

Developments in low-resolution biological X-ray crystallography

Fred Dyda

Address: Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Bethesda, MD 20892, USA

Email: fred.dyda@nih.gov

F1000 Biology Reports 2010, 2:80 (doi:10.3410/B2-80)

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/legalcode>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You may not use this work for commercial purposes.

The electronic version of this article is the complete one and can be found at: <http://f1000.com/reports/b/2/80>

Abstract

Despite the recent substantial technological developments in X-ray crystallography, solving and refining structures at low resolutions remain substantial challenges. Many macromolecular crystals, especially those of large molecules or multicomponent assemblies, diffract X-rays to resolutions that are worse than 3.5Å. This report summarizes several recent advances aiding low resolution crystallographic work.

Introduction and context

While X-ray crystallography can be used to determine molecular structures at atomic resolution in principle, it is often not possible because of limitations in crystal quality. This is especially so in the case of biological macromolecules. Due to their inherent flexibility and the relative sparsity of contacts with which to hold a crystal lattice together, the highest resolution where X-ray diffraction data are measurable is often worse than 3.5Å. The Protein Data Bank [1] currently contains 863 entries at resolutions lower than 3.5Å and their number seems to be increasing rapidly [2].

One might ask why researchers bother to solve structures that clearly will not yield near-atomic level detail rather than focusing on projects where such detail can be revealed? It turns out that a large number of some of the most important questions that structural biology attempts to address involve determining how large macromolecules and their complexes assemble, and understanding how their components interact. Substantial insight into these issues does not necessarily require structures at near-atomic resolution, although if such structures were possible to obtain they would clearly be much more desirable. Given that biological interest is focused on large complexes, the crystallographer often has no choice but to compromise and accept and deal with the inherent experimental limitation of low resolution.

The road to a low resolution crystal structure is a difficult one to navigate, paved with potholes of frustrations and ambiguities, and it is in sharp contrast to what feels like driving on a well-lit multilane highway when resolutions better than 3Å are available; here, progress can be trusted to automated programs that require less and less human intervention as technology develops. The principal difficulty in low resolution structure determination stems from the limited number of independent (unique) X-ray reflections. In principle, in order to determine a structure, one would need at least as many independent reflections as the number of flexible torsions in the molecule or assembly [3], so if diffraction data were available at least to about 5Å, in most cases determination should be possible. (The exact number varies depending on solvent content, as this relates to the sampling density of the molecular transform.) Diffraction data that are limited to low resolution also tend to be weak, and therefore the error in the intensity estimates could be high. Furthermore, data might be affected substantially by systematic errors such as radiation decay because experimenters tend to expose their crystals to large X-ray doses in order to get stronger diffraction. Since the experimental information is very limited compared to the number of degrees of freedom, redundancy is minimal or none, and the intensity errors are especially detrimental. In addition, data quality can be further compromised if the final data set is the result

of merging partial sets collected on a number of crystals that may be only approximately isomorphous.

Major recent advances

In some cases, some relatively trivial measures have been used to improve data quality. For example, in order to detect reflections beyond 4.5 Å in the case of crystals of the SIV (simian immunodeficiency virus) gp12 envelope glycoprotein, it was important to make sure that a small beamstop was placed close to the crystal and to move the detector 400 mm back from it [4]. This served to minimize the background, as the diffraction limit is basically a signal to noise issue; most of the noise (which is mainly from background) is contributed by the diffuse scattering from the sample and from air scattering of the direct beam. Large crystal-to-detector distances are especially helpful at those synchrotron beamlines where the beam has very small crossfire. In addition, there is the problem of the series termination errors that give rise to ripples next to real density features (see e.g., Minichino *et al.* [5] and references therein). Also, diffraction that is not isotropic, with diffraction limits that are dependent on the direction of the scattering vector, occurs frequently. This situation can be helped by ellipsoidal truncation and anisotropic scaling that can, for example, be done on the UCLA (University of California, Los Angeles) web server [6,7] or the CCP4 (Collaborative Computational Project No. 4) program SCALEIT. The effects of radiation decay in the data sets can be alleviated by applying the so-called zero dose correction [8], provided that each unique reflection was measured a number of times, which may be difficult to achieve in low symmetry space groups.

The combination of the above sorts of factors will almost always lead to electron density maps that are noisy and lacking in detail. In the 4–5 Å resolution range, α -helices appear as tubes and β -sheets as walls of density with no indication where the individual strands might run. Indeed, for the latter, the hydrogen bonding between β -sheet strands is notorious for causing confusion in tracing the path of the polypeptide chain. Of course, even to see this much, some kind of phase information is needed in addition to the intensity data. Unless the macromolecular assembly is mainly made up of α -helical domains, in this resolution range a complete *de novo* structure determination without known three-dimensional structures of its components and domains is extremely challenging and in many cases might be impossible [9]. An important exception is the class of cases where high-order non-crystallographic symmetry is available, such as in the case of spherical or cylindrical viruses.

