Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Quasi-Laue neutron-diffraction study of the water arrangement in crystals of triclinic hen egg-white lysozyme

Cécile Bon,[a,b] Mogens S. Lehmann[b]* and Clive Wilkinson[c]

[a]Université Joseph Fourier, F-38400 Saint Martin d'Hères, France, [b]Institut Laue–Langevin, F-38042 Grenoble, France, and [c]EMBL Outstation, F-38042 Grenoble, France

Received 22 October 1998
Accepted 29 March 1999

PDB Reference: triclinic hen egg-white lysozyme, 1lzn.

Triclinic crystals of lysozyme, hydrogen–deuterium exchanged in deuterated solvent, have been studied using neutron quasi-Laue techniques and a newly developed cylinder image-plate detector. The wavelength range employed was from 2.7 to 3.5 Å, which gave 9426 significant reflections [$F \geq 2\sigma(F)$] to a resolution limit of 1.7 Å. The deuteration states of the H atoms in the protein molecule were identified, followed by an extensive analysis of the water structure surrounding the protein. The final R factor was 20.4% ($R_{\rm free}$ = 22.1%). In total, the 244 observed water molecules form approximately one layer of water around the protein with far fewer water molecules located further away. Water molecules covering the apolar patches make tangential layers at 4–5 Å from the surface or form C—H···O contacts, and several water-molecule sites can be identified in the apolar cavities. Many of the water molecules are apparently orientationally disordered, and only 115 out of the 244 water molecules sit in mean single orientations. Comparison of these results with quasi-elastic neutron scattering observations of the water dynamics leads to a picture of the water molecules forming an extended constantly fluctuating network covering the protein surface.

1. Introduction

Much of the biological activity of proteins takes place in an aqueous environment, and solvent effects are frequently invoked in order to understand protein stability, function, interactions and internal dynamics. A full comprehension of the role of these mechanisms clearly depends on the available information, both structural and dynamic, and one of the techniques of choice is single-crystal diffraction, which can supply time- and space-averaged atomic positions for the crystalline part of the solvent, which has sufficient long-range order to give a coherent diffraction signal.

Atomic positions can be determined using either X-rays or neutrons, but while X-rays will predominantly give information about electron-rich atoms (non-H atoms), hydrogen – and especially the deuterium isotope – is as visible as most other atoms when neutrons are used. It is, therefore, quite obvious that for the detailed study of specific H atoms or water molecules, neutrons should be preferred as radiation, possibly preceded by a high-resolution X-ray investigation of the other parts of the structure.

Unfortunately, macromolecular crystallography employing neutrons is hampered by the inherent low flux of even the best neutron source and also by the small size of most protein crystals. In a number of studies performed so far (*e.g.* Schoenborn, 1969; Kossiakoff & Spencer, 1980, 1981; Wlodawer & Sjölin, 1983; Wlodawer *et al.*, 1984; Savage *et al.*,

1987; Lehmann & Stansfield, 1989; Roth *et al.*, 1989; Cheng & Schoenborn, 1990; Eisenstein *et al.*, 1991; Timmins *et al.*, 1992; Kossiakoff *et al.*, 1992; Bouquiere *et al.*, 1993; Langan *et al.*, 1999), it has been shown that much information can be obtained employing the special nature of neutron-scattering cross sections.

Improvements of the neutron-scattering method have therefore been proposed, employing either a large band of long-wavelength neutrons from a steady-state reactor cold source (Wilkinson & Lehmann, 1991; Schoenborn, 1992*a*) or the white beam of a spallation source (Schoenborn, 1992*b*). These Laue or quasi-Laue techniques increase the number of simultaneously active reflections, and should thus lead to a gain in measurement speed.

The choice of a cold rather than a thermal steady-state source is due to the large divergence of the neutron beam and the large crystal dimensions required, which lead to Laue-spot sizes on the detector which are typically in the millimetre range. The only way to separate the individual peaks is therefore to use the very long wavelength neutrons from a cold source, where the wavelength typically ranges from just under 3 Å to around 8 Å. A consequence of this choice is a large diffraction range, with the highest order data being observed at 2θ scattering angles of 150° or more, and the detector surface must consequently be very extended. A cylindrical detector was therefore built in a collaboration between the EMBL outstation in Grenoble and the Institut Laue–Langevin (Cipriani *et al.*, 1996), and recently a first report of the use of this detector for the study of tetragonal lysozyme crystals from hen egg-white to a resolution of 2 Å has appeared (Niimura *et al.*, 1997).

Many of the atomic or near-atomic resolution studies undertaken previously (and cited above) have concentrated on the location of one or a few H atoms essential for an enzymatic process, while fewer have dealt with the complex problem of the water arrangement in the cell, which covers the whole range from perfect order to complete randomness (see, for example, Cheng & Schoenborn, 1990; Kossiakoff *et al.*, 1992; Bouquiere *et al.*, 1993). With the advent of this new technique and detector, higher resolution data should be obtainable, and this can help in analysing both the well ordered water molecules and those that are either disordered or undergo constant rearrangement. In this paper, we report the study of the water structure of triclinic hen egg-white lysozyme using data to a resolution of 1.7 Å.

2. The detector, the neutron beam and the measurement technique

The large detector is based on neutron-sensitive image plates. These are used for X-rays (Amemiya *et al.*, 1988; Templer, 1991) and are thin flexible sheets coated with a layer of photostimulable phosphor. The material is most commonly BaFBr doped with Eu^{2+} ions, and when this is irradiated with X-rays or γ-rays, electrons liberated by ionizing Eu^{2+} to Eu^{3+} are trapped in states just below the conduction band to form colour centres. The decay time of the trapped electrons is long, so the radiation falling on the plate can be stored for later read-out by photo-stimulation with visible laser radiation. A photomultiplier is used to record the emitted blue light as a function of position, and an additional erasure is performed afterwards with an intense source of visible light.

A neutron-to-γ-ray/X-ray/electron converter makes the image plates sensitive to neutrons. Gadolinium is a good candidate with a very strong (n,γ) resonance to thermal neutrons (Rauch *et al.*, 1967; Backlin *et al.*, 1982) producing both a wide range of keV γ-rays and a cascade of conversion electrons. Both can stimulate the phosphor. The conversion process is best performed by mixing a gadolinium compound into the phosphor, in which case the conversion electrons play an important role. As the electrons have a short range in the material, this can give a resolution of better than 100 µm (Rausch *et al.*, 1992; Cipriani *et al.*, 1994; Hofmann & Rausch, 1995).

