Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The determination of protonation states in proteins

H. U. Ahmed,[a] M. P. Blakeley,[b]‡ M. Cianci,[c]§ D. W. J. Cruickshank,[a] J. A. Hubbard[d] and J. R. Helliwell[a,c]*

[a]School of Chemistry, Brunswick Street, The University of Manchester, Manchester M13 9PL, England, [b]EMBL Grenoble Outstation, BP 181, 38042 Grenoble CEDEX 9, France, [c]STFC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England, and [d]Computational, Analytical and Structural Sciences, GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, England

‡ Present address: ILL, 6 rue Jules Horowitz, BP 156, 38042 Grenoble, CEDEX 9 France.
§ Present address: PETRA III Project, EMBL Outstation, DESY Laboratory, Hamburg, Germany.

Correspondence e-mail: john.helliwell@manchester.ac.uk

© 2007 International Union of Crystallography
Printed in Denmark – all rights reserved

The protonation states of aspartic acids and glutamic acids as well as histidine are investigated in four X-ray cases: Ni,Ca concanavalin A at 0.94 Å, a thrombin–hirugen binary complex at 1.26 Å resolution and two thrombin–hirugen–inhibitor ternary complexes at 1.32 and 1.39 Å resolution. The truncation of the Ni,Ca concanavalin A data at various test resolutions between 0.94 and 1.50 Å provided a test comparator for the 'unknown' thrombin–hirugen carboxylate bond lengths. The protonation states of aspartic acids and glutamic acids can be determined (on the basis of convincing evidence) even to the modest resolution of 1.20 Å as exemplified by our X-ray crystal structure refinements of Ni and Mn concanavalin A and also as indicated in the 1.26 Å structure of thrombin, both of which are reported here. The protonation-state indication of an Asp or a Glu is valid provided that the following criteria are met (in order of importance). (i) The acidic residue must have a single occupancy. (ii) Anisotropic refinement at a minimum diffraction resolution of 1.20 Å (X-ray data-to-parameter ratio of ∼3.5:1) is required. (iii) Both of the bond lengths must agree with the expectation (*i.e.* dictionary values), thus allowing some relaxation of the bond-distance standard uncertainties required to ∼0.025 Å for a '3σ' determination or ∼0.04 Å for a '2σ' determination, although some variation of the expected bond-distance values must be allowed according to the microenvironment of the hydrogen of interest. (iv) Although the $F_o - F_c$ map peaks are most likely to be unreliable at the resolution range around 1.20 Å, if admitted as evidence the peak at the hydrogen position must be greater than or equal to 2.5σ and in the correct geometry. (v) The atomic B factors need to be less than 10 Å^2 for bond-length differentiation; furthermore, the C═O bond can also be expected to be observed with continuous $2F_o - F_c$ electron density and the C—OH bond with discontinuous electron density provided that the atomic B factors are less than approximately 20 Å^2 and the contour level is increased. The final decisive option is to carry out more than one experiment, *e.g.* multiple X-ray crystallography experiments and ideally neutron crystallography. The complementary technique of neutron protein crystallography has provided evidence of the protonation states of histidine and acidic residues in concanavalin A and also the correct orientations of asparagine and glutamine side chains. Again, the truncation of the neutron data at various test resolutions between 2.5 and 3.0 Å, even 3.25 and 3.75 Å resolution, examines the limits of the neutron probe. These various studies indicate a widening of the scope of both X-ray and neutron probes in certain circumstances to elucidate the protonation states in proteins.

Received 6 March 2007
Accepted 19 June 2007

PDB References: Ni,Ca concanavalin A, 2uu8, r2uu8sf; thrombin–hirugen, 2uuf, r2uufsf; thrombin–hirugen–gw473178, 2uuj, r2uujsf; thrombin–hirugen–gw420128, 2uuk, r2uuksf; (Mn,Ca) concanavalin A, 2yz4, r2yz4sf.

1. Introduction

Hydrogen atoms can play a major role during protein function. They are involved in enzymatic redox reactions that involve the transfer of protons. The mechanism by which serine proteases hydrolyse peptide bonds (substrates) involves acylation and deacylation in the triad active site. This catalytic process has been extensively debated with regard to the residues that are involved in the process, in particular the chemical base. Using neutron protein crystallography, Kossiakoff & Spencer (1980) identified that His57 in trypsin was protonated (as deuterium). The location of H atoms (as deuterium) *via* neutron protein crystallography has also revealed the protonation in the active site of lysozyme and in aspartic proteinase as well as the preferred orientation of methyl groups in crambin and the ordered solvent patterns for concanavalin A (reviewed by Tsyba & Bau, 2002). In order to understand structure–function relationships and for the rational three-dimensional design of drugs, knowledge of the protonation states of protein, ligands and also water molecules is ideally required.

The number of H atoms exceeds the total number of C, N, O and S atoms in an enzyme. In high-resolution structures, the detailed structural information improves our understanding of enzyme mechanisms. Ultrahigh-resolution X-ray structures (typically ∼0.9–1.0 Å) have allowed the determination of atom positions with very high accuracy, including the positions of H atoms. In the X-ray structure of human aldose reductase solved at an exceptional 0.66 Å resolution, 54% of the possible H atoms in the whole protein were observed in $F_o - F_c$ maps contoured at 2.5σ (Howard *et al.*, 2004). The X-ray structure of the serine protease subtilisin was solved at 0.78 Å resolution with 65% of the H atoms modelled in the final $F_o - F_c$ map contoured at 2σ; 14% of them appeared in the map contoured at 3σ (Kuhn *et al.*, 1998). H atoms have also been observed at 1.10 Å resolution: in elastase, His57 was singly protonated on $N^{\delta1}$ as indicated by the $F_o - F_c$ map contoured at 2.5σ (Würtele *et al.*, 2000). In terms of methods, Sheldrick (1990) has described the minimum diffraction resolution required to solve a structure by direct methods as being 1.20 Å. This 'atomic resolution' limit is based on the atomicity principle, where electron-density peaks in the Fourier synthesis corresponding to individual atoms do not overlap. As the average bond length within crystallized molecules is ∼1.2–1.50 Å, the requirement of non-overlapping atoms is thus often fulfilled. Atomic resolution has further been proposed to require that more than 50% of the reflections in the outer shell of such diffraction data should have $I > 2\sigma(I)$ and should have an R_{merge} on intensity of <25%.

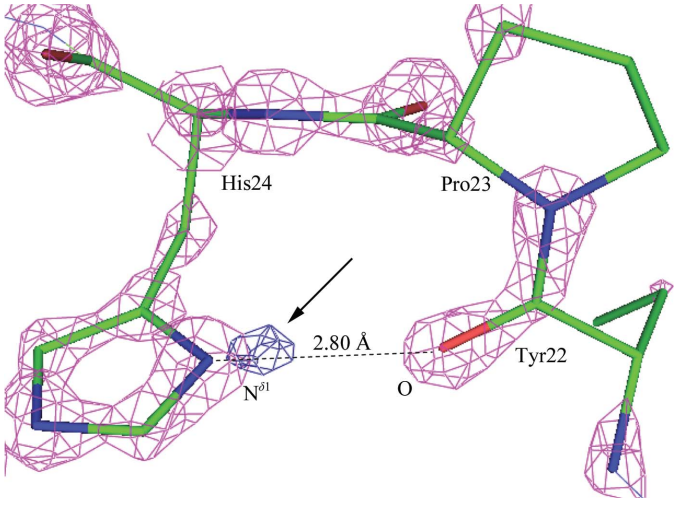

Figure 1
$2F_o - F_c$ electron-density map section (magenta, 4σ) showing His24 $N^{\delta1}$ forming a hydrogen bond with Tyr22 main-chain O. The $F_o - F_c$ electron-density map (blue, 3σ, marked with an arrow) indicates the H atom on the $N^{\delta1}$ atom.

