Cholesterol oxidase: ultrahigh-resolution crystal structure and multipolar atom model-based analysis

Bartosz Zarychta, Artem Lyubimov, Maqsood Ahmed, Parthapratim Munshi, Benoît Guillot, Alice Vrielink and Christian Jelsch

Cholesterol oxidase: ultrahigh-resolution crystal structure and multipolar atom model-based analysis

Bartosz Zarychta,[a,b] Artem Lyubimov,[c] Maqsood Ahmed,[a,d] Parthapratim Munshi,[a,e] Benoît Guillot,[a] Alice Vrielink[f] and Christian Jelsch[a]*

[a]Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2), CNRS, UMR 7036, Institut Jean Barriol, Faculté des Sciences et Technologies, Université de Lorraine, BP 70239, 54506 Vandoeuvre-lès-Nancy CEDEX, France, [b]Department of Chemistry, Opole University, ul. Oleska 48, 45-052 Opole, Poland, [c]Howard Hughes Medical Institute, Stanford, CA 94305-5432, USA, [d]Department of Chemistry, The Islamia University of Bahawalpur, Bahawalpur, Pakistan, [e]Department of Chemistry and Center for Informatics, School of Natural Sciences, Shiv Nadar University, NH91, Tehsil Dadri, Gautam Buddha Nagar, Uttar Pradesh 201 314, India, and [f]School of Chemistry and Biochemistry, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. *Correspondence e-mail: christian.jelsch@univ-lorraine.fr

Examination of protein structure at the subatomic level is required to improve the understanding of enzymatic function. For this purpose, X-ray diffraction data have been collected at 100 K from cholesterol oxidase crystals using synchrotron radiation to an optical resolution of 0.94 Å. After refinement using the spherical atom model, nonmodelled bonding peaks were detected in the Fourier residual electron density on some of the individual bonds. Well defined bond density was observed in the peptide plane after averaging maps on the residues with the lowest thermal motion. The multipolar electron density of the protein–cofactor complex was modelled by transfer of the ELMAM2 charge-density database, and the topology of the intermolecular interactions between the protein and the flavin adenine dinucleotide (FAD) cofactor was subsequently investigated. Taking advantage of the high resolution of the structure, the stereochemistry of main-chain bond lengths and of $C{=}O \cdots H{-}N$ hydrogen bonds was analyzed with respect to the different secondary-structure elements.

1. Introduction

Cholesterol oxidase (ChOx) is a 55 kDa secreted bacterial enzyme that binds the flavin adenine dinucleotide (FAD) cofactor (reviewed in Vrielink & Ghisla, 2009). Two main forms of the enzyme have been identified (for a review, see Sampson & Vrielink, 2003). In form I the FAD cofactor is noncovalently bound to the protein, whereas in form II the FAD group is covalently bound to His121. The noncovalent forms of the flavoprotein are members of the GMC (glucose, methanol, choline) oxidoreductase family of flavoenzymes (Cavener, 1992), in which two residues, His447 and Asn485, which have been shown to be involved in substrate oxidation, are semi-conserved (Kass & Sampson, 1998; Yue *et al.*, 1999; Yin *et al.*, 2001).

This water-soluble enzyme is mainly used in serum cholesterol assays. Cholesterol oxidases are bifunctional enzymes that catalyze two reactions in a single active site. The first is the oxidation of cholesterol to cholest-5-en-3-one and the second is its isomerization to cholest-4-en-3-one (Stadtman *et al.*, 1954; Flegg, 1973). The oxidation of the hydroxyl group on the steroid substrate requires a FAD cofactor that is concomitantly reduced (Fig. 1). The cofactor is re-oxidized by the reduction of molecular oxygen to generate hydrogen peroxide. The presence of hydrogen peroxide can be detected through a
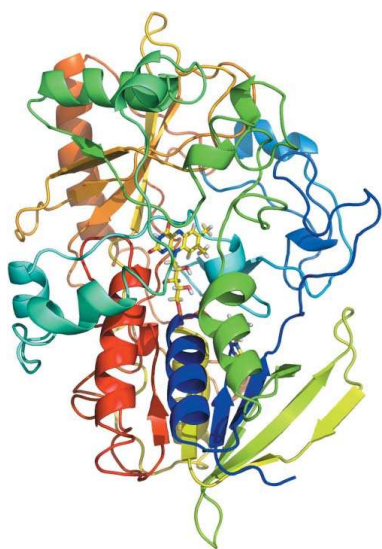
colorimetric assay, which forms the basis of clinical assays for the determination of serum cholesterol.

ChOx-utilizing bacteria can be classified into two types: nonpathogenic and pathogenic. Nonpathogenic bacteria, for example streptomycetes and the fast-growing mycobacteria, can utilize cholesterol as their carbon source and up-regulate the expression of cholesterol oxidase in the presence of cholesterol (Av-Gay & Sobouti, 2000). Pathogenic bacteria, for example *Rhodococcus equi*, require cholesterol oxidase for infection of the host macrophage; cholesterol also regulates the expression of the enzyme in these organisms (Navas *et al.*, 2001). The role of the enzyme in pathogenesis is assumed to be the alteration of the physical structure of the lipid membrane by converting cholesterol into cholest-4-en-3-one (Shen *et al.*, 1997).

ChOx also possesses insecticidal properties against *Coeloptera* larvae, which are agricultural pests, and is being developed for use in agricultural crop treatment (Corbin *et al.*, 1998, 2001). The crystal structure of the ChOx protein has been determined at increasing resolutions, under different pH conditions and as different mutant forms (Yue *et al.*, 1999; Lario *et al.*, 2003; Lario & Vrielink, 2003; Lyubimov *et al.*, 2006, 2007, 2009). In all of these reports different aspects of the enzyme structure were explored to understand its catalytic mechanism. It has been reported that the dioxygen molecule reaches the active site *via* a tunnel (Lario *et al.*, 2003). This tunnel, lined with hydrophobic residues, is visible in the type II enzyme at 1.7 Å resolution (Coulombe *et al.*, 2001), whereas in the type I enzyme, in which the cofactor is noncovalently bound to the protein, the tunnel becomes clear only at 0.95 Å resolution (Lario *et al.*, 2003). In the former case, the tunnel is gated by Arg477, whereas in the latter case Asn485 gates the active-site end of the tunnel. Asn485 has also been shown to play a role in stabilizing the reduced flavin cofactor through a C—H···π interaction with the isoalloxazine moiety (Yin *et al.*, 2001; Chen *et al.*, 2008). Various residues close to the active site exist in multiple conformations which are proposed to be involved in the gating of the tunnel. The residues lining the tunnel are hydrophobic in nature (Lario *et al.*, 2003).

The study of the electrostatic potential of molecules is highly important as it provides an opportunity to understand the specificity of protein–ligand binding and also helps in understanding the phenomenon of molecular recognition. Electron density at the subatomic level can now almost routinely be observed for small-molecule crystals studied at ultrahigh resolution, mainly owing to the state-of-the-art equipment available to contemporary crystallographers. In contrast, for larger molecules (including proteins), electron density can be modelled at lower resolution through the application of electron-density transferability principles (Muzet *et al.*, 2003; Guillot *et al.*, 2008; Dominiak *et al.*, 2009; Johnas *et al.*, 2009).

In this method, the electron-density parameters studied from small molecules are stored in the ELMAM2 database of multipolar atoms (Domagała *et al.*, 2011, 2012). Upon database transfer, chemical atom types are identified and their electron-density parameters are retrieved. Derived molecular

properties such as electrostatic potential can subsequently be computed within the transferability approximation. This method is particularly useful in the case of proteins, as they rarely diffract to subatomic resolution. Only about 20 of the protein structures currently reported in the PDB (Bernstein *et al.*, 1977; Berman, 2000) have been refined to resolutions at or higher than 0.75 Å. The topology of covalent bonds and hydrogen bonds in the hen egg-white lysozyme active-site structure at 0.65 Å resolution have recently been analyzed. The Bader (1991) quantum theory of atoms in molecules was applied to electron-density maps obtained from ELMAM2 database multipole transfer (Held & van Smaalen, 2014).

In the present study, the structure of cholesterol oxidase is refined using synchrotron data measured to 0.74 Å resolution. This is the highest resolution data set that has been achieved for this protein to date and is exceptional for a protein of this size (more than 500 residues). The high-resolution data yield good-quality electron-density maps in which it is possible to precisely locate many of the H atoms using difference Fourier
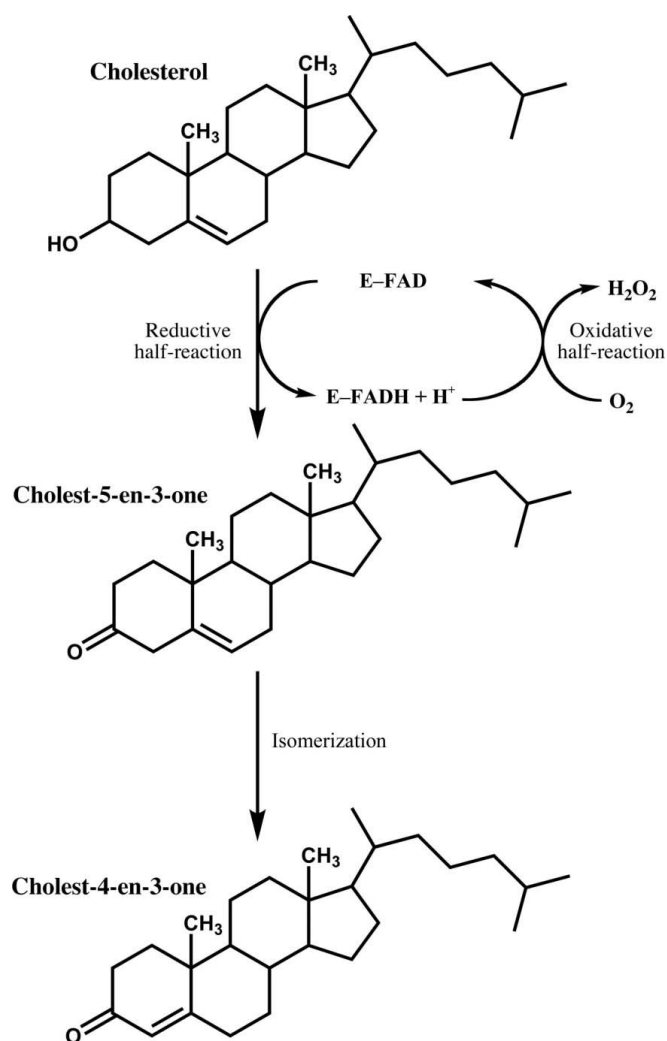

Figure 1
Scheme showing the oxidation and isomerization reactions of cholesterol oxidase.

maps, including the hydrogen-bonding networks connecting water molecules. The structure displays an average atomic thermal motion that is too high compared to the requirements for charge-density refinement. Conversely, in human aldose reductase a large number of atoms had $B_{eq} < 4$ Å^2, which enabled a charge-density refinement of the polypeptide main chain using chemical equivalence constraints (Guillot *et al.*, 2008). The precision of atomic coordinates, including H atoms, however, supports the application of the transferability principle towards the ChOx model. Therefore, an electron-density model of ChOx was constructed based on the ELMAM2 multipolar atoms database (Domagała *et al.*, 2011, 2012). This led to a computation of the derived electrostatic potential at the interacting surface between the protein and the FAD cofactor. Furthermore, the topological properties of protein–cofactor interactions as well as the intramolecular interactions of the FAD could be quantified. The insights obtained from this work provide an overview regarding the protein–FAD cofactor recognition and electrostatic interaction.

