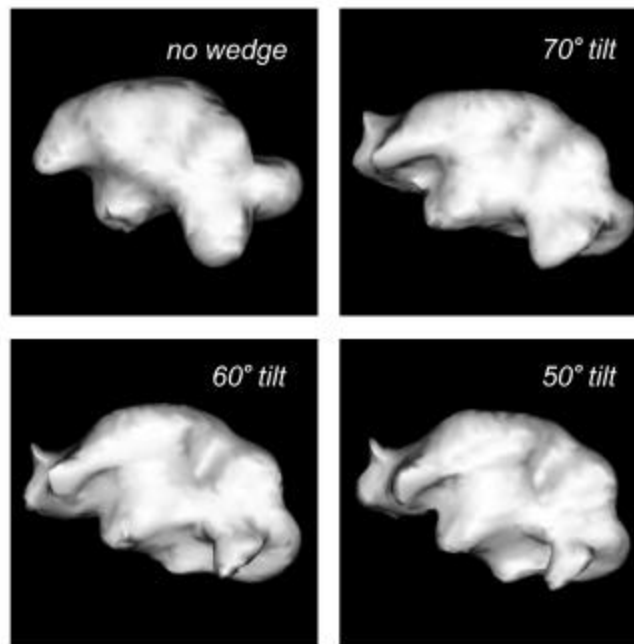

Figure 1: Effect of missing wedges on the density of a myosin molecule segmented from a calculated actomyosin map at 25Å resolution.
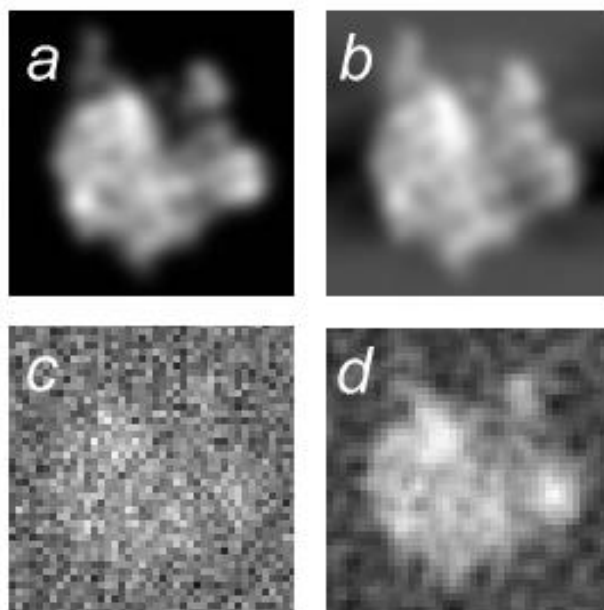
Figure 2: A slice of the calculated 70S ribosome density at 25Å resolution. a. Error-free map. b. Map with a 40º wedge, emulating a ±70º tilt series. c. Map with noise added (0.2 SNR). d. Noisy map after convolution with a Gaussian kernel.

5.19Å. At resolution lower than 50Å, the docking appears to break down completely and the best few solutions have rms values of greater than 90Å. There is little effect of noise (after Gaussian kernel convolution), the docking accuracy degrades only little (<1Å) and the break-down point remains at about 50Å even if the SNR is as low as 0.1. Also, a moderate wedge of

40º (±70º tilt series) has little influence on the docking accuracy (<1Å). However, larger wedges do degrade the docking accuracy considerably and shift the break-down point towards higher resolution.

Figure 3: A slice of the calculated actomyosin map. The two actin strands are visible with myosin molecules attached. a. Error-free map. b. Map with a 40º wedge, emulating a ±70º tilt series. c. Map with noise added (0.1 SNR). d. Noisy map after convolution with a Gaussian kernel.

Figure 4: Docking accuracy (rms in Å) as a function of resolution for segmented modules of the 50S ribosomal subunit (a) and actin-bound myosin in the rigor state (b). The magenta lines correspond to error-free densities calcula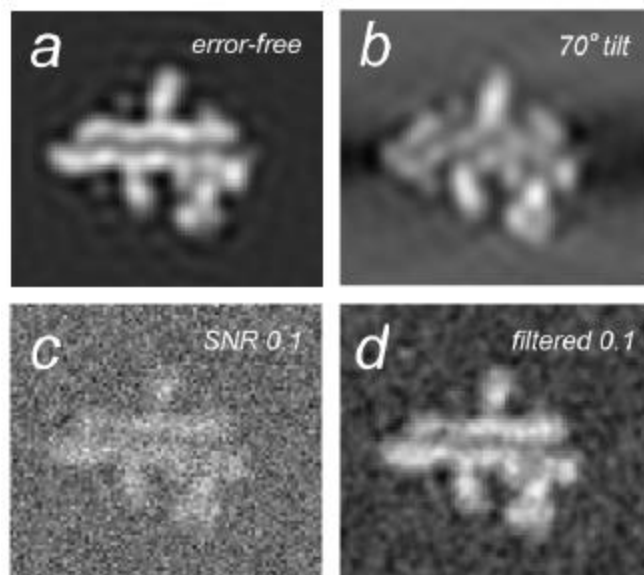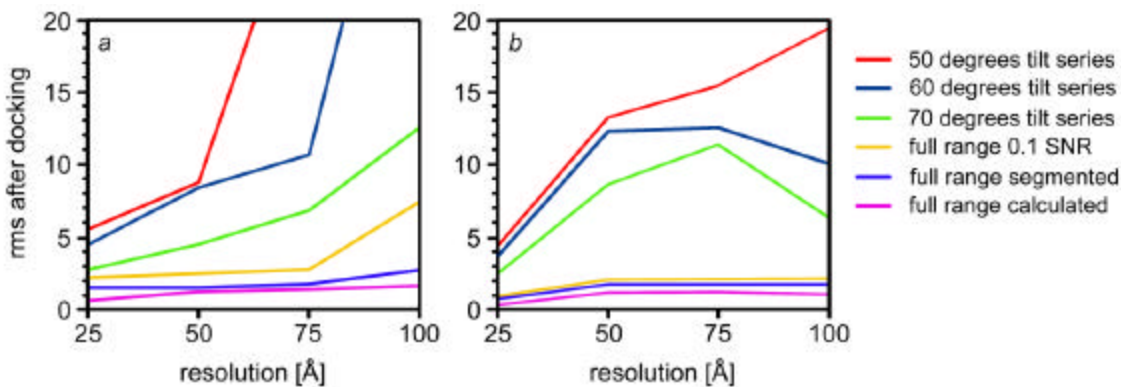ted from the module's atomic structure for reference. The blue lines correspond to segment maps, where the maps were segmented from error-free maps of the whole complex. The orange lines correspond to segment maps, where the maps where segmented from a map with 0.1 SNR and after convolution with a Gaussian kernel. The green, dark-blue and red lines correspond to segment maps that were segmented at 25Å resolution from maps of the whole complex with different missing wedges.

The docking of the 50S atomic model into the 50S segments of the ribosome maps is summarized in Figure 4a, the docking of the myosin atomic model into the myosin segments of the actomyosin maps is summarized in Figure 4b. Docking accuracy of the models into plain calculated maps (magenta lines) is shown for reference and is better than 1Å rms at 25Å resolution and better than 2.5Å even at 100Å resolution. The penalty for using the segmentation is very small (blue lines; <1.5Å) and the addition of noise (0.1 SNR and Gaussian kernel convolution; orange) degrades performance very little for actomyosin (<1Å) and not too much for the ribosome (<1Å down to 75Å, <5Å at 100Å). Missing data has the worst effect on docking performance on the segmented data. However, a moderate wedge of 40º (±70º tilt series; green) results in rms values better than 13Å over all resolution ranges with rms values of around 3Å at 25Å resolution. At this resolution larger wedges give still a reasonable accuracy for both structures (<6Å).

Conclusions

Our calculations indicate that even with the current limitations in resolution (50-100Å) and quality of electron tomograms a meaningful docking that gives near atomic or residue-level information may well be achievable, especially if the docking is combined with advanced segmentation algorithms. While the SNR and the resolution seem to be of no major concern, the artifacts introduced by the missing wedge problem appear to be a more serious limitation. This problem may in part be solved by more complete data collection strategies such as double-tilt tomography (Penczek et al., 1995). The docking could also be improved by incorporating a wedge model directly into the docking process. Currently, COAN is automatically correcting for spherical errors in reciprocal space such as those introduced by the microscope's contrast transfer function (Volkmann et al., 2000). This concept should be readily extendible to handle non-spherical error models. If the anticipated resolution of 20-40Å can be achieved, reliable modular docking without the use of segmentation should be possible. Thus, the tools to build an atomic model of the cell from cellular tomograms and their molecular components appear to be within reach.