The detector lay-out (Cipriani *et al.*, 1996) is shown in Fig. 1. The sample is placed in the middle of the cylinder and the beam is orthogonal to the axis. The beam enters the detector through a small hole on one side and exits diametrically opposite to be trapped by a beam stop. The diffracted neutrons pass through the aluminium cylinder to be absorbed in the image plate which is placed at the outside of the cylinder. After exposure, this is read off much like an Edison phonograph – the cylinder is rotated at high speed while the reader slowly moves from one side to the other. Finally, any remaining signal is erased using a lamp placed at the side of the detector (not shown in the figure).

The detector efficiency varies with the wavelength and the density of gadolinium, and efficiencies from 20–70% can be attained (Rausch *et al.*, 1992; Niimura *et al.*, 1994; Hofmann & Rausch, 1995). At the time of the detector construction neutron image-plates were not yet commercially available, but

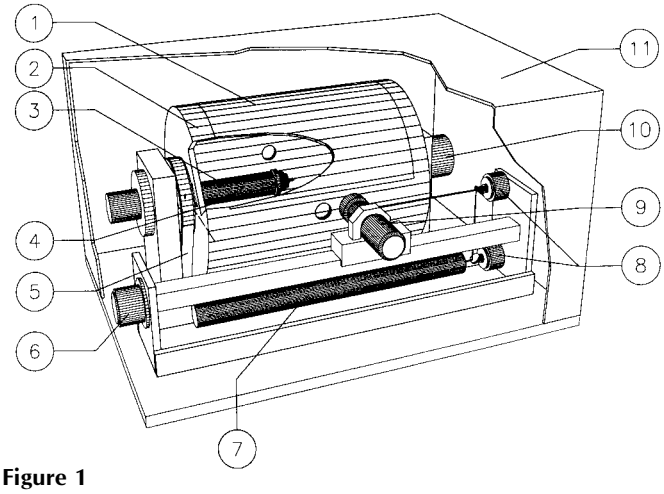

Figure 1
Schematic diagram of the cylinder detector (Cipriani *et al.*, 1996). 1, image plate on drum. 2, drum. The diameter is 31.8 cm, which gives a circumference of 1 m. The length is 40 cm. 3, sample holder. 4, crystal. 5, transmission belt to drive drum. The motor is under the table. 6, carrier for reading head with photomultiplier. 7, He–Ne laser. 8, mirrors for bringing the laser light to the reader head. 9, reader head with photomultiplier (parts 7, 8 and 9 were later replaced by a single unit including a diode laser). 10, encoder for drum rotation. 11, cover.

some specimens fabricated by Fuji Films (Niimura *et al.*, 1994) were made available. They contain a 25% molar ratio of Gd_2O_3 and the layer of photostimulable phosphor has a thickness of 140 μm.

The dimension of the instrument was based on the observation that in the case of neutrons the effect of divergence will define the size of the Laue spots and thus be the dominating factor in setting the peak-to-background ratio. This is a consequence of the large divergence of a beam coming from a neutron guide. For a Ni-coated neutron guide, the full width at half height of the beam divergence is 0.0093 radians for 3 Å neutrons, and the spot size for a point crystal will thus be approximately $2 \times 0.0093 \times L$, where L is the sample-to-detector distance. As soon as this value is larger than the real crystal size, little is gained in peak-to-background ratio by moving the detector further away, and the condition for L is therefore

$$2 \times 0.0093 \times L > r,$$

where r is the dimension of the crystal. Assuming r to be at most 2 mm, we find $L > 100$ mm. The actual choice of $1000/(2\pi) = 159$ mm for the cylinder radius is, therefore, very suitable.

A total of four neutron-sensitive image plates of dimensions 200×400 mm were used to cover 80% of the detector surface. A strip of 20 cm all along the cylinder in the back-scattering region was left without image plates and this has been used for tests of other plates. The pixel size was set to 200×200 μm, and rotating the drum at 800 rev min^{-1} leads to a total read-out time of 3 min, giving a total of 8 Mpixels of data per recording.

The use of long-wavelength neutrons is, as discussed above, dictated by the large spot size on the detector. However, the full range of neutrons coming from the cold source is too large for most unit cells, and a limited band is therefore normally obtained by using a multilayer monochromator (Høghøj *et al.*, 1996), which can be designed for any λ and wavelength range,

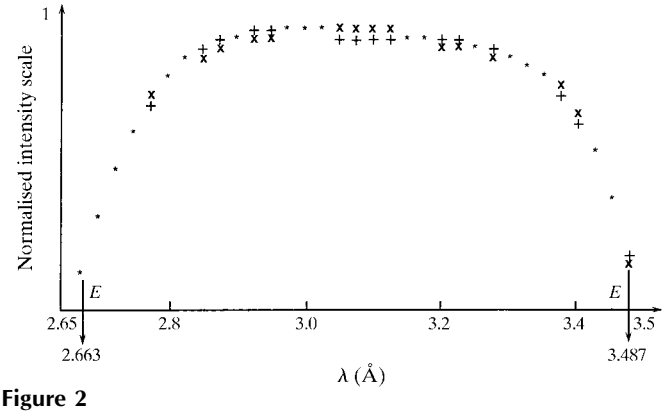

Figure 2
The wavelength-distribution curve. The values given are the normalization function obtained when the data are scaled together using the program *LAUENORM* from the *CCP*4 Laue suite (Campbell, 1995). The curve combines the effects of incoming flux and detector response, but as the latter is basically constant over the range in question, the curve is a good measure of the wavelength distribution. x, observed points; +, fitted points using *LAUENORM*; *, the two coinciding. The two arrows show the limiting values of λ.

$\Delta\lambda/\lambda$. In the present experiment $\Delta\lambda/\lambda$ was chosen to be 24% centred around 3.08 Å, and a measure of the wavelength distribution is given in Fig. 2. One advantage of this 'quasi-Laue' technique is that the number of multiplets is much reduced compared with normal Laue methods and is typically a few percent. Another advantage of this approach is that it removes the neutrons which tend to increase the background on the detector. This particularly applies to very long wavelength neutrons, which will give rise to few reflections, but which can add considerably to the general isotropic background, mainly owing to incoherent scattering from the H atoms in the sample.

The measurement procedure is then simple. Diffraction patterns are recorded for a sufficient number of orientations of the crystal as it is rotated around the cylinder axis. For each crystal setting, the pattern is indexed and integrated using the *CCP*4 Laue suite (Campbell, 1995; Helliwell *et al.*, 1989) modified for cylinder geometry (Campbell *et al.*, 1998) and employing an integration routine based on the minimal $\sigma(I)/I$ method (Wilkinson *et al.*, 1988; Prince *et al.*, 1997). The total set of data is then finally scaled using a wavelength-distribution function which is adjusted to give the best fit between symmetry-related or identical reflections from the various sets of recordings. Normally, there is no reason in this procedure to account for degradation of the crystal during exposure, as it will not suffer from radiation damage. Moreover, the absorption effects are relatively small for neutrons even at very long wavelengths, and a partial correction will be obtained *via* the scaling of related reflections.