2. Materials and methods

2.1. Crystallization

Recombinant *Streptomyces* sp. ChOx was expressed and purified as described previously (Yue *et al.*, 1999). The pure enzyme was concentrated to 6.5 mg ml^{-1} in 10 mM HEPES buffer. Crystals were grown by the hanging-drop vapour-diffusion method in 24-well crystallization plates (Hampton Research, California, USA). The crystallization solution was composed of 9–11% poly(ethylene glycol) (PEG) molecular weight 8000, 75 mM MnSO$_4$, 100 mM sodium cacodylate buffer pH 5.2. Drops consisting of 1.0 μl crystallization solution and 1.0 μl wild-type protein solution were suspended over a well containing 1.0 ml crystallization solution. Drops were streak-seeded using ChOx crystals grown previously. Small crystals appeared overnight and grew to a suitable size for data collection within two weeks.

2.2. Crystallographic data collection and processing

X-ray diffraction data were collected from a total of five crystals on beamlines 9-2 and 11-1 at Stanford Synchrotron Radiation Laboratory (SSRL). Four crystals were used to collect ultrahigh-resolution data (1.0–0.74 Å). A fifth crystal was used to collect three low exposure time sweeps of data; firstly to atomic resolution (∼0.95 Å), then to the maximum attainable resolution (∼0.74 Å) and finally, using an atte-nuated beam, to medium resolution (∼1.8 Å). The low-exposure sweeps were necessary to avoid overloaded reflections at low resolution.

Diffraction images were processed with the *d*TREK* software suite (Pflugrath, 1999). The final data-processing statistics are shown in Table 1. The scaled data were converted to MTZ format using the *CCP*4 software suite (Winn *et al.*, 2011) for initial refinement, and 5% of the reflections were randomly selected as a test set and excluded from all refinement steps. The overall completeness of the diffraction data is 78.8%. The

Table 1
Crystallographic data-collection, data-reduction and structure-refinement statistics using a spherical atom model.

R.m.s.d., root-mean-square deviation. Values in parentheses are for the outermost shell.

Data collection	
Space group	$P2_1$
Unit-cell parameters (Å, °)	$a = 51.24$, $b = 72.92$, $c = 63.01$, $\beta = 105.13$
Temperature (K)	100
Total No. of reflections	3141008
No. of unique reflections	465386
Optical resolution (Å)	0.94
Resolution (Å)	0.74 (0.80–0.74)
R_{merge} (%)	5.2 (64.4)
$\langle I/\sigma(I)\rangle$	10.7 (0.7)
Completeness (%)	78.8 (21.2)
Multiplicity	6.75 (1.15)
Refinement	
R_{work}/R_{free} (all data)	0.116/0.123
No. of atoms (iso/aniso)	4309/5143
No. of occupancies refined	2725
No. of $I(hkl)$ data:No. of parameters ratio	7.2
$F_o - F_c$ map r.m.s. value (e Å^{-3})	0.0066
Isotropic B factors (without H) (minimum/average) (Å^2)	
Protein main chain	5.0/8.3 ± 2.3
Protein side chain	5.1/9.8 ± 2.8
FAD	4.7/5.5 ± 0.6
R.m.s.d, bond lengths (Å)	0.013
R.m.s.d, bond angles (°)	1.595

completeness at very low resolution $d < 12.7$ Å is 93% (see Supporting Information). The optical resolution (Urzhumt-seva *et al.*, 2013) of the diffraction data, computed using *SFCHECK* (Vaguine *et al.*, 1999), is 0.94 Å.

2.3. Structure refinement

A previous model of the enzyme refined with data measured to 0.95 Å resolution (Lario & Vrielink, 2003) was used as the starting model for the current refinement of ChOx. The structure was refined with *PHENIX* (Afonine *et al.*, 2012), using the maximum-likelihood refinement method and anisotropic thermal motion parameters for the non-H atoms. The structure was refined using all reflections with no resolution or $I/\sigma(I)$ cutoff, as recommended by Karplus & Diederichs (2012), although the diffraction data beyond 0.8 Å resolution are incomplete and weak. Bulk-solvent correction and anisotropic scaling were applied. Water molecules were refined automatically by *PHENIX* but were also handled manually; their occupancy factors were refined. The refinement details are listed in Table 1. The structure was not refined using a multipolar atom model because, unlike crambin (Jelsch *et al.*, 2000) and aldose reductase (Guillot *et al.*, 2008), the Fourier residual density did not show systematic visible, nonmodelled, deformation electron density on individual covalent bonds.

2.4. Electron-density transfer and analysis

Electronic properties of macromolecular systems can be studied by the use of transferability even if an experimental charge-density analysis is not possible. The protein model obtained from *PHENIX* was imported into the *MoPro*

software (Jelsch *et al.*, 2005) and the electron-density parameters from the ELMAM2 library (Domagała *et al.*, 2011) were transferred to the structure.

After electron-density transfer, the H-atom positions were set to the standard neutron diffraction H—X bond lengths (Allen *et al.*, 2006). On the basis of the transferred electron-density parameters, the electrostatic potential generated by the cofactor FAD and the active-site region of the protein were calculated with the *VMoPro* software using the *MoProViewer* (Guillot, 2011) graphical interface. Atom types that are not present in the ELMAM2 database, notably the atoms of the pyrophosphate group, were derived from the NAD$^+$ cofactor, which has been studied by charge-density analysis (Guillot *et al.*, 2003).

3. Results and discussion

3.1. Protein structure

A ribbon view of the ChOx protein structure is shown in Fig. 2. The structure was refined to R_{work} and R_{free} values of 11.6 and 12.3%, respectively. The final model contains 499 visible amino acids, a FAD cofactor (Fig. 3) and 897 water molecules. The Ramachandran plot (Ramachandran *et al.*, 1963) of the refined protein structure obtained from *MolProbity* (Chen *et al.*, 2010) is shown in Fig. 4. Residue Val217 is slightly outside the favourable regions of the Ramachandran plot, as was seen in previous crystal structures of ChOx. 98% of the amino acids were observed in favoured regions of the Ramachandran plot and 99.8% were observed in allowed regions. The final refined coordinates and structure
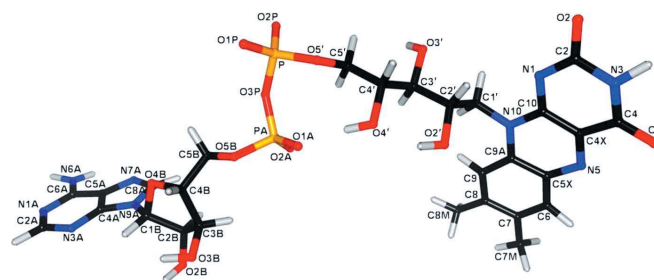
factors have been deposited in the Protein Data Bank (PDB) as entry 4rek.

The $2F_o - F_c$ electron-density map for the FAD cofactor from structural refinement using *PHENIX* is shown in Fig. 5. The well defined FAD cofactor has atoms with B_{eq} factors in the range 4.7–7.8 Å^2. The electron density contoured at 5σ provides a view of the ellipsoidal motion around non-H atoms and gives a qualitative view of the anisotropy of the atomic thermal motions.
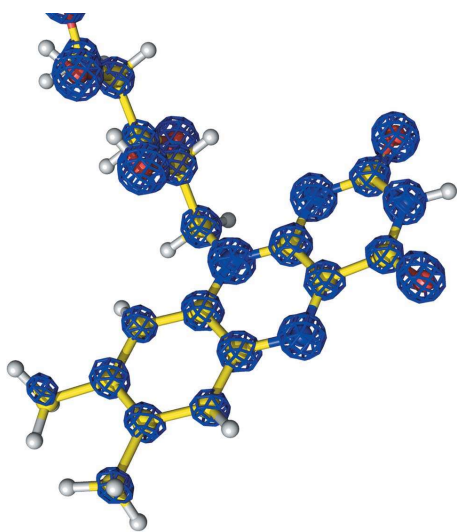
3.2. Protein charge-density model

The very high resolution of the collected data can be related to the B factor obtained from the Wilson plot (Wilson, 1942), $B = 7.3$ Å^2, which is relatively low for a protein. After *PHENIX* refinement based on the spherical atom model, Fourier residual maps were computed to observe non-modelled electron density on the covalent bonds using *MoPro*. A significant number of atoms in cholesterol oxidase have B_{eq} factors as low as 4–7 Å^2. After refinement using all reflections, the Fourier residual map rarely reveals the nonmodelled

Figure 3
Ball-and-stick model of the FAD cofactor shown with the atom-numbering scheme for non-H atoms.

Figure 2
Ribbon cartoon model of the cholesterol oxidase protein structure. The FAD cofactor is shown as a stick model.

Figure 4
Ramachandran plot of the refined ChOx protein structure.

deformation density on covalent bonds. For a few residues with the lowest thermal motion some of the bonding peaks may appear, but they are at the level of random noise (Fig. 6). Bonding electron-density peaks were clearly visible in several protein structures at ultrahigh or atomic resolution: a scorpion toxin at 0.96 Å (Housset *et al.*, 2000; Afonine *et al.*, 2007), crambin at 0.54 Å (Jelsch *et al.*, 2000), RD1 antifreeze protein at 0.62 Å (Afonine *et al.*, 2004), trypsin at 0.8 Å (Schmidt *et al.*, 2003; Afonine *et al.*, 2004), phospholipase at 0.8 Å (Liu *et al.*, 2003; Afonine *et al.*, 2007) and human aldose reductase at 0.66 Å (Guillot *et al.*, 2008). The application of high-order refinement, where only the subatomic resolution reflections are used (typically $d < 0.7$ Å), permits one to highlight the remaining bonding density (Hansen & Coppens, 1978) in Fourier maps computed using only the lower resolution reflections.

To decrease the noise in Fourier residual maps and to highlight the presence of possible bonding deformation density, average maps were also computed in the peptide-bond planes. The peptide moieties were sorted according to increasing B_{eq} factors on the O atom and the maps were averaged over samples of 40 successive peptides. Regrouping the maps by the magnitude of the B_{eq} factors enables the discrepancy in thermal motion between the different maps averaged to be limited. Such averaging of the residual electron density along the polypeptide backbone was performed on crambin at 0.54 Å resolution and resulted in significant bonding density (Jelsch *et al.*, 2000). However, bonding features were also visible on individual covalent bonds in the Fourier maps in the case of crambin.