The two proteins studied here were crystallized at pH 7.0 (thrombin) and at pH 6.5 (concanavalin A), with the chance of seeing both protonated and unprotonated carboxyl residues, including double and single *versus* mixed bond lengths, respectively. The reasons for observation of protonation at pH 6–7 could arise from a perturbed pK_a around a carboxylate group as reported by Engh & Huber (1991) and Harris & Turner (2002). Indeed, Minasov *et al.* (2002) reported the protonation of an active-site glutamic acid residue in the protein TEM-1 β-lactamase (solved at 0.85 Å resolution with X-rays) crystallized at pH 8. The pK_a of Glu and Asp in a protein is approximately 4.5 (Cantor & Schimmel, 1980). Another reason, apart from pK_a effects, for an extended bond length could be the consequence of the oxygen (from the carboxylate group) being involved in hydrogen bonding, *i.e.* its microenvironment is important (see §1.2). The δ^+ hydrogen can be so strongly attracted to the lone pair of an O atom that it is almost as if a coordinate (dative covalent) bond is beginning to be formed. Although not an actual covalent bond, such an attraction is significantly stronger than an ordinary dipole–dipole interaction such that the C–O bond length increases. An example of this can be seen in the atomic resolution analysis of an aspartic proteinase by Erskine *et al.* (2003), in which Fig. 8 shows a COO^- for Asp32 in this protein but with one C—O bond distance lengthened to 1.27 Å by a proximal strongly hydrogen-bonding H atom 2.58 Å away.

The work reported here includes the X-ray structural study and refinement of a human α-thrombin–hirugen binary complex at 1.26 Å resolution, which is tantalizingly close to the possibility of observing the well ordered H atoms. Furthermore, human α-thrombin is a trypsin-like serine protease in which protonation details are of keen interest. Thrombin plays a very important role in the coagulation of blood, in which contact with fibrinogen (the natural substrate) results in the formation of fibrin intermediates that polymerize into a blood clot.

As our X-ray diffraction resolution for the thrombin–hirugen binary complex was 1.26 Å, we initially assumed that we would be unable to unambiguously discriminate a single bond from a double bond or observe $F_o - F_c$ electron-density maps that would confirm the protonation states of the N atoms (N^{δ} and or N^{ε}) in histidines and O atoms ($O^{\delta1}/O^{\varepsilon1}$ or $O^{\delta2}/O^{\varepsilon2}$) of the carboxylate group in acidic residues. To examine this assumption led us to explore Ni,Ca concanavalin A data

(0.94 Å resolution) as a case study for discriminating a single bond from a double bond in carboxylate groups at various truncated data-set resolutions between 0.94 and 1.50 Å resolution. Concanavalin A is a saccharide-binding protein isolated from the jack bean. It is thought to mediate cell–cell interactions by binding to polysaccharide on the cell surface and may also have an antifungal plant-defence role [for a review, see Kalb (Gilboa) & Helliwell, 2001]. Here, we have investigated how the bond lengths of carboxyl C—O bonds change with the truncated diffraction data-set resolutions and examined the minimum resolution required to distinguish a carboxyl group from a carboxylate group. We also assess the visibility of hydrogen on histidines at 0.94 Å using X-rays and 2.20 Å resolution using neutrons. An evaluation of neutron protein crystallography at truncated resolutions is then also made with concanavalin A and reported here.

1.1. Electron-density difference maps ($F_o - F_c$) and the visibility of H atoms: general remarks and specific examples from the current work

The electron density at a point (xyz) is given by

$$\rho(xyz) = (1/V)\sum_{hkl} F(hkl)\exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)].$$

For good electron-density maps the resolution and data completeness are two of the most important criteria. Low-resolution reflections, which have strong intensity ($I \propto |F|^2$), make a large contribution to the map. High-resolution reflections with their weaker intensity provide fine detail. Therefore, in order to obtain good electron-density maps one must capture the high-resolution reflections whilst retaining the low-resolution data. Reflections that are not measured will contribute noise to the density map and as a result possibly create ambiguities in parts of the structure; in other words, there is a risk of being wrong.

H atoms are difficult to determine by X-ray diffraction because they only contain one electron and therefore diffract X-rays very weakly. However, at atomic and ultrahigh resolution H atoms start to appear in $F_o - F_c$ electron-density maps. This is true provided that the H atoms are relatively static, *i.e.* very well ordered, such as polypeptide main-chain H atoms (Dauter, 2003). Clearly, the bonding situation is also important: in a covalent bond the electron on the H atom is drawn away towards its bonding partner, exacerbating the challenge to X-rays as the probe. Fig. 1 clearly shows a protonation site visible in the $F_o - F_c$ map of His24 $N^{\delta1}$ obtained from the X-ray crystal structure of Ni,Ca concanavalin A at 0.94 Å resolution reported here (PDB code 2uu8).

For thrombin, we illustrate the problem of relying purely on $F_o - F_c$ electron-density maps at resolutions that are 'only' approaching atomic. Figs. 2(*a*) and 2(*b*) show the electron density ($2F_o - F_c$; magenta, 3.5 r.m.s.) and $F_o - F_c$ (blue, 2.5 r.m.s.) maps of the catalytic triad residues Asp102 and His57 for the 1.26 Å thrombin–hirugen structure (Fig. 2*a*) and the 1.32 Å ternary complex structure with the inhibitor gw473178 (Fig. 2*b*). These $F_o - F_c$ electron-density maps offer no significant detail, therefore the focus has to be on the relevant bond lengths, which we now investigate in detail.

1.2. Bond-length differences and protonation states

1.2.1. Aspartic and glutamic acids. Besides the use of difference electron-density maps as key indicators of protonation states from X-ray studies, the other possibility for analysis that will confirm the protonation states of acidic residues and histidines is to compare the bond-length differences. High-resolution data play an important role, as the X-ray data-to-parameter ratio in refinement will be high and so the positions of the atoms can be much better defined along with the anisotropic atomic displacement parameters (ADPs). The model will also contain lower estimated standard uncertainties. Borthwick (1980) discusses the carboxylate group (COO^-) bond length and angle correlations and uses 1.25 Å for the delocalized C—O bond distance and 1.21 and 1.31 Å, respectively, for the C═O and C—OH bond distances in the carboxylic group (COOH). These distances can be influenced *via* the atomic environment. Two examples free from such effects are now described. The first is *N*-acetylglycine (Donohue & Marsh, 1962), in which the C═O and C—OH bonds are 1.219 (6) and 1.319 (6) Å, respectively. Another such