Because of the very simple cylinder geometry, care must be taken to orient the crystal so that all of the asymmetric part of reciprocal space is explored, and in the case of triclinic crystals this normally requires measurements to be performed for two different crystal orientations with respect to the rotation axis.

It should also be noted that the general background arising from inelastic and incoherent scattering processes is much higher in Laue and quasi-Laue diffraction techniques at steady-state reactor sources than for step-scanning monochromatic methods. This arises from the simple fact that the intensity of a given reflection does not increase when the wavelength range is increased, but the general background will increase with the increased flux on the sample, thus affecting the peak-to-background ratio. There are therefore good reasons, as given above, for a careful choice of $\Delta\lambda/\lambda$ but also, in the case of neutron scattering, the incoherent signal from hydrogen is very important. Depending on the wavelength, the scattering cross section of hydrogen can be as large as 80×10^{-24} cm^2, which for a hydrogen-containing protein crystal will correspond to an attenuation coefficient of around 3 cm^{-1}. For a 1 mm thick sample, approximately 25% of the beam hitting the crystal will consequently be scattered over the detector as incoherent scattering background.

As the incoherent scattering from deuterium is small, the best remedy is, therefore, to express the protein in a deuterated media (Gamble *et al.*, 1994), and if this is not possible, to use hydrogen–deuterium exchange either by crystal growth in D_2O solvents or *via* H_2O/D_2O exchange of the mother liquor.

Such an *a posteriori* exchange of hydrogen with deuterium will however only apply to the water molecules and the accessible 'acidic' protons of the protein, *i.e.* H atoms bound to nitrogen or oxygen or, exceptionally, to carbon. Many H atoms cannot be exchanged, notably those bound to aliphatic C atoms, and as a rule this kind of 'deuteration' only reduces the background by a factor of approximately two.

3. Crystal and experimental data

The space group is $P1$ with unit-cell parameters a = 27.28, b = 32.04, c = 34.27 Å, α = 88.8, β = 108.8 and γ = 111.6°; M_r = 14.6 kDa for 129 residues. There are 967 H atoms, 612 C atoms, 193 N atoms, 185 O atoms and 10 S atoms in the protein. The solvent part holds approximately 310 water molecules, and ten negative counter ions are required, of which at least five are nitrate groups, while the remainder are most likely to be chloride ions, further nitrate ions or acetates. The neutron-scattering lengths used in the structure refinements were H, −3.742; D, 6.674; C, 6.648; N, 9.36; O, 5.805; S, 2.847 fm.

4. Crystal preparation, hydrogen–deuterium exchange and data collection

The large crystals used in the experiment were obtained by the batch technique. Equal amounts of a 1.0% solution of hen egg-white lysozyme (Boehringer, thrice recrystallized) and a 2.8% solution of $NaNO_3$ in 50 m*M* acetate buffer (pH 4.7) were mixed and left in large closed plastic vials. The growth was initiated in 1984, and the crystals developed to a suitable size within several years. After this they stayed unperturbed and do not seem to have suffered from degradation with time considering the Debye–Waller factors, which compare well with observations from previous structural work.

Of the exchangeable H atoms in the crystals, approximately 50% were replaced by deuterium six months prior to the measurements. This was performed *in situ* after the crystals had been mounted in the capillary, by successive replacements of the mother liquor in the crystal capillary with fresh D_2O buffer solutions.

Two crystals, both with volumes of 2 mm^3, but with different orientations, were used for the measurements in order to completely cover reciprocal space. For the first crystal, a rotation of 180° was executed in steps of 15° for two orientations of the crystal separated by a tilt of 30° on the goniometer head, while for the second crystal a single range of 180° was measured, this time with steps of 18°. The average measuring time for a frame was 10 h, giving a total measurement period of two weeks.

Data were indexed, integrated and averaged as described above, including both sets of crystal data in the analysis. Significant reflections were observed to a limit of 1.5 Å, but at this resolution most intensities were weak, so eventually only data to 1.7 Å were included in the analysis. In all, 46061 reflections were integrated, which after averaging gave 8976 independent structure amplitudes with $F > 2\sigma(F)$. The internal agreement factor for the intensities of reflections with wavelength differences of less than 0.1 Å was 11.4%, and the completeness of the data was better than 83%. Owing to the relatively long rotation steps used in the measurement, some low-order reflections were missing, and these (450 in all) were obtained from earlier data of lower resolution recorded with lower counting times (Oleinek, 1995; Cipriani *et al.*, 1996), giving in all 9426 reflections for the structure analysis.

Several attempts were also made to collect high-resolution data on non-deuterated crystals, but none of the measurements were very successful. This is a consequence of both the lower signal for the hydrogen-containing crystals, as the scattering length of hydrogen is smaller than for deuterium, and the higher background arising from the incoherent scattering of all the extra H atoms. In the best cases the limiting resolution was around 2.0 Å, and these data have, therefore, not been used in the analysis.

5. Structure analysis and location of the water molecules

A first complete analysis was carried out starting from the X-ray structure at 1.97 Å resolution and 295 K obtained by Ramanadham *et al.* (1990), and was divided into two steps, namely the correct location of all H and D atoms of the molecule followed by the determination of the water structure.

All H atoms were placed using the location of the non-H atoms and the routines available in *X-PLOR* (Brünger *et al.*, 1987), including an energy minimization for non-unique positions, and the potentially exchangeable H atoms – all those bound to nitrogen and oxygen – were given the scattering length of deuterium. No water molecules were included, and in a first series of refinements individual isotropic temperature factors were adjusted, followed by a refinement of the protein as a rigid body.

At this point, the 5% of the reflection data assigned for the estimation of the free R factor (Brünger, 1992) was removed from the calculation of the R value and the scattering densities. There then followed a detailed study of the neutron-scattering ($F_o - F_c$) difference density maps using the visualization program *O* (Jones *et al.*, 1991), in order to identify the exact state of deuteration throughout the protein molecule. The structure-factor phases used for this calculation were based on the refined structure, but without including the potential D atoms. Each of these sites was then investigated using maps contoured at the 3σ and -2σ level, respectively, where σ is the r.m.s. fluctuation of the map. D atoms appear as peaks in the former map, while non-exchanged H atoms appear as peaks in the latter map owing to the negative scattering length of hydrogen. There was no difficulty in assigning the deuteration state for all the atoms in question. A total of 214 out of the 268 potentially exchangeable H atoms had been exchanged, and although partial deuteration might occur, all the H/D atoms were given a site-occupation factor of one.