The bonding density is clearly visible in the average map, with the lowest B_{eq} factors on the C═O, C─C$^\alpha$ and C─N bonds. The maps (Fig. 7) show decreasing bonding density as the thermal motion increases.
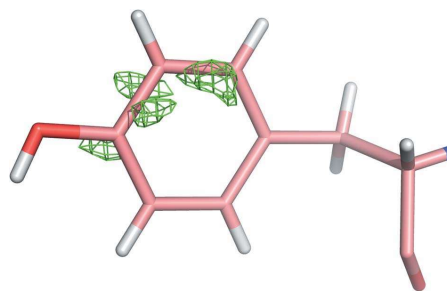
In a helical pseudo-octapeptide, it was observed that the Fourier residual bonding electron density is poorer towards the extremity of a phenyl side chain owing to increased thermal motion (Jelsch *et al.*, 1998). Such results are also generally observed in small molecules at ultrahigh resolution, where the atoms with larger thermal motion ($B_{eq} > 2$ Å^2) show poor and attenuated residual electron density on the covalent bonds and lone pairs. Therefore, comparison of residual electron densities for the same chemical group in different environments of a small molecule has to be carried out with care, as thermal motion effects can be much larger than the polarization itself. The refined static electron densities are also of poor quality when the thermal motion of atoms is higher than $B_{eq} = 2$ Å^2 and do not enable the observation of polarization differences. In protein structures, the thermal motion is too high to enable the refinement of individual atomic electron-density parameters. When the electron density is obtained from ELMAM2 database transfer, an average atom polarization is applied.

3.3. Electrostatic potential

The electrostatic potential (ESP) derived from the multipolar atom database ELMAM2 (Domagała *et al.*, 2011) was computed at the interacting surface of the ChOx protein and the FAD ligand. Such database transfer necessitates the modelling of all H atoms of the molecular structure.

The ESP mapped over the FAD molecule surface is shown in Fig. 8(*a*). The FAD prosthetic group generates an overall negative electrostatic potential that complements a mainly positive potential on the inner cavity of the protein surface (Figs. 8*b* and 8*c*). The total charge on the selected protein region obtained after database transfer was +2.3 e, which complements the −2.0 e charge on the FAD molecule well. The two multipolar ESPs generated by the protein and by the FAD show a nice complementarity, which can be seen in Fig. 8.

The ESP generated by *AMBER*03 (Case *et al.*, 2008) point charges was also computed and is shown in Fig. 8(*c*). The correlation coefficient between the ELMAM2 and *AMBER*03 ESP values on the molecular surface is relatively low at 0.05, but the two ESPs show qualitatively similar features in Fig. 8. The average *AMBER* ESP on the FAD molecular surface is −0.03 e Å^{-1}, while the average ELMAM2 ESP is 0.32 e Å^{-1}. This is owing in part to a different global charge of the residues located within 5 Å of the FAD cofactor: +1.0 e for

Figure 5
$2F_o − F_c$ electron density on the isoalloxazine ring system of the flavin cofactor at the 5σ level. The electron density shown at the high sigma level is generated mostly by the core electron shell located near the nuclei. The ellipsoidal shapes of the density therefore provide a picture of the anisotropic thermal motion of the atoms.

Figure 6
Residual Fourier map on residue Tyr195 drawn at the 2σ contour level.

AMBER and +2.3 e for ELMAM2. In addition, an ESP derived from the total electron density surrounding the nuclei is, by definition, more positive than that generated by point charges: for example, a spherical neutral atom generates a positive potential in space, while a point-charge model yields a zero potential.
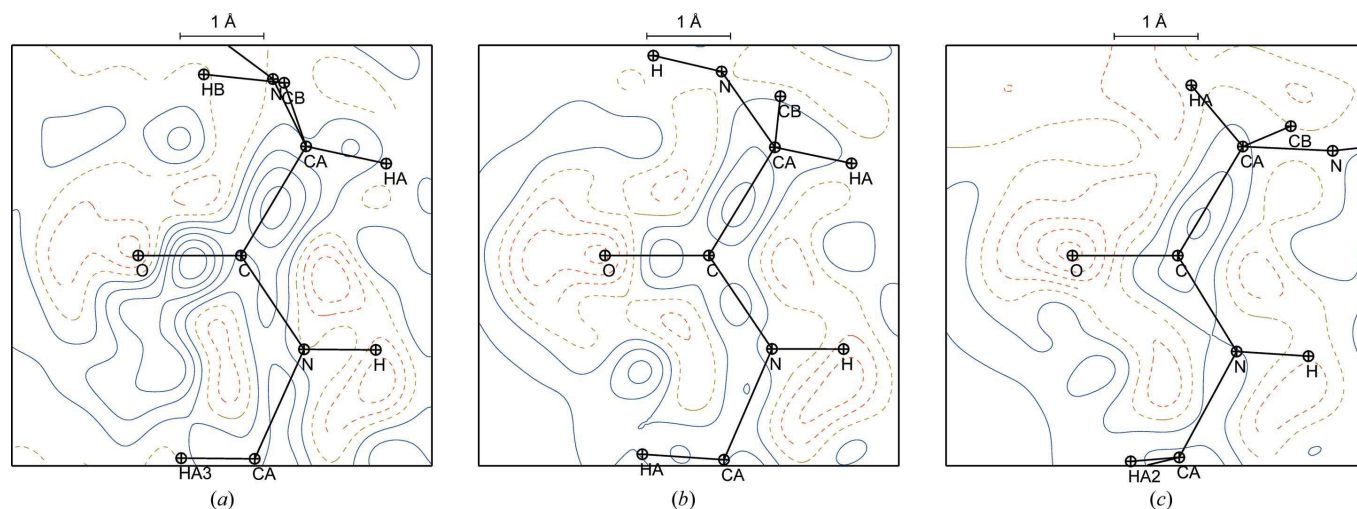
Figure 7
Average $F_o - F_c$ residual electron-density maps in the peptide-bond plane. (*a*) The 40 residues with lowest B_{eq} factor on the carbonyl O atom. Average $B_{eq} = 5.73$ Å^2, $B_{min} = 5.1$ Å^2, $B_{max} = 6.0$ Å^2. (*b*) The 40 next residues with a higher *B* factor. Average $B_{eq} = 6.2$ Å^2, $B_{min} = 6.0$ Å^2, $B_{max} = 6.4$ Å^2. (*c*) 40 residues with a *B* factor between $B_{min} = 7.4$ Å^2 and $B_{max} = 7.8$ Å^2 (from the 161st to the 200th residues with *B* factors in ascending order), average $B_{eq} = 7.6$ Å^2. Contour levels are at ±0.02 e Å^{-3}. Positive densities are shown as blue solid lines and negative densities are shown as red dashed lines. All maps were computed with the O atom at the origin; the carbonyl C atom defines the *x* direction and the map plane is defined by the triplet (O, C, N). Only the O atoms are strictly superposed in the different maps. No geometric compensation was applied to the maps to obtain perfect superposition for the C and N atoms. The geometric deviations are small within the peptide moiety (0.01 Å for C atoms, for instance).
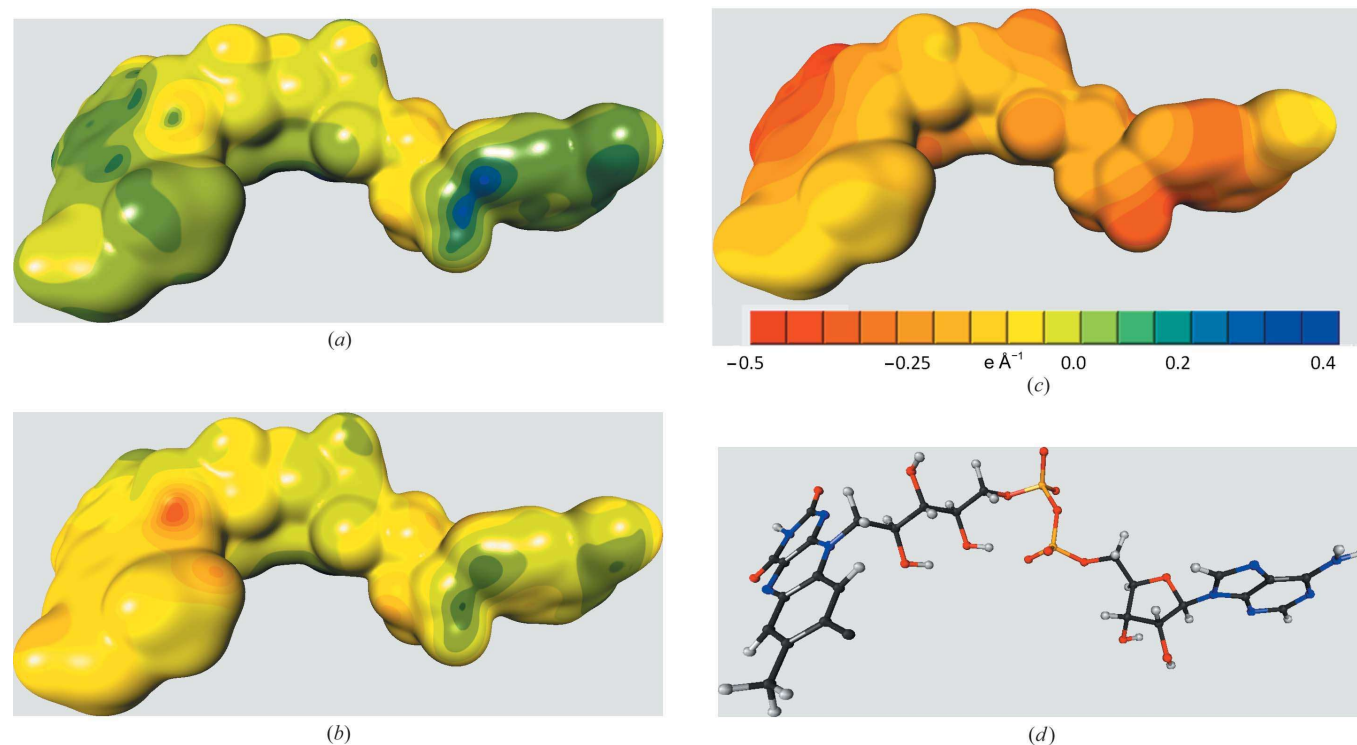
Figure 8
View of the FAD molecular surface coloured according to the electrostatic potential (ESP). (*a*) ESP derived from an ELMAM2 multipolar model generated by the protein active-site residues (all residues but FAD with an atom adjacent within 5 Å to the FAD molecule were selected). (*b*) The same as (*a*) but derived from *AMBER*03 point charges. (*c*) ELMAM2 ESP generated by the FAD cofactor. (*d*) Orientation of the FAD cofactor. In (*a*), (*b*) and (*c*), ESP is mapped onto the 0.01 e Å^{-3} total electron-density isosurface.

Table 2
Topological properties at the CPs of the interactions between the protein and the FAD ligand.

d_{12} is the distance between the two atoms, $\rho(\mathbf{r}_{cp})$ is the electron density at the critical point and $\nabla^2\rho(\mathbf{r}_{cp})$ is the Laplacian. Only the strongest hydrogen bonds ($D_e \geq$ 20 kJ mol^{-1}) are shown; the whole list is given in the Supporting Information. G^{cp}, V^{cp} and E^{cp} are the kinetic, potential and total electronic energies (Abramov, 1997) at the critical point, respectively; D_e is the estimated dissociation energy (Espinosa & Molins, 2000). Angle relates to the $\angle D-H\cdots A$ angle value, where is D and A are the donor and acceptor atoms, respectively.