Acknowledgements

References

1. Baumeister, W., Grimm, R., and Walz, J. (1999). Electron tomography of molecules and cells. Trends Cell Biol *9*, 81-85.
2. Böhm, J., Frangakis, A. S., Hegerl, R., Nickell, S., Typke, D., and Baumeister, W. (2000). From the cover: toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. Proc Natl Acad Sci U S A *97*, 14245-14250.

3. Chen, L. F., Blanc, E., Chapman, M. S., and Taylor, K. A. (2001). Real space refinement of acto-myosin structures from sectioned muscle. J Struct Biol *133*, 221-232.
4. Chiu, W., McGough, A., Sherman, M. B., and Schmid, M. F. (1999). High-resolution electron cryomicroscopy of macromolecular assemblies. Trends Cell Biol *9*, 154-159.
5. Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A. W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Quart Rev Biophys *21*, 129- 228.
6. Frank, J. (2001). Cryo-electron microscopy as an investigative tool: the ribosome as an example. Bioessays *23*, 725-732.
7. Grimm, R., Singh, H., Rachel, R., Typke, D., Zillig, W., and Baumeister, W. (1998). Electron tomography of ice-embedded prokaryotic cells. Biophys J *74*, 1031-1042.
8. Jiang, W., Baker, M. L., Ludtke, S. J., and Chiu, W. (2001). Bridging the information gap: computational tools for intermediate resolution structure interpretation. J Mol Biol *308*, 1033-1044.
9. Kühlbrandt, W., and Williams, K. A. (1999). Analysis of macromolecular structure and dynamics by electron cryo-microscopy. Curr Opin Chem Biol *3*, 537-543.
10. McEwen, B. F., and Frank, J. (2001). Electron tomographic and other approaches for imaging molecular machines. Curr Opin Neurobiol *11*, 594-600.
11. McIntosh, J. R. (2001). Electron microscopy of cells: a new beginning for a new century. J Cell Biol *153*, F25-32.
12. Nicastro, D., Frangakis, A. S., Typke, D., and Baumeister, W. (2000). Cryo-electron tomography of neurospora mitochondria. J Struct Biol *129*, 48-56.
13. Penczek, P., Marko, M., Buttle, K., and Frank, J. (1995). Double-tilt electron tomography. Ultramicroscopy *60*, 393-410.
14. Perkins, G. A., Renken, C. W., Song, J. Y., Frey, T. G., Young, S. J., Lamont, S., Martone, M. E., Lindsey, S., and Ellisman, M. H. (1997). Electron tomography of large, multicomponent biological structures. J Struct Biol *120*, 219-227.
15. Roseman, A. M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. Acta Crystallogr D Biol Crystallogr *56*, 1332-1340.
16. Rossmann, M. G. (2000). Fitting atomic models into electron-microscopy maps. Acta Crystallogr D Biol Crystallogr *56*, 1341-1349.
17. Saibil, H. R. (2000). Conformational changes studied by cryo-electron microscopy. Nat Struct Biol *7*, 711-714.
18. Stahlberg, H., Fotiadis, D., Scheuring, S., Remigy, H., Braun, T., Mitsuoka, K., Fujiyoshi, Y., and Engel, A. (2001). Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. FEBS Lett *504*, 166-172.
19. van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys *33*, 307-369.
20. Volkmann, N. (in press). A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. J Struct Biol.
21. Volkmann, N., and Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. J Struct Biol *125*, 176-184.
22. Volkmann, N., Hanein, D., Ouyang, G., Trybus, K. M., DeRosier, D. J., and Lowey, S. (2000). Evidence for cleft closure in actomyosin upon ADP release. Nat Struct Biol *7*, 1147-1155.
23. Wriggers, W., and Chacon, P. (2001). Modeling tricks and fitting techniques for multiresolution structures. Structure (Camb) *9*, 779-788.
24. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 A resolution. Science *292*, 883-896.

RECONCILING SHAPE WITH STRUCTURE: MORPHOMETRIC STRATEGIES FOR MULTI-RESOLUTION FLEXING

Willy Wriggers, Stefan Birmanns*, Florence Tama, Julio Kovacs, and Pablo Chacón; *Central Institute for Applied Mathematics, Research Centre Jülich, Germany; Department of Molecular Biology, Mail Code TPC-6, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; URL: http://www.scripps.edu/wriggers

Abstract
In the past decade significant progress was made in the development of 'rigid-body' docking techniques for assembling molecular fragments into large sub-cellular complexes. Nevertheless, it appears that the components of sub-cellular machines undergo a wide range of motions, and shapes observed at low resolution often deviate from the known atomic structures. For example, RNA polymerase structures exhibit considerable flexing motions that cannot be explained by rigidly docked sub-units alone. What is required for the modeling of such motions is a combination of global shape constraints based on the low-resolution data with a local modeling of atomic interactions. Ideally, such a multi-scale modeling approach should involve both a reduced number of refinement parameters to avoid over-fitting, as well as an accurate sampling of atomic interactions to maintain the molecular stereochemistry. For the modeling we build deformable models of biological structures at reduced complexity that retain much of the mechanical properties of biological structures. The elastic models yield a decomposition of the predicted motion into vibrational normal modes and are amenable to interactive manipulation with virtual reality and molecular graphics software.

Introduction
The study of many diseases has led to the realization that future advances in modern biology and medicine will depend on an understanding of fundamental cellular processes, most of which involve the actions and interactions of large biomolecular assemblies [1]. Processes in the cell that may constitute targets for drug design and delivery involve dynamic interactions among ordered biomolecules in a highly organized way. Large assemblies of proteins and polynucleotides with specific active sites act both as templates and catalysts in the functional interaction with ligand molecules. Examples currently studied in our group are the actomyosin complex in muscle [2], motor proteins in concert with cytoskeletal filaments [3], the ribosome complex [4], virus capsids [5,6], and the RNA polymerase complex [7]. Such systems have been called 'macromolecular machines' [1] to characterize the complexity and precision of their highly coordinated moving parts. Ideally, such large structures are visualized at an atomic level of detail, but low- to medium-resolution structural data in the 5-30Å resolution range has become increasingly ubiquitous. Within the past decade, several powerful biophysical techniques such as EM [8], X-ray or electron crystallography [9], and small-angle X-ray scattering [10] have emerged that routinely produce 3D structural information at a reduced spatial resolution. These methods are capable of yielding low-resolution density maps under a wide range of biochemical conditions that allow atomic structures of components to be fitted and docked [2,11-13]. To advance our understanding of cellular function it is essential that we are able to characterize the structure and functionally relevant motions of these systems in a computationally efficient manner. As described in the following, we develop advanced algorithms for the modeling and simulation of large biomolecular systems at various resolution scales.

Docking of Fragments: State of the Art
Perhaps the oldest fitting criterion is the (globally normalized) cross-correlation coefficient C. If deformations are to be modeled, it is possible to approximate conformational changes by a sequential fitting of individual domains that are deemed rigid. The method has been adopted by a large number of authors [14-20] and a number of computer programs are readily

available [15,17,20,21]. One maximizes C with respect to the translational and rotational degrees of freedom (DOF) and thereby minimizes the least-squares density discrepancy of two structural data sets.

One important aspect of the globally normalized correlation is that its calculation can be significantly sped up by the use of fast Fourier transform (FFT) techniques [22-25]. The calculation of C in direct space is very expensive, since it requires M^2 multiplications for each translation, where M is the total number of voxels. However, C can be expressed in Fourier space as the product of the structure factors. The reverse FFT of the structure factors is significantly faster than the direct space calculation, because we need to perform only two transformations (each of them scale as M log M). We are developing separate FFT-accelerated fast matching methods for the translational and for the rotational search space. The latter problem is solved mathematically by combining FFT with spherical harmonics (see our poster by Kovacs *et al.*).

