The benefits of atomic resolution

Zbigniew Dauter*†, Victor S Lamzin§ and Keith S Wilson*‡

After a long gestation, the elucidation of the crystal structures of proteins at atomic resolution is now maturing. The use of such data for both refinement and structure solution is advancing apace. The necessary technology is generally available, in terms of data collection, computing hardware and software. The structures appearing in the literature mainly relate to demonstration projects on native proteins. The importance of these alone is already obvious. Biologically significant results, in terms of ligand complexes and prosthetic groups, are just starting to emerge.

Addresses
*Department of Chemistry, University of York, Heslington, York YO1 5DD, UK
†e-mail: dauter@yorvic.york.ac.uk
‡e-mail: keith@yorvic.york.ac.uk
§European Molecular Biology Laboratory, c/o DESY, Notkestrae 85, D-22603 Hamburg, Germany; e-mail: victor@embl-hamburg.de

Current Opinion in Structural Biology 1997, **7**:681–688

http://biomednet.com/elecref/0959440X00700681

Abbreviations

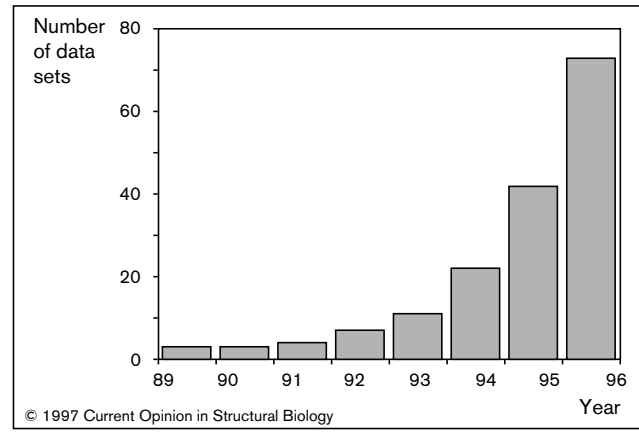
ADP	atomic displacement parameter
AR	atomic resolution
ARP	automated refinement for proteins
rmsd	root mean square deviation

Introduction

A couple of years ago, when a previous review on this topic was published [1], only a few structures of macromolecules were available at atomic resolution (AR). Data were known to have been collected for several others, but they were in process of refinement. It is sad, albeit perhaps unavoidable given the need to develop sound protocols for handling the detailed analysis of such large systems, that many of them have still the same 'not yet published' status. During the past years, the rate of recording AR structures has accelerated (Figure 1), and a number of analyses have been published. Especially, at the IUCr Congress in Seattle, it became clear that we should expect an explosion of AR structural analyses. Almost every issue of *Acta Crystallographica D* now contains papers describing a detailed analysis of a protein at AR or methodological advances in this area.

Let us define once more what is meant by AR. The data should extend to at least 1.2 Å, with 50% or more of the theoretically measurable reflections in the outer range having $I > 2\sigma(I)$ as suggested by Sheldrick [2] for a structure potentially to be solved using conventional direct methods. The advantages of AR data are not only quantitative, in that they lead to models with lower

Figure 1



The rapid increase in the number of atomic resolution data sets collected at the European Molecular Biology Laboratory, DESY, Hamburg. This is only a representative set: AR data are also being recorded at other synchrotron sites.

estimated standard uncertainties, but qualitative in that new features can be identified. Individual anisotropic atomic displacement parameters (ADPs) provide the best example.

The protocols used in refinement of protein models against X-ray data at 'ordinary' resolution, rarely exceeding 1.5 Å, are well established. These include several different approaches and options, such as minimization with geometrical restraints, use of energy terms, and noncrystallographic symmetry. Such protocols are not completely appropriate for handling AR structures, and the application of different procedures, comparable with those for small molecules, needs to be investigated. AR refinement still poses problems arising from the substantial disordered solvent component of the crystals and the inherent static/thermal disorder in parts of the protein itself.

We address the advantages of atomic resolution in protein structure analysis and present the most recent accomplishments in this field.

Data

Generating a complete list of reflection indices is necessary, but not sufficient! Without significant intensities, these are not informative. It is impossible to overstate the importance of data quality and completeness [3]. These are not 'hard' criteria. As a guide, the data in the outer shell should have a merging R factor of about 30%, corresponding to a $I/\sigma(I)$ ratio of ~2. This depends on the space group symmetry and the multiplicity of

the measurements. It is typical to get a lower R_{merge} in P1 than in I432 [4]. In practice, the $I/\sigma(I)$ ratio is a better criterion than R_{merge} [5], provided that the $\sigma(I)$ is correctly estimated. This is important both in data evaluation and refinement where $\sigma(I)$ is used to derive optimum weighting schemes. The estimation of valid $\sigma(I)$ is a nontrivial procedure. We have confidence in the error estimates used in the programs DENZO [6] and MOSFLM [7]. Nevertheless, this requires further investigation and analysis.

Synchrotron sources, efficient 2D detectors and cryogenic freezing are vital for AR data. Almost all recent structures are from frozen crystals, which give considerable increase in resolution and crystal lifetime. Carrying out the study on a single crystal is always preferable to avoid the problems of merging data from different samples.