Residue No.	Protein	FAD	d_{12} (Å)	Angle (°)	$\rho(\mathbf{r}_{cp})$ (e Å^{-3})	$\nabla^2\rho(\mathbf{r}_{cp})$ (e Å^{-5})	G^{cp} (kJ mol^{-1} bohr^{-3})	V^{cp} (kJ mol^{-1} bohr^{-3})	E^{cp} (kJ mol^{-1} bohr^{-3})	D_e (kJ mol^{-1})
Glu40	OE2	HO2A–O	1.76	172.5	0.300	1.2	64	−99	−35	49
Phe487	N–H	O2	1.90	174.5	0.226	1.8	59	−97	−38	49
Met122	O	HN3–N	1.94	147.2	0.179	2.0	53	−90	−36	45
Glu40	OE1	HO3A–O	1.84	162.9	0.250	1.3	54	−84	−30	42
Gly115	N–H	O1A	1.99	170.0	0.184	1.5	46	−75	−29	38
Gly475	N–H	O2P	2.00	169.8	0.178	1.5	45	−73	−28	36
Met122	N–H	O4	2.10	161.2	0.142	1.4	73	−60	−23	30
Val250	O	H61A–N	2.06	148.6	0.126	1.4	34	−59	−23	30
Gly114	C–HA2	O3B	2.34	108.0	0.082	1.3	29	−45	−21	25
Val250	N–H	N1A	2.24	153.5	0.125	1.1	30	−48	−18	24
Asn119	N–HD22	O2′	2.20	171.5	0.108	1.1	27	−43	−16	22
Asn119	C–HB2	N5	2.38	155.5	0.099	1.1	26	−43	−17	21

3.4. Topological analysis of FAD–protein hydrogen bonds

The presence of intermolecular interactions can be accurately quantified by performing a topological analysis of electron-density distributions based on the quantum theory of atoms in molecules (QTAIM) approach (Bader, 1991). The presence of a (3, −1) saddle critical point (CP) along with a bond path between the two atoms is an indication of the presence of an interaction. This approach is extensively used to quantify intermolecular interactions in small-molecular systems (Mallinson *et al.*, 2003; Munshi & Row, 2005). The CPs, like the electrostatic potential, can be calculated on the basis of the transferred electron-density parameters. Abramov (1997) has proposed the evaluation of the local electronic kinetic energy density $G(\mathbf{r}_{cp})$ (in kJ mol^{-1} bohr^{-3}) from the total electron density $\rho(\mathbf{r}_{cp})$ at the CPs of closed-shell interactions,

$$G(\mathbf{r}_{cp}) = \frac{3}{10}(3\pi^2)^{2/3}\rho^{5/3}(\mathbf{r}_{cp}) + \frac{1}{6}\nabla^2\rho(\mathbf{r}_{cp}). \quad (1)$$

The local form of the virial theorem relates the Laplacian to both the local electronic kinetic energy density $G(\mathbf{r}_{cp})$ and the local electronic potential energy density $V(\mathbf{r}_{cp})$ (Bader, 1991; Espinosa & Molins, 2000),

$$V(\mathbf{r}_{cp}) = \frac{1}{4}\nabla^2\rho(\mathbf{r}_{cp}) - 2G(\mathbf{r}_{cp}). \quad (2)$$

The estimated values of the interaction dissociation energies D_e (in kJ mol^{-1}) can be obtained from the properties at the CPs (Espinosa & Molins, 2000) using the equation

$$D_e = -\frac{1}{2}a_0^3 V_{cp}, \quad (3)$$

where a_0 is the Bohr radius and V_{cp} is the value of the potential energy density at the CP (Espinosa & Molins, 2000).

In total, the FAD cofactor forms 66 intermolecular interactions with the active-site resides of the protein. All of these interactions along with their topological properties are tabulated in Supplementary Table S1 and the most relevant (strongest) interactions are shown in Table 2. The contacts include C—H$\cdots$O, C—H$\cdots\pi$, N—H$\cdots$O, O—H$\cdots$O, N—H$\cdots$N and H$\cdots$H interactions. FAD atoms O5′, O4′, O3′ and O1A form linear hydrogen bonds, while the O5B, O4B, O4, O2′, O2 and O1P atoms form bifurcated hydrogen bonds to neighbouring residues. The O3B atom is trifurcated and the O2P and O2B atoms each form four hydrogen bonds. The interactions are found to be directional in nature and generally point towards the electron lone-pair regions of the O atoms. All of the N atoms of FAD, except for N3 and N9A, participate in intermolecular interactions with the surrounding residues. The N1A atom is a trifurcated acceptor and the N3A atom is a bifurcated acceptor, while the N1, N5, N6A, N7A and N10 atoms each form a unique hydrogen bond.

Residues forming strong interactions with the FAD molecule are shown in Fig. 9. Val250 forms two strong hydrogen bonds to FAD; namely N6A—H61A$\cdots$O=C and N—H$\cdots$N1A bonds with estimated dissociation energies of 29 and 24 kJ mol^{-1}, respectively (Table 2).

The carboxylate OE1 and OE2 atoms of Glu40 form two O$\cdots$H—O-type hydrogen bonds to two sugar hydroxyl groups (HO3A and HO2A) of the FAD ligand. The resulting double O$^{-}\cdots$H—O interaction pattern can be identified as a synthon. These hydrogen bonds are almost linear (straight O—H$\cdots$O angle). The corresponding dissociation energies are 49 and 42 kJ mol^{-1}, respectively. For reference, the dissociation energy for the O—H$\cdots$O^{-} interaction was estimated to be in the range 60–120 kJ mol^{-1} as reported by Meot-Ner & Sieck (1986). These interactions belong to the four strongest ($D_e >$ 40 kJ mol^{-1}) hydrogen bonds between the protein and the FAD molecule.

Additionally, Gly114 also binds to the O3A atom. This CA—HA2$\cdots$O3B contact is only a 'weak' C—H$\cdots$O hydrogen bond, for which the interaction energy is normally in the range 4–10 kJ mol^{-1} (Desiraju, 1991). The estimated dissociation energy is high (25 kJ mol^{-1}) and the contact distance of 2.34 Å is well below the sum of van der Waals radii ($r_O + r_H = 1.52 + 1.2 = 2.72$ Å; Bondi, 1964), indicating that the
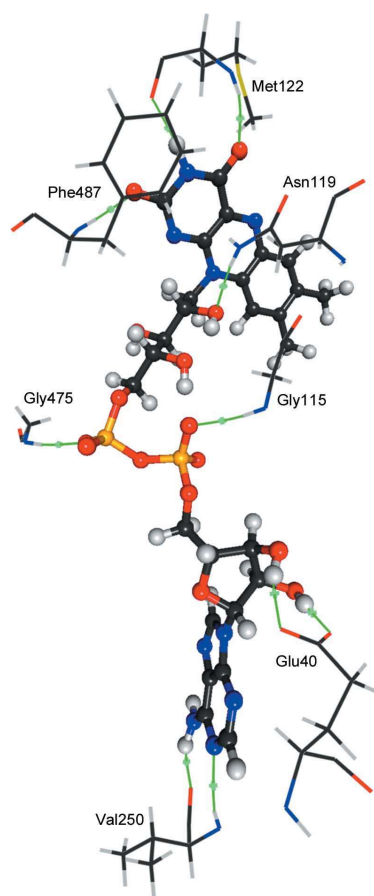
interaction is strongly involved in the anchoring of the FAD molecule in the active site.

Gly115 and Gly475 each form nearly linear strong N—H···O hydrogen bonds ($\angle$N—H···O $\simeq$ 170°) to the O3B, O1A and O2P atoms of FAD, respectively, and the H···O distances are short. The dissociation energies are quite high (25, 37 and 36 kJ mol^{-1}), as the energy of hydrogen bonds between neutral moieties varies over a wide range from less than 4 kJ mol^{-1} to greater than 35 kJ mol^{-1} (Subramanian & Zaworotko, 1994). These hydrogen bonds play a key role in the conformation of the bound FAD molecule as they are connected to the phosphate moiety.

From a chemical point of view, the important part of the FAD molecule is the isoalloxazine three-ring system. This portion of the molecule plays a crucial role in the redox chemistry of the enzyme. The moiety is involved in four strong hydrogen bonds to Asn119 (one hydrogen bond), Met122 (two hydrogen bonds) and Phe487 (one hydrogen bond) (Figs. 9 and 10). The last three hydrogen bonds appear to strongly influence the deformation of the isoalloxazine moiety from its planar conformation.

An aliphatic hydrogen bond is formed between Asn119 (CB—HB2) and the N5 atom of the FAD. The dissociation

energy owing to this hydrogen bond is computed to be 21 kJ mol^{-1}, which places this 'weak' C—H···O interaction among the list of strongest hydrogen bonds. Nevertheless, this hydrogen bond involves the essential N5 atom of FAD, which is the site of electron transfer to the cofactor during the redox reaction.

The main chain (atoms N and O) of Met122 forms two N—H···O-type hydrogen bonds to the FAD. In one case, the carbonyl O atom of Met122 acts as acceptor; in the other case, the O4 atom of FAD plays the role of acceptor.

In addition to influencing the geometry of the FAD iso-alloxazine moiety, the energies and the involvement of backbone atoms in these hydrogen bonds cause a steric hindrance to residue Phe487. This residue has the only significantly deviating C$^{\beta}$ position in the structure, $\delta = 0.284$ Å, based on analysis using *MolProbity* (Chen *et al.*, 2010). Fig. 10 illustrates the theoretical ideal geometry (in red) of Phe487. If the side-chain geometry of Phe498 is regularized using *Coot* (Emsley *et al.*, 2010), strong clashes occur with Met122 and Ala123 (distances between non-H atoms of less than 1.5 Å). More-over, the main-chain NH group of Phe487 is also involved in a very short N—H···O hydrogen bond to the FAD isoalloxazine, displaying a short distance and an almost straight angle (H···O2 = 1.90 Å and $\angle$N—H···O2 = 174.7°). The strength of this interaction results in a dissociation energy of 49 kJ mol^{-1}.

The N—H···O hydrogen bond between the Asn119 side chain and the O2′ atom of FAD is also worth mentioning as its dissociation energy is high (22 kJ mol^{-1}) and the N—H···O geometry is almost linear. All of the strongest hydrogen bonds shown in Table 2 have H···O distances which are smaller by 0.34 to 0.96 Å than the sum of van der Waals radii (2.72 Å).