Unlike the rms deviation that is commonly used for comparing atomic structures, the correlation measure is not very sensitive to positional or orientational changes [13] and exhibits relatively broad distributions of possible solutions sets [15]. We proposed to alter the functional form of the compared densities by applying filters that enhances the contours in the data sets [8,26]. For example, a Laplacian operator, well known in image processing [27], assigns positive densities to the surface contours and negative densities to the interior volumes. This approach effectively maximizes both surface and volumetric overlap when C is maximized [26]. Tests using synthetic low-resolution density models indicated that the difference in score between the correct fit and spurious fits increases by 35-50% relative to correlation-based docking. At present we evaluate a variety of direct-space, Fourier-space and Wavelet-Transform filters in terms of their docking performance (see our poster by Chacón *et al.*).

Flexing Requires a Reduced Representation of Biomolecular Structure
In the presence of induced fit, rigid-body docking alone gives a sub-optimal alignment of the known structure of a fragment with the low-resolution data. For example, the structure of the ribosomal protein EF-G exhibits a striking conformational change on the 70S ribosome involving motions of three protruding domains [4,28]. The flexible modeling of such motions requires a reduced number of refinement parameters to avoid over-fitting (there are many more degrees of freedom in an atomic structure than there are independent pieces of information in an EM map [8]). Here, we briefly describe our medium-resolution ('mesoscopic') tessellation of a structure at reduced complexity:

We implemented artificial neural networks for the representation of structures at a reduced level of detail. Topology-Representing Networks (TRN) [29-31] are typically applied to solve important information processing and clustering tasks in artificial intelligence, such as robot control [30,32]. A consequence of the network formation is a discrete representation of a probability density function $P(\vec{r})$ of input signals $\vec{r}$, using a finite number of so-called 'landmark points' $\vec{w}_i$, $i = 1, \ldots, N$. Figure 1b,c presents a TRN tessellation of the 3D density of a virus capsid. The TRN algorithm distributes the landmarks ('neurons') over a 3D biological data set according to the probability density $P(\vec{r})$, yielding a reduced representation of the data (Fig. 1c). Proximity relationships ('lateral neural connections') can be learnt from the distribution of landmarks: The 'competitive Hebb rule' [29,31] constructs connections between adjacent neurons, if the connection is, at least partially, covered by the density distribution $P(\vec{r})$. The algorithm is based on the mathematical theory of Delaunay triangulation [33]. The competitive Hebb rule creates a 'skeleton' of connected features that are useful for flexible fitting and for creating mechanical models (see below). It can be shown [34] that the landmarks $\vec{w}_i$ correspond to the centers of mass of the $P(\vec{r})$-weighted neighborhood regions or 'Voronoi cells' (Fig. 1a) [35].
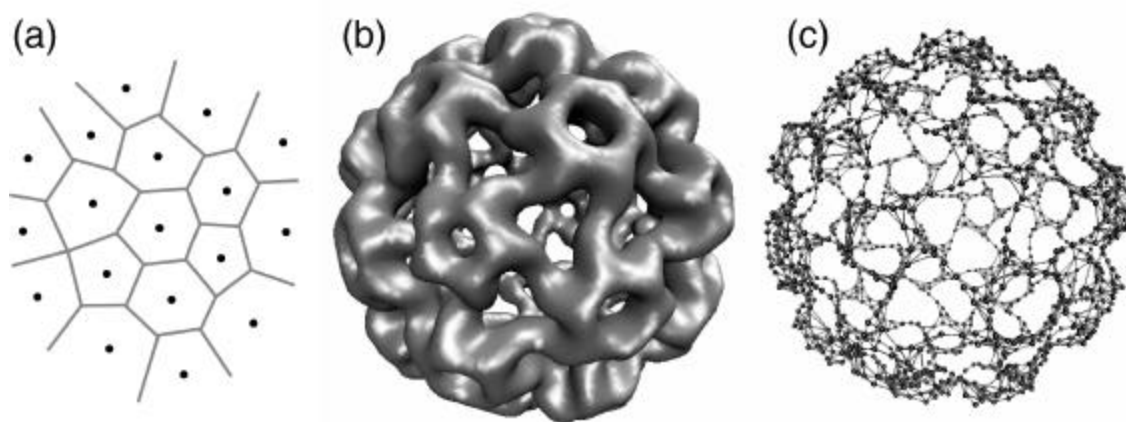
Fig. 1: Functionality of Voronoi tessellation, TRN, and competitive Hebb rule. (a) This idealized 2D system illustrates the basic concepts of TRNs. Tessellation of 2D space into 18 'neighborhood' subregions or Voronoi cells, $V_i = \left\{ \vec{r} \in {}^{\circ\,3} \middle| \left\| \vec{r} - \vec{w}_i \right\| \leq \left\| \vec{r} - \vec{w}_j \right\| \; \forall j = 1, \dots, N \right\} \; i = 1, \dots, N$. (b) The swollen form of cowpea chlorotic mottle virus at 23Å resolution [36]. (c) Tessellation of the density with the TRN method using 1380 landmarks (shown as beads). The landmark connectivities were constructed using the competitive Hebb rule (see text).

Flexible Docking with TRNs

Rigid-body docking with TRNs [10], as implemented in our Situs docking package [13,29,37], laid the groundwork for the development of a flexible docking technique that brings deviating features of multi-resolution structures into register [4,38] if the atomic structure in one conformation is known. In such situations, the deviating atomic structure is moved towards the EM density by forcing the corresponding landmarks $\vec{w}_i$ in both data sets to coincide. This is done in a molecular dynamics refinement of the atomic structure where a quantity equivalent to the landmark discrepancy forms a global penalty that is imposed by distance constraints (see [38] for details) while preserving the moved structure at the local level.

The major advantage of TRNs is computational speed. Both calculation of the TRN and docking by the reduced representation can be carried out within seconds of compute time. In contrast, the exhaustive search of all rigid-body degrees of freedom (see above) can take many hours even when FFT-accelerated. The accuracy that can be achieved in flexible and rigid-body docking with simulated (noise-free) data is one order of magnitude above the nominal resolution of the EM map, or better [38].

One of the open questions in flexible docking is how to maintain the stereochemical quality of a fitted structure [39], since any over-fitting to noise-induced vector displacements would compromise the quality of the atomic model. In a recent improvement to our flexible fitting algorithm, inter-vector distances along the connected polypeptide chain are constrained. The resulting vector skeletons (distance-constrained vectors) eliminate the longitudinal degrees of freedom that are deemed inessential for the flexible docking, while permitting lateral flexibility. Thereby, the skeleton-based fitting approach provides additional robustness against the effects of noise and experimental uncertainty (see Fig. 2 in [8]). Figure 2a,b demonstrates that the deformations due to flexible fitting may be considerable.

Physics-Based Deformable Models of Biological Structures

Traditional molecular modeling techniques such as molecular mechanics [40], molecular dynamics [41], or Monte Carlo [42] are able to describe structural fluctuations and transitions of biomolecules at the atomic level, but such calculations are very time-consuming and an adequate sampling of conformational space is feasible only for systems comprising a few thousand atoms [43,44]. However, it is possible to obtain useful information on the dynamics, long-range coupling, and elastic properties of polynucleotides without requiring atomic resolution [45-47]. These facts suggest the importance of developing mesoscopic models for large biomolecular assemblies to go beyond the size that can be handled at atomic detail.

Normal modes [48,49] are an interesting alternative to molecular dynamics as a tool to study the motion of large assemblies. Traditionally, normal modes analysis (NMA) is applied to an atomic structure where the atomic interactions are

described by a standard molecular dynamics force field. One of our main goals is to reduce the spatial resolution of elastic models of large biomolecular assemblies below the single residue level. This suggestion is rooted in the observation that atomic motions corresponding to low-frequency normal modes are not localized [50]. Hence, it seems reasonable to expect that a sparse estimation of the displacement field using a spatial resolution of 5-10Å will reproduce the displacements well. After generating such a sparse estimation, displacements can be extended to full space by spline-interpolation. Moreover, a mesoscopic description of the dynamics can be applied to both atomic structures and 3D image reconstructions from EM, as it is independent of the resolution of the underlying data. Consequently, we use the landmarks that can be computed with the TRN algorithm (Fig. 1) for a reduced representation of atomic structures or low-resolution data from EM.

