feature articles

Neutron protein crystallography: beyond the folding structure of biological macromolecules

Nobuo Niimura[a]* and Robert Bau[b]

[a]Institute of Applied Beam Science, Graduate School of Science and Engineering, Ibaraki University, 4-12-1 Naka-Narusawa, Hitachi, Ibaraki 316-8511, Japan, and [b]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA. Correspondence e-mail: niimura@mx.ibaraki.ac.jp

Neutron diffraction provides an experimental method of directly locating H atoms in proteins, a technique complementary to ultra-high-resolution X-ray diffraction. Three different types of neutron diffractometers for biological macromolecules have been constructed in Japan, France and the USA, and they have been used to determine the crystal structures of proteins up to resolution limits of 1.5–2.5 Å. Results relating to H-atom positions and hydration patterns in proteins have been obtained from these studies. Examples include the geometrical details of hydrogen bonds, the role of H atoms in enzymatic activity, CH_3 configuration, H/D exchange in proteins and oligonucleotides, and the dynamical behavior of hydration structures, all of which have been extracted from these structural results and reviewed. Other techniques, such as the growth of large single crystals and a database of hydrogen and hydration in proteins, are described.

1. Introduction

The three-dimensional structure determinations of biological macromolecules such as proteins and nucleic acids by X-ray crystallography have improved our understanding of many of the mysteries involved in biological processes. At the same time, these results have clearly reinforced the commonly held belief that H atoms and water molecules around proteins and nucleic acids play a very important role in many physiological functions. However, since it is very hard to identify H atoms accurately in protein molecules using X-ray diffraction alone, a detailed discussion of protonation and hydration sites can only be speculated upon so far. In contrast, it is very well known that neutron diffraction provides an experimental method of directly locating H atoms.

Neutron protein crystallographic data were first collected by Schoenborn (1969), who later determined the positions of the H atoms of myoglobin (Phillips & Schoenborn, 1981). Neutron diffraction studies of several proteins were subsequently carried out by several authors (Kossiakoff & Spencer, 1980; Wlodawer & Sjolin, 1982; Mason et al., 1984). However, since it takes weeks or months to complete such structure determinations, neutron protein crystallography has not become a common technique in the field of structural biology, in sharp contrast to X-ray crystallography.

The development of a neutron imaging plate (NIP) was a breakthrough event in neutron protein crystallography (Niimura et al., 1994) for reactor neutron sources. The first application of the NIP was a structure determination of tetragonal hen-egg-white lysozyme (Niimura et al., 1997) using a quasi-Laue diffractometer, LADI (Helliwell & Wilkinson, 1994), at the Institut Laue–Langevin (ILL) in Grenoble, France. After that, we constructed high-resolution neutron diffractometers dedicated to biological macromolecules (BIX-3, BIX-4), which use a monochromated neutron beam as well as the NIP at the Japan Atomic Energy Agency (JAEA) (Tanaka et al., 2002; Kurihara, Tanaka, Refai Muslih et al., 2004). More recently, the LANSCE time-of-flight ^{3}He detector for neutron protein crystallography has started producing results (Schoenborn & Langan, 2004). Thus, these three technical developments have greatly improved the capability of neutron protein crystallography: namely, the time needed to measure data, the diffraction resolution attained and the molecular-weight ceiling reachable. As a result, the field has enjoyed a significant resurge in recent years.

Unfortunately, neutron protein crystallography still remains a severely limited technique, but hopefully things will improve substantially in the future. The current development of 'next generation' spallation neutron sources, such as J-PARC (the Japanese Proton Accelerator Research Complex) in Japan and SNS (the Spallation Neutron Source) in the USA, as well as new developments and improvements at existing sources [e.g. LADI-3 at the ILL and the proposed LMX++ instrument at ISIS TS2 (target station 2) in England], will enable several more powerful protein crystallographic instruments to be installed. In some of the aforementioned new spallation sources, a gain in neutron intensity of almost two orders of magnitude is expected. At that point, the use of neutron diffraction is expected to greatly expand the field of structural biology.

The general subject of neutron protein crystallography has been reviewed by several authors (Wlodawer, 1982; Schoenborn, 1985; Kossiakoff, 1985; Helliwell, 1997; Niimura, 1999; Tsyba & Bau, 2002; Niimura *et al.*, 2006). Also of potential interest to readers are articles describing the synergy and complementarity between neutron diffraction and ultra-high-resolution X-ray diffraction (Hazemann *et al.*, 2005). In this paper, we will summarize selected results regarding H-atom positions and hydration in proteins, obtained mainly using the two BIX-type diffractometers at JAEA.

2. Diffractometer and crystallization for neutron protein crystallography

2.1. The BIX-type diffractometers and other currently available instruments

The neutron imaging plate (NIP) consists of a neutron converter material, Gd, which captures neutrons and emits secondary charged particles, which are absorbed by a photostimulated luminescence (PSL) material (BaFBr:Eu^{2+}). The neutron capture efficiency at a neutron wavelength of 1 Å is about 80%, and a cylindrical shape of the detector is easily obtained with the NIP since it is a large flexible (plastic) sheet. The high spatial resolution (less than 0.2×0.2 mm) of the NIP has allowed us to reduce the distance between the sample and the detector to 200 mm, which contributes to an overall decrease in size of the diffractometer (Niimura *et al.*, 1994). After developing a novel practical neutron monochromator (a
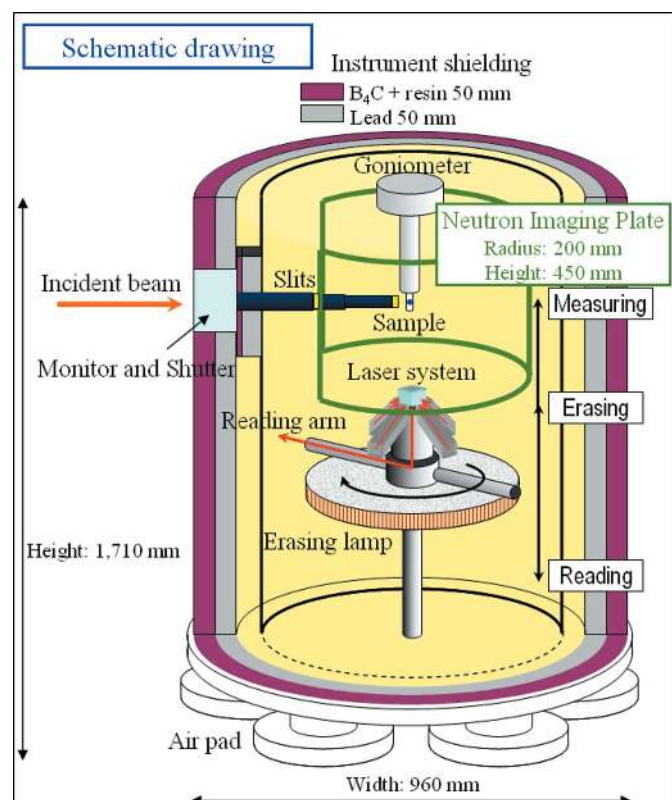
set of elastically bent perfect Si plates), two BIX-type diffractometers (BIX-3 and BIX-4) (Tanaka *et al.*, 2002; Kurihara, Tanaka, Refai Muslih *et al.*, 2004) dedicated to protein crystallography were constructed at the JAEA. The most characteristic and novel design of BIX-type diffractometers is the overall vertical arrangement of the main components of the diffractometer (Fig. 1). This enables a compact design to be realized, and consequently a higher flux of neutrons is obtained because of the close proximity of the sample to the monochromator (*i.e.* to the neutron source). The BIX-3 and BIX-4 instruments are conceptually similar to the LADI diffractometer at the ILL in Grenoble, France, which uses a Laue geometry equipped with a NIP and a broad spectral range to maximize the neutron flux on the sample (Helliwell & Wilkinson, 1994; Myles *et al.*, 1998), except that LADI is based on a horizontal configuration. In contrast, the PCS (Protein Crystallographic Station) diffractometer (Schoenborn & Langan, 2004) located at the LANSCE neutron spallation source in Los Alamos National Laboratory (New Mexico, USA) is unique in that it is the first instrument designed to use a pulsed neutron source for macromolecular crystallography.