The FAD molecule presents many internal conformational degrees of freedom. The FAD structure is significantly twisted within the cholesterol oxidase structure (Fig. 11) and its conformation is stabilized by both intramolecular hydrogen bonds and hydrogen bonds to the protein. Rotations around

Figure 9
Stereographic view of the residues forming the strongest intermolecular interactions, in terms of the highest values of the dissociation energy at the CPs ($D_e \geq 20$ kJ mol^{-1}), with the cofactor FAD. The CPs are shown as green circles and the bond paths are shown in green.
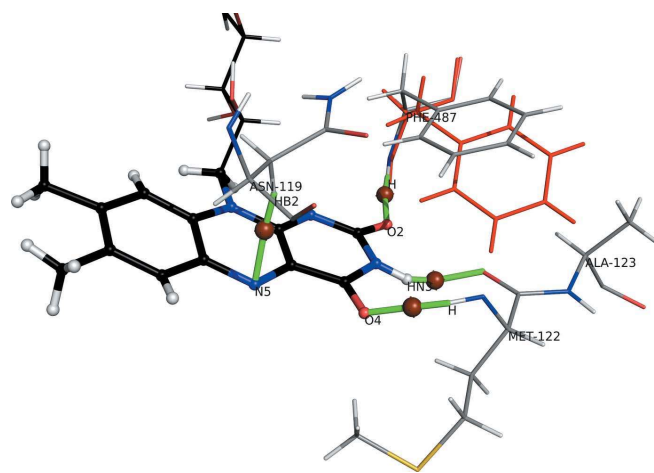

Figure 10
Residues forming the strongest intermolecular interactions with the FAD isoalloxazine moiety. The CPs are shown as brown spheres and the bond paths are shown in green. The theoretical ideal geometry of the Phe487 side chain is shown in red.

covalent bonds involving the central pyrophosphate moiety lead to a FAD conformation in which the molecule forms a zigzag chain, with the pyrophosphate being oriented at almost 90° with respect to the nucleotide long axis. While the protein active-site topology guides the conformation of the bound FAD ligand, a number of intramolecular interactions which further stabilize the folded conformation are observed.

3.5. Stereochemistry of hydrogen bonds between main-chain atoms

The C=O$\cdots$H—N-type interactions are the most common hydrogen bonds found in proteins and constitute an important driving force involved in the formation and stabilization of α-helices and β-sheets (Chothia, 1984). They are of special importance for understanding protein structure, function, folding and stability (Bolen & Rose, 2008). Moreover, hydrogen-bond analysis in protein three-dimensional structures provides essential information for modelling and protein structure prediction. Most hydrogen bonds in proteins are between main-chain atoms; an average proportion of 68% was reported by Stickle et al. (1992).

In this study, we focus on the C=O$\cdots$H—N type of hydrogen bonds formed between main-chain atoms and quantify their conformational preferences within α-helices, β-sheets and 'other' regions. The residues of the protein which are neither in an α-helix nor in a β-sheet secondary structure are referred to as 'other'. C=O$\cdots$H—N-type interactions involving moieties in side chains such as, for instance, the amide group of Asn and Gln or NH groups in Trp, Arg or His are not considered. The stereochemistry of hydrogen bonds in proteins, including side chains, has been reviewed by Baker & Hubbard (1984).

A total of 220 (117 in α-helices, 42 in β-sheets and 61 'other') N—H$\cdots$O bonds are found in the protein structure. In α-helices, hydrogen bonds occur typically between N—H and C=O groups separated by four amino acids. However, some additional hydrogen bonds occur in α-helices between atoms separated by three residues. Therefore, within α-helices, the hydrogen bonds have been divided into two subgroups: $i{\rightarrow}i + 3$ (notated $i + 3$) and $i{\rightarrow}i + 4$ (notated $i + 4$), where i refers to
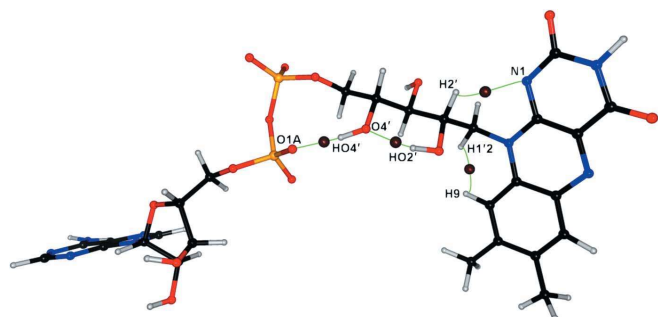
the residue number; the numbers of hydrogen bonds are 32 and 85, respectively.

The statistical distribution of the occurrences of O$\cdots$H distances in C=O$\cdots$H—N hydrogen bonds for α-helical, β-sheet and 'other' parts of the protein is depicted in Fig. 12. The average O$\cdots$H distances are 2.11 (2), 2.18 (2) and 2.46 (3) Å in β-sheets and in $i + 4$ and $i + 3$ α-helix hydrogen bonds, respectively. The number in parentheses is the uncertainty on the average value (r.m.s.d./$N^{1/2}$). The corresponding standard deviations within the samples are 0.14, 0.17 and 0.19 Å, respectively. Statistically, hydrogen bonds have a tendency to be shorter within β-sheets compared with α-helices and the distance distribution is also less spread out. The canonical $i + 4$ hydrogen bonds within α-helices are also significantly shorter than the $i + 3$ hydrogen bonds. This is related to the generally less favourable hydrogen-bond directionality of $i + 3$ hydrogen bonds, which form $\angle$C—O$\cdots$H angles mostly between 114 and 90°. Another reason is that $i + 3$ hydrogen bonds often occur in addition to $i + 4$ hydrogen bonds in bifurcations and they have a tendency to be sterically hindered. As observed in Fig. 12, the maximal frequency of C=O$\cdots$H—N hydrogen bonds occurs for $d_{O\cdots H}$ distances around 2.0 Å for β-sheet and $i + 4$ hydrogen-bond types in α-helices, while the peak is around 2.6 Å for the $i + 3$ type. The 'other' hydrogen bonds show average O$\cdots$H distances of intermediate value, $\langle d_{O\cdots H}\rangle = 2.20$ (3) Å, with a similar spread of values (r.m.s.d. = 0.16 Å). In a study of six protein crystal structures at atomic resolution, Liebschner et al. (2011) found similar trends for O$\cdots$H distances in α-helices: 2.03 (16) and 2.22 (19) Å for $i + 4$ and $i + 3$ hydrogen bonds, respectively.

When the O$\cdots$N distances are considered, the average values found are 2.92 (2) Å in β-sheets and 2.98 (2) and 3.11 (2) Å in $i + 4$ and $i + 3$ α-helix hydrogen bonds, while the r.m.s.d. values are 0.12, 0.14 and 0.12 Å, respectively. These
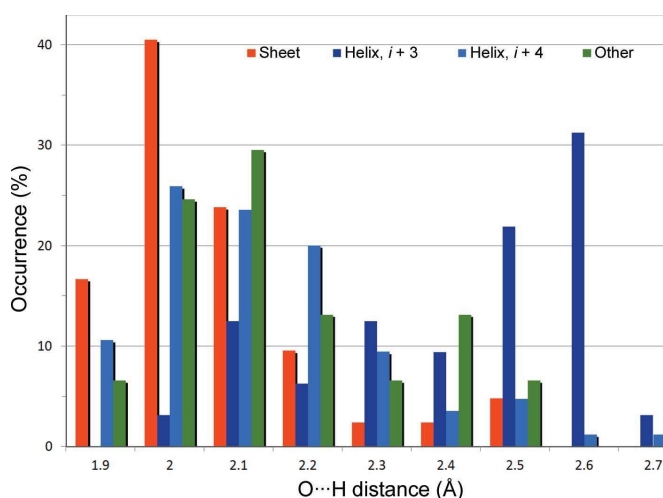

Figure 11
Conformation of the FAD cofactor in the ChOx binding site with internal hydrogen bonds shown. The CPs of hydrogen bonds are shown as brown spheres and the bond paths are shown in green.

Figure 12
Occurrences of O$\cdots$H distances in C=O$\cdots$H—N hydrogen bonds for α-helices, β-sheets and 'other' parts of the protein. Hydrogen bonds within α-helices were subdivided into $i{\rightarrow}i + 3$ and $i{\rightarrow}i + 4$ interactions, where i denotes the residue number. The helix/sheet/other classification refers to the O atom.

distances are in accordance with those presented in a previous study (Koch *et al.*, 2005), which showed that the mean hydrogen-bond length is d_{ON} = 2.94 (5) Å for parallel β-sheets and 2.94 (3) Å for antiparallel β-sheets and the mean d_{ON} value is 2.99 (2) Å for α-helices, whereas the median d_{ON} values of Thomas *et al.* (2001) are 2.93 Å for β-sheets and 3.00 Å for α-helices.

The second most characteristic feature of hydrogen bonds is their directionality, which can be analysed through precise descriptions of the $\angle$C=O$\cdots$H, $\angle$C=O$\cdots$N and $\angle$N—H$\cdots$O angles. Stereochemical analyses of hydrogen bonds in the crystals of small molecules are carried out using H-atom positions. An experimental and theoretical charge-density analysis combined with a statistical survey of more than 500 000 crystal structures revealed that hydrogen bonds show strong directionality towards the electron lone pairs of O atoms, especially in strong hydrogen bonds (Ahmed *et al.*, 2013). Even fine differences in the positions of the lone pairs found between alcohol and phenol oxygen acceptors have been found to influence the position of donor H atoms.

In protein crystal structures at usual resolutions, H atoms are not visible in the electron-density maps; therefore, authors generally carry out stereochemical analyses on the C, O and N atom positions (Koch *et al.*, 2005; Thomas *et al.*, 2001). In the present ultrahigh-resolution crystal structure of cholesterol oxidase, most of the H atoms in NH groups are visible in the electron-density maps. In other cases, the position of the amide H atom can be placed according to stereochemical rules such as the standard N—H distance, which is known from neutron diffraction studies, and the assumption of a planar

peptide. In a survey of the peptide ω angles in the Cambridge Structural Database (CSD) of small molecules (Allen, 2002), MacArthur & Thornton (1996) observed up to a 6° deviation from planarity.