Reduction of the mean ADP allows more accurate definition of atomic positions, which is crucial in areas of biological interest — active-site residues or bound ligands. The structure of the protein at ambient cellular temperature is what we wish to know, however, and this differs from the cryogenic structure. On freezing to ~100K, significant changes do occur, especially in the solvent and at the protein/solvent interface. The biological significance of these changes needs thorough investigation. An excellent review on cryogenic freezing has been published [8]. Two AR structures published recently by Hope (a pioneer in this field) and coworkers [9,10] are of interest.

Direct and related methods for structure solution

Traditional direct methods are based on three properties; positivity; atomicity; and atoms of equal height, which gives a simple geometric relation between the electron density and its square. This leads to mathematical relations between the structure factors, placing restrictions on phases once amplitudes are known. The relations are sufficiently powerful to provide a straightforward solution of the phase problem for small structures. The lack of AR data has precluded the application of direct methods to proteins, and one of the aims of attaining such data is to remove this barrier.

Unfortunately, the statistical probabilities of the triple product phase relationships (by far the most important for large structures) are inversely proportional to the square root of the number of atoms. For proteins, the individual probabilities are extremely low, and extending a small set of starting phases to generate the complete set has proved extremely difficult. In addition, criteria for identifying correct phase sets are generally not available.

With AR data, the structures of small proteins (up to 500 atoms) containing one or a few heavy atoms have proven to be tractable with conventional direct methods (the real size barrier for conventional direct methods is only ~200 atoms, and only a handful of larger structures have been solved in this way). The structures of avian pancreatic peptide [11] and rubredoxin [12] have both been solved. New structures have not been solved at the rate we anticipated, partly because alternative methods using the presence of the heavy atoms have been exploited. In addition, many such structures are solved by molecular replacement. We expect that *ab initio* phasing of AR (or even lower resolution) structures by maximum likelihood will evolve [13].

A second set of approaches is being developed in parallel in several laboratories, involving a combination of direct methods or Patterson algorithms with real space density selection of atoms. One of these is 'Shake-and-Bake' in the program SnB [14]. This has been applied successfully to several systems [15] and has recently been reviewed [16]. The largest of these systems is scorpion neurotoxin with 624 atoms at 0.96 Å [17•].

Several results using SHELXD-97 ('Half-Baked') have been presented by the Sheldrick group at a recent NATO Advanced Study Institute on 'Direct Methods for Solving Macromolecular Structures', Erice, May 1997. The largest structure solved using *ab initio* methods (i.e. no user intervention) is HiPIP with an Fe_4S_4 cluster, comprising two molecules of 84 residues in the asymmetric unit, 1260 protein atoms and 300 waters, at 1.2 Å resolution by E Parisini. The largest structure without heavy atoms is triclinic lysozyme, with 1001 protein atoms and 200 waters at 0.95 Å. The largest unknown structure solved is hirustasin by I Uson, with data collected by P Mittl to 1.2 Å but very noisy beyond 1.4 Å, with ~600 unique atoms. The largest structure solved by 'conventional' direct methods remains gramicidin A at 0.86 Å, with 317 unique atoms, solved by Langs [18].

Another approach is automated refinement for proteins (ARP) [19]. This is not a direct method in the strict sense, but an automated method for the extension of initial sets of phases or partial structures. ARP becomes extremely powerful with AR data. It can extend a single Fe model for rubredoxin to the complete structure [20]. A recent application is the solution of cytochrome *c-553* from the position of the Fe defined from the anomalous Patterson synthesis at 0.97 Å by S Benini, S Ciurli, WR Rypniewski and KS Wilson (Erice, 1997).

Methods such as ARP and other 'direct' or automated phasing procedures would greatly benefit from a good set of starting phases. Introduction of anomalous or MAD phases into the starting set certainly enhance their performance, and such combinations of experimental and direct methods will be of great use. Several advances in these approaches were reported at the CCP4 meeting in York in January 1997 [21].

Refinement

The great advantage of AR lies in the ratio of experimental observations to refined parameters. At 1 Å resolution, there are about five observations for every parameter, even with an anisotropic model. Provided the data are complete and significant, least-squares or maximum-likelihood refinement is well determined. Errors for each individual atomic parameter can be estimated rigorously by inversion of the least-squares matrix. Coordinate errors are as low as 0.02 Å for carbon atoms and down to 0.002 Å for heavier atoms. Errors are larger for the disordered or flexible parts of the structure.

Small-structure refinement programs are, in principle, capable of handling protein anisotropic models but require the implementation of an extensive set of restraints. SHELXL [22••] has been extensively adapted to treat the special properties of macromolecules. These particularly relate to disorder, the need to impose restraints on some regions, a model for the solvent continuum and appropriate techniques of dealing with solvent. We expect other programs, for example REFMAC [23•], to be used in coming years. Refinement, including AR models, was discussed at the CCP4 meeting in York, in April 1995 [24••].