As an example, we demonstrate how essential motions can be extracted from low-resolution EM maps in the absence of a known atomic structure. Figure 2 compares the lowest-frequency mode from mesoscopic NMA with the motions from the flexible fitting of *T. aquaticus* RNA polymerase (RNAP) to a low-resolution map of *E. coli* RNAP [7]. The differences between the crystal and the EM isoforms (Fig. 2b) can be attributed to crystal packing effects and reveal a closing of the RNAP jaws relative to the EM data. Interestingly, the lowest-frequency mode shows a very similar closing and opening motion of the jaws (Fig. 2c). The Figure exemplifies how one could use NMA to predict functionally relevant motions from a single low-resolution structure.
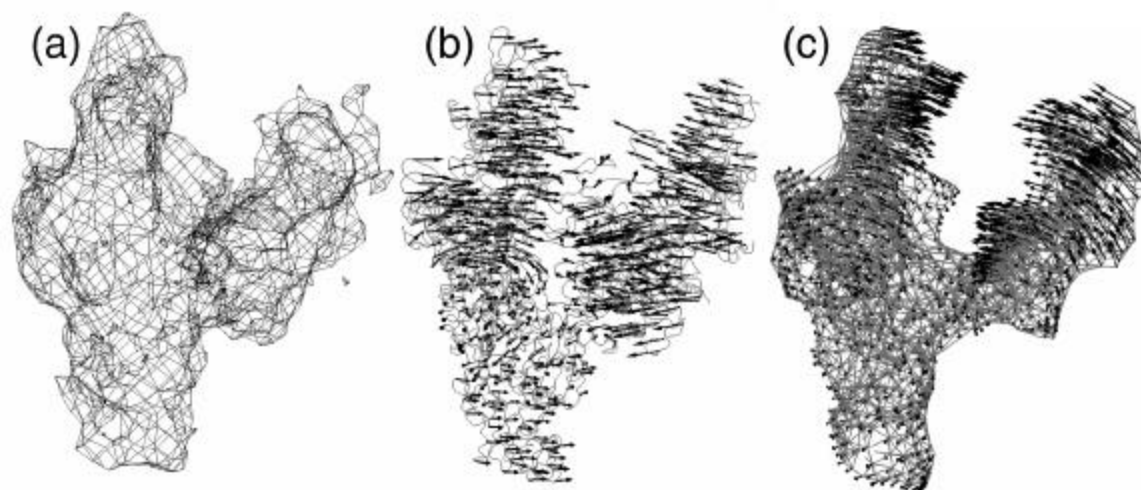

Fig. 2: Comparison of functionally relevant motions of RNA polymerase (RNAP) with mesoscopic NMA. (a) Low-resolution EM map of *E. coli* RNAP as described [7]. (b) Displacements from flexible fitting of the *T. aquaticus* crystal structure to the EM density [7]. The arrows point from the flexibly fitted structure (backbone trace) to the original crystal conformation (not shown). (c) Mesoscopic NMA (lowest-frequency mode) of the EM map in (a) using 1500 TRN landmarks.

We parametrize our deformable models based on the TRN tessellation (Fig. 1c) such that the NMA-derived motions optimally approximate the motions of structures sampled at atomic detail. The goal is to maintain continuity of parameters at all scales. We expect to be able to investigate not only large protein assemblies of known structure (such as actin filaments, muscle cross-bridges, and coiled coils), but also complexes for which only low-resolution data is available, such as RNA polymerase (Fig. 2) and a variety of virus coat proteins [6].

Molecular Graphics, Virtual Reality, and Force-Feedback

One of the challenges in computational structural biology is to enable the efficient use and inter-operation of a diverse set of techniques to simulate, analyze, model, and visualize the complex architecture and interactions of macromolecular systems. Our latest development effort will permit scientists to build models, perform docking of atomic and volumetric data, visualize

results from template convolution, and perform morphing and warping (flexible docking) interactively within a single computational environment. We are developing a 3D graphics extension for Situs, termed 'SenSitus', that is capable of supporting virtual reality (VR) devices such as stereo glasses, 3D trackers, and force-feedback (haptic) devices (Fig. 3). In addition to automated fitting, microscopists have a need to evaluate and manipulate docking models interactively 'by eye'. 3D capabilities and the 'physics of touch' offer tangible benefits for modelers who wish to explore a variety of docking situations in a VR environment (Fig. 3a).
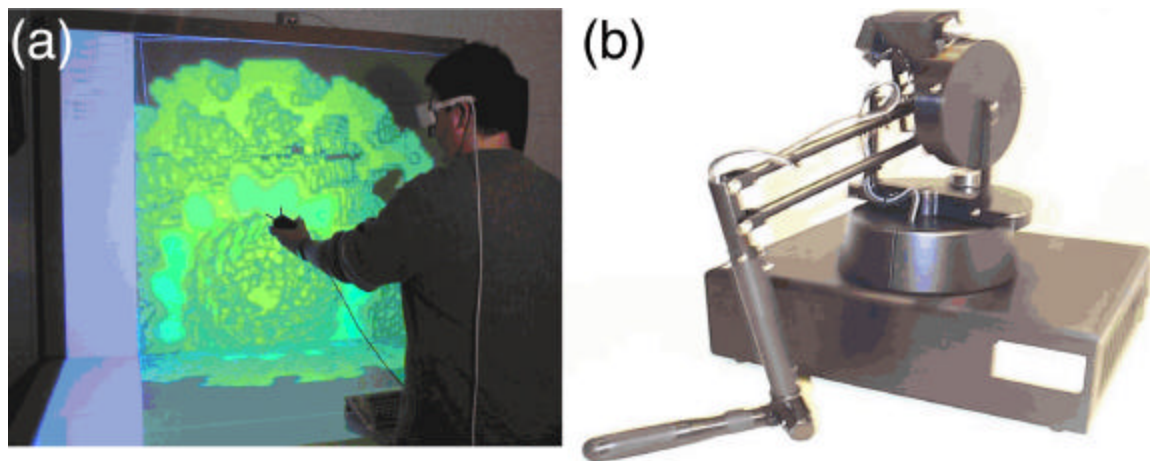

Fig. 3: (a) Visualization of a microtubule surface in an immersive VR environment using volslice3d [21], an earlier prototype of SenSitus. (b) The 6DOF Phantom force feedback device from SensAble Corp.

The force feedback device (Fig. 3b) not only enables the user to position the structure in the 3D space, but also directs the fitting to the next suitable location. The device measures a user's hand position and exerts a precisely controlled force on the hand. Our software supports this by calculating forces according to the correlation coefficient of density maps and crystallographic data. The high sampling frequency required for force feedback (refresh rate > 1kHz) is achieved by means of the TRN algorithm (Fig. 1) that reduces the complexity of the data representation to manageable levels. This technique is very useful, as it facilitates the detection of possible fitting locations and simplifies the fine positioning of the structure.

SenSitus and the interactive force feedback fitting will be demonstrated in our poster booth at the meeting (see our poster by Birmanns & Wriggers). The software is already fully functional as a docking tool. Ultimately, we will integrate all of our advanced flexing functionality and the efficient simulation of multi-resolution data by normal modes analysis (Fig. 2). The interactive flexing technology is currently in its infancy but it will undoubtedly gain in importance and popularity in the near future when more structures become available that require an induced fit of their components.

Bibliography

1. B. Alberts. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, 92:291-294, 1998.
2. I. Rayment, H.M. Holden, M. Whittaker, C.B. Yohn, M. Lorenz, K.C. Holmes, and R.A. Milligan. Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, 261:58-65, 1993.
3. R.D. Vale and R.A. Milligan. The way things move: Looking under the hood of molecular motor proteins. *Science*, 288:88-95, 2000.
4. W. Wriggers, R.K. Agrawal, D.L. Drew, J.A. McCammon, and J. Frank. Domain motions of EF-G bound to the 70S ribosome: Insights from a hand-shaking between multi-resolution structures. *Biophys. J.*, 79:1670-1678, 2000.

5. N.H. Olson, P.R. Kolatkar, M.A. Oliveira, R.H. Cheng, J.M. Greve, A. McClelland, T.S. Baker, and M.G. Rossmann. Structure of a human rhinovirus complexed with its receptor molecule. *Proc. Natl. Acad. Sci. USA*, 90:507-511, 1993.

6. J.J. Rux and R.M. Burnett. Spherical viruses. *Curr. Opinion Struct. Biol.*, 8:142-149, 1998.

7. S.A. Darst, N. Opalka, P. Chacón, A. Polyakov, C. Richter, G. Zhang, and W. Wriggers. Conformational flexibility of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. USA*. In press.