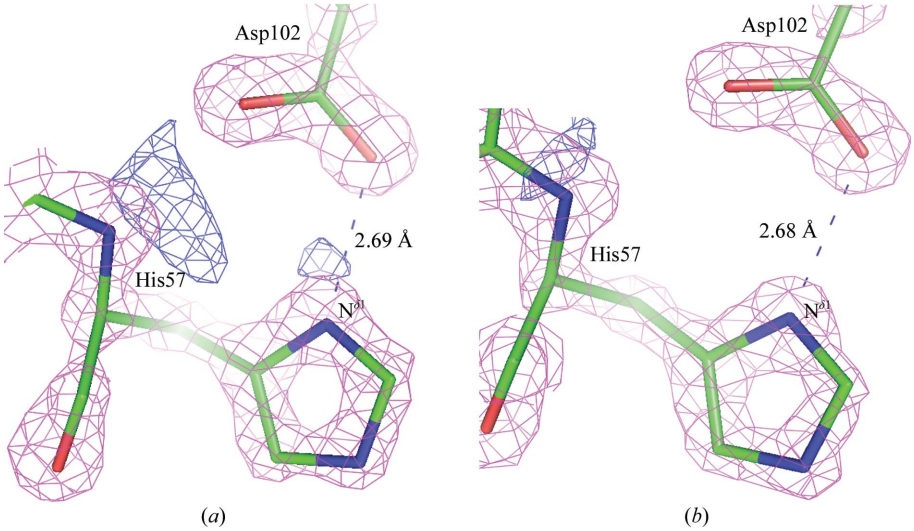

Figure 2
(*a*) The 1.26 Å thrombin–hirugen binary complex triad active site with His57 and Asp102. The $2F_o - F_c$ electron-density map is shown in magenta (3.5 r.m.s.) and the $F_o - F_c$ map (2.5 r.m.s.) is visible on the amide N and $N^{\delta1}$ of His57. However, when contoured at 3 r.m.s. the $F_o - F_c$ map peak is no longer visible on $N^{\delta1}$. (*b*) The 1.32 Å thrombin–hirugen–gw473178 ternary complex triad active site with His57 and Asp102. The $2F_o - F_c$ electron-density map is shown in magenta (3.5 r.m.s.) and the $F_o - F_c$ map peak in blue (2.5 r.m.s.) is visible below the amide N (slightly offset) only.

case that is free of intermolecular effects is where gas-phase electron diffraction has been used. Oxalic acid dihydrate studied in the crystal and in the gas phase (Ahmed & Cruickshank, 1953; Nahlovska *et al.*, 1970) yielded C=O and C—OH bond distances of 1.187 (22) and 1.285 (12) Å for the crystal and 1.208 (1) and 1.339 (2) Å in the gas phase, respectively. Zobel *et al.* (1992) undertook a charge-density study on oxalic acid dihydrate at 15 K, yielding respective bond lengths of 1.224 and 1.288 Å, *i.e.* a difference of only 0.064 Å: is this a consequence of low temperature, charge-refinement methods or some other reason? With these variations in mind, one approaches the fact that there are differences in dictionary defined distances: Engh & Huber (1991) quote 1.249 (19) Å for the COO^- bond distance and 1.208 (23) and 1.304 (22) Å for the C=O and C—OH bond distances, respectively. The Cambridge Structural Database (CSD) embodied in *SHELXL*-97 and *REFMAC* uses 1.249 Å for COO^-. In terms of a protein example, Deacon *et al.* (1997) made a precise determination with standard uncertainty (s.u.) values for Asp28 in concanavalin A, yielding COOH double-bond and single-bond distances from the carbon of 1.170 (9) Å and 1.324 (10) Å, respectively, while its hydrogen-bonded partner the unprotonated Glu8 had bond distances from the C atoms to its O atoms of 1.253 (9) and 1.275 (9) Å, respectively; thus, it was shown that the difference (0.083 Å) between the double-bond distance (1.170 Å) and the shortest COO^- distance (1.253 Å) was significant (s.u. of 0.0127 Å on 0.083 Å). Fig. 3 shows the current CSD values. It should be emphasized that the precisions of these distances are optimal if the relevant atomic *B* factors are as low as possible, as this is a predominant driver of the precision obtainable. The spread of values also includes chemical causes that depend on microenvironment effects, as referred to above. The quoted values for concanavalin A above are for a 100 K cryo-temperature structure and therefore the atoms involved have *B* factors much improved over room temperature.

1.2.2. Histidine. The histidine bond-length differences around the imidazole ring show only small variations and small differences upon protonation (*e.g.* Edington & Harding, 1974, and the Cambridge Structural Database; Fig. 4). The imidazole ring can exist as two tautomers (Fig. 4*a*), with a H atom on either $N^{\delta1}$ or $N^{\epsilon2}$, which can further complicate analyses using bond distances or electron-density maps. It is also possible for His to be fully protonated, *i.e.* a H atom is bonded to both N atoms (Fig. 4*b*).

1.3. Using neutron diffraction with H atoms replaced by D atoms

D atoms usually exchange with the H atoms that are attached to N atoms and O atoms. The scattering length of deuterium for neutrons is positive and is similar to those of carbon, nitrogen and oxygen; thus, D atoms are visible at more modest diffraction resolutions, *e.g.* 2.00–2.50 Å, including even the more mobile H atoms (as deuteriums), when using neutrons compared with X-rays. The D atoms in bound D_2O molecules are much more efficiently determined using neutrons than the H atoms in bound H_2O are determined using X-rays (Habash *et al.*, 2000; Deacon *et al.*, 1997). This approach is reviewed in Blakeley *et al.* (2004). A single oxygen density peak is mostly observed for water molecules using X-ray data, providing only locational and not orientational information and therefore only a very partial picture of the hydrated protein surface. Occasionally, X-rays do reveal water H atoms (see Plate 2b of Deacon *et al.*, 1997). Here, we also investigate in detail the determination of protonation states in proteins using neutron data, scrutinizing the diffraction resolution required with neutrons.

2. Methods

2.1. Crystallization

2.1.1. Thrombin–hirugen binary and ternary (*i.e.* with inhibitor) complexes. The activation and purification of human α-thrombin was performed using a modified procedure of Ngai & Chang (1991). The crystallizations were performed using a modification of the method of Skrzypczak-Jankun *et al.* (1991). An approximately tenfold excess of hirugen was prepared in phosphate buffer and added to human α-thrombin. Crystals were obtained using the sitting-drop vapour-diffusion method with 2 + 2 µl drops and the following conditions: 28% PEG 4K, 600 m*M* NaCl, 100 m*M* HEPES pH

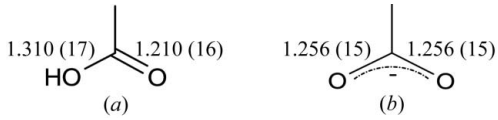

Figure 3
Bond lengths of (*a*) a carboxylic group and (*b*) the delocalized carboxylate group obtained from the Cambridge Structural Database (CSD; September 2006). The search in (*a*) was for entries involving any carboxylic acid side chain to an α-carbon and in (*b*) for entries with a delocalized carboxylate group.

