Biochimica et Biophysica Acta 1854 (2015) 258–268

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

You are lost without a map: Navigating the sea of protein structures

Audrey L. Lamb [a,*], T. Joseph Kappock [b], Nicholas R. Silvaggi [c,**]

[a] Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, United States
[b] Department of Biochemistry, Purdue University, West Lafayette, IN 47907, United States
[c] Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, Milwaukee, WI 53211, United States

ARTICLE INFO

Article history:
Received 8 December 2014
Accepted 22 December 2014
Available online 29 December 2014

Keywords:
Protein structure
X-ray crystallography
Molecular models
Electron density maps
Atomic coordinate files
Atomic displacement parameters

ABSTRACT

X-ray crystal structures propel biochemistry research like no other experimental method, since they answer many questions directly and inspire new hypotheses. Unfortunately, many users of crystallographic models mistake them for actual experimental data. Crystallographic models are *interpretations*, several steps removed from the experimental measurements, making it difficult for nonspecialists to assess the quality of the underlying data. Crystallographers mainly rely on "global" measures of data and model quality to build models. Robust validation procedures based on global measures now largely ensure that structures in the Protein Data Bank (PDB) are largely correct. However, global measures do not allow users of crystallographic models to judge the reliability of "local" features in a region of interest. Refinement of a model to fit into an electron density map requires interpretation of the data to produce a single "best" overall model. This process requires inclusion of most probable conformations in areas of poor density. Users who misunderstand this can be misled, especially in regions of the structure that are mobile, including active sites, surface residues, and especially ligands. This article aims to equip users of macromolecular models with tools to critically assess *local* model quality. Structure users should always check the agreement of the electron density map and the derived model in all areas of interest, even if the global statistics are good. We provide illustrated examples of interpreted electron density as a guide for those unaccustomed to viewing electron density.

1. Introduction

Advances in crystallization, data collection, and computers have made macromolecular crystal structures commonplace. Biochemists, medicinal chemists, chemical biologists and many others have come to rely on macromolecular structural data as never before, and it has become routine to read, write, and review manuscripts that contain crystal structures. Furthermore, advances in the field have made it possible for scientists with limited training in crystallography to determine protein structures. Thus, even scientists with no formal background in crystallography need to know how to critically evaluate these complex experiments. While it has been noted recently that poorly determined structures have a negative impact on the drug design community [2], the focus here is on how to avoid the improper use of well-determined structural models. The first step is to understand how crystallographic models are made.

Every atom in the repeating unit of a crystal (the unit cell) contributes to the intensity of *every reflection* in the diffraction pattern. The measured intensity for each diffraction spot is the result of scattering from the *entire model*. Particular data points cannot be associated with specific parts of a model. For example, there is no "metal spot" in data collected from a metalloprotein crystal; the metal contributes to the intensity of every reflection (see Box A for a description of the crystallographic experiment). While crystallographic statistics reported in structure papers provide numerical indications of the overall quality of the diffraction data (for an excellent review, see [3]), these do not report on how well-determined individual parts of a model are. The Protein Data Bank (PDB)[1] has recently adopted a new structure report format that gives a graphical representation of how a given model compares with others in the PDB in terms of five statistical measures of model quality [4–7]. These reports are based on the excellent work of numerous leaders in the field of X-ray structure determination [6]. As good as these reports are, they are focused on the global quality of the structure.

Even in the best cases, there are areas of the electron density map that are poorly defined (Fig. 1). Thus, even a crystal structure that is based on high quality diffraction data and was carefully and competently built and refined will have local areas of the model that are less reliable than the rest. Very often, these regions are on the surface of a protein, and for most users, will not be important in drawing conclusions about molecular structure and function. Of course, if one