2.2. Crystallization

One major disadvantage of neutron macromolecular crystallography is that large crystals are needed: currently, the volume of samples studied is normally larger than 1 mm^3, unless they are fully deuterated (Hazemann *et al.*, 2005). Usually such a large single crystal is difficult to grow. However, like others (Saridakis *et al.*, 1994; Sauter *et al.*, 1999; Budayova-Spano *et al.*, 2007), we have found that one rational way to find the proper conditions to grow large single crystals is to establish a detailed crystallization phase diagram first. Fig. 2 shows a typical protein crystallization phase diagram, where
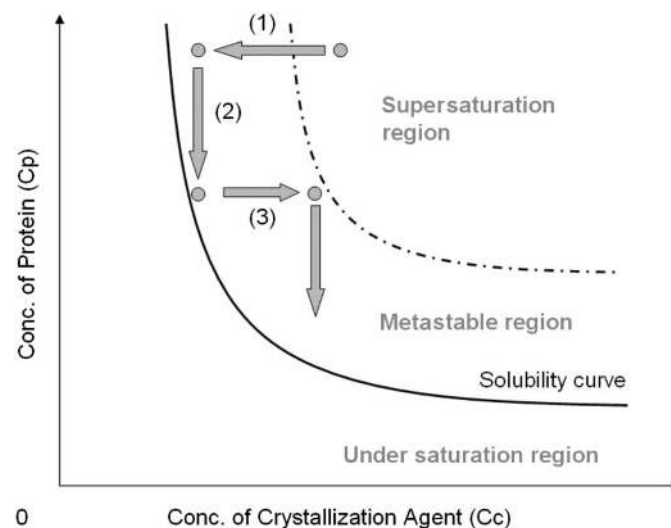

Figure 1
A schematic view of the BIX-type diffractometer.

Figure 2
A typical general crystallization phase diagram. The principle of crystallization of a large crystal is illustrated by the sequence of gray arrows.

feature articles

Cp and Cc are the protein and precipitant concentrations, respectively. The phase diagram is classified into three regions: the unsaturated, metastable and supersaturated regions. Generally speaking, the best large single crystals can be grown under metastable conditions close to the solubility boundary (the curved solid line in Fig. 2) but nucleation cannot occur in that region. Normally, crystals grow easily in the super-saturated region but the problem is that one usually gets several very tiny crystals if the crystallization starts directly in the supersaturated region. To turn this into a smaller number of larger crystals, we decrease the Cc value to enter in the metastable region [left-pointing arrow labeled (1) in Fig. 2], which causes most of the smaller crystals to dissolve and this allows a few survivors to become seed crystals. This is repeated back and forth a few times until one seed crystal is left. This seed crystal is then allowed to continue to grow slowly in the metastable region. As the crystal grows, the protein concentration remaining in solution decreases simultaneously [downward arrow labeled (2) in Fig. 2] and reaches the solubility curve (solid line) and crystal growth stops. Then Cc is increased to the right boundary [right-pointing gray arrow labeled (3)] of the metastable region and the crystal growth resumes. Using a systematic way of evaluating the quality of crystals grown at various points in such a phase diagram, we have successfully grown large single crystals of three macro-molecular samples: (a) cubic porcine insulin (Maeda et al., 2004), (b) dissimilatory sulfite reductase D (Chatake et al., 2003) and (c) a DNA oligomer (Arai et al., 2002); all of which were successfully used in neutron protein data collection. An evaluation process involving the application of a 'relative Wilson plot' method was developed, involving quick X-ray data collections on selected crystalline samples, to evaluate the quality of crystals prior to neutron data collection (Arai et al., 2004).

3. Hydrogen and Hydration in Proteins Data Base (HHDB)

The results from neutron protein crystallography have been accumulating significantly in recent years and we have realized that it is becoming increasingly important to deposit this information in a publicly accessible site so that these data can be retrieved more efficiently. We have created a Hydrogen and Hydration in Proteins Data Base (HHDB) that catalogs all H-atom positions in biological macromolecules and in hydration water molecules that have been determined thus far by neutron macromolecular crystallography. This allows important structural information such as the geometry of hydrogen bonds to be abstracted and categorized for the first time in the field of structural biology.

The HHDB provides us with a graphic interface for visualizing all types of interactions involving H atoms, such as hydrogen bonds (defined by $X-H \cdots Y$, where X is a hydrogen donor and Y is a hydrogen acceptor). For example, one type of plot (Fig. 3) allows the user to visualize the distribution of the acceptor atoms (Y) as a function of the $H \cdots Y$ distance and the supplementary $X-H \cdots Y$ angle (180° minus the actual angle). From these results, it is very surprising that in proteins there are very few acceptor atoms which form truly collinear $X-H \cdots Y$ hydrogen bonds, in contrast to the situation in small inorganic and organic compounds in which linear hydrogen bonds are quite common. In the HHDB, the user has the option of easily changing the cut-off hydrogen-bonding values (such as 2.7, 2.6 or 3.1 Å) if he or she wishes, and we found that the non-linearity of hydrogen bonds is not sensitive
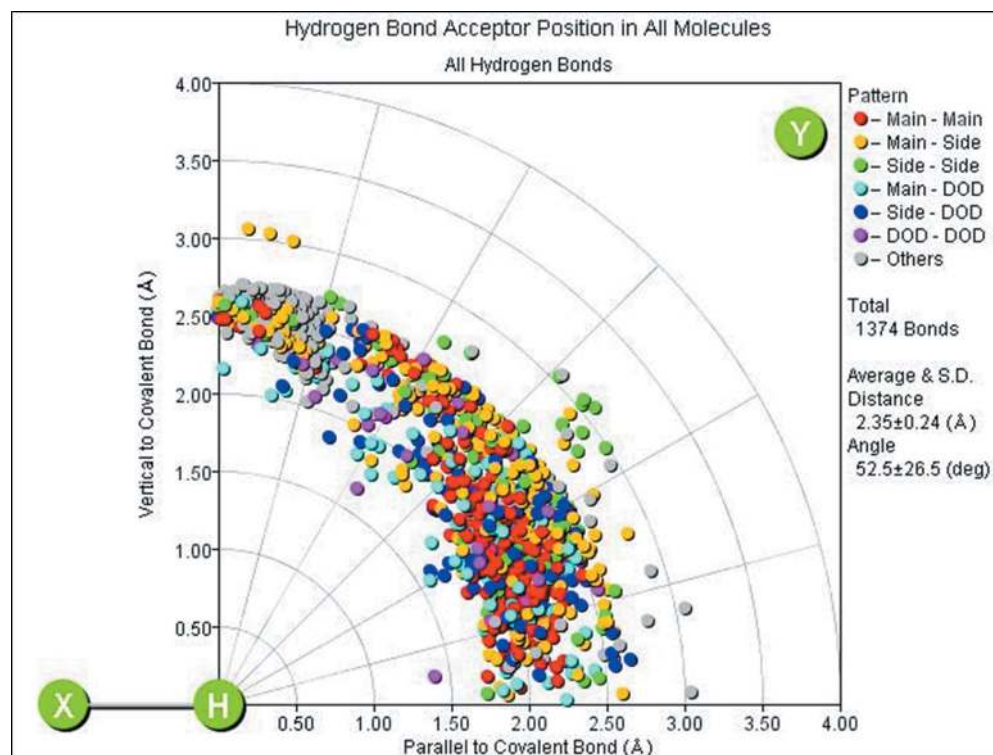

Figure 3
Schematic statistics of hydrogen bonds. Acceptor positions (Y) are searched and plotted in the first zone. In this example, hydrogen-bond lengths are defined by $X \cdots Y$ values less than 2.7, 2.6 and 3.1 Å for the acceptor atoms, N, O and S, respectively, and hydrogen-bond angles are defined to be the supplement of 180°. These cut-off values (2.7, 2.6 and 3.1 Å) are just defaults and users can modify them. Note the very surprising result that in proteins there are very few acceptor atoms which form truly collinear $X-H \cdots Y$ hydrogen bonds, in contrast to the situation in small inorganic and organic compounds in which linear hydrogen bonds are quite common. (Note: the word 'others' on the right-hand side of this plot refers to entities other than protein residues or water molecules, such as heme groups, substrates, counterions etc.)

to what the cut-off values are. The strengths of hydrogen bonds obtained in X-ray protein crystallography have often been assigned using the $X \cdots Y$ length as a criterion, but in our opinion this assumption may be somewhat risky, because this plot suggests that the usual argument that $X-H \cdots Y$ hydrogen bonds are approximately linear may not be valid in proteins.

The HHDB depository has been developed by cooperation between the Japan Science and Technology Agency (JST) and the JAEA under a database project funded by the JST. It is available *via* the website http://hhdb01.tokai-sc.jaea.go.jp/HHDB/. From the HHDB, we have provided an Internet link to the Protein Data Bank (PDB) and the protein code and the nomenclature used in the two databases are the same.