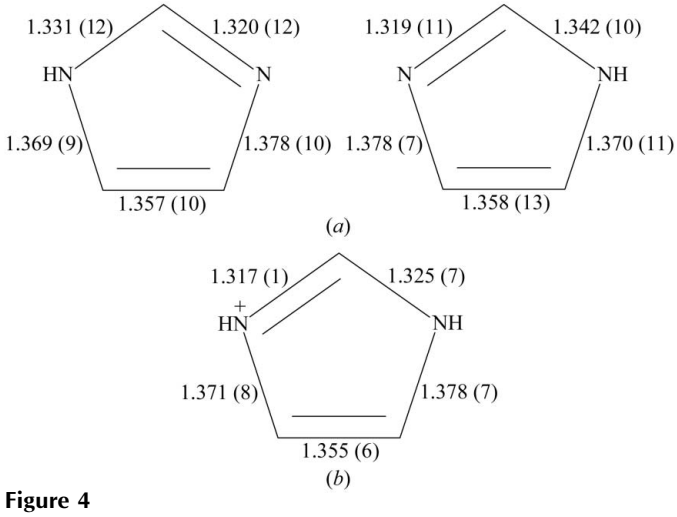

Figure 4
(*a*) Bond lengths of a histidine imidazole ring from the current CSD. As in Fig. 3, the CSD search allowed any α-carbon substituents. (*b*) Bond lengths of a protonated histidine imidazole ring from the CSD (September 2006).

Table 1
Summary of diffraction and refinement data statistics.

The models were initially refined against *F* using *REFMAC*5 and then finally refined against F^2 using *SHELXL*-97, apart from the thrombin–hirugen–gw473178 complex where both refinements were against *F* as this was well behaved; we note that ice-ring annuli and missing data were pronounced for this data set. For the full neutron refinement of Mn concanavalin A, *CNS* was used for refinement against *F*. Values in parentheses are for the highest resolution shell.

	X-ray diffraction data				Neutron diffraction data
	Ni,Ca concanavalin A	Thrombin–hirugen	Thrombin–hirugen –gw473178	Thrombin–hirugen –gw420128	(Mn,Ca) concanavalin A
PDB code	2uu8	2uuf	2uuj	2uuk	2yz4
Space group	*I*222	*C*2	*C*2	*C*2	*I*222
Unit-cell parameters (Å, °)					
a	88.87	70.27	70.00	70.35	89.38
b	86.05	71.29	71.97	71.44	87.31
c	61.54	72.04	72.19	72.12	63.05
β		100.00	100.63	100.03	
Resolution (Å)	33.23–0.94	49.39–1.26	35.78–1.32	18.55–1.39	500–2.2
$R_{\rm merge}$	0.093 (0.33)	0.065 (0.51)	0.075 (0.36)	0.067 (0.34)	0.21 (0.31)
Completeness (%)	89.22 (54.37)	94.72 (93.00)	86.93 (99.99)	95.91 (92.40)	67.2 (46.9)
$\langle I/\sigma(I)\rangle$	11.84 (1.98)	7.6 (1.5)	12.75 (2.32)	14.9 (4.1)	6.8 (3.1)
Multiplicity	6.0 (2.5)	3.8 (2.9)	5.7 (3.4)	3.9 (3.3)	6.2 (4.0)
No. of unique reflections	136027 (11781)	89355 (12770)	71295 (11087)	67018 (8517)	8511 (861)
R factor (%)	12.0	14.6	15.3	11.8	28.1
$R_{\rm free}$ (%)	14.2	19.4	21.2	17.3	31.3
Ramachandran core/additional (%)	90.9/9.1	86.9/13.1	88.2/11.8	86.9/13.1	88.9/11.1
Cruickshank's DPI (Å)	0.017	0.041	0.056	0.047	—
No. of atoms in protein	1896	2291	2320	2316	3564†
Mean *B* factor of protein (Å^2)	14.6	20.6	22.4	21.2	26.4†
No. of atoms in hirugen molecule	—	94	94	94	—
Mean *B* factor of hirugen (Å^2)	—	48.3	50.6	44.9	—
Bond-distance r.m.s. deviation (Å)	0.018	0.034	0.012	0.061	—
Bond-angle r.m.s. deviation (°)	2.514	2.721	2.277	2.689	—
No. of atoms in inhibitor	—	—	28	33	—
Mean *B* factor of inhibitor (Å^2)	—	—	15.8	16.2	—
No. of water molecules	277	344	335	345	83
Mean *B* factor of water molecules (Å^2)	30.2	31.4	30.5	32.8	26.5‡
Mean *B* factor for all side-chain atoms (Å^2)	17.2	32.7	34.7	32.2	26.6†
Mean *B* factor for all main-chain atoms (Å^2)	12.1	18.5	20.4	25.5	25.8†
Mean *B* factor for all side chains and waters (Å^2)	23.7	32.0	32.6	32.5	26.6†
Other atoms	[Ni^{2+}, Ca^{2+}]	[Na^{+}, Ca^{2+}]	[Na^{+}, Ca^{2+}]	[Na^{+}, Ca^{2+}]	[Mn^{2+}, Ca^{2+}]
Total No. of atoms	2185	2733	2799	2790	3810†
Mean *B* factor for all atoms (Å^2)	16.7	23.0	24.6	23.4	26.4†

† Includes H and D atoms. ‡ D_2O molecules.

7.0 at 293 K. The crystal habit is cuboid, with dimensions 200 × 200 × 100 µm. For the thrombin–hirugen–inhibitor ternary complexes, a thrombin–hirugen crystal was harvested in 28% PEG 4K, 600 m*M* NaCl, 100 m*M* HEPES pH 7.0 and then soaked for 24–48 h at 277 K with the small-molecule inhibitor (100–200 µm dissolved in 100% DMSO). (The precise chemical nature of the inhibitors gw473178 and gw420128 are not relevant to the matters discussed here.)

2.1.2. Ni,Ca concanavalin A. The purification and crystallization of saccharide-free metal-substituted concanavalin A has been described by Kalb (Gilboa) *et al.* (1988). The Ni,Ca concanavalin A crystals grew to average dimensions of 300 × 300 × 300 µm; they were smaller than other metal-substituted concanavalin A crystals. MPD was mixed with the mother liquor at 25% as a cryoprotectant as used by Deacon *et al.* (1997). The very large crystals of native (Mn,Ca) concanavalin A for the room-temperature neutron study were grown by Dr A. J. Kalb (Gilboa) by batch dialysis as described previously by Greer *et al.* (1970); in brief, crystals were grown at room temperature and pH 6.5 from protein solutions in dialysis bags immersed for several weeks in 0.1 *M* $NaNO_3$, 0.05 *M* Tris–acetate, 1 m*M* $MnCl_2$, 1 m*M* $CaCl_2$ in D_2O.