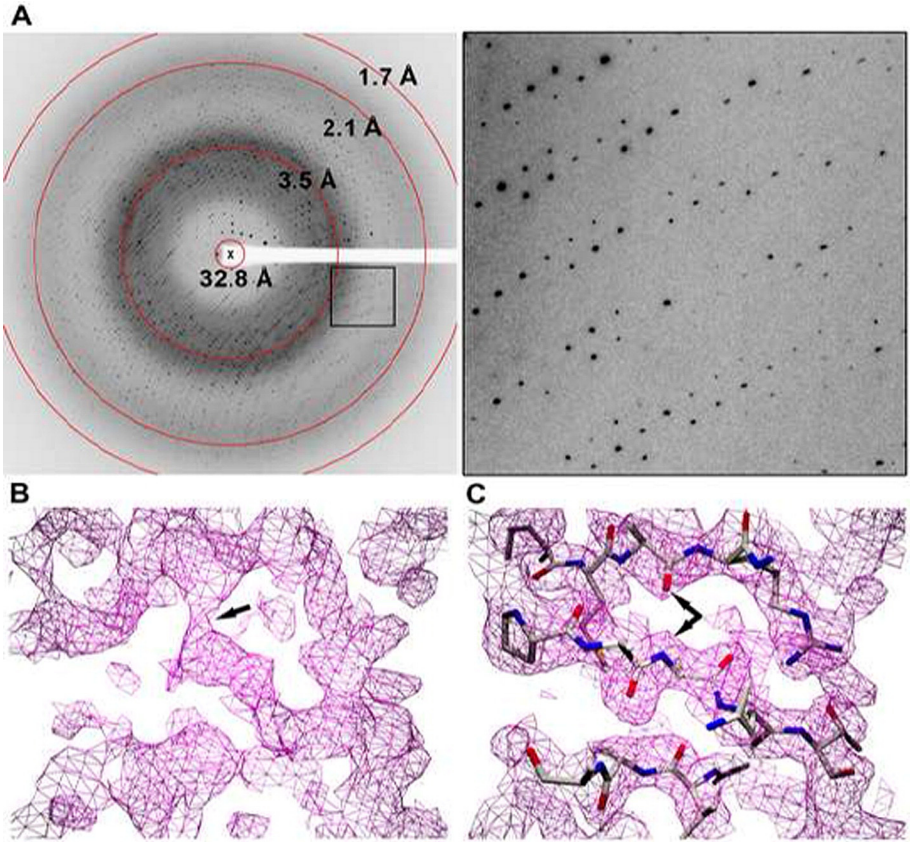
* Corresponding author. Tel.: +1 785 864 5075.
** Corresponding author. Tel.: +1 414 229 2647.
E-mail addresses: lamb@ku.edu (A.L. Lamb), silvaggi@uwm.edu (N.R. Silvaggi).

[1] The current Protein Data Bank is a cooperation of three different organizations, RCSB PDB, PDBe, and PDBj which all contribute entries to the wwPDB (wwpdb.org) [1].

http://dx.doi.org/10.1016/j.bbapap.2014.12.021
1570-9639/

Box A
From X-ray dataset to finished model.

The figure below highlights the steps in X-ray data collection and refinement. A single oscillation image (panel A) is obtained by rotating the crystal through a small rotation angle while it is illuminated by X-rays. Hundreds of these images comprise a data set that completely samples the entire three-dimensional diffraction pattern. The resolution of the data increases from the center to the edge of the image. The highest resolution where the diffraction spots still have measurable intensity gives some idea of the resolution of the data set, about 1.7 Å in this example. The diffuse grey ring near 3.5 Å is background scattering from solvent surrounding the crystal in the sample holder.
In order to calculate an electron density map, a crystallographer requires both the amplitudes of the diffracted X-ray waves *and* their relative phase angles. The amplitudes are measured as the intensities of the diffraction spots in the experiment, but the phase information is lost. This is the crystallographic phase problem. The missing phase information can be obtained from using the structure of a homologous protein (molecular replacement) or by a number of experimental methods involving incorporation of heavy atoms (*e.g.* Hg, Se) into the ordered array of the crystal. There are a number of excellent introductory and advanced texts that provide excellent explanations of phasing methods[2]. However the initial estimates of the phases are obtained, they typically have large errors, and the resulting electron density maps are relatively noisy and ill-defined (Panel B). Once this imperfect electron density map is calculated, the process of building a crystallographic model begins. A macromolecular crystallographer working on a new structure begins with either a molecular replacement model that likely contains significant portions that need to be rebuilt, or an empty map into which they build the polypeptide chain from scratch or using an automated algorithm [48–50]. In either case, the initial model is never an optimal match to the electron density. The initial model is iteratively altered to improve its fit to the electron density by refining some or all atomic parameters (Panel C). When adjustments to the model no longer improve the phase estimates, refinement is stopped and the model is said to be finished.



[2] For readers interested in a more comprehensive explanation of diffraction physics and the X-ray crystallographic experiment, the authors recommend these outstanding texts, ranked in approximate order of difficulty:

1) Rhodes, Gale. *Crystallography Made Crystal Clear.* Academic Press, New York, 2000.
2) Blow, David. *Outline of Crystallography for Biologists*. Oxford University Press, New York, 2002.
3) Glusker, Jenny P. with Lewis, Mitchell and Rossi, Miriam. *Crystal Structure Analysis for Chemists and Biologists.* Wiley-VCH, New York, 1994.
4) Rupp, Bernhard. *Biomolecular Crystallography*. Garland Science, New York, 2010.

is interested in protein–protein interactions, these regions are relevant. One's interests determine which parts of the electron density map to inspect.

Regions of the electron density map that are poorly defined due to mobile, disordered sections of the polypeptide frequently have important functions. For example, an enzyme may adopt multiple conformations