4. Results

4.1. Positions of H atoms in macromolecules

In JAEA, we have already analyzed the neutron structures of several proteins and two DNA oligomers: myoglobin (Ostermann *et al.*, 2002), rubredoxin (Kurihara, Tanaka, Chatake *et al.*, 2004), a rubredoxin mutant (Chatake *et al.*, 2004), hen-egg-white lysozyme (Niimura *et al.*, 1997; Maeda *et al.*, 2001), human lysozyme, porcine insulin (Maeda *et al.*, 2004; Ishikawa *et al.*, 2007), dissimilatory sulfite reductase D (DsrD) (Chatake *et al.*, 2003), and oligomers of *B*-DNA (Arai *et al.*, 2005) and Z-DNA (Chatake *et al.*, 2005). For example, in the rubredoxin mutant, 96% of H atoms have contours more negative than the -1.0σ level, while all D atoms have positive contours higher than 1.5σ, which means that practically all of the H and D atoms of the protein were directly observable at that resolution (1.6 Å). Even for neutron experiments carried out at medium resolution, most of the H/D atoms can be found. For example, in the neutron analysis of the protein DsrD, carried out at 2.7 Å resolution, about 60% of the H/D atoms of the molecule could be located (Chatake *et al.*, 2003). The positions of H atoms covalently bonded to C atoms can be calculated stereochemically based on the coordinates of C and N atoms determined by high-resolution X-ray crystal structure
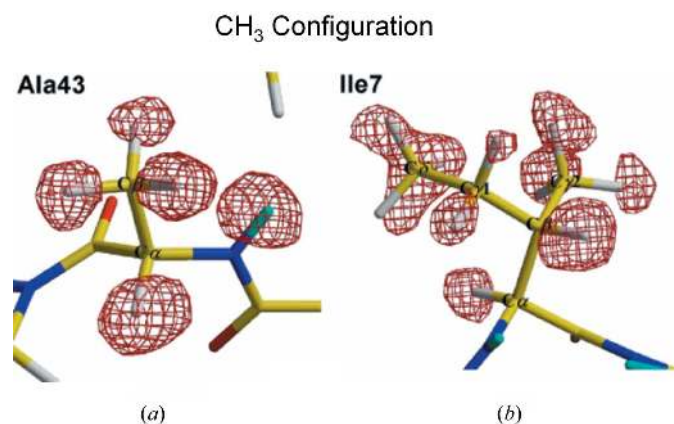
analysis, but the methyl H atoms can sometimes be a significant distance from their predicted positions because of free rotation around C—C bonds. Consequently, if their precise positions are required, they should be determined by neutron diffraction experiments. Fig. 4 shows examples of some methyl H atoms in rubredoxin determined from our neutron diffraction data (Kurihara, Tanaka, Chatake *et al.*, 2004).

4.2. Methyl CH₃ group configuration

In gaseous ethane, the C—C bond rotates freely but a staggered conformation is most stable and an eclipsed conformation is most unstable. In proteins, CH_3 group conformations in some amino acid residues such as alanine, valine, leucine and isoleucine have not been much discussed so far because the H atoms of these CH_3 groups were not identified well. Our high-resolution neutron diffraction analyses of myoglobin and other proteins have provided this information. As shown in Fig. 4, the H atoms of alanine and isoleucine in rubredoxin are clearly identified. These positions have been deposited in the HHDB. All the CH_3 group configurations in amino acid residues are extracted and discussed as follows. Fig. 5(*a*) shows CH_3 group conformations as Newman projections and the dihedral angles of the CH_3 group are shown as curved arrows. Atoms H1, C1 and C2 and atoms $R1$, C2 and C1 form two planes and the dihedral angle between the two planes defines the conformation of the CH_3 group as being either staggered or eclipsed. In the myoglobin case, 92 CH_3 groups have been identified, their dihedral angles were calculated and their distribution is shown in Fig. 5(*b*). It is found that most of the CH_3 groups belong to the stable staggered conformations (angles close to 60°), but several % of them appear to have values close to eclipsed conformations
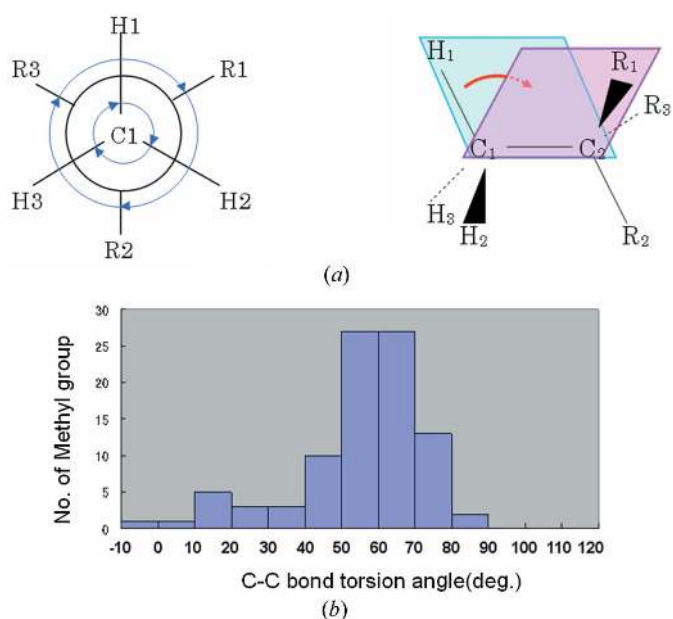

Figure 4
$|F_o| - |F_c|$ omit nuclear-density map of the H atoms around the residues (*a*) Ala43 and (*b*) Ile7 of wild type rubredoxin.

Figure 5
(*a*) CH_3 group conformations depicted by Newman projection formula, and (*b*) the distribution of torsional angles of the CH_3 group in myoglobin.

(defined by angles close to 0°). The structural mechanism which allows H atoms to approach the higher-energy eclipsed conformations has been analyzed in the case of myoglobin, but these deviations may very well be due to steric interactions from neighboring atoms.

4.3. Hydrogen bonds

4.3.1. General comments. Hydrogen bonds clearly play important roles in countless biological processes. The strengths of hydrogen bonds are intermediate between those of covalent bonds and van der Waals forces: they are directional and form different types of networks under various conditions. Along with other investigators, Baker & Hubbard (1984) have extensively discussed hydrogen bonds in globular proteins, using H-atom positions calculated using the atomic coordinates derived from high-resolution protein X-ray data (Baker & Hubbard, 1984). It was pointed out that, while most of the H-atom positions in a protein can be reliably predicted (for example, those of C—H bonds), H-atom positions could not be uniquely defined for most O—H and S—H bonds, as

well as some N—H bonds. In some cases, (a) the orientation of the H atoms (e.g. alcoholic protons such as those of Ser, Thr and Tyr) may not be known, in other cases (e.g. Asp, Glu, His), the questions are (b) the extent of protonation (e.g. is the carboxylate group of an Asp residue protonated or not?) and (c) if protonated, which of the two competing sites (e.g. the two O atoms of Asp and Glu or the two N atoms of His) is the protonated one? These are often the most interesting protons as far as chemical reactivity is concerned. In their conclusion, Baker & Hubbard (1984) stressed the necessity of high-resolution neutron diffraction studies. In this article, we will discuss several examples in which questions of this type have been answered through single-crystal neutron work.