What do such refinements reveal? The high degree of accuracy and anisotropic representation of ADP are usually accompanied by a 5% drop in R factor and R_{free}, which leads to a decrease in the noise level in Fourier syntheses. This allows the modelling of features normally unidentifiable in macromolecular structures. Flexible regions can be modelled with confidence with partial occupancies. Hydrogen atoms ride at calculated positions: the majority of them can also be seen in the difference Fourier maps. At the other extreme, the positions of heavier atoms such as the Fe_4S_4 clusters in ferredoxins can be defined at an accuracy matching that of inorganic model compounds. Water (solvent) molecule occupancies can be estimated in parallel to their ADPs and alternative hydrogen-bond networks analyzed.

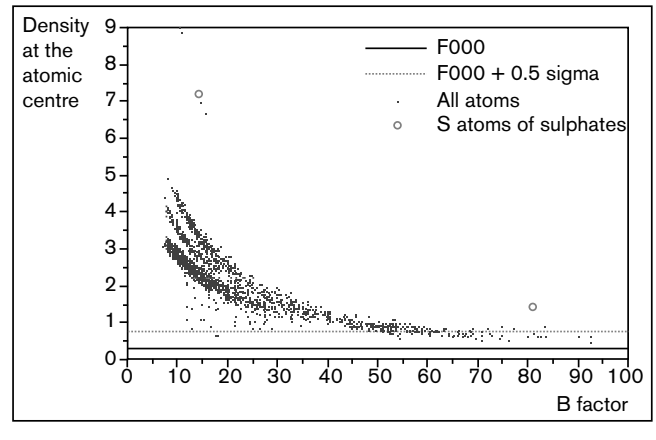
The analysis of water structure and a brief discussion of structures solved recently is provided in the next sections. More extended descriptions of protocols for refining AR structures have been provided by several authors [1,22••,25•,26••].

Water structure

Water is the curse and blessing of the protein crystallographer. Crystals contain about 50% solvent, which means that the protein experiences an environment not very different from that *in vivo*, and that ligands can diffuse through the water channels to bind to the protein. The disadvantage is the disorder, which is hard to model and is to a large extent responsible for the weakness of the high-resolution data. In addition, the sensitivity of protein crystals to X-rays is largely due to the effects of secondary damage by the diffusion of ions and radicals in these channels. The latter problem is diminished by cryogenic data collection.

The disorder problem is by no means eliminated by freezing. There is a first shell of waters around the surface, many but not all of which have close to unit occupancy. Part of this shell and certainly the next shells have only partial occupancy, reflecting the presence of overlapping hydrogen-bonding networks. Further away from the surface of the protein, the situation descends into disorder with no individual solvent sites visible. The model for the disordered region is currently based on Babinet's principle [27]. This models a uniform sea of density away from the ordered region and a contrast-matching transition at the interface. That such a model is reasonable is reflected in the substantial reduction in R factor at low resolution. The R factor still remains surprisingly high, however, often more than 30% in the lowest resolution ranges. The physical significance of this remains unclear and needs further investigation.

Figure 2



A plot of density at the atomic centre versus B factor for RNase Sa. There is a clear distinction between the C, N and O atoms, highlighted by the three curves, which is especially significant at low ADP. Reproduced with permission from [25•].

The 'ordered' region of the solvent also remains a problem, and this one factor that is delaying the rapid completion of the refinement of AR structures. We need at least a semiobjective means of identifying the huge number of water sites and their occupancies. To do this using subjective computer graphics analysis is not tractable if the increasing number of structures is to be satisfactorily analyzed and compared. Global indicators such as R_{free} are not of great help in defining the meaningfulness of individual low occupancy/high occupancy ADP water sites. For the highest resolution structures, a plot of density at the atomic centre against ADP value [25•] is promising for occupancy evaluation (Figure 2).

One point has become very clear: the low-resolution data are vital. These underpin the contrast between protein and solvent regions, and, even if the R factor is high, contribute most importantly to the electron density maps. Omitting data below 10 Å (or by some mispractitioners, below 6 Å) lowers the R factor (which is not the object of a crystallographic analysis — the best structure is more important) at the expense of map quality. Paradoxically, for the best solvent model, both the lowest resolution data and AR data are required.

Structure validation and libraries

AR structures potentially provide sufficiently accurate structural models to evaluate whether the target libraries used for chemical and stereochemical parameters, and the relative weights accorded these parameters are appropriate. The chemical bond lengths and angles are generally based on those defined by Engh and Huber [28] or modified implementations [29,30]. This library is used by all major refinement packages. The stereochemical restraints or force fields used vary and can often be adjusted at the whim of the user. This is prone to abuse. The small number of AR structures has already made a contribution. For ribonuclease Sa at 1.1 Å [25•] and cutinase at 1.0 Å [26••], detailed analyses of stereochemistry were carried out. Wilson and coworkers [31••] have analyzed a set of eight AR structures. While the sample number is small, several clear conclusions have already emerged.

First, the importance of accurate calibration of wavelength and detector position is evident, as ignoring this leads to errors of magnitude comparable with the accuracy of derived interatomic distances.