2.2. Data collection and processing

Table 1 summarizes the data statistics. Our strategy with thrombin involved firstly undertaking as high a resolution X-ray crystallographic data collection as possible at the Daresbury SRS on one of its strongest intensity beamlines, namely SRS beamline 10 (Cianci *et al.*, 2005), including consideration of the use of cryoprotectants. The latter did not lead to the best diffraction resolution; direct freezing in a 100 K nitrogen-gas stream produced the best diffraction resolution (1.26 Å for thrombin–hirugen and 1.32 Å for thrombin–hirugen–gw473178). Presumably, the 28% PEG 4K in the crystallization solution proved to be a good cryoprotectant without further modification. As we were tantalizingly close to 1.20 Å resolution using SRS beamline 10

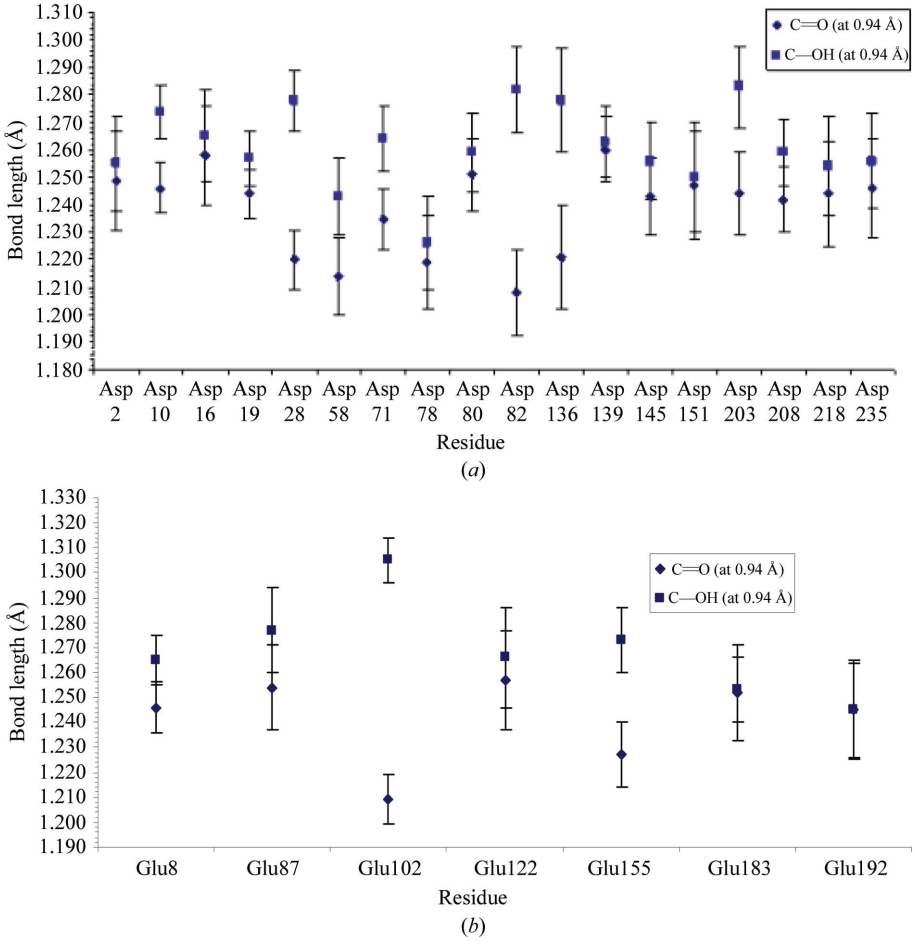

Figure 5
(*a*) Carboxyl-group bond lengths (restrained) in aspartic acids for Ni,Ca concanavalin A at 0.94 Å resolution. Asp28 and Asp82 show protonation. Asp136 is marginal. (*b*) Carboxyl-group bond lengths (restrained) in glutamic acids for Ni,Ca concanavalin A at 0.94 Å resolution. Glu102 shows good indication of protonation. Glu8 is involved in a hydrogen bond with Asp28 and therefore is not protonated, as is also clearly indicated.

(supplementary material *A* Figs. 1–3[1]), we further attempted data collection at the yet more intense APS SBC Sector undulator ID19. This led to a data set for thrombin–hirugen–gw473178 with an overall increased $\langle I/\sigma(I)\rangle$ (supplementary material *A* Fig. 2, curve labelled THR-hir-gw473178_APS_19ID), but the data resolution limit was restricted (1.35 Å) owing to spot splitting. This data was not merged with the SRS data because of non-isomorphism. For Ni,Ca concanavalin A, as these crystals are intrinsically small (as reported above), the APS SBC Sector undulator ID19 was used, although preliminary diffraction tests of pre-frozen Ni,Ca concanavalin A crystals were made on SRS beamline 10 (prior to shipping in a dewar to APS). Supplementary material *A* Figs. 4–6 show the diffraction processing statistics for the Ni,Ca concanavalin A APS SBC ID19 data along with the CHESS multipole wiggler native (Mn,Ca) concanavalin A data statistics (Deacon *et al.*, 1997) for ease of comparison; the statistics for the native (Mn,Ca) concanavalin A (Price, 1999) are similar but have somewhat higher completeness than those reported in Deacon *et al.* (1997), especially between 0.9 and 0.94 Å [*i.e.* for the data in Price (1999) the completeness was 83.4% overall and 33.6% in the last shell (0.92–0.94 Å), $R_{\rm merge}$ was 3.5% (44.4% in the last shell), the total number of unique reflections was 138 735 and $\langle F/\sigma(F)\rangle$ at the edge was 2.0].

2.2.1. Thrombin–hirugen binary complex. X-ray data were collected at SRS Daresbury on beamline 10 (Cianci *et al.*, 2005) with a wavelength of 0.98 Å using a CCD MAR 165 detector. Two sets of data were collected: a slow pass with 30 s exposure per 1° image (200 images) and a quick pass with 2 s exposure per 1° image (200 images). The crystal was cryocooled using a liquid-nitrogen gas stream set at 100 K. Data processing was performed using *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1997; see supplementary material *A* Figs. 1–3 for full data statistics). The resolution where $R_{\rm merge}$ crossed 0.20 was 1.43 Å and the resolution where $\langle I/\sigma(I)\rangle$ crossed 2.0 was 1.32 Å. Since there has been extensive debate about resolution limits, for example on the CCP4 bulletin board, we have preserved the limit as 1.26 Å, *i.e.* where $\langle I/\sigma(I)\rangle$ crossed 1.5 [$\langle F/\sigma(F)\rangle = 3$].

2.2.2. Thrombin–hirugen–gw473178 ternary complex. X-ray data collection to 1.32 Å from thrombin–hirugen–gw473178 was performed at SRS Daresbury on beamline 10 (Cianci *et al.*, 2005) with a wavelength of 1.37 Å using a CCD MAR 165 detector. Two sets of data were collected: a slow pass with 80 s exposure per 1° image (200 images) and a quick pass with 20 s exposure per 1° image (200 images). The crystal was cryocooled using a liquid-nitrogen gas stream set at 100 K. Ice rings were present on the diffraction images, but data collection was continued as it was of high overall quality. Data processing was performed using *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1997). The resolution where $R_{\rm merge}$ crossed 0.20 was 1.45 Å. The resolution where $\langle I/\sigma(I)\rangle$ crossed 2.0 was 1.26 Å, but since there was splitting of the diffraction spots evident at the edge of the pattern this led us to use the data to 1.32 Å.

2.2.3. Thrombin–hirugen–gw420128 ternary complex. X-ray data collection to 1.39 Å from the thrombin–hirugen–gw420128 complex resolution was performed at the APS in Chicago on beamline 17ID (IMCA CAT) with a wavelength of 1.00 Å using an ADSC Quantum 210 detector. The crystal was cryocooled using a liquid-nitrogen gas stream set at 100 K. Data processing was performed using *HKL*-2000 (Otwinowski

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: FW5132). Services for accessing this material are described at the back of the journal.