4.3.2. Non-linear hydrogen bonds. In the neutron analysis of the α-helices of myoglobin, the linearity of the $X-H\cdots Y$ hydrogen bonds has been analyzed. Fig. 6(a) shows the statistics of hydrogen bonds $N-H\cdots O=C$ of the α-helices of myoglobin as listed in the HHDB. Fig. 6(b) shows the geometrical configuration of the $N-H\cdots O=C$ hydrogen bonds, where the distance between H and O atoms is r and the NHO and COH angles are θ_1 and θ_2, respectively. Figs. 6(c)
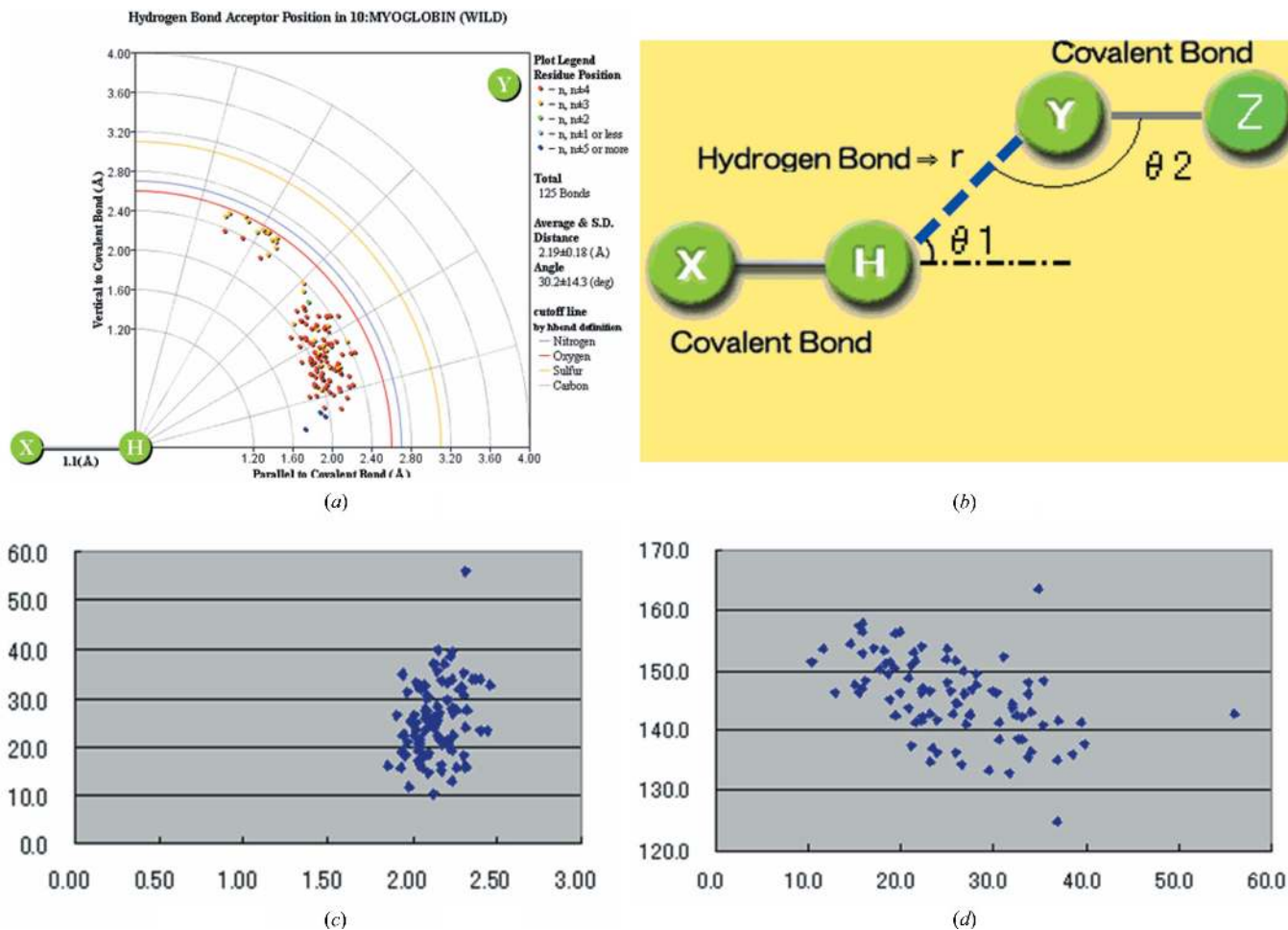

Figure 6
Hydrogen-bond configuration in the α-helix of myoglobin. (a) Schematic statistics of hydrogen bonds $N-H\cdots O=C$ realized in the α-helix of myoglobin provided by HHDB, (b) the geometrical configuration of the $N-H\cdots O=C$ hydrogen bonds, where the length between H and O is r and the angle NHO and COH are θ_1 and θ_2, respectively, (c) and (d) the correlation between r and θ_1 and between θ_1 and θ_2, respectively.

and 6(*d*) show the correlation between *r* and θ_1, and between θ_1 and θ_2, respectively. We have already commented on the non-linearity of the actual N—H···O bond angle θ_1 (Figs. 3 and 6*a*). The significance of Fig. 6(*d*), which shows several H—O═C angles θ_2, is that these angles deviate significantly from the value of 120° normally expected for the lone-pair electrons of an sp^2-hybridized carbonyl O atom. Once again, it seems that the steric constraints imposed by the packing of atoms in a protein molecule, in this case an α-helix, necessitates considerable deviation normally found in small-molecule structures.

4.4. H/D exchange

In most neutron experiments, in order to reduce background scattering from H atoms (which have a large incoherent scattering cross section), protein solutions either are subjected to H_2O/D_2O exchange prior to crystallization or crystals are soaked in D_2O after crystallization from H_2O solution. This allows D atoms to be substituted for many of the H atoms, not only in the solvent water molecules but also at sites containing 'exchangeable' H atoms (mostly H atoms of O—H bonds, as well as most of the N—H bonds exposed to the solvent).

In general, the H atoms of C—H bonds are not easily replaced by D atoms. However, two exceptional examples of H/D exchange were found in a protein structure and in an oligomeric DNA structure.

(*a*) *Histidine*. The first exception involves the $C(\varepsilon_1)$—H bond in the imidazole ring of histidine residues. This happens to be the most acidic C—H bond in amino acids (Matsuo *et al.*, 1972), suggesting the possibility that it might participate in H/D exchange. We have recently confirmed this *via* neutron diffraction: that the $C(\varepsilon_1)$—H bond of histidine does in fact
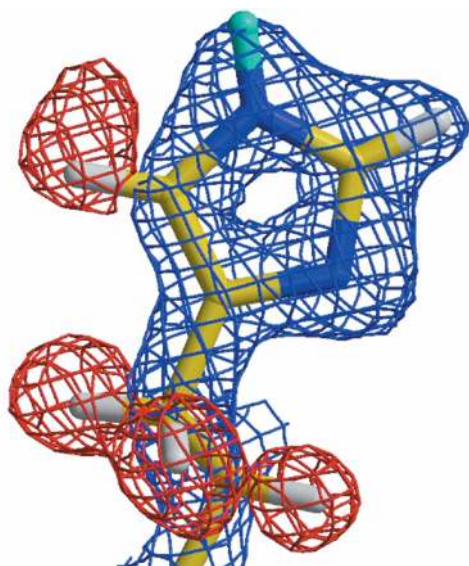

Figure 7
A neutron $|F_o| - |F_c|$ Fourier map around His97 in myoglobin. Note the fact that the $C(\varepsilon1)$—$H(\varepsilon1)$ bond of this imidazole ring has largely been deuterated (blue contours), while the other C—H group of the ring (upper left of the diagram) still has a red (negative) peak corresponding to an undeuterated H atom.

become deuterated in myoglobin (Ostermann & Parak, 2005). Fig. 7 shows a neutron density map around His97 in myoglobin. One clearly sees positive (blue) neutron density contours near $C(\varepsilon_1)$, indicating that the $C(\varepsilon_1)$—H bond is now primarily a C—D bond. In contrast, the other C—H group in the ring (the one on the left in Fig. 7) still retains a negative (red) peak, clearly indicating that it has not been deuterated.

(*b*) *Guanine*. Another example of H/D exchange involving a C—H bond was found in the guanine bases of d(CGCGCG)$_2$, a hexameric duplex form of Z-DNA (Chatake *et al.*, 2005). A neutron $2|F_o| - |F_c|$ Fourier map of one of the six G–C pairs is shown in Fig. 8. The key point we wish to emphasize here is that the C8—H8 bond of guanine is similar to the $C(\varepsilon_1)$—H bond of histidine, in the sense that both are located between two N atoms of a five-membered ring. Previous proton-NMR studies had suggested that the C8—H8 bond can be deuterated in Z-DNA (Brandes & Ehrenberg, 1986), and the present neutron analysis provides direct experimental evidence confirming this. Almost all the guanine bases in the crystal of d(CGCGCG)$_2$ showed evidence of a substantial amount of H/D exchange: an average of 75% of the H8 atoms in this duplex were found to have been replaced by D atoms (Chatake *et al.*, 2005).

