

Atomic Resolution Protein Crystal Structures

ZBIGNIEW DAUTER

EMBL Hamburg, Germany
and Dept. of Chemistry, University of York, England

In small molecule X-ray crystallography, the term "resolution" is rarely used; it is assumed that the data are collected to the maximum 2θ angle achievable on the diffractometer with copper radiation which corresponds to an interplanar spacing of 0.8 Å. Protein crystals in general do not diffract as well as small molecules due to their much larger size and the high content of disordered solvent. Traditionally, macromolecular structures solved at 3 Å or less are called low resolution, up to 2 Å-medium and beyond 2 Å-high resolution. About five years ago, only a few

small proteins (like crambin, rubredoxin, insulin) were known to diffract to beyond 1.2 Å resolution, which can be termed "atomic" [1].

In the last few years, important advances have taken place which have enabled the collection of atomic resolution diffraction data on crystals of larger proteins. The most important factors were the existence of fast detectors, such as imaging plate and CCD scanners, the use of cryogenic cooling and the availability of very high intensity synchrotron beam lines.

The amount of data is proportional to the cube of resolution, and going from 2 to 1 Å, there are 8 times more reflections to collect, integrate and merge. Moreover, these very high resolution intensities are usually weak. Therefore it is important to use fast, sensitive and accurate detectors together with efficient integration software.

The experience of the last few years shows that it is possible to shock-freeze most protein crystals and this technique has become routine. The lifetime of the crystal in the beam is prolonged and it is possible to collect all data from an individual crystal. Any slight non-isomorphism would make it impossible to merge 1 Å data from more than one crystal. Successfully frozen crystals usually diffract better than at ambient temperature. As an example, the triclinic crystals of liver alcohol dehydrogenase diffract to 1.0 Å when frozen but not further than to 1.7 Å at 4°C.

The most important prerequisite for the atomic resolution data to be collected is the availability of synchrotron beam lines, based on wigglers or undulators, providing very strong and highly collimated X-radiation. Such beam lines now exist in most synchrotron sites and new ones are being constructed.

During atomic data collection, one should take every precaution that the data are as complete as possible, especially in the lowest resolution range, and that they are accompanied by well estimated standard uncertainties. This reduces the errors during refinements and ensures that Fourier maps show meaningful features even at very low density levels. It is advisable to use F^2 (intensities) rather than F (amplitudes) for

**You can make no better choice
for an Undulator or Wiggler
than **STI**OPTRONICS**

- ✓ Permanent Magnet/
Hybrid/Electromagnet
- ✓ Principal Radiation Source for the
Advanced Photon Source
(APS-Undulator A)
- ✓ Delivered Over 30 Devices
- ✓ Developed Scalable 10-meter
Hybrid Device (NISUS)
- ✓ Developed Undulator Shimming
- ✓ Developed Enhanced
Wedged-Pole Geometry

Contact: Dr. Kem E. Robinson
2755 Northup Way
Bellevue, WA 98004-1495
206-827-0460 Fax: 206-828-3517
magnetics@stiopt.com

Circle No. 135

refinement, without any omissions and including intensities estimated as negative. If R free cross-validation is employed, all reflections should be used in the final refinement cycles.

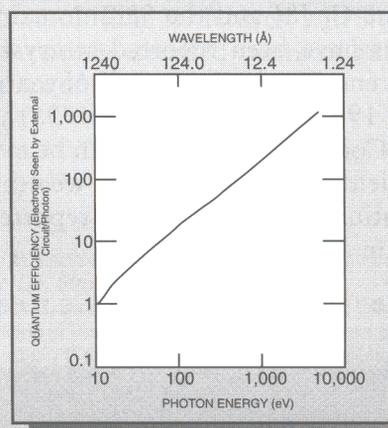
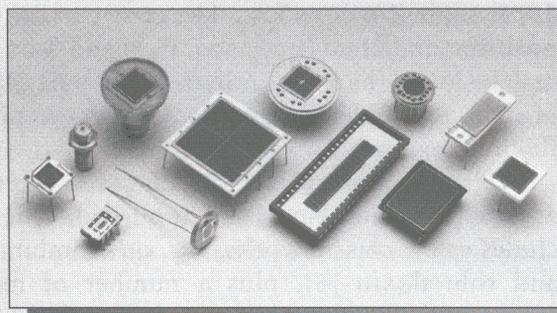
Completeness of the data is particularly important (again, both at high and low resolution limits) if they are to be used to solve the unknown structures by direct or Patterson search methods. Such methods work successfully for metalloproteins; hopefully ever larger structures will be solved by direct methods in the near future [2,3].