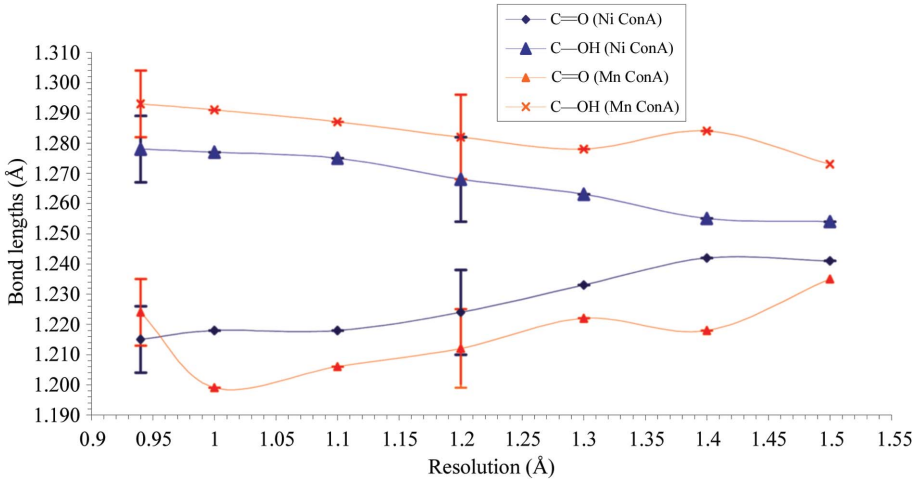

Figure 6
Asp28 carboxyl-group bond lengths (restrained) *versus* resolution for Ni,Ca concanavalin A and native (Mn,Ca) concanavalin A refined using *SHELXL*-97. As expected, with restraints on the dictionary 1.249 Å bond length dominates from ∼1.35 to 1.50 Å resolution. This is true in both cases, but Ni,Ca concanavalin A 'behaves' better.

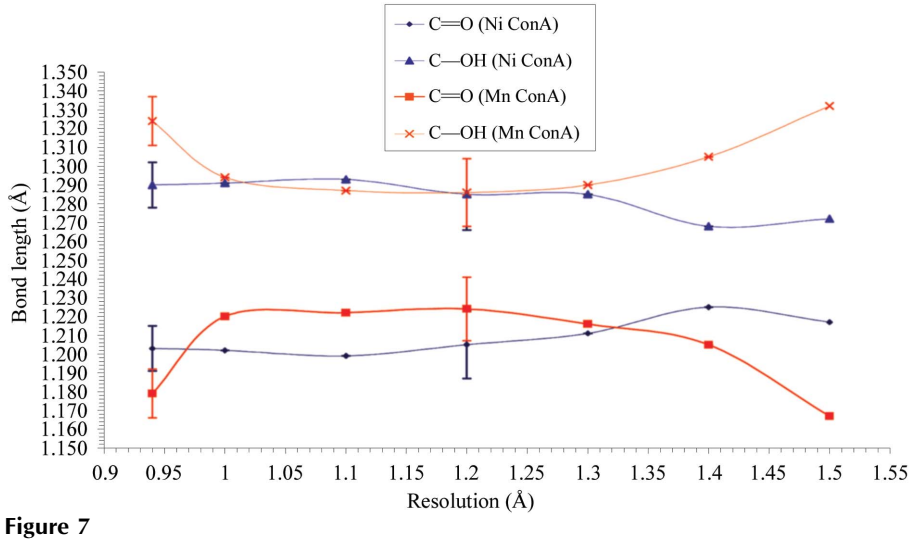

Figure 7
Asp28 carboxyl-group bond lengths (unrestrained) *versus* resolution for Ni,Ca concanavalin A and native (Mn,Ca) concanavalin A refined using *SHELXL*-97. With restraints off, the stability of these bond distances as the protonation indicator is good in both cases to 1.30 Å. Between 1.30 and 1.50 Å resolution the two cases behave differently.

& Minor, 1997; see supplementary material *A* Figs. 1–3 for full data statistics). The resolution where $R_{\rm merge}$ crossed 0.20 was 1.46 Å and the detector geometric edge and distance from the crystal had set the resolution limit as 1.39 Å [$\langle I/\sigma(I)\rangle$ = 4.1]. [The resolution where $\langle I/\sigma(I)\rangle$ crossed 2.0 was 1.30 Å after extrapolation.]

2.2.4. Ni,Ca concanavalin A. Data collection was performed at APS in Chicago on beamline 19ID (SBC) with a wavelength of 0.67 Å using an ADSC Quantum 315 detector. Three crystals of Ni,Ca concanavalin A were used to collect a full data set. The crystals were cryocooled using a liquid-nitrogen gas stream set at 100 K; the cryoprotectant was MPD, as described in Deacon *et al.* (1997). Diffraction data processing was performed using *HKL*-2000 (Otwinowski & Minor, 1997; see supplementary material *A* Figs. 4–6 for full data statistics). The resolution where $R_{\rm merge}$ crossed 0.20 was 1.11 Å. The resolution where $\langle I/\sigma(I)\rangle$ crossed 2.0 was 0.95 Å.

2.2.5. Neutron data collection from native (Mn,Ca) concanavalin A at room temperature. The selected crystal (∼7 × 3 × 1 mm) was mounted in a quartz capillary with a small amount of mother liquor in order to prevent the crystal from drying out during data collection. A restricted narrow bandpass ($\delta\lambda/\lambda$ = 25%) of wavelength range 2.9–3.9 Å was selected in order to reduce background scatter using a multi-mirror band-pass filter comprising 40 Si-crystal mirrors, each with 748 alternating 74–90 Å thick Ti and Ni layers. Neutron Laue data were collected at the Institut Laue–Langevin 'LADI-1' diffractometer in 8° step intervals about the rotation axis of the instrument (φ). The exposure time per image was 12 h; 5 min were needed to scan the image plate and 3 min to erase it. A total of 30 images were collected; thus, the entire data set was collected in ∼15 d, illustrating the efficiency of Laue data-collection methods for a protein of quite high molecular weight compared with most previous neutron protein crystallography structural studies.[2] The neutron Laue data were processed using the Daresbury Laboratory software (Campbell, 1995; Helliwell *et al.*, 1989) modified for the cylindrical detector geometry and with the polarization correction removed (Campbell *et al.*, 1998). The orientation of each crystal was determined using the auto-indexing routines in the *LAUEGEN* program and integration of each Laue reflection was performed. The program *LSCALE* (Arzt *et al.*, 1999) was used to derive the wavelength-normalization curve using the intensities of symmetry-equivalent reflections measured at different wavelengths. No account was made for crystal damage since neutrons do not induce detectable radiation damage and no explicit absorption corrections were applied. The internal agreement factor from *LSCALE* for the intensities of reflections with wavelength differences of less than 0.1 Å was 17.8% to 2.2 Å. These data (52 759 observations) were merged using *SCALA* (Evans, 1997), with the final data set comprising 8512 unique reflections to 2.2 Å resolu-

[2] Exposure times on 'LADI-3' are ∼2.5 times less *pro rata* owing to an improved readout efficiency.

tion, with an overall R_{merge} of 21.3%. This diffraction resolution is a significant improvement over the previous room-temperature neutron structures of native concanavalin A of 2.7 and 2.4 Å (Habash *et al.*, 1997, 2000). The resolution where R_{merge} crossed 0.20 was 3.9 Å and that where $\langle I/\sigma(I)\rangle$ crossed 2.0 is at ∼2.0 Å (by extrapolation). $\langle I/\sigma(I)\rangle$ at 2.2 Å is 3.1.