The maps were also used to find five nitrate ions and to ensure the correct location of the amide nitrogen and carbonyl

oxygen of the asparagine and glutamine side chains, as these two are of quite similar size in the X-ray study while very different for neutrons. A general inspection of all the H atoms showed that the hydrogen bound to His15 $C^{\varepsilon 1}$ had been exchanged, proving this atom to be slightly acidic (Derewenda & Derewenda, 1995), while no other H atoms bound to carbon had been exchanged. An inspection of the carboxyl groups of aspartic and glutamic acid confirmed the protonation states deduced from the pK_a values observed by NMR analysis (Delepierre *et al.*, 1987) and the pK of the mother liquor (4.7), notably concerning Glu35 and Asp52 which were protonated and deprotonated, respectively, in agreement with the proposed enzymatic mechanism of lysozyme (Phillips, 1966).

Finally, a refinement of isotropic temperature factors was carried out with the new assignment of D and H atoms. No attempts were made to refine the site-occupation factors of these atoms because of the high correlation between the temperature factor and the site-occupation factor, but the correct state of deuteration was confirmed from an inspection of the final temperature factors for these atoms, none of which showed abnormally large values.

At this stage the R factor was 26.2%, while the free R factor was 26.0%. However, a full refinement of all atomic positions of the protein using *X-PLOR* quickly led to strongly diverging R values, and a model including an overall description of the water using solvent flattening and various weighting schemes did not fare any better. The reason for this is partly the large scattering density of the solvent, which is more than twice as large as the scattering density of the protein, and partly the high number of parameters to be refined. While the 'X-ray molecule' contains 967 non-H atoms, another 1001 H and D atoms have to be included for refinement of the neutron data. The risk of over-refinement is therefore inherent even at 1.7 Å resolution. Indeed, the X-ray analysis will always give a better structure for the non-H atoms, simply because the ratio of the number of reflections to the number of parameters is so much better. The usefulness of the neutron data is clearly in elucidating that which is not visible with X-rays, and any full refinement of the neutron data will only lead to a less precise structure. This will be worsened on inclusion of the water molecules, so in the continued analysis of the structure the non-H atoms of the protein were kept at the positions from the X-ray determination and only the isotropic temperature factors were refined.

The modelling of the solvent was performed in successive steps using σ_A-weighted maps (Read, 1986) at 1.7 Å resolution and following the recommendations given by Kleywegt & Jones (1997). $F_o - F_c$ difference maps were used at the 3σ level, which is a strong criterion, as the fluctuations of the scattering density are larger than in the X-ray case, ranging from the negative H atoms in the protein to the strong signal from the deuterated solvent. The density contained both peaks which revealed complete water molecules, with the location of the D atoms clearly visible, and single peaks. However, the latter were located at distances from each other (and to the full water molecules or to the protein) which as a rule agreed with normal hydrogen-bond dimensions. They were therefore also assumed to represent water molecules, possibly in a disordered state owing to several overlapping local hydrogen-bond arrangements, as for example found in the water structure of B_{12} (Savage, 1987), or (much less likely) rotating completely freely. In such a case both the O and the H atoms will be spread over a region which is a fraction of the water-molecule dimension, but because of the strong scattering of both deuterium and oxygen, the site will still be observable in the map as a broader and lower peak.

In the first cycles, only complete water molecules were included in the iterative procedure. During this process a number of single peaks developed the appearance of complete water molecules, at which stage they were added to the refinement. Many others continued to be single peaks; if they remained stable during several steps they were eventually included as a single O atom. No site-occupation factors were refined, as these are expected to be strongly correlated with the temperature parameters (Kundrot & Richards, 1987*a*). Instead, at each stage of refinement the best overall site-occupation factors for complete water molecules, for the single-peak sites and for hydration in apolar cavities were found manually by scanning in the range 0–1 and then selecting the values which gave the best $R_{\rm cryst}$ and $R_{\rm free}$. Finally, at the last stage of refinement the values for complete and incomplete water sites were both found to be 0.65, while the occupancy for the molecules near the apolar cavities was determined to be 0.40.

In all, six cycles of water placement were carried out, but when the analysis was completed, the very high resolution X-ray structure to 0.95 Å at 295 K (Walsh *et al.*, 1998) became available and the full procedure was repeated from scratch. In this second round, the assigned D atoms were found to be identical to the first analysis, but the agreement factor improved to 24.9% ($R_{\rm free}$ = 24.6%). The full water molecules were the same, but the slightly improved agreement allowed the location of a few more partial water sites. In all 115 water molecules and 129 incomplete sites were identified, and the final R value was 20.4% ($R_{\rm free}$ = 22.1%). The small difference between the two R values is mainly a consequence of the small number of parameters varied in these refinements, in which the non-H atoms were fixed from the X-ray analysis. Indeed, quite large differences might occur if all atoms are refined, as seen, for example, in the neutron analysis of tetragonal lysozyme (Niimura *et al.*, 1997).

6. The water structure and other molecules

The total set of water molecules found, complete and partial, give a nearly complete first water layer, while far fewer molecules, 12% in all, were located further away from the protein.

The 115 complete water molecules, which compare well in number with the 110 strongly bound molecules found by NMR measurements (Bourret & Parello, 1984), are mainly located directly on the protein surface, particularly at the most ordered parts near the polar and charged residues. An example is given in Fig. 3. In these areas the water packing is

particularly tight, in agreement with observations by small-angle scattering, which show the presence of a high-density water region in the vicinity of the protein (Svergun *et al.*, 1998). The atoms have temperature parameters between 5 and 54 Å^2, with a mean value of 35 Å^2, which vary according to the number of hydrogen bonds between the protein and the water molecule, the best ordered of these molecules being those bound directly to the main chain.

In all, eight clusters of complete molecules can be identified, which all start from a water molecule strongly bound to the protein surface. Six of these clusters hold 3–4 molecules, while two have six members each. Beyond the first water level the temperature factors are all around 40 Å^2, showing that the order drops on moving away from the protein surface. A few full molecules are only held by C—H···O contacts or by hydrophobic interactions, and an example of this is shown in Fig. 4. These molecules have temperature factors of 50 Å^2, comparable with the water ordering in the second water layer.

The 129 incomplete water molecules identified have temperature parameters with a mean value of 30 Å^2. These are mainly found near the more disordered parts of the protein, beyond the first hydration layer or near the apolar part of the surface, and in this last group the thermal parameters are all 40–45 Å^2. Six of these molecules are localized in apolar cavities, as discussed below.