The angle distribution within C=O$\cdots$H—N hydrogen bonds was therefore analysed (Figs. 13, 14 and 15) and revealed a number of discrepancies between the different secondary-structural elements. The most favoured regions for the H-atom position in α-helices and β-sheets is 154–180° (Fig. 13), which does not correspond to the commonly accepted direction of the electron lone pairs (120°). Furthermore, this is not the favoured region in the crystal structures of small molecules; Wood *et al.* (2008) found that the most energetically favourable angles are in the range 127–140°. The average $\angle$C=O$\cdots$H angle for β-sheets is 151° and is close to that found for i + 4 hydrogen bonds in α-helices (151°). By contrast, i + 3 hydrogen bonds show their highest frequencies very far from the most energetically favourable $\angle$C=O$\cdots$H angles, with an average value of 107°. Two different maxima are clearly visible (Fig. 13) in the ranges 143–180 and 90–114° for i + 4 and i + 3 hydrogen bonds, respectively. Hydrogen bonds obey special geometric constraints in secondary-structure elements of proteins, which make them deviate from ideal geometry (Baker & Hubbard, 1984). As α-helices are known to be the most constrained backbone structures in peptides and proteins (Baker & Hubbard, 1984), this result appears to be sensible. These results are comparable with those of the study of Liebschner *et al.* (2011), in which the average $\angle$C=O$\cdots$H value was found to be 149 (8) and 110 (11)° for i + 4 and i + 3 hydrogen bonds in α-helices, respectively.
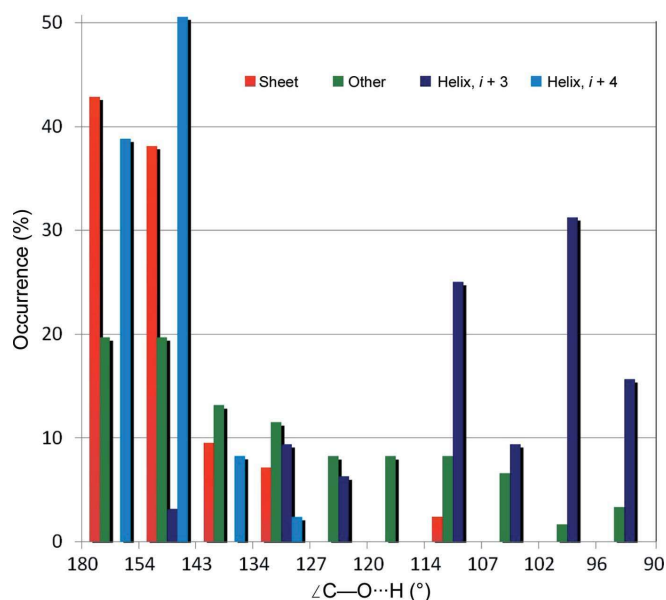
Figure 13
Percentages of hydrogen bonds found in α-helices ($i{\rightarrow}i$ + 3 and $i{\rightarrow}i$ + 4), β-sheets and 'other' parts of the protein in different intervals of $\angle$C=O$\cdots$H. The hemisphere $\angle$C=O$\cdots$H = 90–180° was divided into 10° intervals. The angle interval limits were at 90° + arccos(0.1), 90° + arccos(0.2), $\dots$, 90° + arcos(0.9) in order to have successive crowns with the same solid angle. The percentages are then representative of the frequency of hydrogen bonds in a given direction.
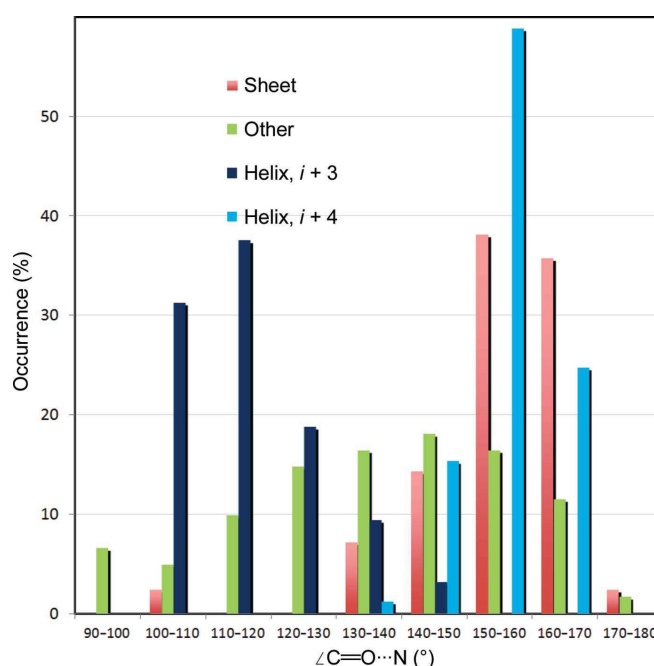
Figure 14
Percentage distribution of $\angle$C=O$\cdots$N as found in the C=O$\cdots$H—N hydrogen bonds for α-helices, β-sheets and 'other' parts of the cholesterol oxidase protein.

Table 3
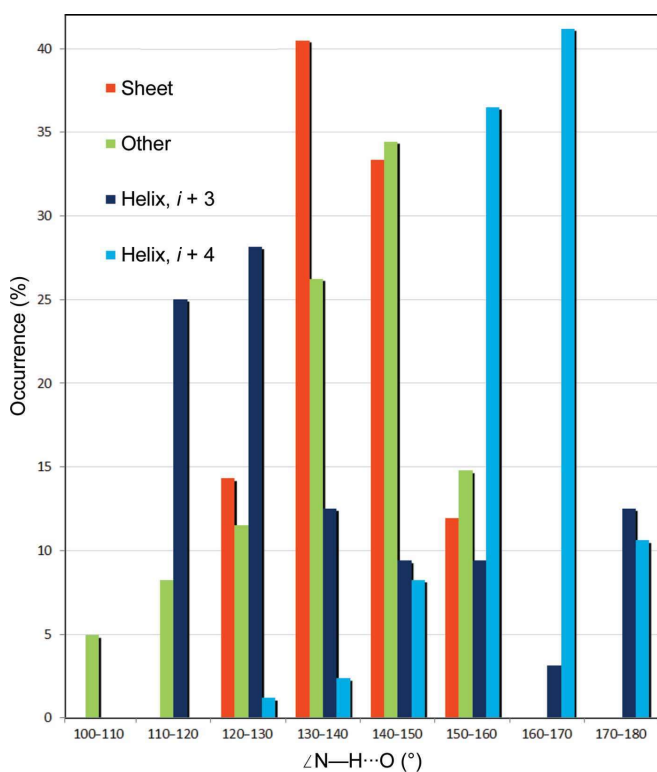Average values of $\angle C{=}O{\cdots}H$ and $\angle C{=}O{\cdots}N$ angles in the secondary-structure elements.

The root-mean-square deviation for the samples is also given.

Secondary structure	$\angle C{=}O{\cdots}H$ (°)		$\angle C{=}O{\cdots}N$ (°)	
	Mean (e.s.d.)	R.m.s.d. (Å)	Mean (e.s.d.)	R.m.s.d. (Å)
β-Sheet	151 (2)	12	154 (2)	12
α-Helix				
$i + 4$	151 (1)	7	155 (1)	6
$i + 3$	107 (2)	14	115 (1)	9
Other	134 (3)	20	136 (3)	21

The conformations closest to the ideal angle values are realised for the 'other' hydrogen bonds, which are not involved in these secondary-structure elements. For 'other' hydrogen bonds, the average $\angle C{=}O{\cdots}H$ value is 134° and is the closest to the average angle of 139.0 (8)° found for $C{=}O{\cdots}H{-}N$ bonds in the CSD (Wood *et al.*, 2008). In the statistical analysis of the CSD by Ahmed *et al.* (2013), it was found that for hydrogen bonds involving $C{=}O$ carbonyl moieties in organic molecules, the highest propensity of hydrogen localization was towards the oxygen electron lone-pair directions and was close to the sp^2 hybridization plane. Occurrences within ±13° of the electron lone-pair direction at 120° are however not frequent in ChOx, except for $i + 3$ α-helix hydrogen bonds. As shown in Fig. 13, the majority of $\angle C{=}O{\cdots}H$ angles for β-sheet and canonical $i + 4$ α-helix hydrogen bonds are in the range 150–180°. The stereo-
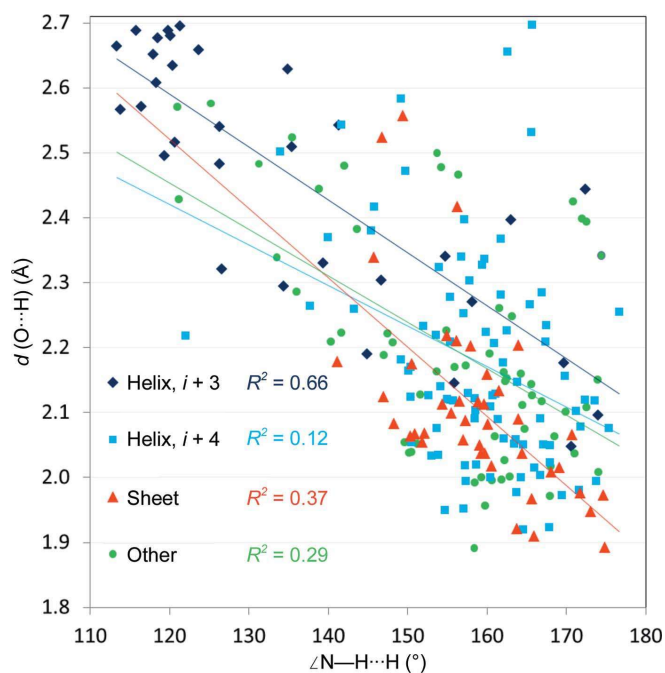
chemistry of hydrogen bonds in protein–ligand environments closely resembles that observed for small-molecule crystal structures, as verified by Klebe (1994). However, this is not the case for β-sheet and $i + 4$ α-helix main-chain hydrogen bonds, which are constrained by the global architecture of these secondary-structural elements.

The mean values of $\angle C{=}O{\cdots}N$ are 155, 115, 154 and 136° for $i + 4$ and $i + 3$ α-helix, β-sheet and 'other' hydrogen bonds respectively (Table 3, Fig. 14). The average angle for $i + 3$ hydrogen bonds is the lowest and there are frequent occurrences only for low angles between 101 and 143°. As expected, the $i + 4$ α-helix hydrogen bonds mostly have directions close to linear for $\angle C{=}O{\cdots}N$ (in the range 154–180°). As a consequence of the α-helix geometry, the $C{=}O$ and N atoms are not far from having a linear geometry.

β-Sheets show similar $\angle C{=}O{\cdots}N$ geometry as $i + 4$ α-helix hydrogen bonds. 'Other' hydrogen bonds have a much wider and flatter spectrum of angles; presumably, the distribution is based on the specific interactions rather than being constrained by the helical or sheet structure.

As the values $\angle C{=}O{\cdots}N$ and $\angle C{=}O{\cdots}H$ are partly correlated, some similar tendencies are found for the two angles analysed with respect to the four types of hydrogen bonds ($i + 3$ and $i + 4$ α-helix, β-sheet and other).

It has been well documented that $D{-}H{\cdots}A$ (where is D is a donor and A is an acceptor) hydrogen-bonded interactions are believed to have a statistical preference for linearity based on database studies (Wood *et al.*, 2008). The definitions of hydrogen bonds in the literature generally require $\angle D{-}H{\cdots}A$ angles larger than a given value (90–120°; Arunan *et al.*, 2011). The percentage distribution of $\angle N{-}H{\cdots}O$ angles (Fig. 15) shows that the majority of hydrogen bonds of

Figure 15
Percentage distribution of $\angle N{-}H{\cdots}O$ as found in the $C{=}O{\cdots}H{-}N$ hydrogen bonds for α-helices, β-sheets and 'other' parts of the ChOx protein.

Figure 16
Scatterplot of $\angle N{-}H{\cdots}O$ angles *versus* $H{\cdots}O$ distances.

Table 4
Statistics for some covalent bonds in the ChOx protein.

Distances are divided into α-helix ($i + 3/i + 4$), β-sheet and 'other' categories. Bifurcated ($i + 3$, $i + 4$) hydrogen bonds on the carbonyl acceptor were classified as $i + 3$ hydrogen bonds. Disordered atoms of ChOx were excluded. C=O and C—N bond distances were excluded for C atoms with $B_{eq} > 10$ Å^2. Similarly, C$^\alpha$—C and C$^\alpha$—N bonds were excluded for C$^\alpha$ atoms with $B_{eq} > 10$ Å^2. In the category 'All', only disordered atoms were excluded. The 'Jaskolski proteins' line refers to the study by Jaskolski *et al.* (2007) of very high resolution $d < 0.8$ Å protein structures and atoms with $B < 40$ Å^2. The 'Jaskolski CSD' line refers to the same article in which bond lengths were collected from structures in the CSD with $R < 5\%$. The *PHENIX* target value is from the Engh & Huber (1991) stereochemical dictionary, which was built using appropriate chemical fragments found in the CSD.