Second, although the Engh and Huber [28] set seems generally valid, small details may change if a larger number of structures are analyzed. The present library does not address the problem of different protonation states of moieties such as carboxylates.

Third, there is a deviation of many peptide angles from planarity by up to 20° with an rmsd of ~6°. This reflects real deviations of the peptide from planarity, in keeping with the distribution for small peptides from the Cambridge Structural Database [32]. Analysis of about 200 'good' structures from the PDB at resolution less than AR showed a much narrower distribution with an rmsd of ~3°, suggesting too tight restraints were conventionally applied for ψ [33].

Fourth, the angles in the sidechains are much more tightly distributed around the preferred rotamer conformations in the AR structures. This is a useful validation tool, as these angles are not (usually) restrained during the refinement.

Finally, a comparable conclusion comes from the Ramachandran plot [34], where residues in the AR structures cluster very tightly in the 'allowed' regions of (ϕ,ψ) conformational space, suggesting that the preferred regions should be redefined. The Ramachandran plot remains the best validator of overall fold correctness. Even for a protein with essentially no secondary structural elements, such as rubredoxin [35•], at AR, the Ramachandran plot has all residues clustered in the allowed core regions.

Examples

During the past two years, several AR structures have been published (Table 1). These still are largely restricted to small protein, often metalloproteins. Structures of several more representative proteins with molecular mass greater than 20 kDa have been reported at various meetings. We will mention a few examples and will try to emphasize relevant biological implications. This is biased towards research at Hamburg, but AR data are being recorded at other synchrotron sites.

Cutinase

Cutinase is a 22 kDa enzyme responsible for the degradation of cutin. Its structure had previously been refined to 1.6 Å [36]. The 1.0 Å data allowed refinement with an anisotropic model to a final R factor of 9.4% [26••]. The reduction of the mean coordinate error by one order of magnitude to 0.021 Å compared with the 1.6 Å analysis produced substantial improvements in map and model quality.

Table 1

A selection of recently published atomic resolution structures.

Protein	Asymmetric unit (kDa)	V_M (Å^3 Da^{-1})	Resolution (Å)	Reference
BPTI*	6.5	1.9	1.1	[9]
Concanavalin A	25.0	2.4	1.2	[10]
RNase Sa	21.1	2.3	1.15	[25•]
Cutinase	22.0	2.0	1.0	[26••]
Rubredoxin Cp	6.0	2.2	1.1	[35•]
Pheromone Er-1	4.4	1.5	1.0	[38••]
Crambin	4.7	1.8	0.83	[41]
γB-crystallin	21.0	1.8	1.18	[43]
p56^{lck} SH2 domain	12.0	2.1	1.0	[44]
α-conotoxin	1.6	1.4	1.1	[45]

*BPTI, bovine pancreatic trypsin inhibitor.

Extensive modelling of disorder became possible for about ten residues, one of these in the active site. Important information has been gained on the flexibility of the active-site flap. Most importantly, the protonation state of the active-site histidine is clear: there is excellent density for this hydrogen. Cambillau and coworkers [26••] emphasize another aspect of AR structures: it is possible to distinguish C, N and O atoms from the density and ADPs alone, giving absolute definition of orientation for histidine, asparagine and glutamine.

Cutinase is the largest AR structure published to date. It has allowed a thorough analysis of the protein stereochemistry. Some potential deviations in bond distances from the target values [28] have been discussed. The reliability of the observed deviations is validated by means of the associated estimated error. This provides a further example that will contribute to the challenge of finding new and more appropriate parameters for protein stereochemistry.

Two cytochrome structures

In 1995, Sheldrick and coworkers [37] reported that the structure of a novel cytochrome c_6 had been solved, using Patterson methods and SHELXS peak enhancement, at 1.2 Å resolution [37]. This was the first *ab initio* structure determination of a new protein. Cytochrome c_6 gave the largest drop in R_{free}, 8%, observed to date on going anisotropic. The refined model provides the most accurate view of a haem group in a protein. If we are to understand and probe the effect of local environments on the properties of the haem prosthetic group, we must have structural data at the highest possible resolution. The derived haem distances agree very well with those available from the Cambridge Structural Database.

Recently, in Hamburg, another novel cytochrome structure has been determined, that of cytochrome *c-553*, by S Benini, S Ciurli, WR Rypniewski and KS Wilson (Erice, 1997). The crystal was frozen and data extended to 0.97 Å. The position of the iron atom was determined from the anomalous Patterson synthesis, and the complete structure was developed in a series of steps using ARP. This is the first structure of a cytochrome from Gram-positive bacteria. The high quality of the structure makes it possible to examine the basis of the unusually low reduction potential of *c-553*, and in particular the mechanism of the stabilization of the 3+ charge on the Fe ion.

Pheromone Er-1 from *Euplotes raikovi*

The pheromone Er-1 from *E. raikovi*, a 40 amino acid peptide, has provided diffraction data to 1.0 Å and is extremely tightly packed in the cell with a specific packing volume (V_M) of 1.53 $Å^3\,Da^{-1}$. The structure solution has been attempted using a barrage of techniques: molecular replacement (MR), direct phasing with 'Shake-and-Bake' (SnB), isomorphous replacement and multiple-wavelength anomalous scattering [38••]. The structure was actually first solved using MR with considerable difficulty because of the tight packing. It was subsequently the first structure to be solved using SnB without direct human intervention.