There is no indication of the existence of several completely different overlapping water networks in the solvent region, as this would have led to a multi-peak disorder arrangement with very many unphysically short distances between the water sites. An analysis of the hydrogen bonds involving all the water molecules gives about 800 hydrogen bonds, and only around 50 of these have unusually short O—H···O or N—H···O bonds (in the range 2.1–2.2 Å), while around 40 have very short C—H···O contacts. As the majority of these contacts involve incomplete water molecules where the peak is a mean over the possible water locations, such a number of deviations seems acceptable. This of course does not exclude the existence of many relatively similar competing water networks which mainly differ in their hydrogen-bond arrangements. Indeed, the large number of partly ordered molecules indicates the capacity of many water molecules to occupy several orientations while remaining on the same site. This is especially remarkable in the string arrangements found, of which one is shown in Fig. 5. In these arrangements alternatively ordered and disordered water molecules are often observed, suggesting a mode for fast proton transfer, for example by a flip-flop mechanism (Jeffrey & Saenger, 1991).

A comparison with the 139 water molecules determined in the 0.95 Å resolution X-ray study (Walsh *et al.*, 1998) shows that 86 of those correspond – within a maximum shift of 0.5 Å – to full molecules in this study, while 19 agree with the location of incomplete/disordered water molecules. It therefore seems that the very high resolution X-ray study mainly identifies ordered water molecules, while the lower resolution

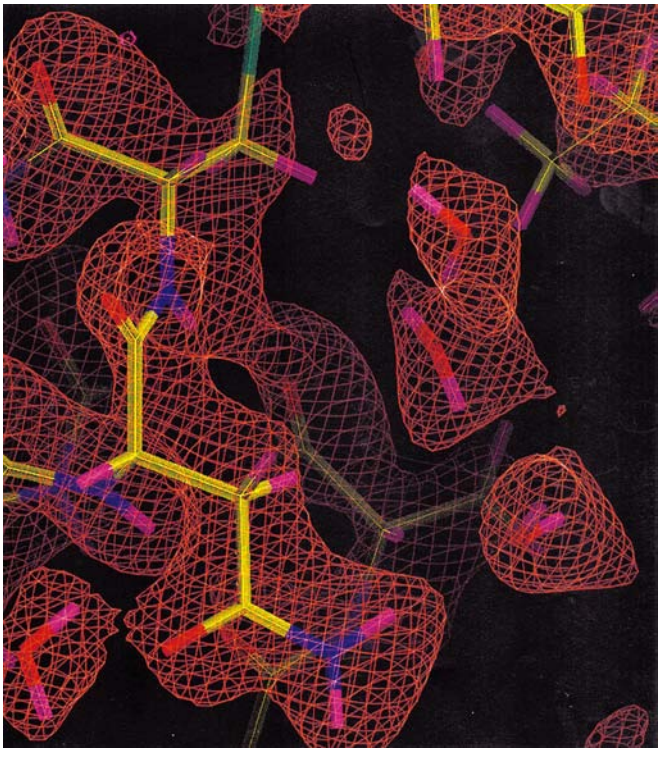

Figure 3
Water molecules bound to the protein surface. $2F_o - F_c$ map, omitting the water molecules and contoured at σ level. Carbon, yellow; oxygen, red; nitrogen, blue; sulfur, green; hydrogen/deuterium, pink.

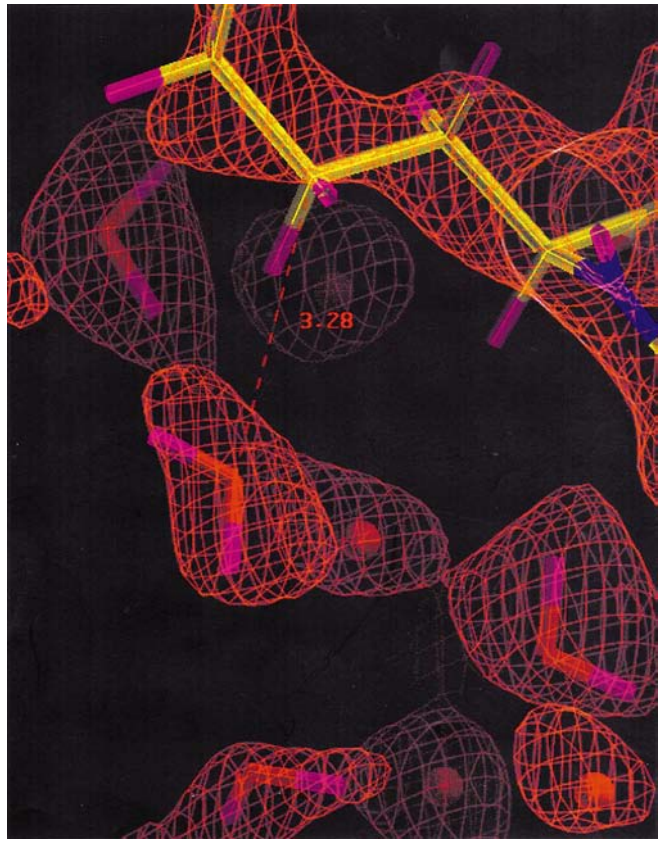

Figure 4
A water molecule held in place by C—H···O contacts only. $2F_o - F_c$ map, omitting the water molecules and contoured at the σ level. Atom colour codes as in Fig. 3.

neutron study of 1.7 Å also shows a good number of disordered molecules, in agreement with the high scattering power of the deuterated water molecule.

The five nitrate groups found were at the same locations as in the X-ray study. However, a sixth disordered molecule determined in the X-ray analysis had been taken for a disordered water molecule. Other possible ions are sodium and acetate ions from the crystallization solvent and the chlorine ions present in the lyophilized lysozyme which serves as starting material for the crystallization. No acetates were found, and it is difficult to distinguish a single-atom ion from a disordered water molecule. One case was found, however, of an atom near the negative part of helix B of the protein, adjacent to the peptide carbonyl group of residue 36, which remained constant all through the refinement procedure, and where the surrounding water molecules were arranged in a solvation shell. This made it a tentative site for a sodium ion. The distance to the carbonyl O atom is short, 1.4 Å, below the expected value of about 2.2 Å, but this can be explained by the small neutron-scattering length of sodium of 3.6 fm, which makes an exact location difficult and gives an estimated σ of more than 0.3 Å.