Bond	No.	$\langle d \rangle$ (e.s.d.) (Å)	R.m.s.d. (Å)
C=O			
Helix $i + 4$	85	1.2378 (9)	0.009
Helix $i + 3$	32	1.2415 (17)	0.010
Sheets	42	1.2304 (11)	0.008
Other	59	1.2334 (13)	0.010
All	469	1.2338 (5)	0.010
PHENIX target	—	1.231 (20)	—
Jaskolski proteins	—	1.234	0.013
Jaskolski CSD	480	1.231	0.009
C—C$^\alpha$			
Helix $i + 4$	83	1.5258 (8)	0.008
Helix $i + 3$	32	1.5242 (17)	0.010
Sheets	40	1.5211 (11)	0.007
Other	59	1.5247 (10)	0.008
All	455	1.5233 (4)	0.008
PHENIX target	—	1.525 (21)	—
Jaskolski proteins	—	1.527	0.013
Jaskolski CSD	146	1.523	0.011
C—N			
Helix $i + 4$	85	1.3302 (7)	0.006
Helix $i + 3$	32	1.3293 (11)	0.006
Sheets	38	1.3312 (9)	0.006
Other	59	1.3320 (10)	0.008
All	459	1.3313 (3)	0.007
PHENIX target	—	1.329 (14)	—
Jaskolski proteins	—	1.334	0.013
Jaskolski CSD	348	1.332	0.008
N—C$^\alpha$			
Helix $i + 4$	83	1.4561 (7)	0.006
Helix $i + 3$	32	1.4553 (12)	0.007
Sheets	40	1.4534 (11)	0.007
Other	59	1.4530 (13)	0.010
All	457	1.4549 (4)	0.008
PHENIX target	—	1.458 (19)	—
Jaskolski proteins	—	1.454	0.012
Jaskolski CSD	231	1.455	0.007

cholesterol oxidase are not far from being straight (angle of >154°). The average $\angle$N—H$\cdots$O values for β-sheets and $i + 4$ α-helices are not far from 180° and are comparable at 158 (1) and 158 (1)°, respectively, as a consequence of the geometry of these secondary-structure elements. 'Other' $\angle$N—H$\cdots$O angles reach a similar mean value of 156 (2)°, while for $i + 3$ α-helices the value is much lower at 136 (4)°.

Three of the four types of hydrogen bonds ($i + 4$ α-helix, β-sheet and other) show almost the same trend in the occurrence graph, with the only strong frequencies being at geometries $\angle$N—H$\cdots$O $\simeq$ 180°. The $i + 3$ α-helix hydrogen bonds show a distinct behaviour and the geometry is not as directional as for the other types, as there are two maxima of occurrences around $\angle$N—H$\cdots$O angles of 180° and 120°. The $i + 3$ hydrogen bonds are bifurcated; they occur in addition to the canonical $i + 4$ hydrogen bonds in α-helices. The two oxygen acceptors in bifurcated hydrogen bonds must point in distinct directions; therefore, the additional $i + 3$ hydrogen bonds are often required to have an $\angle$N—H$\cdots$O angle far from linearity.

Based on a statistical study of the Cambridge Structural Database (Allen, 2002) and *ab initio* intermolecular inter-action energy calculations, Wood *et al.* (2008) have analysed the geometry of D—H$\cdots A$ interactions. Interactions with D—H$\cdots A$ angles in the range 120–140° are observed to have substantially reduced stabilization energies and angles below 120° are generally unlikely to correspond to significant inter-actions.

28% of the $i + 3$ interactions (Fig. 15) have D—H$\cdots A$ values lower than 120° and were not considered as hydrogen bonds but rather as van der Waals interactions. The majority (54%) of the $i + 3$ interactions in α-helices have $\angle$N—H$\cdots$O values lower than 127°; although these hydrogen bonds are polar, their presence is mostly a consequence of the local α-helical structure arrangement with strong $i + 4$ hydrogen bonds rather than a driving force. This is in accordance with Fig. 12, in which most of the $i + 3$ hydrogen bonds show long O$\cdots$H distances (>2.5 Å) at almost van der Waals contact distances (2.72 Å).

The two-dimensional scatter plot of $\angle$N—H$\cdots$O angles *versus* H$\cdots$O distances (Fig. 16) shows a clear global rela-tionship between distance and angle; the average H$\cdots$O distance decreases as the $\angle$N—H$\cdots$O angle becomes closer to 180°. For all types of hydrogen bonds, there is a clear dependence between the two values, as previously observed by Wood *et al.* (2008) for small-molecule structures in the CSD. Strong hydrogen bonds with $d_{O\cdots H} < 2.2$ Å are only found at $\angle$N—H$\cdots$O angles larger than 140°. The α-helix $i + 3$ hydrogen bonds show peculiar characteristics, with a large number of interactions occurring at large H$\cdots$O distances and low $\angle$N—H$\cdots$O values in the range 110–130°. The linear fits show weak to moderate correlation between the $\angle$N—H$\cdots$O and $d_{O\cdots H}$ values, except for α-helix $i + 3$ interactions. The fitted linear curves of $d_{O\cdots H}$ *versus* $\angle$N—H$\cdots$O however still indicate that H$\cdots$O distances have a tendency to be higher for α-helix $i + 3$ interactions compared with the other three types.

3.6. Stereochemistry of main-chain bonds

To analyse the source of differences in C=O bond lengths, the structural dependences along the polypeptide main-chain of ChOx were investigated in the different secondary-structural elements. Bond lengths were restrained to the *PHENIX* target definitions of interatomic distances, with an allowed standard deviation of 0.02 Å. Deviations of distances from the target values were still observed as the diffraction data are at atomic resolution, and have been analyzed.

Table 4 shows the distribution of C=O bond-length occurrences along with other statistical factors. The mean C=O bond lengths for 'All' and 'Other' hydrogen bonds are similar and reflect the value of 1.234 Å reported by Jaskolski *et al.* (2007) for very high resolution protein structures. In the

case of aldose reductase, higher values of 1.236 (8) and 1.235 (9) Å were found using a multipolar and an aspherical atom model, respectively.

The bond length reported for the *PHENIX* target [1.231 (20) Å], which originates from statistics on the CSD, is the same as that reported for the CSD study of Jaskolski *et al.* (2007). This value is slightly lower, but comparable within uncertainty, to that reported for the protein study of Jaskolski *et al.* (2007). The discrepancy was interpreted by Jaskolski and coworkers as reflecting the systematic involvement of the main-chain carbonyl groups of proteins in similar hydrogen-bonding interactions.
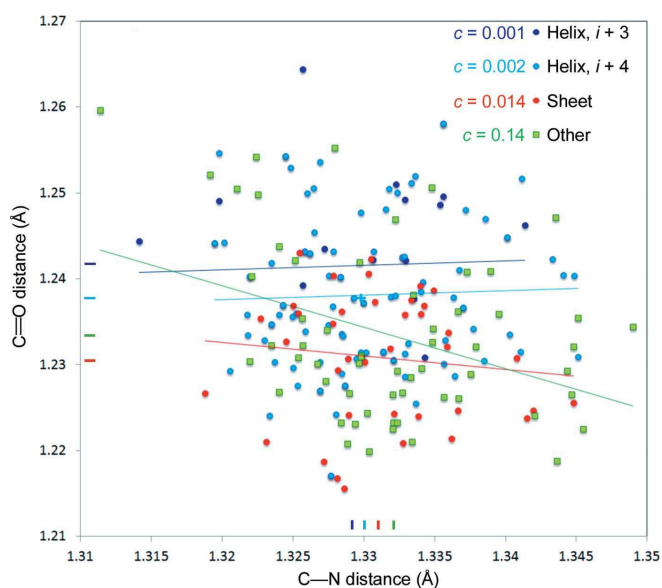
Figure 17
Scatter diagram of the C—N *versus* C=O bond lengths in the ChOx structure.
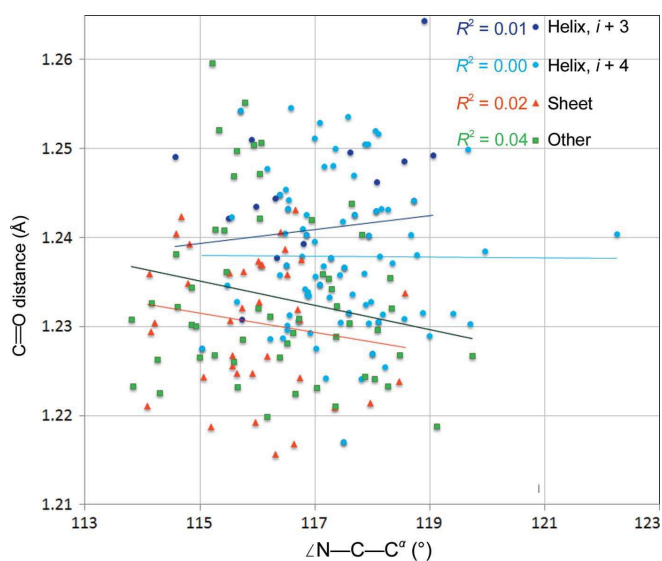
Figure 18
Scatterplot of the C=O distances *versus* the $\angle C^\alpha$—C—N angles. The linear fits are shown to highlight the average values of distances (despite correlation coefficients of close to zero between the two variables).

Significant discrepancies were evident between the mean C=O bond lengths of the different secondary-structural elements in ChOx. The mean C=O bond length for β-sheets is noticeably shorter among the values listed for the secondary-structural elements, as observed by Lario *et al.* (2003) for ChOx at a lower resolution. In contrast, for α-helices the mean C=O bond length for the $i + 3$ type of hydrogen bonds is slightly longer than that of the conventional $i + 4$ type of hydrogen bonds.

Table 4 also lists the statistical values for the other covalent bonds of the main-chain peptide. The r.m.s.d. of the C—C^α distances in the different samples is about 0.01 Å. However, some discrepancies are noteworthy.

Unlike the C=O bond lengths, the three other covalent bond lengths have average lengths which are higher in the α-helix $i + 4$ type of hydrogen bond compared with the $i + 3$ type. Overall, the mean bond lengths obtained for cholesterol oxidase are comparable with the other statistical values reported in Table 4. With the exception of C—C^α, the average main-chain bond distances in cholesterol oxidase lie between the 'high-resolution protein' values reported by Jaskolski *et al.* (2007) and the *PHENIX* target restraints.

The shortest mean value for the C—C^α bond lengths are observed in the case of β-sheets; the other distances in cholesterol oxidase are very close to the *PHENIX* target restraints. The shortest C—N bond lengths are found for α-helix residues; this may be related to the long C=O bonds found for this secondary-structure element. For N—C^α bond lengths, the shortest mean value is observed for 'other' and β-sheet residues.