4.5. Protonation states of certain amino acid residues

4.5.1. Protonation state of the Nπ (Nδ) and Nτ (Nε) atoms of histidine. The charges of various amino acid side chains depend on the pH: for example, at a high pH (low acidity conditions), carboxylic acids tend to be negatively charged (deprotonated) and amines uncharged (unprotonated). At a low pH (high acidity), the opposite is true. The pH at which exactly half of any ionized amino acid is charged is called the pK_a of that amino acid. These pK_a values of such ionizable amino acid side chains are tabulated in standard textbooks. However, whether a certain amino acid side chain in a protein is charged or not cannot be estimated from standard pH values measured from isolated amino acids in solution because inside a protein the pH may vary significantly depending on the local environment surrounding the particular amino acid residue. The electrically charged states of the amino acid residues are very important in understanding the physiological function of the protein, the interaction between ligands and proteins, molecular recognition, structural stability and so on. Let us take one example: normally, the standard pK_a value of
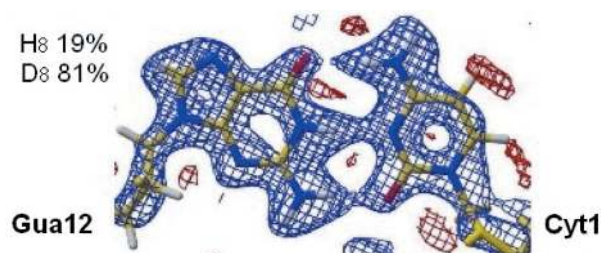

Figure 8
A neutron $2|F_o| - |F_c|$ Fourier map of one of the G–C pairs among six G–C pairs in d(CGCGCG)$_2$, a hexameric duplex form of Z-DNA.
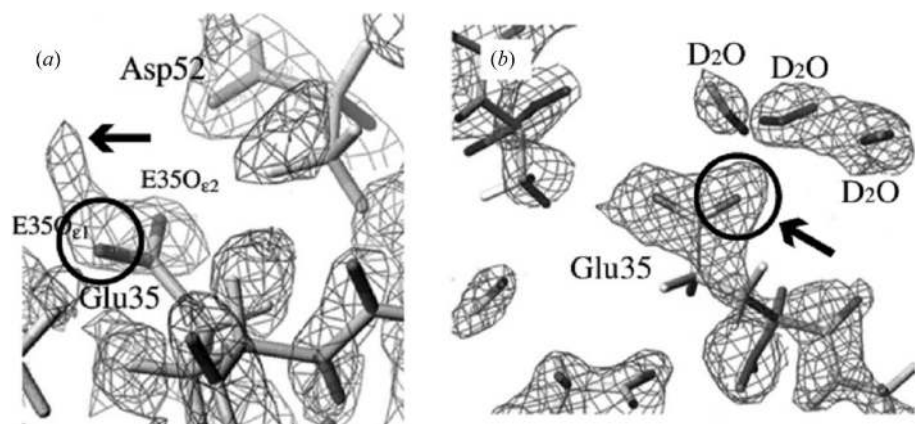
Figure 10
$2|F_o| - |F_c|$ nuclear-density map around the carboxyl group of Glu35 of lysozyme at (a) pD = 4.9 and (b) pD = 7.0. Note the pronounced 'finger' of nuclear density (see arrow) protruding upwards from atom $E35O_{\varepsilon 1}$ on the left (pD = 4.9) and (b) the lack of such a feature on the carboxylate group on the right (pD = 7.0), that is, the O atom has lost its proton. This suggests that residue E35 of lysozyme is in its protonated (deuterated) form under more active (acidic) conditions. [Note: in both plots, atom $E35O_{\varepsilon 1}$ is circled.]
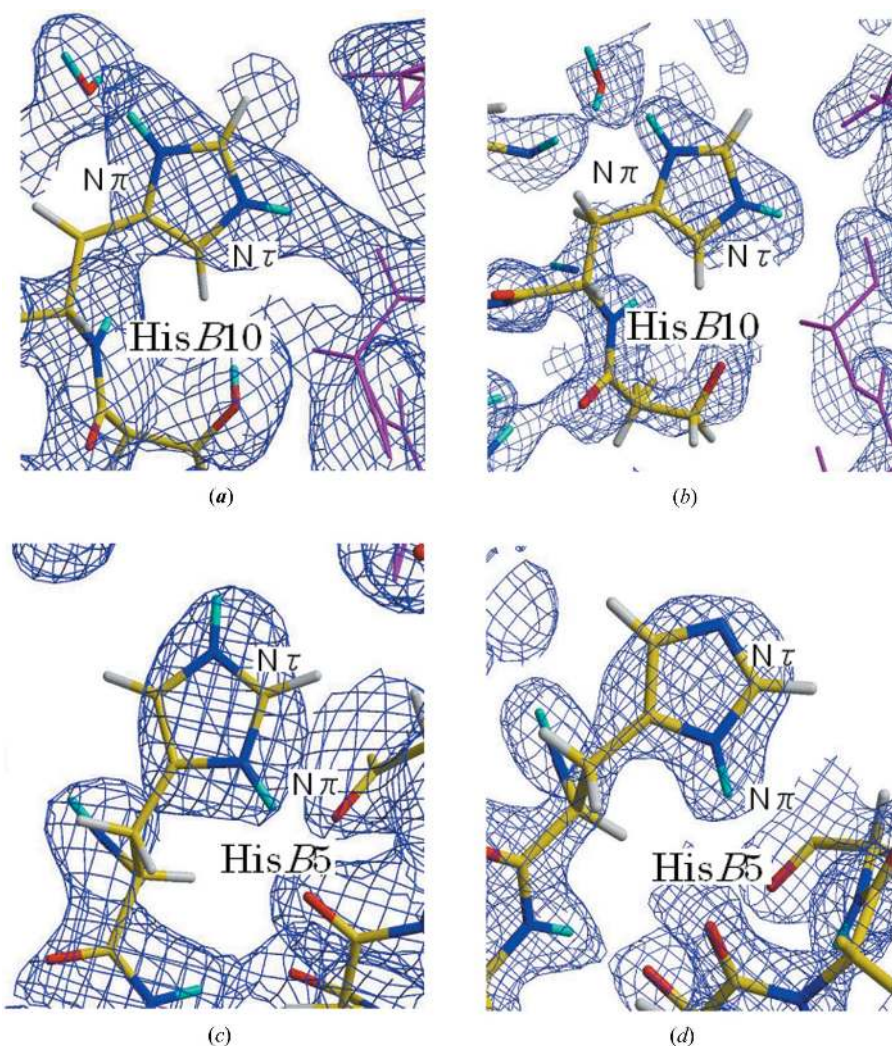

Figure 9
$2|F_o| - |F_c|$ positive nuclear density map of (a) HisB10 at pD 6.6, (b) HisB10 at pD 9.0, (c) HisB5 at pD6.6 and (d) HisB5 at pD 9.0.

histidine is about 6.0. In insulin, there are two histidine amino acid residues (both on chain B) and it is of interest to know whether at pD 6.6 they are protonated or deprotonated since pD 6.6 is clearly close to the borderline pK_a value. Moreover, it is also important to know which of the two N atoms (Nπ, Nτ) in the imidazole ring of histidine are ionized in order to discuss the physiological function of insulin. [In fact, HisB10 is a zinc ion-binding site (Smith et al., 2003).] X-ray structural studies of insulin at various pH's have reported that a conformational change of HisB10 is likely to occur at pH's higher (i.e. more basic) than pH 9, but that of HisB5 is expected to be uncharged (Gursky et al., 1992; Diao, 2003). Neutron diffraction experiments of porcine insulin (crystallized at pD 9.0 and 6.6) were performed at room temperature (Maeda et al., 2004; Ishikawa et al., 2007). Figs. 9(a) and 9(b) show the positive neutron density map of HisB10 at pD 6.6 and pD 9.0, respectively, and we have found that both Nπ and Nτ are protonated. On the other hand, the pH dependence of protonation of HisB5 is different from HisB10. Fig. 9(c) shows the positive neutron-density map of the imidazole ring of HisB5 at pD 6.6 in which both Nτ and Nπ are again found to be protonated. However, at pD = 9.0 the Nτ atom is not protonated (Fig. 9d) and only the Nπ atom is protonated. Thus, one can hypothesize that the environments of HisB10 and HisB5 are distinctly pH-dependent.