8. W. Wriggers and P. Chacón. Modeling tricks and fitting techniques for multi-resolution structures. *Structure*, 9:779-788, 2001.

9. R.M. Glaeser. Electron crystallography: Present excitement, a nod to the past, anticipating the future. *J. Struct. Biol.*, 128:3-14, 1999.

10. W. Wriggers and P. Chacón. Using *Situs* for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *J. Appl. Cryst.*, 34:773-776, 2001.

11. R.A. Milligan. Protein-protein interactions in the rigor actomyosin complex. *Proc. Natl. Acad. Sci. USA*, 93:21-26, 1996.

12. W. Chiu, R.M. Burnett, and R.L. Garcea. *Structural Biology of Viruses*. Oxford University Press, New York, 1997.

13. W. Wriggers, R.A. Milligan, and J.A. McCammon. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.*, 125:185-195, 1999.

14. P.L. Stewart, S.D. Fuller, and R.M. Burnett. Difference imaging of adenovirus: Bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO J.*, 12:2589-2599, 1993.

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

16. E. Nogales, M. Whittaker, R.A. Milligan, and K.H. Downing. High-resolution model of the microtubule. *Cell*, 96:79-88, 1999.

17. G.J. Kleywegt and T.A. Jones. Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Cryst. D*, 53:179-185, 1997.

18. B.L. de Groot, J.B. Heymann, A. Engel, K. Mitsuoka, Y. Fujiyoshi, and H. Grubmüller. The fold of human aquaporin 1. *J. Mol. Biol.*, 300:987-994, 2000.

19. P. Meurer-Grob, J. Kasparian, and R.H. Wade. Microtubule structure at improved resolution. *Biochemistry*, 40:8000-8008, 2001.

20. W. Jiang, M.L. Baker, S.J. Ludtke, and W. Chiu. Bridging the information gap: Computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.*, 308:1033-1044, 2001.

21. K. Cowtan. Modified phase translation functions and their application to molecular fragment location. *Acta Cryst. D*, 54:750-756, 1998.

22. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, and I.A. Vakser. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, 89:2195-2199, 1992.

23. H.A. Gabb, R.M. Jackson, and M.J. Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, 272:106-120, 1997.

24. J.G. Mandell, V.A. Roberts, M.E. Pique, V. Kotlovyi, J.C. Mitchell, E. Nelson, T. Tsigelny, and L.F. Ten Eyck. Protein docking using continuum electrostatics and geometric fit. *Protein Engineering*, 14:105-113, 2001.

25. I. A. Vakser, O. G. Matar, and C. F. A. Lam. Systematic study of low-resolution recognition in protein-protein complexes. *Proc. Natl. Acad. Sci. USA*, 96:8477-8482, 1999.

26. P. Chacón and W. Wriggers. Multi-resolution contour-based fitting of macro-molecular structures. *J. Mol. Biol.*, In press.

27. J. C. Russ. *The Image Processing Handbook*. CRC Press, 3rd edition, 1998.

28. R.K. Agrawal, A.B. Heagle, P. Penczek, R.A. Grassucci, and J. Frank. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nature Struct. Biol.*, 6:643-647, 1999.

29. W. Wriggers, R.A. Milligan, K. Schulten, and J.A. McCammon. Self-organizing neural networks bridge the biomolecular resolution gap. *J. Mol. Biol.*, 284:1247-1254, 1998.

30. H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, New York, 1992.

31. T. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7:507-522, 1994.
32. M. Zeller, R. Sharma, and K. Schulten. Motion planning of a pneumatic robot using a neural network. *IEEE Control Systems Magazine*, 17:89-98, 1997.
33. B. Delaunay. Sur la spère vide. *Bulletin of the Academy of Sciences USSR*, VII:793-800, 1934.
34. T. Kohonen. *Self-Organizing Maps*. Springer Verlag, Berlin, 2nd edition, 1995.
35. M.M. van Hulle. *Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organisation*. John Wiley and Sons, 2000.
36. J.A. Speir, S. Munshi, G. Wang, T.S. Baker, and J.E. Johnson. Structures of the native and the swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, 3:63-78, 1995.
37. V.E. Galkin, A. Orlova, N. Lukoyanova, W. Wriggers, and E.H. Egelman. ADF stabilizes an existing state of F-actin and can change the tilt of F-actin subunits. *J. Cell Biol.*, 153:75-86, 2001.
38. W. Wriggers and S. Birmanns. Using Situs for flexible and rigid-body fitting of multi-resolution single molecule data. *J. Struct. Biol.*, 133:193-202, 2001.
39. W.J. Rice, H.S. Young, D.W. Martin, J.R. Sachs, and D.L. Stokes. Structure of Na^+, K^+ -ATPase at 11-Å resolution: Comparison with Ca^{2+} -ATPase in E^1 and E^2 states. *Biophys. J.*, 80:2187-2197, 2001.
40. M.J. Packer and C.A. Hunter. Sequence dependent DNA structure: The role of the sugar phosphate backbone. *J. Mol. Biol.*, 280:407-420, 1998.
41. J.A. McCammon and S.C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, 1987.
42. A.R. Leach. *Molecular modelling: Principles and applications*. Addison Wesley Longman, Essex, UK, 1996.
43. M.A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100(7):2567-2572, 1996.
44. J.B. Clarage, T. Romo, B.K. Andrews, B.M. Pettitt, and G.N. Phillips. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA*, 92:3288-3292, 1995.
45. Y. Yang, I. Tobias, and W.K. Olson. Finite element analysis of DNA supercoiling. *J. Chem. Phys.*, 98:1673-1686, 1992.
46. T.P. Westcott, I. Tobias, and W.K. Olson. Elasticity theory and numerical analysis of DNA supercoiling: An application to DNA looping. *J. Phys. Chem.*, 99:17926-17935, 1995.
47. N. Bruant, D. Flatters, R. Lavery, and D. Genest. From atomic to mesoscopic description of the internal dynamics of DNA. *Biophys. J.*, 77:2366-2376, 1999.
48. D.A. Case. Normal mode analysis of protein dynamics. *Curr. Opinion Struct. Biol.*, 4:285-290, 1994.
49. D.A. Case. Normal mode analysis of biomolecular dynamics. In W.F. van Gunsteren, P.K. Weiner, and A.J. Wilkinson, editors, *Computer simulation of biomolecular systems, Vol. 3*, pages 284-301, Dordrecht, Netherlands, 1997. Kluwer Academic Publishers.
50. D. Lin, A. Matsumoto, and N. Go. Normal mode analysis of a double-stranded DNA dodecamer d(CGCGAATTCGCG). *J. Chem. Phys.*, 107:3684-3690, 1997.

COMBINING ELECTRON MICROSCOPIC WITH X-RAY CRYSTALLOGRAPHIC STRUCTURES

Michael G. Rossmann, Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392, USA

Introduction

Cryo-electron microscopy (cryoEM) image reconstructions of biological macromolecules or macromolecular assemblies are rapidly improving, both in quality and in the attainable resolution (Böttcher *et al.*, 1997; Conway *et al.*, 1997; Chiu *et al.*, 1999; Mancini *et al.*, 2000; Zhou *et al.*, 2000). This has made it possible to study objects that are unlikely to crystallize at the equivalence to near atomic resolution by combining the lower resolution cryoEM results of these molecular complexes with higher resolution crystal structures of the component parts.

Comparison of a proposed structure with an observed cryoEM density can be performed either in reciprocal (Kolatkar *et al.*, 1999) or real (Cheng *et al.*, 1995; Grimes *et al.*, 1997; Volkmann & Hanein, 1999; Wriggers *et al.*, 1999; Kikkawa *et al.*, 2000; Roseman, 2000; Rossmann, 2000) space. Real space has the benefit of being able to define more easily the volume within which the comparison is to be made, whereas reciprocal space has the advantage of utilizing a more readily definable search or refinement target. The procedure described here considers the impact of symmetry on the cryoEM density and the imposition of restraints provided by additional information, such as the location of glycosylation sites or knowledge about the function of specific amino acids. These algorithms have been incorporated into the EMfit program, which is freely available from the author.

Because of the restrictions in the suggested space for this presentation, no actual examples (figures or tables) are given here. However, it is hoped to be able to give these as a poster or as supplemental text during the time of the meeting. Furthermore, much of this material (with examples) is in press (Rossmann *et al.*, 2002).