Rubredoxin and ferredoxin

Rubredoxin was the first protein to which a true least-squares refinement was applied [39]. It remains a paradigm for accurate protein structure analysis. The function of rubredoxin remains to be fully defined, but the role of the polypeptide chain is to impose nonaqueous chemistry on the FeS_4 cluster and fine tune its electron-transfer redox properties. An analysis of the *Clostridium pasteurianum* Fe-substituted (1.1 Å) and Zn-substituted (1.2 Å) rubredoxin has been published recently [35•]. The accuracy of the metal-cluster distances, with bond errors of ~0.003 Å, parallels that available for model compounds. The analysis is currently being followed up by a study of the oxidized and reduced forms of the Fe wild-type protein.

A 0.95 Å analysis at 110K of the ferredoxin from *Clostridium acidiurici* reveals the geometries of the two Fe_4S_4 clusters with an unparalleled accuracy (Z Dauter, KS Wilson, LC Sieker, J Meyer, J-M Moulis, unpublished data). The problem is to relate the small but statistically highly significant distortions of the clusters from ideal symmetry, and the nature of their surroundings, to the redox potentials of this and other related ferredoxins. These results put the onus on theorists to rationalize the redox properties in terms of the environment of the clusters in these two metalloproteins.

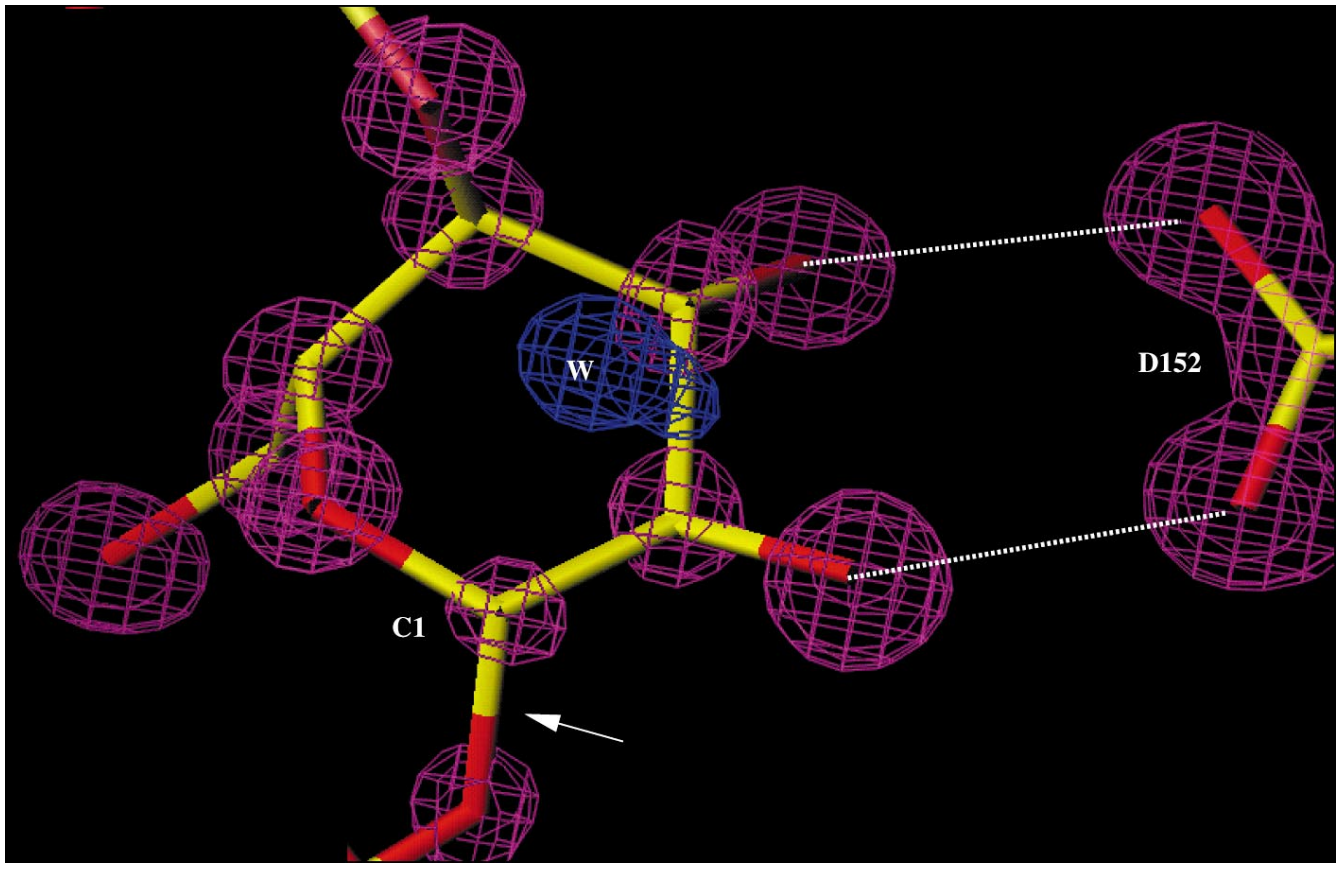
Cellulase

D Guerin and PM Alzari (personal communication) have determined the structure of a complex of bacterial endoglucanase CelA with a D-glucosyl residue bound to subsite –1 of the active site (Figure 3). The high-resolution structure of this complex, between an inactive mutant of the endoglucanase with cellopentaose, is being refined at 0.94 Å resolution. The R factor is below 10% for all (~200,000) independent reflections. The conformation of the sugar in the –1 site is extremely well defined and is distorted from the normal $^4C_1$ chair. Above the sugar ring, a water molecule is well positioned to serve as a nucleophile to hydrolyze the glycosidic linkage. This crystal structure will contribute to the elucidation of the details of the reaction mechanism and will serve as an excellent model system to analyze protein–carbohydrate interactions and the nature of the conformational changes during the formation of the enzyme–substrate complex.

Triclinic lysozyme

This is another system that has been known to diffract to AR for many years. Walsh *et al.* (M Walsh, TR Schneider, LC Sieker, Z Dauter, VS Lamzin, KS Wilson, unpublished data) have refined the room temperature and cryogenic structures at 0.95 Å to R factors below 10%. The solvent structures reveal differences between the frozen and ambient temperature crystals. The protonation states of

Figure 3



The complex of bacterial endoglucanase CelA, with a D-glucosyl residue bound to subsite –1 of the active site (DM Guerin, PM Alzari, private communication). The density for the water molecule (W) positioned to serve as a nucleophile is shown in purple. The scissile glycosidic linkage is marked with an arrow. Figure courtesy of DM Guerin and PM Alzari.

the carboxylate groups were amenable to analysis; this involved a combination of direct observation of the difference Fourier peaks and the C–O bond lengths.

Protein receptor–phosphate complexes