At atomic resolution the number of measured reflections exceeds the number of refined parameters several times. Even in case of anisotropic refinement, with 3 positional and 6 anisotropic displacement parameters, at 1.0 Å the number of observables is more than 5 times higher than the number of parameters, assuming about 40% of solvent in the protein crystal. That, in principle, allows refinement of the protein model without stereochemical restraints. In practice this is true for most of the protein chain; however in every structure there are always highly flexible parts which have to be restrained even at the highest resolution. These are usually long, charged side chains at the protein surface, sometimes parts of main chain loops, larger ligands (e.g. oligopeptides or oligosugars) and poorly defined water molecules.

Although almost any macromolecular refinement program used for lower resolution analyses can be used with atomic resolution data, the refinement of protein models at ultra-high resolution requires a specially careful approach. Most of the default restraints and parameter values must be re-evaluated. The program particularly suited for such work, and written with macromolecular as well as small molecular applications in mind, is SHELXL [4]. It allows a comprehensive refinement of atomic models and offers a broad choice of options. One can use either a conjugate gradient algorithm, faster but requiring more cycles, or a full- or block-matrix method, slower and more computer memory demanding but having a larger radius of convergence and requiring less cycles. The full matrix algorithm (standard for small molecules) is the only one giving proper estimates of the uncertainty of individual atomic parameters and of values dependent on them, like bond lengths or angles. SHELXL offers a rich choice of restraints

ABSOLUTE

XUV SILICON PHOTODIODES



TYPICAL QUANTUM EFFICIENCY OF
AXUV SILICON PHOTODIODES

Additional Features

- No Change in QE with 100 Mrad (Si) 124 eV Photons
- 8 Decades of Linearity
- May be Operated Without Bias
- Cryogenically and UHV Compatible
- With Integrated Bandwidth Limiting Filters
- 0.27 A/W Responsivity from 100 eV to 12 keV
- Excellent for Low Energy Electrons and Ions Detection

IRD

INTERNATIONAL RADIATION DETECTORS INC.

2545 West 237th Street, Unit I • Torrance, California 90505-5229
(310) 534-3661 • FAX (310) 534-3665 • email: ird@kaiwan.com

DEVELOPED IN COLLABORATION WITH NIST, NIH, LLNL, LANL, NCAR

Circle No. 136

of both stereochemical and atomic displacement (thermal) parameters.

At about 1 Å most protein models can be refined anisotropically to an R factor of about 10%; some down to 7%. The errors of atomic positions and bond lengths for C, N and O atoms may be lower than 0.02 Å in the good part of the protein molecule ranging to 0.05 Å in more flexible regions.

Not very many macromolecular structures at atomic resolution have been published. This includes a few classic works, e.g. on crambin [5] and rubredoxin [6], plus a number of more recent reports, e.g. ribonuclease Sa [7], cytochrome C₆ [8] and the SH2 domain [9]. More studies have been reported at crystallographic conferences in the last couple of years (ECM-16, Lund, 1995; ACA, Montreal, 1995) and after the IUCr Congress in Seattle it can be expected that this field will boom and a wealth of atomic resolution structures will be reported and analyzed in detail.

Some of the papers on atomic resolution structures report the detailed refinement of models known previously from lower resolution studies and some, in addition, report the solution of unknown structures *ab initio* using direct methods. The application of direct methods to larger structures advanced considerably in recent years, mainly due to the work of groups in Buffalo (Shake'n'Bake) and Goettingen (SHELXS). Currently it seems that structures with up to 500 atoms have a chance to be solved by this approach, especially if they contain heavier atoms. Herbert Hauptman formulated the optimistic opinion that with the next 5–10 years, larger structures (above 1,000 atoms) will be solved by direct methods, possibly also at somewhat lower than atomic resolution. The direct methods approach is of particular importance for oligopeptides, which are too small to develop well defined secondary structure, therefore resist the molecular replacement approach and generally do not bind heavy atoms [10, 11, 12].

Osmic: Multilayer Optics for Synchrotron Applications

*Improve your flux/performance
for applications ranging from
50 eV to 100 keV*

- Multilayer monochromators
- Supermirrors for broader energy range
- Kirkpatrick-Baez microfocus optics
- Spherical and aspherical imaging optics
- Brewster-polarizers
- Custom optical designs

visit our website at www.osmic.com



1788 Northwood Drive ▪ Troy, MI 48084-5532
800-366-1299 ▪ 810-362-1290 ▪ FAX 810-362-4043

Manufacturers of



OVONYX™
Multilayers

Circle No. 139

The size of the protein studied at atomic resolution range from the oligopeptides (sometimes with multiple copies in the asymmetric unit) with 30–40 amino acids to larger proteins. Data to 1.0 or beyond have been collected on crystals of P1 lysozyme (15 kDa), bacterial trypsin (19 kDa), ribonuclease Sa (2×10 kDa), cutinase (20 kDa), gB-crystallin (21 kDa), fungal trypsin (22 kDa), savinase (27 kDa), concanavalin a (25 kDa), xylanase (35 kDa), neuraminidase (42 kDa), up to LADH (2×37 kDa).