CryoEM Map Scaling

Calibration of electron microscope magnification is not very accurate, in general, and can be off by as much as ±5% in some cases. More certain results can be achieved by including objects of accurately known size in the specimen. However, that cannot always be done easily or such data might be unavailable. An alternative method is to compare the cryoEM map with a related X-ray crystallographically determined map (Rossmann, 2000).

Determining the Orientations that Give the Best Fits of an Atomic Model into a CryoEM Map

Let the atomic positions of a model be given by (X,Y,Z), or, in vector notation, by $\mathbf{X}$, relative to an orthogonal coordinate system. Let the center of gravity of the atoms be at $\mathbf{S}$. Let the rotation matrix required to place the model into the "reference" density be $[E]$, where the nine elements of this matrix are most conveniently expressed in terms of Eulerian angles (Rossmann & Blow, 1962). Then,

$$\mathbf{X'} = [E]\mathbf{X} + \mathbf{d}, \tag{1}$$

where $\mathbf{X'}$ are the coordinates of the model in the map and $\mathbf{d}$ is a translation vector. Let $\mathbf{S'}$ be the approximate target position in the EM map for placement of the model's center of gravity. It follows that

$$\mathbf{S'} = [E]\mathbf{S} + \mathbf{d}$$

and, hence,

$$\mathbf{d} = \mathbf{S'} - [E]\mathbf{S}. \tag{2}$$

In many cases, particularly in the study of icosahedral viruses, the cryoEM density has been generated with the assumption that the structure has specific symmetry properties (532 in the case of an icosahedron). The

placement of an atomic object at a specific reference position with an assumed orientation will require the generation of other symmetry-related objects. By analogy with crystallography, symmetry that has been imposed in the EM reconstruction can be considered as being "crystallographic symmetry", whereas other symmetry operations (e.g. those resulting from quasi-symmetry in spherical virus structures) are analogous to "non-crystallographic symmetry" (NCS).. It will be necessary to test the fit of the known structure to the density at each NCS-related site. Let the reference molecule or subunit be reproduced by L crystallographic and M NCS operations, given by $[R_m]$ ($m = 1,LM$) about a central point $\mathbf{C} = \mathbf{C'}$. Thus, using (1) and (2),

$$\mathbf{X}_m = [R_m][E](\mathbf{X} - \mathbf{S}) + [R_m](\mathbf{S'} - \mathbf{C}) + \mathbf{C} . \tag{3}$$

The quality of fit (*sumf*) can be measured as the average value of the density at all M independent, NCS-related atomic positions using either all atoms or only the $C_.$ atoms of a protein. The position of an atom will not necessarily be at a grid point. Hence, the density will need to be interpolated. If there are N atoms in the search, then the normalized quality of fit is given by

$$sumf = \left\{ 100.0 \cdot \sum_M \sum_N \left[\rho(\mathbf{X}_{m,n}) \right] \right\} \Big/ \left(M \cdot N \cdot \rho_{\max} \right), \tag{4}$$

where $\bullet_{\max}$ is the highest density in the map.

The sensitivity for finding a good fit can be enhanced by counting the number of atoms (*neg*) in each of the M molecules per NCS asymmetric unit, expressed as a percentage (p_{neg}) of the total number of atoms ($M \cdot N$) that occur in low or negative density, where

$$p_{\text{neg}} = \left(100.0 \cdot \sum_M neg \right) \Big/ \left(M \cdot N \right). \tag{5}$$

In essence, this is putting more emphasis on keeping atoms out of low density than on insisting on fitting atoms to the highest possible density values. A good fit will be one that maximizes the sum of sampled density and minimizes the number of atoms in low density.

In addition, the number of steric clashes should be minimized. Clashes are defined as the number of atoms (*nclash*) in one subunit that come to within a specified distance ($\bullet$) of any atom in a neighboring subunit. This can be expressed as a percentage (p_{clash}) of the number of atoms per subunit, where

$$p_{\text{clash}} = \left(100.0 \cdot \sum_M nclash \right) \Big/ \left(M \cdot N \right). \tag{6}$$

If the search structure contains all main chain and side chain atoms, then the minimal approach distance should be set to about 3.4 Å or, if only $C_.$ atoms are available (each representing an amino acid), then an approach distance of about 7 Å is reasonable.

Other criteria can also help to establish the best fit by recognizing specific features in the cryoEM map. Examples are glycosylation sites on the atomic model whose positions in the cryoEM map had been established previously (Kolatkar *et al.*, 1999; Pletnev *et al.*, 2001) and knowledge that a certain part of the atomic model is required to dock to a specific region of another structure whose location in the map is already known (He *et al.*, 2001). A set of restraining positions can be identified on the atomic model that corresponds to a set of target positions in the map. Preference can be given to those fits that produce the lowest r.m.s. or mean distances ("*avgdist*") between restraining and target positions. Conversely, it may be required that regions in the map need to be avoided as they are known to be occupied by another structure (Pletnev *et al.*, 2001). A suitable criterion to control this situation is to minimize (*near*) the number of atoms in the atomic model that approach the occupied region to less than a specified distance.

Combining the Various Criteria for Determination of the Best Fit

It is necessary to establish a single criterion of fit (R_{crit}) from the diverse considerations that contribute to a reasonable fit of the atomic model to the cryoEM density. The various criteria mentioned above have a variety of different dimensions ranging from units proportional to electrons per atomic site (*sumf*), to lengths

(*avgdist*), and to being dimensionless (*neg, nclash, near*). It is possible, however, to put each criterion onto a comparable dimensionless scale by expressing each as a ratio of its value (*v*) with its standard deviation, $\sigma(v)$, from its mean computed over the random set of orientations and positions encountered in the general three-dimensional orientation search (see below). Then, the average of these various criteria is given by

$$R_{\text{crit}} = \sum \omega_i \left[(v_i - <v>) / \sigma_i(v) \right] / \sum \omega_i , \qquad (7)$$

where the sums are taken over the *i* independent criteria and ω_i are the weights to be placed on each. Normally, the weights would be unity, assuming that each has an equal importance in finding the best fit. But if, for instance, more weight is to be placed upon fitting the carbohydrate sites to their target positions in the map than on maximizing the fit of the model to the density, then the weight on *avgdist* should be increased to, say, 3.0 instead of 1.0.

The Search Process
A complete three-dimensional search of the angles that define [*E*], using the relationship given by (3), can be computed to make sure that all likely fits of the model to the density have been found. This search can be performed conveniently using Eulerian angles (Rossmann & Blow, 1962) covering all unique orientations and saving the best (usually between 25 and 100) results for further refinement. Because of the considerable computing time required for a three-dimensional search, a good strategy is to use fairly large angular intervals for the initial search and to refine subsequently the best retained fits. About 10° increments ($\delta°$) are suitable for cryoEM maps at 20 Å resolution (*res*) with an object that has a mean diameter (*dia*) of 40 Å, in accordance with the formula

$$\delta = (res * 180) / (3 * \pi * dia). \qquad (8)$$

Refinement of the Best Search Results
The top (usually between 25 and 100) best results from the general search could be refined by either a least-squares procedure or a systematic or "*climb*" search (Rossmann, 2000). The *climb* procedure consists of investigating one parameter (usually there are three rotational and three translational parameters) at a time using a set of preset increments. The first parameter is increased or decreased to attain the best fit. This procedure is then repeated for the second parameter and so forth. When the last parameter has been adjusted, the search is returned to the first parameter until no further change has occurred in any of the parameters. All the parameter increments are then divided by 10, and the same procedure is repeated. The process is considered to have converged when the parameter increments have decreased to an insignificant value, say 0.25° in angles and 0.5 Å in translations, when searching a 20 Å resolution map (see eq. 8).

The selected top results are likely to cluster around the larger peaks of the search space. Hence, the refinement of the best results will converge to only a few different end points. These can often be reduced further because some, that may appear to have rather different angles, are in reality essentially the same as a consequence of degeneracy in the Eulerian angles.

Additional Criteria to Check the Best Structural Candidates
There usually are a number of additional data that can verify the uniqueness of the structure. For instance, the quality of fit into the cryoEM density of different structural domains or different quasi-symmetry (NCS)-related molecules should be similar. A further check can be made to determine whether all the larger pixels have been interpreted once all components of the complex (such as a virus) have been fitted into the density.