7. Hydrophobic contacts and apolar cavities

The lysozyme molecule has a high percentage of hydrophobic surface, approximately 41% (Lee & Richards, 1971), and the water molecules covering these areas are fixed *via* C—H···O contacts or organized in a layer according to the description given by Finney *et al.* (1993) for the water surrounding tetramethyl ammonium in solution: as the water molecules cannot form hydrogen bonds with the surface of the hydrophobic part of this molecule, they connect mainly by inter-water hydrogen bonds, creating a tangential water layer at a distance of 4–5 Å from the surface, as shown in Fig. 6. This layer is a blend of full and partial water molecules and the temperature parameters are 40–50 Å^2, *i.e.* higher than the thermal motion parameters elsewhere on the surface. Finney *et al.* (1993) noticed that the water cages surrounding tetramethyl ammonium in solution are not more ordered than the free liquid, and at first sight this seems to contradict the present observation. It should be kept in mind, however, that the hydrophobic patches are surrounded by surface regions which will bind to water, thus creating an ordered layer, and it is quite likely that this order should 'spread' into the water molecules covering the hydrophobic patches.

While cavities with polar surfaces will often show ordered water structure, it has been difficult to identify water molecules in apolar cavities. Recently, however, NMR studies seem to indicate the presence of water in a number of cases (Wolfenden & Radzica, 1994; Ernst *et al.*, 1995; Matthews *et al.*, 1995; Zhang & Hermans, 1996), and Kossiakoff *et al.* (1992) located two water molecules in one apolar cavity of trypsin by neutron crystallography.

Lysozyme has three apolar cavities with accessible volumes of 11 Å^3 (defined by Trp28, Ala31, Leu6, Met105 and Trp108), 40 Å^3 (Leu6, Met12, Leu17, Ile55, Leu56, Ile88, Ser91 and Val92) and 13 Å^3 (Leu17, Tyr20, Trp28, Ala95, Lys96 and Val99), calculated using a probe radius of 1.4 Å and standard van der Waals radii, and a detailed study revealed that cavities one and two held two and four sites, respectively, while there was nothing near the third cavity.

The two sites in cavity one have temperature factors of 43 and 46 Å^2, and while one of the molecules is complete and bound to the carbonyl of Trp28, the other water is incomplete and only 2.5 Å away from C$^{\delta 2}$ of Leu56. The peak observed may therefore mainly arise from one of the D atoms rather than the average water molecule. Similar short distances have also been observed by Kossiakoff *et al.* (1992) for trypsin.

In the second cavity, one molecule is complete while the three other water sites only show a single peak. Again the temperature parameters were high, in the range 42–49 Å^2. The full molecule is bound by hydrogen bonding to the carbonyl

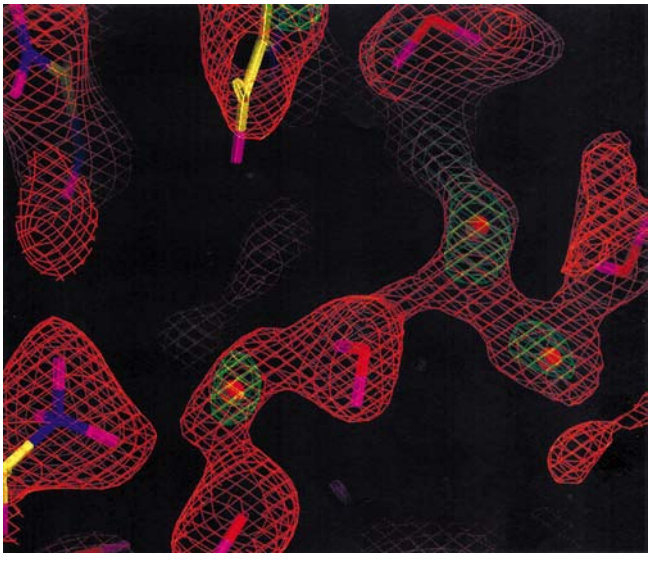

Figure 5
A string arrangement. Orange map: $2F_o - F_c$, omitting full water molecules and contoured at 2σ level. Green map: $F_o - F_c$, contoured at the 2.5σ level. Atom colour codes as in Fig. 3.

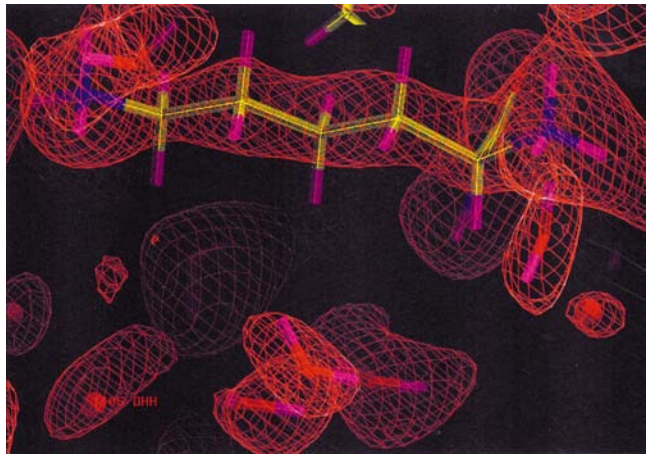

Figure 6
Water molecules arranged along the apolar part of Lys1. $2F_o - F_c$ map, omitting the water molecules and contoured at the σ level. Atom colour codes as in Fig. 3.

group of Ser91 and is probably identical to the water observed in the structure of tetragonal lysozyme (Kundrot & Richards, 1987*b*) located very near the amide group of Val92, while none of the other sites seem to have been identified in the X-ray maps. The three incomplete water molecules are only held *via* C—H···water contacts and, as in the case discussed above, the distances to the C atoms are very short, being below 2.5 Å. It is therefore tempting to speculate that the D atom, rather than the oxygen, is actually located nearest the surface, and that it is held in place by van der Waals forces and C—H···H-type contacts. In this case, the rest of the water molecule could be much more disordered and therefore invisible. This would explain why these molecules are not found in the room-temperature X-ray analysis.

NMR analysis of lysozyme (Otting *et al.*, 1997) has also shown the presence of disordered water in the two first cavities but nothing in the third, in agreement with the present work, and although a complete NMR assignment was not possible, there is agreement between the attributed peak and the partial sites found in this study. It therefore seems that neutron maps can be very useful in the study of apolar or hydrophobic cavities, although the short distances between the solvent peaks and the protein clearly shows that the mode of interaction between the two is still under question.

8. Structural disorder and dynamics

Following conventional wisdom, the advantage of neutron scattering is its ability to locate H atoms, for example in water molecules. At first it was therefore quite surprising that so many of the H/D atoms in the water molecules were not traceable. On the other hand, it was quite clear that the observed single peaks, all significantly above the noise level of the difference maps, were laid out in a manner and with inter-peak distances which showed the imprint of a water network. The most likely interpretation is, therefore, that we observe a sum of overlapping nearly identical water networks, where the water molecules are located most of the time in reasonably well defined sites, but where they have several local hydrogen-bond schemes to choose between. In this way, hydrogen disorder becomes just as valuable information as hydrogen order, as it can help to understand how and where the water molecules are most likely to move and thus help in unravelling the dynamics of the solvent.