4.5.2. Proteins with enzymatic activity and results from other macromolecules. The enzymatic activity of lysozyme, a saccharide-cleaving enzyme, is maximal at pH 5 and is lower at pH 7. It has been postulated that, at pH 5, the carboxyl group of Glu 35 is protonated, and that it is this proton which is transferred to the O atom of the bound substrate (an oligosaccharide) during the hydrolysis process. During the reaction, another acidic residue, Asp52, is postulated to remain in its dissociated (anionic) state (Phillips, 1966). To elucidate the role of H

atoms in this reaction, neutron diffraction experiments of hen-egg-white lysozyme have been carried out, using crystals which have been grown at different acidities. The results, collected at different instruments and different times, illustrate this difference; specifically, pD = 4.9 (data collected on BIX; Maeda *et al.*, 2001) and pD = 7.0 (data collected on LADI; Niimura *et al.*, 1997). The $2|F_o| - |F_c|$ nuclear-density maps around the carboxyl group of Glu35 are shown at pD 4.9 (Fig. 10*a*) and at pD 7.0 (Fig. 10*b*), respectively. At pD 4.9, the Fourier map shows a positive region (arrow in Fig. 10*a*) extending beyond (*i.e.* attached to) the position of the O atom of the carboxyl group labeled $E35O_{\varepsilon 1}$, suggesting that this carboxyl O atom (circled) is protonated at pD 4.9 (Fig. 10*a*). On the other hand, at pD 7.0, it can be seen that this residue is deprotonated (circled atom in Fig. 10*b*), and in the place formally occupied by an H atom there is a water molecule instead. The observation that the Glu35 catalytic site is deprotonated at pD 7.0 (Fig. 10*b*) rationalizes why lysozyme has significantly reduced activity at neutral conditions. Thus, our results suggest that Glu35 is the site of the enzymatically active proton that is subsequently transferred to the O atom of

the carbohydrate substrate during the hydrolysis process. Incidentally, our results at the less acidic pH (7.0) (Niimura *et al.*, 1997) are consistent with the conclusions of an earlier neutron investigation, as described in a paper by Mason and co-workers many years ago (Mason *et al.*, 1984).

Results from other groups, of course, deserve a mention. Other published structures include (*a*) fungal endothiapepsin (an aspartate protease) complexed with a transition-state analog based on neutron data collected using LADI (Cooper & Myles, 2000; Coates *et al.*, 2001), (*b*) the enzyme xylose isomerase (XI) at high resolution (*ca* 2.0 Å) using data acquired from the PCS diffractometer in Los Alamos (Hanson, 2004). In particular, (*c*) aldose reductase (an enzyme of paramount importance in diabetes) (Hazemann *et al.,* 2005) deserves special mention because even an ultra-high-resolution X-ray data set (0.66 Å) was unable to supply information on the catalytically active H atoms in the active site of this important enzyme. And as in the case of (*d*) concanavalin A (Habash *et al.*, 2000), aldose reductase has been the subject of careful comparison between neutron diffraction (Hazemann *et al.*, 2005) and ultra-high-resolution X-ray results (Howard *et al.*, 2004). Other results include (*e*) cytochrome P450cam (LADI), (*f*) amicyanin (PCS), (*g*) dihydrofolate reductase (DHFR), (*h*) protocatechuate 3,4-dioxygenase (3,4-PCD), a Trp repressor, (*i*) a DNA-binding protein (Daniels *et al.*, 2003) (data collected at the LADI diffractometer at ILL) and (*j*) the W3Y mutant of rubredoxin from *Pyrococcus furiosus* (Li *et al.*, 2004) (data collected at the PCS diffractometer at LANSCE). Because of lack of space, it will not be possible to discuss these structure determinations in this review, and the reader is referred to the original publications for more details.

4.6. Water molecules of hydration

4.6.1. Classification. We have categorized observed water molecules into the following classes based on their appearance in Fourier maps (Chatake *et al.*, 2003; Niimura *et al.*, 2004): (1) triangular shape, (2) ellipsoidal stick shape, and (3) spherical shape, and in some cases molecules of the second category (ellipsoidal stick shapes) can be further subclassified as either (2*a*) short or (2*b*) long. We found that this classification conveniently reflects the degree of disorder and/or dynamic behavior of a water molecule. Full details are given in the original publi-
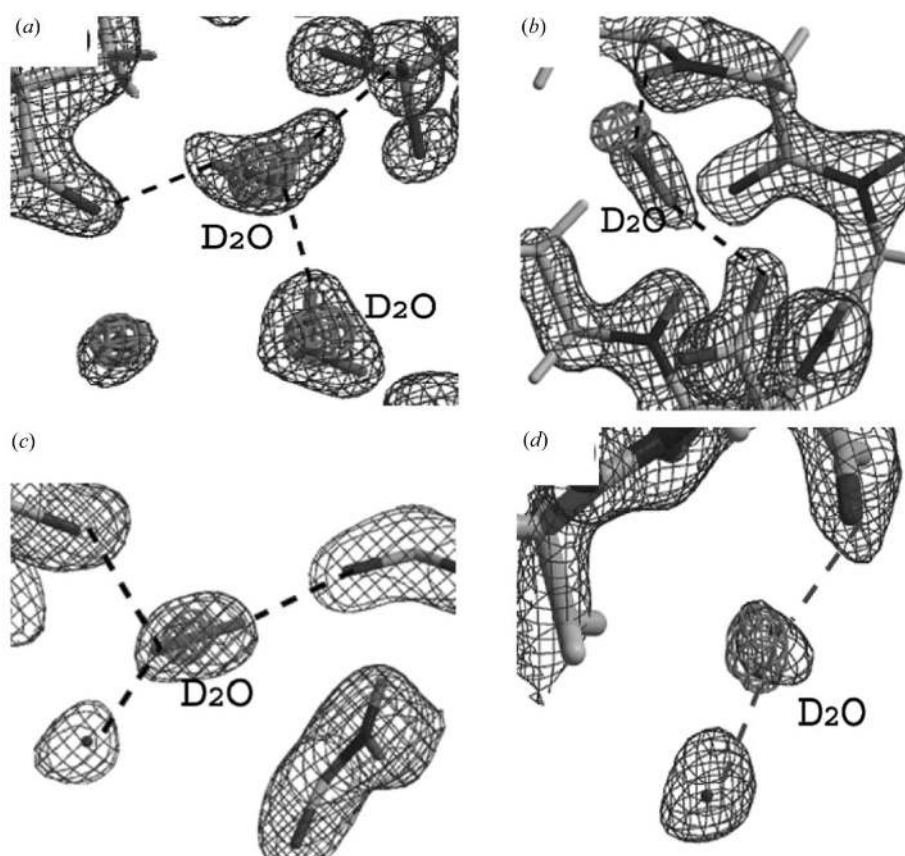

Figure 11
$2|F_o| - |F_c|$ neutron Fourier maps of water molecules of hydration for a rubredoxin mutant and and a myoglobin. Examples shown are peaks having: (*a*) a triangular shape, (*b*) a short ellipsoidal shape, (*c*) a long ellipsoidal shape and (*d*) a spherical shape. In these figures, the central peak (inner contours) embedded in the middle of each water contour corresponds to the O-atom position derived from X-ray data. Note that, from the outer contours, in (*a*) all atoms of the central D_2O molecule are visible, whereas in the other diagrams only some of the solvent atoms are visible in the neutron maps: O and D atoms in (*b*), two D atoms in (*c*), and only O atoms in (*d*).

feature articles

cations (Chatake *et al.*, 2003) but the essential conclusions are summarized in Fig. 11. A triangular shape indicates a completely ordered water molecule, with all three atoms located (Fig. 11*a*), while a short ellipsoidal shape indicates that the O—D bond is visible while the second D atom is disordered (Fig. 11*b*). A long ellipsoidal contour (which is rarely observed) shows two ordered D atoms with the central O atom disordered (Fig. 11*c*), while the very common spherical shape (Fig. 11*d*) corresponds to a completely disordered water molecule. In an X-ray map, all four types of water molecule would appear simply as spherical peaks, and hence it is apparent that neutron maps are informative about hydration structure, especially concerning the water molecules near the surface of a protein, which are much more likely to be ordered.

4.6.2. Oligomeric DNA duplexes. It has long been suspected that the structure and function of a DNA duplex can be strongly dependent on its degree of hydration. In order to investigate the hydration structure of small oligomeric segments of duplex B-DNA (Arai *et al.*, 2005) and Z-DNA (Chatake *et al.*, 2005), neutron diffraction studies have been carried out on such samples using the BIX-3 and BIX-4 single-crystal diffractometers at JAEA.

(A) The B-DNA decamer d(CCATTAATGG)₂

Previous X-ray studies of oligomeric B-DNA have shown that the hydration pattern in the minor groove, especially those near A–T-tract sequences, is relatively well ordered (Wing *et al.*, 1980; Tereshko *et al.*, 1999; Minasov *et al.*, 1999). In those X-ray studies, the observed hydration structure was inferred from the network of O atoms of the water molecules. Goodsell *et al.* (1994) have described the 2.3 Å resolution X-ray structure of the d(CCATTAATGG)₂ duplex and have discussed the implications of B-DNA bending at T–A sites. In that paper, the hydration structure in the minor groove of d(CCATTAATGG)₂ has also been described, in which hydrogen bonds from water to DNA were assumed to exist for all defined O···O distances less than 3.5 Å. From our 3.0 Å neutron study (Arai *et al.*, 2005), the observed water network in the minor groove of d(CCATTAATGG)₂ is shown in Fig. 12. In many cases, the orientations of water molecules can be deduced from the positions of the D atoms, and it was observed that the DNA–water interactions are actually quite complicated: the complexity of the hydration pattern in the minor groove is derived from the extraordinary variety of orientations of the water molecules (Fig. 12).