It is increasingly the case, indeed now very often, that the three-dimensional structures of the components or

fragments of the molecule or assembly in question are already available, and this opens up an avenue towards generating useful initial phase information. Modern automated molecular replacement (MR) programs such as Phaser [10] or AMoRe [11] can generate good solutions even when the search model is a relatively small fraction of the total scattering mass. Also, programs such as Phaser allow an ensemble of search models to be used, thus widening the radius of convergence of MR – an important limitation when only one search model is used. However, for any MR approach to provide meaningful phase information, a large fraction of parts of the assembly has to be known three-dimensionally and the fragment structures should not change much upon assembly formation. In favorable cases, a simple difference Fourier calculated with the MR-based model phases can reveal interesting and previously unseen parts of the assembly [12].

Even when MR is able to place the fragments, it remains extremely desirable to have some experimental phase information. This may come from a selenomethionine (Se-Met) multi-wavelength anomalous diffraction (MAD) experiment, or from heavy atom soaks with, for example, the Ta₆Br₁₂ cluster [13], which is especially suited for low resolution work on large assemblies. The heavy atom substructure should then be solvable with phases computed from the MR solution. Optimizing the heavy atom substructure and subsequent density modification can be done with a variety of programs, such as Sharp/Solomon [14,15], Solve/Resolve [16], and others.

Beyond providing direct phase information, these approaches can also independently verify that things are proceeding well, as the substructure obtained with the MR phases must be the same (except for a possible origin shift) as the ones obtained independently (e.g., by the combined Patterson-Direct methods approach implemented in ShelxD [17]). The huge advantage of either a SAD (single-wavelength anomalous diffraction) or a MAD data set based on Se-Met or Br-dU (bromodeoxyuridine – if there is nucleic acid in the structure) is that the heavy atom substructure could provide further positioning information for the domains or fragments in favorable cases. However, their phasing power is often limited at low resolution. A very important aspect of even modest quality experimental phases is that they are free of model bias.

Model bias, which is more serious at low resolution, is perhaps the greatest caveat in crystallography because the placed model, in the absence of experimental phases, is the only source of phase information. As phase information dominates maps, even an incorrectly placed

or inappropriate model (or both) will inevitably show up in its own density to some extent when a map based on model phases is calculated. In addition to the importance of experimental phases, it is not possible to emphasize how important the exploitation of real space redundancies (non-crystallographic symmetry) is, or – if multiple crystal forms are available – how important it is to attempt multi-crystal averaging [4]. These effectively improve the inherently poor data-to-parameter ratio but assume that the geometrical relationship between the related domains or molecules can be established.

In the past, after placing known fragments into their place and perhaps some rigid body refinement, not much more optimization could be done. However, there have recently been a number of important technical advances in this area. One of these is B-factor sharpening [18], which involves the application of a negative B-factor to the diffraction data set. This increases the highest resolution reflections in the set and can give rise to more detail-rich maps (e.g., visible side chains) and it is especially useful if experimental phases are available. Care is needed in applying this as the weak highest resolution reflections also have the highest errors and it is likely that by increasing their contributions the overall noise of the map will increase as well. The optimum choice is the negative of the pseudo Wilson B-factor of the diffraction data [3]. It is also very important to have a reliable bulk solvent model and to correct for data anisotropy. Previous procedures that have worked well when high resolution data were available displayed unstable results for low resolution sets. New grid search-based iterative parameter optimizations of the bulk solvent model such as the ones implemented in the newer versions of CNS (Crystallography and NMR [nuclear magnetic resonance] system) [19] and Phenix [20] have successfully overcome this problem.

Quite clearly, any attempt to do molecular model refinement at resolutions poorer than 3.5Å has to have stronger and additional restraints applied to the structure. Explicit restraints of secondary structure, typically through some kind of H-bonding potential, are very useful. An exciting recent development is the incorporation of known three-dimensional structures of homologues of the assembly investigated through incorporation of a deformable elastic network (DEN) potential into the target function used in torsion angle dynamics [21]. DEN allows restrained but still large-scale deviations from a high(er) resolution reference structure and this, in principle, overcomes the main limitation of previous refinement protocols.