Globally for the whole protein, no significant correlation (+4%) is observed between the $\angle C^\alpha$—C—N and C=O bond lengths (Fig. 18). Nevertheless, the diagram for the secondary-structural elements reveals different trends. The correlation between the two geometric values is still small in each separate sample; however, the ($\angle C^\alpha$—C—N, C=O) points occupy different regions of the scatter plot. Residues in α-helices have generally large $\angle C^\alpha$—C—N and C=O values, while 'others' and those in β-sheets have low values. The average $\angle C^\alpha$—C—N value is lowest in β-sheets (115.9 ± 1.1°) and highest in α-helices (117.3 ± 1.1°)

For the α-helix $i + 3$ type of hydrogen bond, a weak positive correlation (+1%) is observed, while for β-sheets and 'others' the correlation is found to be slightly negative (−2 and −4%, respectively). For α-helix $i + 4$ residues, the C=O bond lengths seem to be constant with the variation in the $\angle C^\alpha$—C—N angle.

Electronic effects within the peptide bonds play meaningful roles in hydrogen bonding within protein structures. The polarization of the C=O bonds is also influenced by other interactions within the protein and the solvent. An influential contribution of the imine resonance form to the geometry, which allows the O atoms to form stronger hydrogen bonds, could be expected. This is confirmed by the correlation found between the C=O and C—N bond lengths. It was found by Esposito *et al.* (2000) that in several protein structures at atomic resolution and in the CSD the correlation is strictly

negative. This is found to be valid in the structure of cholesterol oxidase (Fig. 17), with a correlation of 14% and a negative slope of −0.48 between the C=O and C—N distances when only the 'other' regions are considered. However, correlation is found for α-helix and β-sheet residues. Thus, the scatter plot only confirms for 'other' regions that the resonance of the peptide bond unit is shifted towards the imine form (short N—C bond) as the C=O distance increases. Similarly, C^{α}—C—N angles and C=O distances are found to be unrelated in α-helices and β-sheets; only a very weak correlation (4%) occurs for 'other' residues (Fig. 18).

4. Conclusion

It is very rare for a protein, especially of around 500 residues, to diffract to subatomic resolution, although with the improvement in data-collection, processing and refinement methods such structures are becoming more feasible. Despite these ultrahigh-resolution structures, it is still a significant challenge to study the electrostatic and stereochemical properties of proteins. However, these challenges can be overcome, in part, by using the transferability principle. The principle, in which a better refined model is obtained and the electrostatic properties can be studied on the basis of the transferred multipolar electron-density parameters, has already been discussed (Zarychta et al., 2007; Domagała et al., 2011, 2012; Dittrich et al., 2005; Volkov et al., 2004). The diffraction of macromolecular structures to atomic or lower resolution no longer poses an obstacle for the study of the electrostatic properties using a multipolar atom model.

Here, a relatively large protein is studied at subatomic resolution. The refinement statistics and the Ramachandran plot show that the refined model is acceptable. To study the electrostatic potential of the active site and to investigate the interactions between the protein and the ligand, experimental electron-density database transfer was taken into account. The results obtained on the basis of database transfer show that the active site has an overall positive electrostatic potential, which is complemented by an overall negative electrostatic potential of the cofactor. The numerous hydrogen bonds found between the ligand and the protein are in agreement with the specificity of this class of proteins for the FAD cofactor.

References

Abramov, Y. A. (1997). Acta Cryst. A53, 264–272.
Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. & Urzhumtsev, A. (2007). Acta Cryst. D63, 1194–1197.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). Acta Cryst. D68, 352–367.
Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). Acta Cryst. D60, 260–274.
Ahmed, M., Jelsch, C., Guillot, B., Lecomte, C. & Domagała, S. (2013). Cryst. Growth Des. 13, 315–325.
Allen, F. H. (2002). Acta Cryst. B58, 380–388.
Allen, F. H., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (2006). International Tables for Crystallography, Vol. C, 1st online ed., edited by E. Prince, pp. 790–811. Chester: International Union of Cystallography.
Arunan, E., Desiraju, G. R., Klein, R. A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., Kjaergaard, H. G., Legon, A. C., Mennucci, B. & Nesbitt, D. J. (2011). Pure Appl. Chem. 83, 1637–1641.
Av-Gay, Y. & Sobouti, R. (2000). Can. J. Microbiol. 46, 826–831.
Bader, R. (1991). Chem. Rev. 91, 893–928.
Baker, E. N. & Hubbard, R. E. (1984). Prog. Biophys. Mol. Biol. 44, 97–179.
Berman, H. M. (2000). Nucleic Acids Res. 28, 235–242.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). J. Mol. Biol. 112, 535–542.
Bolen, D. W. & Rose, G. D. (2008). Annu. Rev. Biochem. 77, 339–362.
Bondi, A. (1964). J. Phys. Chem. 68, 441–451.
Case, D. A. et al. (2008). AMBER10. University of California, San Francisco, USA.
Cavener, D. R. (1992). J. Mol. Biol. 223, 811–814.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). Acta Cryst. D66, 12–21.
Chen, L., Lyubimov, A. Y., Brammer, L., Vrielink, A. & Sampson, N. S. (2008). Biochemistry, 47, 5368–5377.
Chothia, C. (1984). Annu. Rev. Biochem. 53, 537–572.
Corbin, D. R., Grebenok, R. J., Ohnmeiss, T. E., Greenplate, J. T. & Purcell, J. P. (2001). Plant Physiol. 126, 1116–1128.
Corbin, D. R., Greenplate, J. T. & Purcell, J. P. (1998). HortScience, 33, 614–617.
Coulombe, R., Yue, K. Q., Ghisla, S. & Vrielink, A. (2001). J. Biol. Chem. 276, 30435–30441.
Desiraju, G. R. (1991). Acc. Chem. Res. 24, 290–296.
Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). Acta Cryst. A61, 314–320.
Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). Acta Cryst. A68, 337–351.
Domagała, S., Munshi, P., Ahmed, M., Guillot, B. & Jelsch, C. (2011). Acta Cryst. B67, 63–78.
Dominiak, P. M., Volkov, A., Dominiak, A. P., Jarzembska, K. N. & Coppens, P. (2009). Acta Cryst. D65, 485–499.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.
Engh, R. A. & Huber, R. (1991). Acta Cryst. A47, 392–400.
Espinosa, E. & Molins, E. J. (2000). J. Chem. Phys. 113, 5686–5694.
Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. (2000). Protein Eng. Des. Sel. 13, 825–828.
Flegg, H. M. (1973). Ann. Clin. Biochem. 10, 79–84.
Guillot, B. (2011). Acta Cryst. A67, C511–C512.
Guillot, B., Jelsch, C., Podjarny, A. & Lecomte, C. (2008). Acta Cryst. D64, 567–588.
Guillot, B., Muzet, N., Artacho, E., Lecomte, C. & Jelsch, C. (2003). J. Phys. Chem. B, 107, 9109–9121.
Hansen, N. K. & Coppens, P. (1978). Acta Cryst. A34, 909–921.
Held, J. & van Smaalen, S. (2014). Acta Cryst. D70, 1136–1146.
Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J. C. & Lecomte, C. (2000). Acta Cryst. D56, 151–160.
Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007). Acta Cryst. D63, 611–620.
Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). J. Appl. Cryst. 38, 38–54.

research papers

Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.

Johnas, S. K. J., Dittrich, B., Meents, A., Messerschmidt, M. & Weckert, E. F. (2009). *Acta Cryst.* D**65**, 284–293.

Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.

Kass, I. J. & Sampson, N. S. (1998). *Biochemistry*, **37**, 17990–18000.

Klebe, G. (1994). *J. Mol. Biol.* **237**, 212–235.

Koch, O., Bocola, M. & Klebe, G. (2005). *Proteins*, **61**, 310–317.

Lario, P. I., Sampson, N. & Vrielink, A. (2003). *J. Mol. Biol.* **326**, 1635–1650.

Lario, P. I. & Vrielink, A. (2003). *J. Am. Chem. Soc.* **125**, 12787–12794.

Liebschner, D., Jelsch, C., Espinosa, E., Lecomte, C., Chabrière, E. & Guillot, B. (2011). *J. Phys. Chem. A*, **115**, 12895–12904.

Liu, Q., Huang, Q., Teng, M., Weeks, C. M., Jelsch, C., Zhang, R. & Niu, L. (2003). *J. Biol. Chem.* **278**, 41400–41408.

Lyubimov, A. Y., Chen, L., Sampson, N. S. & Vrielink, A. (2009). *Acta Cryst.* D**65**, 1222–1231.

Lyubimov, A. Y., Heard, K., Tang, H., Sampson, N. S. & Vrielink, A. (2007). *Protein Sci.* **16**, 2647–2656.

Lyubimov, A. Y., Lario, P. I., Moustafa, I. & Vrielink, A. (2006). *Nature Chem. Biol.* **2**, 259–264.

MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.

Mallinson, P. R., Smith, G. T., Wilson, C. C., Grech, E. & Wozniak, K. (2003). *J. Am. Chem. Soc.* **125**, 4259–4270.

Meot-Ner, M. & Sieck, L. W. (1986). *J. Am. Chem. Soc.* **108**, 7525–7529.

Munshi, P. & Guru Row, T. N. (2005). *CrystEngComm*, **7**, 608–611.

Muzet, N., Guillot, B., Jelsch, C., Howard, E. & Lecomte, C. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 8742–8747.

Navas, J., González-Zorn, B., Ladrón, N., Garrido, P. & Vázquez-Boland, J. A. (2001). *J. Bacteriol.* **183**, 4796–4805.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Sampson, N. & Vrielink, A. (2003). *Acc. Chem. Res.* **36**, 713–722.

Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.

Shen, Z., Corbin, D. R., Greenplate, J. T., Grebenok, R. J., Galbraith, D. W. & Purcell, J. P. (1997). *Arch. Insect Biochem. Physiol.* **34**, 429–442.

Stadtman, T. C., Cherkes, A. & Anfinsen, C. B. (1954). *J. Biol. Chem.* **206**, 511–523.

Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). *J. Mol. Biol.* **226**, 1143–1159.

Subramanian, S. & Zaworotko, M. J. (1994). *Coord. Chem. Rev.* **137**, 357–401.

Thomas, A., Benhabiles, N., Meurisse, R., Ngwabije, R. & Brasseur, R. (2001). *Proteins*, **43**, 37–44.

Urzhumtseva, L., Klaholz, B. & Urzhumtsev, A. (2013). *Acta Cryst.* D**69**, 1921–1934.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.

Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.

Vrielink, A. & Ghisla, S. (2009). *FEBS J.* **276**, 6826–6843.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 152.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wood, P. A., Pidcock, E. & Allen, F. H. (2008). *Acta Cryst.* B**64**, 491–496.

Yin, Y., Sampson, N. S., Vrielink, A. & Lario, P. I. (2001). *Biochemistry*, **40**, 13779–13787.

Yue, Q. K., Kass, I. J., Sampson, N. S. & Vrielink, A. (1999). *Biochemistry*, **38**, 4277–4286.

Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* A**63**, 108–125.

electronic reprint