In general, the hand of a cryoEM reconstruction is unknown. An arbitrary decision is made in the early stages of the reconstruction when selecting the relative orientation of the initial particle orientations. Thus, the fitting operation will need to be attempted for both hands. If there is a good differentiation with one hand of the map providing a much better recipient for the X-ray structure, then this is a helpful confirmation of the probable correctness of the fit.

A convenient way of sequentially fitting different components to a map can be achieved by setting those pixels to zero that are within a given distance (e.g. van der Waals radius) of a fitted atom. This produces a "difference" map showing only the density of the uninterpreted map. This procedure has been useful for recognizing the shape and organization of the E2 glycoprotein of Sindbis virus (Zhang *et al.*, manuscript in preparation), whose atomic resolution structure still remains unknown. Similarly, this type of difference map has been used to identify the prM protein in dengue virus (Kuhn *et al.*, 2002).

References

1. Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature (London)* **386**, 88-91.
2. Cheng, R. H., Kuhn, R. J., Olson, N. H., Rossmann, M. G., Choi, H. K., Smith, T. J. & Baker, T. S. (1995). Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell* **80**, 621-630.
3. Chiu, W., McGough, A., Sherman, M. B. & Schmid, M. F. (1999). High-resolution electron cryomicroscopy of macromolecular assemblies. *Trends Cell Biol.* **9**, 154-159.
4. Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature (London)* **386**, 91-94.
5. Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I. & Prasad, B. V. V. (1997). An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron microscopy. *Structure* **5**, 885-893.
6. He, Y., Chipman, P. R., Howitt, J., Bator, C. M., Whitt, M. A., Baker, T. S., Kuhn, R. J., Anderson, C. W., Freimuth, P. & Rossmann, M. G. (2001). Interaction of coxsackievirus B3 with the full-length coxsackievirus-adenovirus receptor. *Nat. Struct. Biol.* **8**, 874-878.
7. Kikkawa, M., Okada, Y. & Hirokawa, N. (2000). 15 Å resolution model of the monomeric kinesin motor, KIF1A. *Cell* **100**, 241-252.
8. Kolatkar, P. R., Bella, J., Olson, N. H., Bator, C. M., Baker, T. S. & Rossmann, M. G. (1999). Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. *EMBO J.* **18**, 6249-6259.
9. Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S. & Strauss, J. H. (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell,* in press.
10. Mancini, E. J., Clarke, M., Gowen, B., Rutten, T. & Fuller, S. D. (2000). Cryo-electron microscopy reveals the functional anatomy of an enveloped virus, Semliki Forest virus. *Mol. Cell* **5**, 255-266.
11. Pletnev, S. V., Zhang, W., Mukhopadhyay, S., Fisher, B. R., Hernandez, R., Brown, D. T., Baker, T. S., Rossmann, M. G. & Kuhn, R. J. (2001). Locations of carbohydrate sites on Sindbis virus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* **105**, 127-136.
12. Roseman, A. M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr. sect. D* **56**, 1332-1340.
13. Rossmann, M. G. (2000). Fitting atomic models into electron microscopy maps. *Acta Crystallogr. sect. D* **56**, 1341-1349.
14. Rossmann, M. G., Bernal, B. & Pletnev, S. V. (2002). Combining electron microscopic with X-ray crystallographic structures. *J. Struct. Biol.,* in press.
15. Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24-31.

16. Volkmann, N. & Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.* **125**, 176-184.
17. Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185-189.
18. Zhou, Z. H., Dougherty, M., Jakana, J., He, J., Rixon, F. J. & Chiu, W. (2000). Seeing the herpesvirus capsid at 8.5 Å. *Science* **288**, 877-880.

MODELING OF MOLECULAR ASSEMBLIES BY SATISFACTION OF SPATIAL RESTRAINTS

Andrej Sali, Rockefeller University, New York, NY, email Sali@rockefeller.edu.

Information about 3D structures of macromolecular assemblies is generally obtained by a variety of experimental and theoretical methods, including X-ray crystallography, electron microscopy, NMR spectroscopy, chemical cross-linking, foot-printing assays, immuno-purification, and different kinds of docking calculations. Therefore, there is a need for a protein structure modeling method that derives the best macromolecular model, at the appropriate resolution, given all of the available experimental data and theoretical considerations.

In general, protein structure modeling is an optimization problem whose solution requires satisfaction of spatial restraints. The approach requires a representation of the modeled system, an objective function consisting of spatial restraints, and an optimizer that produces a model by minimizing violations of the input restraints. We implemented modeling by satisfaction of spatial restraints in the MODELLER program [1,9]. Although the program has been applied primarily to comparative modeling, it may also provide a convenient tool in the modeling of molecular assemblies. First, it is possible to use different descriptions of the system of interest, including points representing atoms, united atoms, residues, arbitrary groups of atoms, or whole proteins. Second, it is possible to include restraints on the symmetry of the assembly as well as restraints on the distances, angles, and dihedral angles defined by two, three and four points respectively. And finally, it is possible to use a variety of optimization protocols to find a set of models that satisfy the input restraints as well as possible.

Our efforts so far are illustrated by two specific examples outlined below; each one is also presented as a poster at this Workshop.

Example 1: Comparative protein structure models for molecular modeling of the yeast ribosome (Eswar et al).
While a combination of cryo-electron microscopy (EM) and X-ray crystallography allows us to determine atomic structures of large macromolecular assemblies, this process is time consuming. Fortunately, the structures of large assemblies and their constituent parts tend to be conserved in evolution, enabling homology or comparative protein structure modeling to enhance the value of low-resolution electron density maps. Due to the progress in structural biology, the structures of the individual subunits of larger assemblies are frequently already known or can at least be modeled relatively accurately by comparative modeling. For example, comparative modeling can currently be used to characterize the 3D structure at the level of a fold for approximately 30% of all domains in the known sequences [1]. This fraction will grow rapidly, by ~10% per year for the next few years, in large part due to structural genomics. Therefore, there is a practical need to improve the use of homologous subunit structures in the modeling of protein assemblies. To illustrate this point, we present here the fitting of homology models for the yeast ribosomal proteins into a low-resolution cryo-electron density map of the whole yeast ribosomal particle.

High-resolution crystallographic structures are available for the archaeal ribosomes, but not for the eukaryotic yeast ribosome. However, there is a 15Å cryo-EM map for the yeast ribosome [2]. While there are species-to-species variations in the ribosome size and protein/RNA ratio, evolutionary conservation of rRNA and ribosomal proteins indicates similar quaternary structure among the archaeal and eukaryotic ribosomes. An automated software pipeline for comparative modeling, ModPipe [3], produced comparative models for 43 out of the 77 proteins of the yeast ribosome, based on their archaeal homologs of known 3D structure [4]. The models were calculated using alignments with sequence identities from 20 to 56% (an average of 32%) and E-values better than 10^{-4}. The modeled fraction of the yeast ribosomal sequences ranges from 34 to 99% (an average of 75%). The models were docked manually into the cryo-EM density with the aid of the program O [5]. Thus, a combination of EM and homology modeling resulted in a partial molecular model of the whole ribosomal particle.

The usefulness of comparative models is limited by their accuracy; similar limitations may also apply to the experimentally determined subunit structures, due to the induced fit. The largest errors result from an incorrect choice of the template structure for comparative modeling, misalignments, errors in loop modeling, and errors in the modeling of rigid body shifts. It is usually possible to generate a set of comparative models that are based on alternate alignments, templates, and domain orientations; some of these models may be more accurate than others. The best subunit models and their

positions in the complex may then be identified by automated docking of the alternate models into the electron density data from electron microscopy or low-resolution X-ray crystallography. Ultimately, the best protein assembly model may be obtained by satisfying simultaneously the homology-derived restraints on the individual subunits and shape restraints on the whole complex.

Example 2: Modeling the three-dimensional structure of the yeast nuclear pore complex (Alber et al.)

Some large assemblies, such as the nuclear pore complex, consist predominantly of subunits whose structures have not yet been defined. The assemblies may be characterized only by very low-resolution information about their overall shape and some protein-protein contacts. In such instances, we can expect to be able to model only the configuration of the proteins in the assembly, not their structures.