A next step should therefore be to pin down these individual networks in order to obtain a complete picture of the different hydrogen patterns and their interrelation, but this is probably only possible for molecules such as vitamin B_{12} (Savage, 1987), where much higher resolution data can be recorded. The ratio of the number of parameters to the number of observations for a molecule the size of lysozyme – which is actually a rather small protein – will always prevent such an analysis. Indeed, as observed in the early part of the analysis, even conventional refinements are not very successful owing to the large number of atoms which are 'seen' by the neutrons.

This of course does not render the neutron analysis of water structure meaningless, as it allows (for the reasons given above) the identification of many more water molecules at relevant biological temperatures, *i.e.* room temperature, than even the very high resolution X-ray studies.

The way to use the neutron-diffraction data is therefore to identify as many water sites as possible, even if they do not hold a complete water molecule, of course taking a maximum of precautions not to include noise peaks or to be guided by what one hopes to see. The resultant picture for triclinic lysozyme is then one which shows extended networks occupying 79% of the solvent volume, corresponding to the 244 identified sites out of the total 310 water molecules in the unit cell.

Of course, it is not possible to judge from diffraction data whether the observed pattern of partial disorder is static with the water networks varying from unit cell to unit cell, or whether it is the result of a constantly fluctuating network driven by the motions of the protein and the diffusion of the ions of the solvent. This can, however, be investigated using other techniques, notably inelastic or quasi-elastic neutron scattering.

Measurements of this kind were therefore carried out on the instruments IN5 and IN6 of the Institut Laue–Langevin using a powder of triclinic crystals. This time the measurements, which will be reported in detail elsewhere (Bon *et al.*, 1999), were performed on non-deuterated material. As mentioned previously, H atoms have a very high cross section for incoherent scattering. The incoherent quasi-elastic scattering signal will consequently only contain information about the single-particle motion of the H atoms, and it is thus very well suited for the analysis of the water-molecule dynamics (Bellissent-Funel *et al.*, 1992).

The measurements showed that less than 5% of the H atoms in the crystal were immobile within the resolution of the instrument, and that 60% of the mobile atoms had motion typical of solvent molecules. This corresponds to the known amount of water in the cell, including the terminal protons of lysine and arginine, which will follow the water dynamics. Because the signals from all H atoms are overlaid in the measurement, the individual motion of the water molecules cannot be followed, but instead two types of behaviour were found and quantified. In one group, the H atoms diffuse within a sphere of 3.6 ± 0.4 Å in jumps of 1.6 Å, five times slower than in bulk water at room temperature, and with a residence time of 11 ps, while the other group exhibits a translation mode about 50 times slower than in water.

The behaviour of molecules in the two groups can now be compared with the structural observations. The hydrogen-jump distance of 1.6 Å agrees well with a reorientational motion of a water molecule, as this is the distance between the two H atoms or between a H atom and the position it would take if it switched to one of the lone pair positions of the same molecule. Likewise, the sphere corresponds approximately to the volume formed by the nearest-neighbour water molecules. The water molecules in this group therefore correspond to the water network, with its fluctuating hydrogen-bond patterns

involving many molecular reorientations and an occasional displacement of a molecule to a neighbouring site. The fraction of water molecules belonging to this group as determined from the quasielastic scattering is $75 \pm 5\%$, in good agreement with the 79% given above for the structural occupation of the networks. The remaining 25%, which undergo translation motion, are unlikely ever to be in the same location, and would correspond to the remaining 60–70 water molecules which were never located. They have only a small space left to 'flow' in, hence their much reduced mobility compared with liquid water.

The quasi-elastic measurements, therefore, underline the dynamic nature of the disorder suggested by the many single-peak water molecules observed in this study, and the result is a picture of the protein covered by a constantly fluctuating water network which rapidly becomes less ordered moving away from the protein.

9. Conclusions

The present study is one of the first which has been carried out on the new diffractometer designed for quasi-Laue neutron work and installed at the Institut Laue–Langevin. It is now routinely scheduled as a service instrument for outside scientists, and this analysis of water structure was part of the commissioning process, both to determine how far – in the best case – the resolution could be pushed, and also how well water structure could be determined.

Most of the studies carried out with neutrons on this new instrument will probably be aimed at localizing a few H atoms indicative of a biological mechanism, but even in this case it is essential to work with the optimal crystallographic fit, and a carefully determined water structure is part of this, particularly for neutrons, where the signal from water is very strong. This work has served to develop most of these procedures.

The analysis of water in itself in protein crystals is also very useful, as proteins and other essential biological molecules are in contact *via* the water medium, and indeed the water content in protein crystals is comparable to that in living cells. It is thus essential to know what the protein surface looks like when covered in water. This study has shown considerable orientational disorder of the water molecules in the first water layer which, in combination with quasi-elastic measurements, leads to the picture of fluctuating networks elaborated above.

The quasi-elastic measurements also show that the lifetime for water molecules in a site is longer than in the standard liquid, but whether this will slow the motion of the larger molecules in contact with the protein is not clear. The capacity of water to reorient itself is apparently present, as shown by the water disorder, and a two-dimensional network surface might well be as 'slippery' as a three-dimensional random water arrangement involving many collisions with free water molecules.

In summary, this study has therefore shown that reasonably high resolution neutron data can be observed in an acceptable span of time, and that the data can be analysed and can help in further understanding the water arrangements near a protein surface, and consequently raising new questions.

We thank Drs José Dianoux, Michel Ferrand, Paul Langan, Sax A. Mason, Dean Myles, Joseph Zaccai and Larry Sieker and Professors Eva Pebay-Peyroula and Marc Bée for many useful discussions during this work. CB is indebted to the French Ministry for Science and Industry for a thesis grant supporting this work.