Two receptor–phosphate complexes have recently been solved at 0.98 and 1.05 Å resolution [40]. A point of biological interest is the presence of a very short hydrogen bond, less than 2.45 Å, characterized by a low-energy barrier. The existence of such an interaction had been previously postulated from studies of model compounds in the gas phase, but the accuracy of the AR analysis is sufficient to provide unequivocal experimental evidence in support of this.

Crambin

Crambin is the protein that has been structurally characterized in the most detail, even though its function remains unclear [41]. During a visit to EMBL Hamburg in 1996, M Teeter recorded data to 0.67 Å resolution. After averaging over 42 peptides, the difference maps clearly showed features characteristic of the deformation of electron density from the spherical atom approximation normally accepted in crystallography: the bonding electrons [42]. Knowledge of fine details of the electronic structure of proteins is the next step towards a deeper understanding of their function and chemical interactions.

Conclusions

Clearly the first surge of 3D structural information, even from a medium-resolution analysis, presents a major breakthrough in rationalizing many of the properties of a protein. Indeed, because of the limitations of crystal quality, many analyses will remain limited to such resolution. This does, however, severely restrict the biological questions that can be answered. To really probe the chemistry of the active site of enzymes, of metal prosthetic groups, and of ligand binding via hydrogen bonds and other weak interactions, the most accurate models possible are required. For small molecules, a coordinate accuracy of ~0.002 Å is routine. With synchrotron radiation, detectors and cryogenic freezing, an accuracy approaching this level is being obtained for a rapidly increasing number of proteins. Structures published to date tend to be small, but data have been collected to AR for systems as large as LADH, neuraminidase and subtilisins.

What advances can be expected? To date, with few exceptions, the AR structures have been of native enzymes or of proteins as pilot studies. The next problems to be addressed relate to protein function and require structures of complexes with substrates and/or inhibitors. AR studies of a series of ligands will be of importance for the development of modelling techniques.

In terms of techniques, more powerful synchrotron radiation sources will make AR data collection more tractable on larger structures from smaller crystals. Major advances can also be expected in the application of direct methods and associated automated procedures for phase extension and model building.

Finally, there will soon be a sufficiently large database of AR protein structures to construct a protein-based library of stereochemical parameters; it may take a little longer until there is a good enough database to suggest corrections to the Engh and Huber [28] set, an allowance for different protonation states and variation of properties with cryogenic freezing.

Acknowledgements

We thank the European Community for partial support of this work through the BIOTECHNOLOGY programme contracts CT92-0524 and CT96-0189. We also acknowledge Christoph Hermes and Roelof v Silfhout from the EMBL Hamburg instrumentation group for their continuous efforts in developing and maintaining the beam lines that have allowed collection of much of the AR data described in this review.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dauter Z, Lamzin VS, Wilson KS: **Proteins at atomic resolution.** *Curr Opin Struct Biol* 1995, **5**:784-790.

2. Sheldrick GM: **Phase annealing in *SHELX-90:* direct methods for larger structures.** *Acta Crystallogr A* 1990, **46**:467-473.