2.3. Protein structure refinement

The X-ray crystal structure refinements were carried out using *REFMAC*5 (Collaborative Computational Project, Number 4, 1994) and *SHELXL*-97 (Sheldrick, 1997). *REFMAC*5 was used initially, apart from the last two cycles of refinement, *i.e.* refining multiple occupancies and anisotropic refinement, which were performed in *SHELXL*-97. The reason for using *SHELXL*-97 is because it allows refinement of multiple occupancies whereas *REFMAC*5 does not. The neutron structural refinement of native (Mn,Ca) concanavalin A was carried out using *CNS* (Brünger *et al.*, 1998). The overall refinement statistics are given in Table 1.

2.3.1. Thrombin–hirugen binary complex. The starting model coordinates for structure refinement were those of a different thrombin–hirugen–inhibitor complex solved at 1.30 Å resolution; these were subject to small random atom shifts using *MOLEMAN*2 (Kleywegt *et al.*, 2001) to allow a new R_{free} reflection subset to be picked. Rigid-body refinement was carried out using *REFMAC*5 followed by four rounds of restrained refinement consisting of ten cycles of both positional and isotropic temperature-factor refinement. This was followed by two rounds of restrained conjugate-gradient least-squares anisotropic refinement using *SHELXL*-97. $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated using *FFT* (Collaborative Computational Project, Number 4, 1994) after each round of refinement and the electron density and the model were then analysed using *O* (Jones *et al.*, 1991). Waters were added in several rounds of refinement using peaks $\geq 3\sigma$ in the $F_o - F_c$ electron-density maps, which were determined by *PEAKMAX* and *WATPEAK* and then checked in *O*, until no more suitable peaks could be found. Visual inspection at various stages allowed the incorporation of double conformations for seven side chains and the fractional occupancy of each conformation was refined in *SHELXL*-97 (Sheldrick, 1997). The electron-density map calculation and model building were performed as mentioned above for all the other crystal structures reported here.

2.3.2. Thrombin–hirugen–gw473178 ternary complex at 1.32 Å. The starting model coordinates for structure refinement were those of the thrombin–hirugen binary complex solved at 1.26 Å resolution (see §2.3.1). Rigid-body refinement was carried out using *REFMAC*5 followed by three rounds of restrained refinement consisting of ten cycles of both positional and isotropic temperature-factor refinement. Reflections from the ice rings were removed in round 5 and round 11. The inhibitor was added in the ninth cycle. Two rounds of restrained conjugate-gradient least-squares anisotropic refinement were made using *SHELXL*-97 (Sheldrick, 1997). Attempts were made to utilize data to 1.28 Å, but the resolution finally selected was 1.32 Å by inspection of the outermost shell statistics.

2.3.3. Thrombin–hirugen–gw420128 ternary complex. The starting model coordinates for structure refinement were those of the thrombin–hirugen binary complex solved at 1.26 Å resolution (see §2.3.1). Rigid-body refinement was carried out using *REFMAC*5 followed by three rounds of restrained refinement consisting of ten cycles of both positional and isotropic temperature-factor refinement. This was followed by two rounds of restrained conjugate-gradient least-squares anisotropic refinement using *SHELXL*-97 (Sheldrick, 1997).

2.3.4. Ni,Ca concanavalin A. The starting model coordinates for structure refinement were those of native (Mn,Ca) concanavalin A (PDB code 1nls; Deacon *et al.*, 1997) refined at 0.94 Å resolution. Rigid-body refinement was carried out using *REFMAC*5 followed by seven rounds of restrained refinement consisting of ten cycles of both positional and isotropic temperature-factor refinement. This was followed by two rounds of restrained conjugate-gradient least-squares anisotropic refinement using *SHELXL*-97 (Sheldrick, 1997). A critical feature of assessing the protonation state *via* the bond distances is to gain an understanding of the standard uncertainties on those distances. Full matrix inversion using *SHELXL*-97 was undertaken and proved stable, *i.e.* it yielded reliable distances and standard uncertainties for the cases of full restraints and the 0.94 Å resolution limit (Fig. 5). The inclusion of the R_{free} reflections at the end of model refinement proved sufficient to stabilize the determination of refined bond distances and standard uncertainties for the

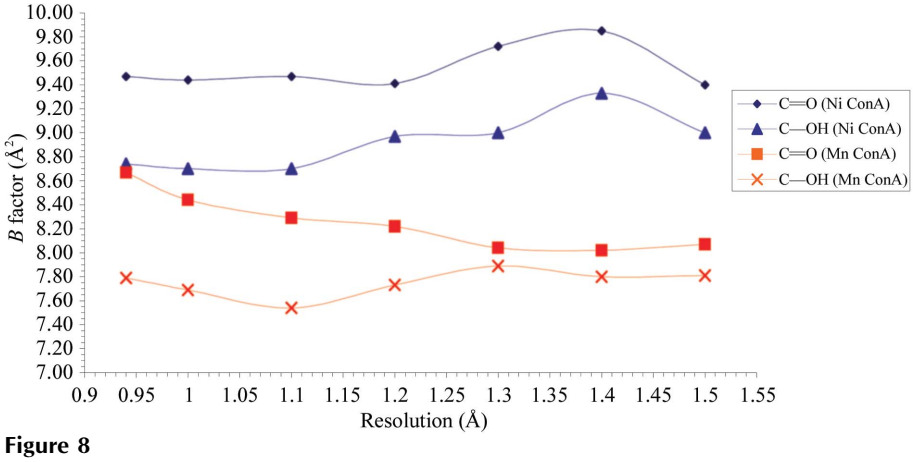

Figure 8
Asp28 *B* factors for O atoms in the carboxylic group (unrestrained) *versus* resolution for Ni,Ca concanavalin A and native (Mn,Ca) concanavalin A refined using *SHELXL*-97. The *B*-factor estimates are stable for both cases up to 1.50 Å. This is certainly a key requirement for the bond lengths investigated in Fig. 7 (above) to be reliable.

1.20 Å Ni,Ca concanavalin A with restraints on (see Figs. 6 and 7 for Asp28 as an example). The stability of this computational procedure for proteins is a concern (Sheldrick, personal communication); nevertheless, standard uncertainties were calculable here and are shown as error bars in Figs. 5(*a*) and 5(*b*). Thus, for example, the Asp28 and Asp82 carboxyl bond distances (from CGLS restrained refinement) and their standard uncertainties (from full-matrix inversion at 0.94 Å resolution) are 1.215 (11), 1.278 (11) Å and 1.208 (16), 1.282 (16) Å, respectively. A carboxyl example from the Glu population of residues is Glu102, whose carboxyl-group bond distances (from CGLS refinement) and standard uncertainties (from full-matrix inversion) are 1.209 (10), 1.305 (9) Å.