We calculated a 3D model of the yeast NPC. The nucleoporin proteins are represented as spheres with the radii estimated from the lengths of their amino acid sequences. The NPC structure was obtained by minimizing violations of spatial restraints obtained from several sources. The restraints included exclusion volume restraints, protein-protein contact restraints extracted from approximately 30 immuno-purification experiments [6,7] and symmetry restraints. Symmetry restraints derived from cryo-electron microscopy [8] reproduce the 8 fold symmetry of the NPC complex. Each of the 8 symmetry units consist of 59 proteins. Starting with random configurations of NPC proteins, many 3D models were calculated by minimizing violations of the input spatial restraints, using a combined method of conjugate gradients and molecular dynamics with simulated annealing, as implemented in the program MODELLER [9]. The final configurations satisfying the input restraints well were clustered. The clusters were analysed to direct new experiments for obtaining maximum spatial information while minimizing the effort, and to address several outstanding questions in the structure-function relationship of the NPC.

Future directions include the use of multiple points to describe multi-domain proteins, perhaps further characterized by the information from the Stokes radius. In addition, we initiated a collaboration with the New York Structural Genomics Consortium that will hopefully result in X-ray structures or comparative models for yeast nucleoporins, to be used directly in the NPC modeling.

References:

1. M.A. Martí-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.

2. Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, Blobel G: Architecture of the Protein-Conducting Channel Associated with the Translating 80S Ribosome. *Cell* 107:361-372, 2001

3. Sanchez R, Sali A: Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc.Natl.Acad.Sci.U.S.A* 95:13597-13602, 1998.

4. Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G, Frank J: Structure of the 80S Ribosome from *Saccharomyces cerevisiae* tRNA- Ribosome and Subunit-Subunit Interactions. *Cell* 107:373-386, 2001.

5. Jones TA, Zou JY, Cowan SW, Kjeldgaard: Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.A* 47:110-119, 1991

6. Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., Chait, B.T. "The Yeast Nuclear Pore Complex: Composition, Architecture, and Transport Mechanism" J. Cell. Biol.148, 635-665, 2000.

7. Chait, B.T, Kipper, J., Rout, M.P., Suprapto, A., Zhang, W., personal communication.

8. Yang, Q., Rout, M.P., Akey, W. "Three-Dimensional Architecture of the Isolated Yeast Nuclear Pore Complex: Functional and Evolutionary Implications" Mol. Cell 1, 223-224, 1998.

9. Sali, A., Blundell, T.L. "Comparative protein modelling by satisfaction of spatial restraints" J. Mol. Biol. 234, 779-815, 1993.

PROBING THE STRUCTURAL AND ENERGETIC BASIS OF PROTEIN-PROTEIN INTERACTIONS.

Barry Honig, Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University,, 630 West 168th Street, New York, NY 10032

The problem of predicting the structure of a complex based on the structures of the individual components, the docking problem, is frequently divided into two steps. One involves a geometric matching of the interacting molecules and the other involves "scoring" the model complexes generated in the first step. Scoring functions can be derived in a number of ways and have been based on the statistical properties of interfaces, e.g. the propensities of certain amino acids to be close to one another, and on physical chemical principles. Docking has been widely applied in the study of protein-small molecule interactions, a problem of direct pharmaceutical relevance. Protein-protein interactions have been studied in this way as well, however the problem of predicting the binding mode for two or more proteins appears to be considerably more difficult than for protein-small molecule interactions. This is due in part to the fact that protein-protein interfaces tend to be more planar than the binding sites of enzymes which are frequently the target of efforts in structure-based drug design. Enzyme binding sites can often be identified as the largest contiguous concave region on a protein surface. This is not the case for protein-protein interfaces.

It has been determined in a number of studies that there is a significant population of charged and polar residues on protein-protein interfaces. Given the fact that the surfaces of proteins are quite polar it is perhaps not surprising that the interfaces are polar as well and that complex formation results in the burial of a significant number of charged and polar residues. However in principle interfaces might have evolved to be nonpolar and hence to be similar to the interiors of the individual monomers. In fact, the contrast between protein-protein complexes and folded proteins, which have large hydrophobic cores and tend not to bury many charged and polar group, is quite striking. The apparent differences in the balance of forces between protein folding and protein-protein association may be made possible in part by the fact that the configurational entropy penalty associated with protein folding is much larger than the loss of translational and rotational entropy associated with complex formation. Thus, the forces that drive protein-protein association need not be as large as those that drive protein folding and, for example, the relative contribution of hydrophobic interactions can be reduced in the former case.

Since, due to desolvation effects, buried charges, ion pairs and hydrogen bonds appear to destabilize protein folding, it might be expected that polar groups buried in protein-protein interfaces oppose complex formation. However, evidence suggests that, besides contributing to the specificity of association, a role generally attributed to polar residues, these residues can play a role in complex stabilization as well. Specifically, it appears that protein interfaces can be designed so as to optimize interactions among charged and polar groups and that these, under certain circumstances, can be larger than the desolvation penalty associated with their burial. This, in turn can produce interactions that are both stabilizing and highly directional and distant dependent, allowing for the remarkable specificity that characterizes the recognition processes involving biological macromolecules.

In order to test our understanding of the physical chemical basis of protein-protein interactions, the methods of continuum electrostatics were used to evaluate the relative binding free energies of docked conformations of the complexes generated from a fast docking algorithm. For each complex both the experimentally determined structure as well as other binding modes generated by the algorithm were considered. In general, it was possible to identify a significant number of conformations that, in terms of geometric complementarity, were as good or even better than the native conformation. However in all cases the native structure was found to have the lowest calculated binding free energy and this was due in large part to the fact that electrostatic free energies were higher in non-native than in native conformations. In the native binding modes, charged groups that were deeply buried were often found to favor complex formation while for non-native binding modes, they were as likely to be stabilizing as destabilizing.

The structural and energetic design of binding interfaces was further characterized in four protein-protein complexes; barnase-barstar, human growth hormone and its receptor, subtype N9 influenza virus neuraminidase

and the NC41 antibody, the Ras Binding Domain (RBD) of kinase cRaf and a Ras homologue Rap1A. In two of the four complexes electrostatics strongly oppose binding (hormone-receptor and neuraminidase-antibody complexes) in one case the net effect is close to zero (barnase-barstar) and in one case electrostatics provides a significant driving force favoring binding (RBD-Rap1A). In order to help understand the wide range of electrostatic contributions that were calculated, the electrostatic free energy was partitioned into contributions of individual charged and polar residues, salt bridges and networks involving salt bridges and hydrogen bonds. Although there is no one structural feature that accounts for the differences between the four interfaces, the extent to which the desolvation of buried charges is compensated by the formation of hydrogen bonds and ion pairs appears to be an important factor. Structural features that are correlated with contribution of an individual residue to stability include partial burial of a charged group in the free monomer, the formation of networks involving charged and polar amino acids, and the formation of partially exposed ion-pairs. The total electrostatic contribution to binding is found to be inversely correlated with buried total and nonpolar surface area.

The problem with using insights such as these in a predictive sense is that it has been difficult to identify physical chemical features that clearly distinguish interfacial regions from the rest of the protein surface. A survey from our lab of all protein-protein interfaces reported in the PDB suggest that sequence conservation patterns may offer the most reliable indicator of the binding region. The incorporation of this information into docking procedures can lead to significant improvement in prediction accuracy. In addition, distinct patches of electrostatic potential, or of nonpolar residues, can in many cases also indicate regions that are important in binding. In cases where structures of different members of a protein family are available, especially if they are complexed with other molecules such as small peptides, a combined physical chemical/bioinformatics approach may prove fruitful.

As an example of such an approach, we have developed a method to predict binding specificity of a sequence of unknown structure based on the structures of complexes of related family member. In a pilot study involving SH2 domains, we calculate the free energy contribution to binding of every residue in the protein for every SH2 domain-peptide complexes of known structure. For each complex we identify the residues that provide the most significant energetic contribution to binding and then map these onto our sequence profile of SH2 domains. A new sequence of unknown specificity is then aligned to this profile thus allowing us to associate sequences of unknown specificity to one of the structurally characterized complexes and, in this way, to predict the binding specificity of a large number of family members. It is essential for this purpose to start with accurate alignments; these can be generated by new methods to align members of a particular protein family that combine both sequence and structural information.

In summary, a wide range of techniques are now available to help construct models of multi-protein complexes based on the structures of the individual components. Focusing on individual protein families may prove to be particularly effective since multiple sequence and multiple structural information can be combined with physical chemical insights to yield family-specific predictive tools.