3. Dauter Z: **Data collection strategy.** *Methods Enzymol* 1997, **276**:326-344.

4. Diederichs K, Karplus PA: **Improved *R*-factors for diffraction data analysis in macromolecular crystallography.** *Nat Struct Biol* 1997, **4**:269-275.

5. Weiss MS, Hilgenfeld R: **On the use of the merging *R* factor as a quality indicator for X-ray data.** *J Appl Crystallogr* 1997, **30**:203-205.

6. Otwinowski Z, Minor W: **Processing of X-ray diffraction data collected in oscillation mode.** *Methods Enzymol* 1997, **276**:307-326.

7. Leslie AGW: **Recent changes to the MOSFLM package for processing film and image plate data.** *CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. Warrington, UK: Daresbury Laboratory; 1991.

8. Garman EF, Schneider TR: **Macromolecular cryocrystallography.** *J Appl Crystallogr* 1997, **30**:211-237.

9. Parkin S, Rupp B, Hope H: **Structure of bovine pancreatic trypsin inhibitor at 125K: definition of carboxyl-terminal residues Gly57 and Ala58.** *Acta Crystallogr D* 1996, **52**:18-29.

10. Parkin S, Rupp B, Hope H: **Atomic resolution structure of concanavalin A at 120 K.** *Acta Crystallogr D* 1996, **52**:1161-1168.

11. Woolfson MM, Yao J-X: **On the application of phase relationships to complex structures. XXX. *Ab initio* solution of a small protein by SAYTAN.** *Acta Crystallogr A* 1990, **46**:409-413.

12. Sheldrick GM, Dauter Z, Wilson KS, Hope H, Sieker LC: **The application of direct methods and Patterson interpretation to high-resolution native protein data.** *Acta Crystallogr D* 1993, **49**:18-23.

13. Bricogne G: **Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives.** *Acta Crystallogr D* 1993, **49**:37-60.

14. Miller R, Gallo SM, DeTitta GT, Khalak HG, Weeks CM: **SnB: crystal structure determination via Shake-and-Bake.** *J Appl Crystallogr* 1994, **27**:613-621.

15. Miller R, DeTitta GT, Jones R, Langs DA, Weeks CM, Hauptman HA: **On the application of the minimal principle to solve unknown structures.** *Science* 1993, **259**:1430-1433.

16. Ealick SE: **Now we're cooking: new successes for Shake-and-Bake.** *Structure* 1997, **5**:469-472.

17. • Smith GD, Blessing RH, Ealick SE, Fontecilla-Camps JC, Hauptman HA, Housset D, Langs DA, Miller R: **The *ab initio* structure determination of a scorpion neurotoxin.** *Acta Crystallogr D* 1997, in press.
The largest structure solved *ab initio* to date using Shake-and-Bake. It contains > 600 atoms.

18. Langs DA: **Three dimensional structure at 0.86 Å of the uncomplexed form of the transmembrane ion channel peptide gramicidin.** *Science* 1988, **241**:188-191.

19. Lamzin VS, Wilson KS: **Automated refinement for protein crystallography.** *Methods Enzymol* 1997, **277**:269-305.

20. Lamzin VS, Wilson KS: **ARP at atomic resolution. *Ab initio* structure determination of rubredoxin.** Abstract O42 of the 4th European Workshop on Crystallography of Biological Macromolecules, 1995 May 21–25, Erice.

21. Wilson K, Davies G, Bailey S, Ashton A (Eds): *Recent Advances in Phasing. Proceedings of the CCP4 Study Weekend.* Warrington, UK: Daresbury Laboratory; 1997.

22. •• Sheldrick GM, Schneider TR: **SHELXL: high-resolution refinement.** *Methods Enzymol* 1997, **277**:319-343.
A definitive review of the SHELXL programme, in particular covering its specific options for macromolecules.

23. • Murshudov GN, Vagin AA, Dodson EJ: **Refinement of macromolecular structures by the maximum-likelihood method.** *Acta Crystallogr D* 1997, **53**:240-255.
A description of the advantages of the maximum-likelihood approach to refinement. REFMAC is one of several programs currently being developed. This paper particularly describes the use of weighted coefficients for map calculations. Maximum-likelihood procedures are expected to make major contributions to AR refinement in the coming years.

24. •• Dodson EJ, Kleywegt G, Wilson, KS: **Report of a workshop on the use of statistical validators in protein X-ray crystallography.** *Acta Crystallogr D* 1996, **52**:228-234.
A general report of the York workshop, summarizing the views of a number of experts in the protein data quality and refinement fields. The importance of good data to the highest resolution and the proper estimation of standard uncertainties is strongly supported.

25. • Sevcik J, Dauter Z, Lamzin VS, Wilson KS: **Ribonuclease from *Streptomyces areofaciens* at atomic resolution.** *Acta Crystallogr D* 1996, **52**:327-344.
One of first AR structures to be published, with 22 kDa (two molecules) in the asymmetric unit. A refinement protocol involving ARP and SHELXL is described. Various novel procedures, such as the plot of density at the atomic centre against temperature factor, are introduced.