References

Amemiya, Y., Matsushita, T., Nakagawa, A., Satow, Y., Miyahara, J. & Chikawa, J. (1988). *Nucl. Instrum. Methods A*, **266**, 645–653.
Backlin, A., Hedin, G., Fogelberg, B., Saraceno, M., Greenwood, R. C., Reich, C. W., Koch, H. R., Baader, H. A., Breitig, H. D., Schult, O. W. B., Schreckenbach, K., von Egidy, T. & Mampe, W. (1982). *Nucl. Phys. A*, **380**, 189–260.
Bellissent-Funel, M. C., Teixeira, J., Bradley, K. F., Chen, S. H. & Crespi, H. L. (1992). *Physica B*, **180**, 740–744.
Bon, C., Dianoux, A. J., Ferrand, M. & Lehmann, M. S. (1999). In preparation.
Bouquiere, J. P., Finney, J. L., Lehmann, M. S., Lindley, P. F. & Savage, H. F. J. (1993). *Acta Cryst.* B**49**, 79–89.
Bourret, D. & Parello, J. (1984). *J. Phys.* **45**, C7255–C7258.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brünger, A. T., Kuriyan, J. & Karplus, K. (1987). *Science*, **235**, 458–460.
Campbell, J. W. (1995). *J. Appl. Cryst.* **28**, 228–236.
Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. (1998). *J. Appl. Cryst.* **31**, 496–502.
Cheng, X. & Schoenborn, B. P. (1990). *Acta Cryst.* B**46**, 195–208.
Cipriani, F., Castagna, J. C., Gabriel, A., Wilkinson, C. & Lehmann, M. S. (1994). *Biophys. Chem.* **53**, 5–14.
Cipriani, F., Castagna, J. C., Wilkinson, C., Oleinek, P. & Lehmann, M. S. (1996). *J. Neutron Res.* **4**, 79–85.
Delepierre, M., Dobson, C. M., Karplus, M., Poulsen, F. M., States, D. J. & Wedin, R. E. (1987). *J. Mol. Biol.* **197**, 111–130.
Derewenda, Z. S. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.
Eisenstein, M., Sharon, R., Berkovitz-Yellin, Z., Gewitz, H. S., Weinstein, S., Pebay-Peyroula, E., Roth, M. & Yonath, A. (1991). *Biochimie*, **73**, 879–884.
Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M. & Clore, G. M. (1995). *Science*, **267**, 1813–1817.
Finney, J. L., Soper, A. K. & Turner, J. Z. (1993). *Pure Appl. Chem.* **65**, 2521–2526.
Gamble, T. R., Clauser, K. R. & Kossiakoff, A. A. (1994). *Biophys. Chem.* **53**, 15–26.
Helliwell, J. R., Habash, J. Cruickshank, D. W. J, Harding, M. M., Greenhough, T. J., Campbell, J. W., Clifton, I. J., Elser, M., Machin, P. A., Papiz, M. Z. & Zurch, S. (1989). *J. Appl. Cryst.* **22**, 483–497.
Hofmann, J. & Rausch, C. (1995). *Nucl. Instrum. Methods A*, **355**, 494–499.
Høghøj, P., Anderson, I. S., Ebisawa, T. & Takeda, T. (1996). *J. Phys. Soc. Jpn*, **65**, Suppl. *A*, 296–298.
Jeffrey, G. A. & Saenger, W. (1991). *Hydrogen Bonding in Biological Structures*, pp. 40–42. Heidelberg: Springer-Verlag.
Jones, A. T., Zou, J.-Z., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 525–545.
Kossiakoff, A. A., Sintchak, M. D., Sphungin, J. & Presta, L. G. (1992). *Proteins*, **12**, 203–222.
Kossiakoff, A. A. & Spencer, S. A. (1980). *Nature (London)*, **288**, 414–416.

Kossiakoff, A. A. & Spencer, S. A. (1981). *Biochemistry*, **20**, 6462–6474.
Kundrot, C. E. & Richards, F. M. (1987*a*). *Acta Cryst.* B**43**, 544–547.
Kundrot, C. E. & Richards, F. M. (1987*b*). *J. Mol. Biol.* **193**, 157–170.
Langan, P., Jogl, G., Lehmann, M. S., Wilkinson, C. & Kratky, C. (1999). *Acta Cryst.* D**55**, 51–59.
Lee, B. K. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Lehmann, M. S. & Stansfield, R. F. D. (1989). *Biochemistry*, **28**, 7028–7033.
Matthews, B. W., Morton, A. G. & Dahlquist, F. W. (1995). *Science*, **270**, 1848–1849.
Niimura, N., Karasawa, Y., Tanaka, I., Miyahara, J., Kahashi, K., Saito, H., Koizumi, S. & Hidaka, M. (1994). *Nucl. Instrum. Methods A*, **349**, 521–525.
Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J.-C., Cipriani, F., Høghøj, P., Lehmann, M. S. & Wilkinson, C. (1997). *Nature Struct. Biol.* **4**, 909–914.
Oleinek, P. (1995). Diplomarbeit E13, Technische Universität München.
Otting, G., Liepinsh, E., Halle, B. & Frey, U. (1997). *Nature Struct. Biol.* **4**, 396–404.
Phillips, D. C. (1966). *Sci. Am.* **215**, 78–90.
Prince, E., Wilkinson, C. & McIntyre, G. J. (1997). *J. Appl. Cryst.* **30**, 133–137.
Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 63–69.
Rauch, H., Grass, F. & Feigl, B. (1967). *Nucl. Instrum. Methods*, **46**, 153–160.
Rausch, C., Bücherl, T., Gähler, R., von Seggern, H. & Winnacker, A. (1992). *SPIE*, **1737**, 255–263.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Roth, M., Lewit-Bentley, A., Michel, H., Deisenhofer, J., Huber, R. & Oesterhelt, D. (1989). *Nature (London)*, **340**, 659–662.
Savage, H. J. F. (1987). *Biophys. J.* **50**, 947–980.
Savage, H. J. F., Lindley, P. F., Finney, J. L. & Timmins, P. A. (1987). *Acta Cryst.* B**43**, 280–295.
Schoenborn, B. P. (1969). *Nature (London)*, **224**, 143–146.
Schoenborn, B. P. (1992*a*). *SPIE*, **1737**, 192–198.
Schoenborn, B. P. (1992*b*). *SPIE*, **1738**, 235–240.
Svergun, D. I, Richards, S., Koch, M. H., Sayers, Z. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.
Templer, R. H. (1991). *Nucl. Instrum. Methods A*, **300**, 357–366.
Timmins, P. A., Poliks, B. & Banaszak, L. (1992). *Science*, **257**, 652–655.
Walsh, M. A., Scheider, T., Sieker, L. C., Dauter, Z., Lamzin, V. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–547.
Wilkinson, C., Khamis, H. W., Stansfield, R. F. D. & McIntyre, G. J. (1988). *J. Appl. Cryst.* **21**, 471–478.
Wilkinson, C. & Lehmann, M. S. (1991). *Nucl. Instrum. Methods A*, **310**, 411–415.
Wlodawer, A. & Sjölin, L. (1983). *Biochemistry*, **22**, 2720–2728.
Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984). *J. Mol. Biol.* **180**, 301–329.
Wolfenden, R. & Radzica, A. (1994). *Science*, **265**, 936–937.
Zhang, L. & Hermans, J. (1996). *Proteins*, **24**, 433–438.