The refinement and analysis of the large amount of structural detail from atomic resolution study takes a long time. That, in part, explains the delay between reporting preliminary results and appearance of the final publications. However, several conclusions can already be formulated. First, even relatively large proteins can diffract to atomic resolution (like LADH with 750 amino acids in the asymmetric unit). That suggests that, in the future, many more atomic resolution studies can be expected. Second, although comprehensive analyses must wait, some trends in these very accurate structures have already been identified. Up to 15% of the amino acid residues exist in alternative conformations, most often serine, threonine and acidic side chains, but sometimes also tyrosine and even tryptophan. Double conformations of the main chain loops have also been found [13]. Ionization states of charged acidic groups can be identified from the corresponding C–O bond lengths. Although the OH hydrogen of the catalytic serine was not found in the difference Fourier map of savinase [14], such studies are feasible. Such subtle features as the electron density corresponding to s and p bonds with the “averaged” peptide group of crambin have been reported. Third, the atomic resolution structures will serve as an improved “library” of stereochemical target values in the refinement and validation of structures at lower resolution. For example, the peptide groups tend to be not completely planar; their spreads ($\sigma=6^\circ$) around the average of 179° reaches up to 20° . Fourth, the solvent structure can be analyzed more accurately; often well defined alternative, partially occupied water sites may be identified.

Atomic resolution studies open up the possi-

bility of identifying and discussing many subtle features which play very important biological roles. Many biological processes involve the transfer of hydrogen or electrons; discussion of such phenomena requires very high accuracy of the structural models, which only atomic resolution can provide. Another example is the discussion of thermophilicity; it seems clear now that there is no single decisive factor responsible for this. Instead, many weak, low energy contributions lead to the global effect. That again requires highly accurate structures for detailed discussion.

There is no doubt that in the near future many new atomic resolution structures will appear, shedding light on the detailed chemical and biological features of proteins. ■

1. Z. Dauter, V.S. Lamzin, K.S. Wilson, “Proteins at Atomic Resolution,” *Curr. Op. Struct. Biol.* **5**, 784-790 (1995).
2. G.M. Sheldrick, Z. Dauter, K.S. Wilson, H. Hope and L.C. Sieker, “The Application of Direct Methods and Patterson Interpretation to High-resolution Native Protein Data,” *Acta Crystallogr.* **D49**, 18-23 (1993).
3. H. Hauptman, “The SAS Maximal Principle,” IUCR Congress, Seattle abstract, No.BL.01, p.C-7 (1996).
4. T.R. Schneider and G.M. Sheldrick, “SHELXL: High-resolution Refinement,” *Meth. Enzym.* **276**, in press.
5. M.M. Teeter, S.M. Roe and N.H. Heo, “Atomic Resolution Crystal Structure of the Hydrophobic Protein Crambin at 130 K,” *J. Mol. Biol.* **230**, 292-311 (1993).
6. K.D. Watenpaugh, L.C. Sieker and L.H. Jensen, “The Structure of Rubredoxin at 1.2 Å Resolution,” *J. Mol. Biol.* **131**, 509-522 (1979).
7. J. Sevcik, Z. Dauter, V.S. Lamzin and K.S. Wilson, “Ribonuclease From *Streptomyces Aureofaciens* at Atomic Resolution,” *Acta Crystallogr.* **D52**, 327-344 (1996).
8. C. Frazao, C.M. Soares et al., “Ab Initio Determination of the Crystal Structure of Cytochrom C_6 and Comparison with Plastocyanin,” *Structure* **3**, 1159-1169 (1995).
9. L. Tong, T.C. Warren et al., “Crystal Structures of the Human p56^{lck} SH2 Domain in Complex with Two Short Phosphotyrosyl Peptides at 1.0 Å and 1.8 Å Resolution,” *J. Mol. Biol.* **256**, 601-610 (1996).
10. E. Pohl, A. Heine et al., “Structure of Octreotide, a Somatostatin Analogue,” *Acta Crystallogr.* **D51**, 48-59 (1995).
11. G. Prive and D. Eisenberg, “Direct Methods Phasing of a 450 Atom Structure,” IUCr Congress, Seattle abstract, MS01.03.05, p. C-17 (1996).
12. I. Karle, “Structure Determination of Helical Hairpin Peptides with up to 30 Residues per Asymmetric Unit,” IUCr Congress, Seattle abstract, MS02.03.01, p. C-51 (1996).
13. Z. Dauter, K.S. Wilson, L.C. Sieker, J.-M. Moulis and J. Meyer, “Zn and Ferubredoxins From *Clostridium Pasteurianum* at Atomic Resolution: The First High Precision Model of a ZnS_4 Unit in a Protein,” *Proc. Natl. Acad. Sci. USA*, at press.
14. Z. Dauter, C. Betzel, K.S. Wilson and K.V.D. Osten, “Subtilisin at Atomic Resolution,” ECM-16 Lund Abstract, M10a-04, p.64 (1995).