Future directions

Given the increasing importance of three-dimensional structures of large assemblies, one can expect further

significant technical advances when dealing with low resolution single crystal X-ray diffraction data. These may include the more robust ways to incorporate electron microscopy or small-angle X-ray or neutron scattering (SAXS or SANS) information as further restraints. Also, as one of the most difficult aspects of low resolution structure determination is the interpretation and building of models into the electron density, it will be interesting to see how far automated processes can be developed to accomplish these often frustrating and ambiguous tasks.

Abbreviations

DEN, deformable elastic network; MAD, multi-wavelength anomalous diffraction; MR, molecular replacement; Se-Met; selenomethionine.

Competing interests

The author declares that he has no competing interests.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH). The author thanks David R Davies, Alison B Hickman, and an anonymous referee for useful comments and suggestions.

References

- Protein Data Bank in Europe (PDBe). [<http://www.ebi.ac.uk/pdbe/>]
 - Karmali AM, Blundell TL, Furnham N: **Model-building strategies for low-resolution X-ray crystallographic data.** *Acta Crystallogr D Biol Crystallogr* 2009, **65**:121-7.
 - Brunger AT, DeLaBarre B, Davies JM, Weis WI: **X-ray structure determination at low resolution.** *Acta Crystallogr D Biol Crystallogr* 2009, **65**:128-33.
 - Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC: **Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein.** *Structure* 2005, **13**:197-211.
 - Minichino A, Habash J, Raftery J, Helliwell JR: **The properties of (2Fo-Fc) and (Fo-Fc) electron-density maps at medium-to-high resolutions.** *Acta Crystallogr D Biol Crystallogr* 2003, **59**:843-9.
 - UCLA Diffraction Anisotropy Server.** [<http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/>]
 - Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D: **Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis.** *Proc Natl Acad Sci U S A* 2006, **103**:8060-5.
- F1000 Factor 7
 Evaluated by Rowena Matthews 20 Jun 2006, Chantal Abergel 22 Jun 2006
- Diederichs K, McSweeney S, Ravelli RB: **Zero-dose extrapolation as part of macromolecular synchrotron data reduction.** *Acta Crystallogr D Biol Crystallogr* 2003, **59**:903-9.

9. DeLaBarre B, Brunger AT: **Considerations for the refinement of low-resolution crystal structures.** *Acta Crystallogr D Biol Crystallogr* 2006, **62**:923-32.
F1000 Factor 6
Evaluated by Fred Dyda 07 Aug 2006
10. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software.** *J Appl Crystallogr* 2007, **40**:658-74.
11. Navaza J: **Implementation of molecular replacement in AMoRe.** *Acta Crystallogr D Biol Crystallogr* 2001, **57**:1367-72.
12. Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, Davies DR: **Structural basis of toll-like receptor 3 signaling with double-stranded RNA.** *Science* 2008, **320**:379-81.
F1000 Factor 9
Evaluated by Peter Artymiuk 05 Jun 2008, Hao Wu 17 Jun 2008
13. Neuweind T, Bergner A, Schneider F, Messerschmidt A, Knablein J: **The suitability of Ta₆Br₁₂(2+)** for phasing in protein crystallography. *Biol Chem* 1997, **378**:219-21.
14. de La Fortelle E, Bricogne G: **Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods.** *Methods Enzymol* 1997, **276**:472-94.
15. Abrahams JP, Leslie AGW: **Methods used in the structure determination of bovine mitochondrial F1 ATPase.** *Acta Crystallogr D Biol Crystallogr* 1996, **52**:30-42.
16. Terwilliger TC, Berendzen J: **Automated MAD and MIR structure solution.** *Acta Crystallogr D Biol Crystallogr* 1999, **55**:849-61.
17. Sheldrick GM: **SHELX applications to macromolecules.** In *Direct Methods for Solving Macromolecular Structures.* Edited by Fortier S. Dordrecht, The Netherlands: Kluwer Academic; 1998, 401-11.
18. Bass RB, Strop P, Barclay M, Rees DC: **Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel.** *Science* 2002, **298**:1582-7.
F1000 Factor 15
Evaluated by Senyon Choe 04 Dec 2002, Petra Fromme 06 Jan 2003, Tracy Palmer 22 Jan 2003
19. Brunger AT: **Version 1.2 of the Crystallography and NMR System.** *Nat Protoc* 2007, **2**:2728-33.
20. Adams PD, Afonine PV, Bunkóczki G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH: **PHENIX: a comprehensive Python-based system for macromolecular structure solution.** *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213-21.
21. Schröder GF, Levitt M, Brunger AT: **Super-resolution biomolecular crystallography with low-resolution data.** *Nature* 2010, **464**:1218-22.
F1000 Factor 9
Evaluated by Matt Redinbo 29 Apr 2010, Fred Dyda 05 May 2010