(B) The Z-DNA hexamer d(CGCGCG)₂

Z-DNA has a zigzag arrangement of the backbone atoms in an unusual left-handed helical arrangement. Originally viewed as a structural oddity, Z-DNA is now believed to play a significant biological role (Wang *et al.*, 1979). The Z-DNA hexameric duplex d(CGCGCG)₂ was investigated by high-resolution X-ray crystallographic analysis (0.6–1.0 Å) and the positions of the O atoms of the water molecules in the hydration shell were determined precisely (Egli *et al.*, 1991). Nevertheless, even at this high resolution the orientational information of water molecules could not be obtained from the X-ray data. We have obtained the 1.8 Å resolution struc-

ture of this duplex by neutron crystallographic analysis (Chatake *et al.*, 2005), which showed 44 water molecules, of which 29 possess a triangular (ordered) shape. The remaining 15 water molecules have a simple spherical (disordered) shape. An interesting relationship was found between the orientational disorder of the water molecules and their locations: almost all water molecules in the minor groove were well ordered in the crystal, whereas about half of the water molecules in the major groove were rotationally disordered. This type of disorder was also observed in the neutron crystal structure of B-DNA discussed in the previous section (Arai *et al.*, 2005) which also showed the waters in the minor groove to be more highly structured than those in the major groove. The complicated hydrogen-bonding networks in the hydration shells of d(CGCGCG)₂ have been discussed in detail (Chatake *et al.*, 2005). Earlier, in section §4.4(*b*), we had pointed out the fact that most of the H atoms attached to the C8 positions of the guanine rings in this molecule had undergone H/D exchange (Fig. 8).

5. 'Next generation' spallation neutron sources

An exciting new development currently under way is the construction of high-intensity state-of-the-art pulsed neutron facilities. These are the Spallation Neutron Source (SNS) at Oak Ridge, Tennessee, USA, and the Japan Proton Accelerator Research Complex (J-PARC) in Ibaraki, Japan. Both
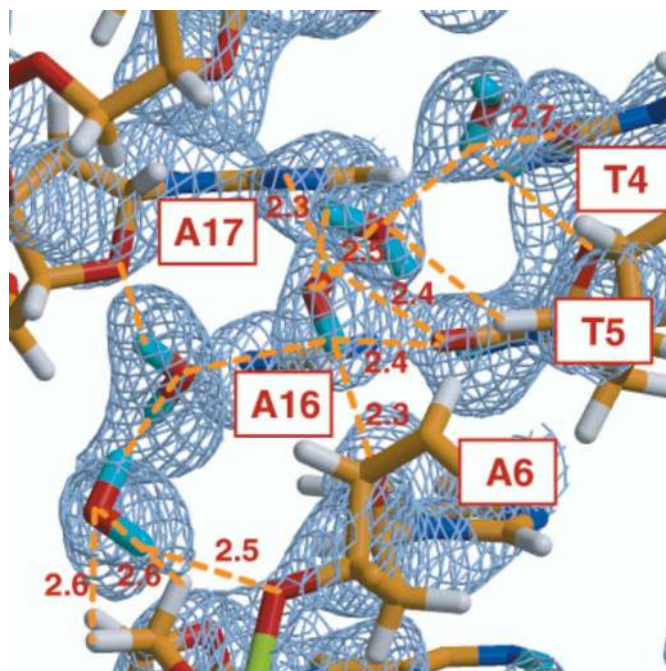

Figure 12
$2|F_o| - |F_c|$ neutron Fourier map (3.0σ level) in the minor groove of the B-DNA decamer d(CCATTAATGG)₂. The bases Thy4, Thy5, Ade6, Ade16 and Ade17 are labeled in this figure. Broken lines indicate the hydrogen bonds. Note that a complicated (but ordered) water network has been observed to be tightly associated with the minor groove. In contrast, many of the water molecules in the major groove (not shown here) are disordered.

SNS and J-PARC promise to deliver neutron intensities that are at least an order of magnitude (or more) higher than those at existing sources: traditional nuclear reactors such as those at ILL in Grenoble (France) and JRR-3 in JAEA (Japan), as well as existing pulsed (spallation) neutron sources such as IPNS in Argonne (USA), ISIS at Didcot (England) and LANSCE at Los Alamos (US).

At both new high-intensity spallation sources, diffractometers for macromolecular crystallography are either being built or planned. At J-PARC, construction of the so-called BIX-P1 instrument has been in progress since 2004 (Tanaka *et al.*, 2006), while at the SNS a corresponding instrument, to be called MaNDi (Macromolecular Neutron Diffractometer) (Schultz *et al.*, 2005) is most probably going to be built. These two diffractometers (BIX-P1 and MaNDi) will differ somewhat in the type of moderator used (this is the material situated between the neutron target source and the sample crystal; its choice determines the intensities and wavelength distribution of the neutrons used in the diffraction experiment). BIX-P1 will use neutrons from a coupled moderator, which has a higher intensity but a wider wavelength range; while the proposal is that MaNDi will use a decoupled moderator, which is less intense but has a sharper distribution. One reason for this difference is that J-PARC has a lower pulsing frequency (25 Hz) than SNS (60 Hz). Overlap between pulses that are separated more widely in time is therefore a less serious problem at J-PARC than the closely spaced pulses at SNS.

Regardless of these technical differences, the implications of the development of SNS and J-PARC on macromolecular research are obvious: if intensities can be increased by an order of magnitude or two, then in principle it will be possible to study crystals that are correspondingly smaller in volume. The consequences of this development will be an enormous dramatic decrease in crystal size, which will in turn have an absolutely dramatic effect on the usefulness of single-crystal neutron diffraction in the field of structural biology. It is hoped that BIX-P1 and MaNDi will both become operational around the period 2009–2010.

6. Conclusions

(I) Activity in the single-crystal neutron analysis of proteins has increased dramatically in recent years: at JAEA (Japan), BIX-type diffractometers equipped with neutron imaging plates (NIP) and the LADI diffractometer at ILL (France) have been successfully used to acquire high-resolution neutron diffraction data (up to 1.5 Å) for the location of H atoms and molecules of hydration in proteins and oligomeric nucleic acids. The PCS instrument at Los Alamos (USA), using a pulsed neutron source, has also begun producing a steady stream of results. In addition, a new improved version of LADI at ILL (called LADI-3) is already operational and has substantially improved data-collection capabilities at Grenoble.

(II) Crystallization phase diagrams are very helpful for systematically determining the optimal conditions for growing large single crystals for neutron biomacromolecular crystal-lography. A new method for determining these phase diagrams, involving the careful control of protein solution volumes *via* a syringe and a dialysis membrane, is being developed. In addition, a rational way to assess the quality of protein crystals, based on the analysis of relative Wilson plots, has been proposed.

(III) A Hydrogen and Hydration in Proteins Data Base (HHDB) that catalogs H atoms and hydration water molecules in proteins, as determined by neutron protein crystal-lography, has been created and is already publicly available.

(IV) Certain structural features related to H atoms have been determined. (*a*) Since almost all the protein H-atom positions can be identified experimentally, the geometrical details of certain types of hydrogen bonds can be visualized; (*b*) since the neutron scattering process distinguishes deuterium from hydrogen, information regarding the H/D exchange behavior of proteins can be obtained and in some cases it has been found to be comparable to the results obtained from NMR spectroscopy; and (*c*), perhaps most importantly as far as mechanistic implications are concerned, the identification of protonation and deprotonation states of certain important amino acid residues can be carried out.

(V) The hydration structure around proteins and the hydration networks around DNA oligomers have been successfully characterized in several outstanding cases. However, unlike the H atoms of the protein itself, it is not always possible to determine the positions of all the D atoms of water molecules. In general, as one gets further away from the protein surface, the water molecules of solvation tend to become more disordered, and less detail becomes observable.

(VI) Future protein crystallographic instruments, such as BIX-P1 and the proposed ManDi diffractometer, at 'next generation' high-intensity spallation (pulsed) neutron sources, promise further gains in performance that will further extend the size and complexity of samples (and, more importantly, reduce crystal sizes) that can be studied by neutron crystal-lography. In addition to that, new developments are occurring at existing neutron sources to build more efficient neutron protein diffractometers (*e.g.* LADI-3 at the ILL). And for target station 2 (TS2) at ISIS, the new proposed initiative of a Large Molecule Neutron Diffractometer (LMX++) will enable new types of neutron protein crystallography experiments to be possible for European researchers as well.