26. Longhi S, Czjzek M, Lamzin V, Nicolas A, Cambillau C:
•• **Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis.** *J Mol Biol* 1997, **268**:779-799.
The largest AR structure published to date. This paper contains a very thorough analysis of cutinase. The authors provide some comparison of the observed parameters with the library values. The mobility of the active-site residues and of the flexible lid are seen to be of direct biological importance.

27. Langridge R, Marvin DA, Seeds WE, Wilson HR, Hooper CW, Wilkins MHF, Hamilton LD: **The molecular configuration of deoxyribonucleic acid. II. Molecular models and their Fourier transforms.** *J Mol Biol* 1960, **2**:38-64.

28. Engh RA, Huber R: **Accurate bond and angle parameters for X-ray protein structure refinement.** *Acta Crystallogr A* 1991, **47**:392-400.

29. Priestle JP: **Stereochemical dictionaries for protein structure refinement and model building.** *Structure* 1994, **15**:911-913.

30. Lamzin VS, Dauter Z, Wilson KS: **Dictionary of protein stereochemistry.** *J Appl Crystallogr* 1995, **28**:338-340.

31. Wilson KS, Butterworth S, Dauter Z, Lamzin VS, Walsh M,
•• Wodak S, Pontius J, Richelle J, Vaguine A, Sander C *et al.*: **Who checks the checkers? Four validation tools applied to eight atomic resolution structures.** *J Mol Biol* 1997, in press.
A detailed analysis of eight AR structures from data collected at EMBL Hamburg. The importance of accurate measurement of cell dimensions is emphasized. Several conclusions regarding improved target distributions and appropriate weights of stereochemical parameters are drawn.

32. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM: **Stereochemical quality of protein structure coordinates.** *Proteins* 1992, **12**:345-364.

33. MacArthur MW, Thornton JM: **Deviations from planarity of the peptide bond in peptides and proteins.** *Protein Eng* 1996, **8**:217-224.

34. Ramachandran GN, Ramakrishnan C, Sasisekharan V: **Stereochemistry of polypeptide chain conformations.** *J Mol Biol* 1963, **7**:95-99.

35. Dauter Z, Wilson KS, Sieker LC, Moulis J-M, Meyer J: **Zinc- and**
• **iron-rubredoxins form *Clostridium pasteurianum* at atomic resolution: a high-precision model of a ZnS_4 coordination unit in a protein.** Proc Natl Acad Sci USA 1996, **93**:8836-8840.
This paper provides perhaps the first AR description of FeS_4 and ZnS_4 clusters in a protein, properly refined with a full anisotropic model. The clusters play the key role in the biological function. The atoms in the cluster are defined to an accuracy about 0.003 Å. This is close to that of inorganic model compounds.

36. Martinez C, De Geus P, Lauwereys M, Matthyssens G, Cambillau C: ***Fusarium solani* cutinase is a lipolytic enzyme with a catalytic serine accessible to the solvent.** *Nature* 1992, **356**:615-618.

37. Frazao C, Soares CM, Carrondo MA, Pohl E, Dauter Z, Wilson KS, Hervas M, Navarro JA, De la Rose MA, Sheldrick GM: ***Ab initio* determination of the crystal structure of cytochrome *c6* and comparison with plastocyanin.** *Structure* 1995, **3**:1159-1169.

38. Anderson DH, Weiss MS, Eisenberg D: **A challenging case**
•• **for protein crystal structure determination: the mating pheromone Er-1 from *Euplotes raikovi*.** *Acta Crystallogr D* 1996, **52**:469-480.
An excellent description of the challenges posed by the structure solution of the tightly packed pheromone Er-1 structure. The advantages and limitations of the various procedures are assessed. The refinement of the structure is described in some detail.

39. Watenpaugh KD, Sieker LC, Jensen LH: **Crystallographic refinement of rubredoxin at 1.2 Å resolution.** *J Mol Biol* 1980, **138**:615-633.

40. Wang Z, Luecke H, Yao N, Quiocho FA: **A low energy short hydrogen bond in very high resolution structures of protein receptor–phosphate complexes.** *Nat Struct Biol* 1997, in press.

41. Stec B, Zhou R, Teeter MM: **Full-matrix refinement of the protein crambin at 0.83 Åand 130K.** *Acta Crystallogr D* 1995, **51**:663-681.

42. Teeter M, Stec B: **Anisotropic vibrational analysis of the small protein crambin.** Abstract MS11.06.05 of the XVII Congress IUCr, 1996, Seattle.

43. Kumaraswamy VS, Lindley PF, Slingsby C, Glover ID: **An eye-lens protein–water structure: 1.2 Å resolution structure of γB-crystallin at 150K.** *Acta Crystallogr D* 1996, **52**:611-622.

44. Tong L, Warren TC, King J, Betageri R, Rose J, Jakes S: **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* 1996, **256**:601-610.

45. Hu SH, Gehrmann J, Guddat LW, Alewood PF, Craik DJ, Martin JL: **The 1.1 Å crystal structure of the neuronal acetylcholine receptor antagonist, α-conotoxin PnIA from *Conus pennaceus*.** *Structure* 1996, **4**:417-423.