(VII) Deuteration facilities are being built at several neutron facilities (some of which, like the one at ILL in France, are already operational). These will further help to reduce the crystal sizes necessary for single-crystal neutron diffraction analysis.

feature articles

References

Arai, S., Chatake, T., Minezaki, Y. & Niimura, N. (2002). *Acta Cryst.* D**58**, 151–153.

Arai, S., Chatake, T., Ohhara, T., Kurihara, K., Tanaka, I., Suzuki, N., Fujimoto, Z., Mizuno, H. & Niimura, N. (2005). *Nucleic Acids Res.* **33**, 3017–3024.

Arai, S., Chatake, T., Suzuki, N., Mizuno, H. & Niimura, N. (2004). *Acta Cryst.* D**60**, 1032–1039.

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

Brandes, R. & Ehrenberg, A. (1986). *Nucleic Acids Res.* **14**, 9491–9508.

Budayova-Spano, M., Dauvergne, F., Audiffren, M., Bactivelane, T. & Cusack, S. (2007). *Acta Cryst.* D**63**, 339–347.

Chatake, T., Kurihara, K., Tanaka, I., Tsyba, I., Bau, R., Jenney, F. E., Adams, M. W. W. & Niimura, N. (2004). *Acta Cryst.* D**60**, 1364–1373.

Chatake, T., Mizuno, N., Voordouw, G., Higuchi, Y., Arai, S., Tanaka, I. & Niimura, N. (2003). *Acta Cryst.* D**59**, 2306–2309.

Chatake, T., Ostermann, A., Kurihara, K., Parak, F. G. & Niimura, N. (2003). *Proteins*, **50**, 516–523.

Chatake, T., Tanaka, I., Umino, H., Arai, S. & Niimura, N. (2005). *Acta Cryst.* D**61**, 1088–1098.

Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2001). *Biochemistry*, **40**, 13149–13157.

Cooper, J. B. & Myles, D. A. A. (2000). *Acta Cryst.* D**56**, 246–248.

Daniels, B. V., Myles, D. A. A., Forsyth, V. T. & Lawson, C. L. (2003). *Acta Cryst.* D**59**, 136–138.

Diao, J. (2003). *Acta Cryst.* D**59**, 670–676.

Egli, M., Willias, L. D., Gao, Q. & Rich, A. (1991). *Biochemistry*, **30**, 11388–11402.

Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1994). *J. Mol. Biol.* **239**, 79–96.

Gursky, O., Li, Y., Badger, J. & Casper, D. L. D. (1992). *Biophys. J.* **61**, 604–611.

Habash, J., Raftery, J., Nuttall, R., Price, H. J., Wilkinson, C., Kalb (Gilboa), A. J. & Helliwell, J. R. (2000). *Acta Cryst.* D**56**, 541–550.

Hanson, B. L. (2004). *Proc. Natl Acad. Sci USA*, **101**, 16393–16394.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Cryst.* D**61**, 1413–1417.

Helliwell, J. R. (1997). *Nature Struct Biol.* **4**, 874–876.

Helliwell, J. R. & Wilkinson, C. (1994). *Neutron and Synchrotron Radiation for Condensed Matter Studies: Applications to Soft Condensed Matter and Biology*, edited by J. Baruchel, J. L. Hodeau, M. S. Lehmann, J. R. Regnard & C. Schlenker, Vol. III, ch. XII. Berlin: Springer Verlag.

Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.

Ishikawa, T., Chatake, T., Ohnishi, Y., Tanaka, I., Kurihara, K., Kuroki, R. & Niimura, N. (2007). *Chem. Phys.* In the press.

Kossiakoff, A. A. (1985). *Ann. Rev. Biochem.* **54**, 1195–1227.

Kossiakoff, A. A. & Spencer, S. A. (1980). *Nature (London)*, **288**, 414–416.

Kurihara, K., Tanaka, I., Chatake, T., Adams, M. W., Jenney, F. E. Jr, Moiseeva, N., Bau, R. & Niimura, N. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 11215–11220.

Kurihara, K., Tanaka, I., Refai Muslih, M., Ostermann, A. & Niimura, N. (2004). *J. Synchrotron Rad.* **11**, 68–71.

Li, X., Langan, P., Bau, R., Tsyba, I., Jenney, F. E., Adams, M. W. W. & Schoenborn, B. P. (2004). *Acta Cryst.* D**60**, 200–202.

Maeda, M., Chatake, T., Tanaka, I., Ostermann, A. & Niimura, N. (2004). *J. Synchrotron Rad.* **11**, 41–44.

Maeda, M., Fujiwara, S., Yonezawa, Y. & Niimura, N. (2001). *J. Phys. Soc. Jpn Suppl. A*, **70**, 403–405.

Mason, S. A., Bentley, G. A. & McIntyre, G. J. (1984). *Neutrons in Biology*, edited by B. P. Schoenborn, pp. 323–334. New York: Plenum Press.

Matsuo, H., Oe, M., Sakiyama, S & Narita, K. (1972). *J. Biochem.* **72**, 1057–1060.

Minasov, G., Tereshko, V. & Egli, M. (1999). *J. Mol. Biol.* **291**, 83–99.

Myles, D. A. A., Bon, C., Langan, P., Cipriani, F., Castagna, J. C., Lehmann, M. S. & Wilkinson, C. (1998). *Physica (Utrecht)*, B**241&243**, 1122–1130.

Niimura, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 602–608.

Niimura, N., Arai, S., Kurihara, K., Chatake, T., Tanaka, I. & Bau, R. (2006). *Cell. Mol. Life Sci.* **63**, 285–300.

Niimura, N., Chatake, T., Kurihara, K. & Maeda, M. (2004). *Cell Biochem. Biophys.* **40**, 351–370.

Niimura, N., Karasawa, Y., Tanaka, I., Miyahara, J., Takahashi, 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.

Ostermann, A. & Parak, F. (2005). *Hydrogen- and Hydration-Sensitive Structural Biology*, edited by N. Niimura, H. Mizuno, J. R. Helliwell & E. Westhof, pp. 87–102. Tokyo: KubaPro.

Ostermann, A., Tanaka, I., Engler, N., Niimura, N. & Parak, F. G. (2002). *Biophys. Chem.* **95**, 183–193.

Phillips, D. C. (1966). *Sci. Am.* **215**, 75–80.

Phillips, S. E. V. & Schoenborn, B. P. (1981). *Nature (London)*, **292**, 81.

Saridakis, E. E. G., Shaw Stewart, P. D., Lloyd, L. F. & Blow, D. M. (1994). *Acta Cryst.* D**50**, 293–297.

Sauter, C., Lorber, B., Kern, D., Cavarelli, J., Moras, D. & Giegé, R. (1999). *Acta Cryst.* D**55**, 149–156.

Schoenborn, B. P. (1969). *Nature (London)*, **224**, 143–146.

Schoenborn, B. P. (1985). *Methods Enzymol.* **114**, 510–529.

Schoenborn, B. P. & Langan, P. (2004). *J. Synchrotron Rad.* **11**, 80–82.

Schultz, A. J., Thiyagarajan, P., Hodges, J. P., Rehm, C., Myles, D. A. A., Langan, P. & Mesecar, A. D. (2005). *J. Appl. Cryst.* **38**, 964–971.

Smith, G. D., Pangborn, W. A. & Blessing, R. H. (2003). *Acta Cryst.* D**59**, 474–482.

Tanaka, I., Kurihara, K., Chatake, T. & Niimura, N. (2002). *J. Appl. Cryst.* **35**, 34–40.

Tanaka, I., Niimura, N., Ozeki, T., Ohhara, T., Kurihara, K., Kusaka, K., Morii, Y., Aizawa, K., Arai, M., Kasao, T., Ebata, K. & Takano, Y. (2006). ICANS-XVII Proceedings, LA-UR-3904, Vol. III, pp. 937–945.

Tereshko, V., Minasov, G. & Egli, M. (1999). *J. Am. Chem. Soc.* **122**, 470–471.

Tsyba, I. & Bau, R. (2002). *Chemtracts*, **15**, 233–257.

Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979). *Nature (London)*, **282**, 680–686.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980). *Nature (London)*, **287**, 755–758.

Wlodawer, A. (1982). *Prog. Biophys. Mol. Biol.* **40**, 115.

Wlodawer, A. & Sjolin, L. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 1418–1422.