2.3.5. Neutron structure refinement of native (Mn,Ca) concanavalin A. The refined X-ray structure at 0.94 Å (PDB code 1nls; Deacon *et al.*, 1997) was used as the starting model for the neutron refinement using *CNS* (Brünger *et al.*, 1998). Neutron scattering lengths (*International Tables for Crystallography*, 1995, Vol. *C*, pp. 384–391) were used and the topology and parameter dictionary files for *CNS* were optimized for neutron scattering. All water molecules were removed and the *B* factors were reset to 30 Å^2. The first step was a 20-cycle rigid-body refinement using all neutron data to 2.2 Å. At this stage, the *R* factor was 34.1% and $R_{\rm free}$ was 34.3%. Next, positional refinement of all the H-isotope atoms of the protein was performed. This improved the *R* factor to 33.0% and $R_{\rm free}$ to 34.1%. After minimization of the protein H-isotope coordinates, *B*-factor refinement for all atoms was performed. This further improved the *R* factor to 31.0% with an $R_{\rm free}$ of 32.7%. $F_{\rm o} - F_{\rm c}$ maps were then further analysed and the D atoms of amino-acid residues with exchangeable H-atom positions (His, Tyr, Thr, Ser, Asp, Glu, Asn and Gln) were included in the structure so that the protonation states for each of the amino-acid residues were determined. Large $F_{\rm o} - F_{\rm c}$ peaks ($\sim 5\sigma$) were observed for the side chains of both Asn104 and Ser215 and the orientations of these side chains were therefore adjusted. Water molecules were included very carefully only after the construction of the entire protein structure. $F_{\rm o} - F_{\rm c}$ maps were calculated and used to identify possible water peaks using the water-picking procedure in *CNS*. Water O atoms were assigned if they satisfied certain distance restraints and had positive $F_{\rm o} - F_{\rm c}$ map σ values of greater than 3.5. The water positions identified were then included in the model and positional and *B*-factor refinement of these was performed. $2F_{\rm o} - F_{\rm c}$ maps were checked in order to verify the fit to the nuclear density. 80 water molecules were finally assigned as full D_2O molecules, *i.e.* with strong positive nuclear density and the characteristic 'boomerang' shape nuclear density. A further four water molecules displayed only spherical density and thus were allocated as single O atoms. The final *R* factor for the model was 28.1% with an $R_{\rm free}$ of 31.2%.

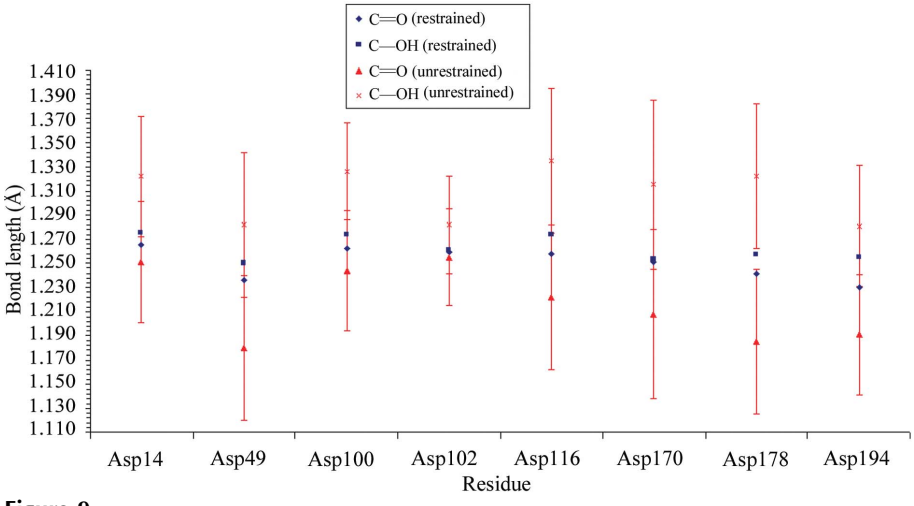

Figure 9
Carboxylate-group bond lengths in aspartates for the thrombin–hirugen binary complex at 1.26 Å resolution. The restrained refinement bond-length values all hold to the target 1.25 Å. Asp49, Asp116, Asp170, Asp178 and Asp194 show an indication of protonation after unrestrained refinement. Asp102 from the catalytic triad is unprotonated.

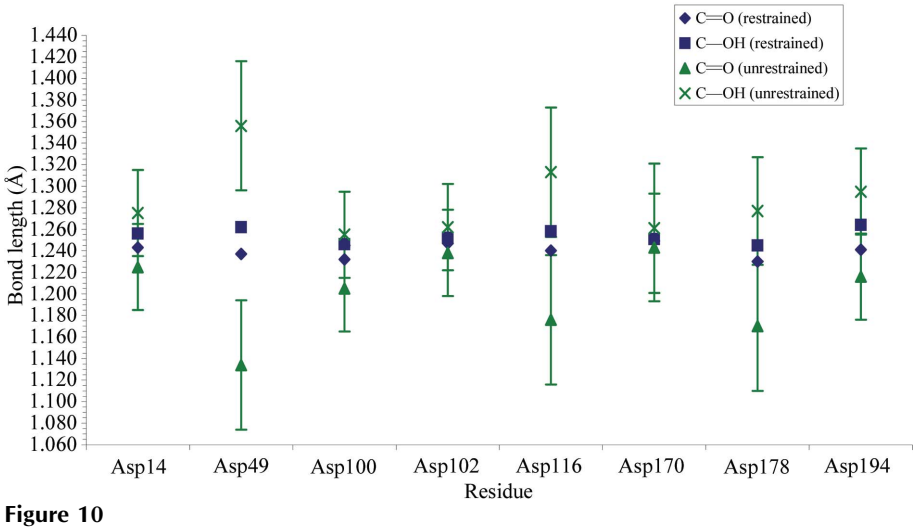

Figure 10
Carboxylate-group bond lengths in aspartates for the thrombin–hirugen–gw473178 ternary complex at 1.32 Å resolution. Asp116 and Asp178 show indication of protonation in agreement with the binary complex at 1.26 Å (Fig. 9). Asp102 from the catalytic triad is indicated to be unprotonated.

2.4. Focusing on Asp28 carboxylic C–O bond lengths in concanavalin A

Asp28 was chosen as a primary test model to observe how truncating the X-ray data resolution affects the

carboxylate bond lengths and to establish key indicators and their behaviour *versus* resolution. This residue was chosen because of its low *B* factors (less than 10 Å^2) and because it has been consistently observed to be protonated previously (Deacon *et al.*, 1997; Price, 1999; Habash *et al.*, 1997, 2000).

We obviously and immediately hypothesize that as the resolution is truncated the C—O bond-length estimates should deteriorate. This is because as we truncate the resolution the X-ray data-to-parameter ratio worsens and therefore the restraints alone will determine the positions of the atoms (Fig. 6). The bond-length restraint held on the carboxylate C—O bond ($O^{\delta1}$—C^{γ}—$O^{\delta2}$ in the COO^- group) is 1.249 Å in *SHELXL*-97 (Sheldrick, 1997) and now also in *REFMAC*5 (Collaborative Computational Project, Number 4, 1994). When the restraints are removed, we expect the bond lengths at higher resolutions to be more precisely determined and therefore exhibit the ideal bond lengths of a single bond (C—O), a double bond (C═O) or an ionized bond (C···O). As the resolution is truncated without the bond-length restraints, the bond-length behaviour needs to be examined carefully to establish if this may be a viable key indicator at even 'moderate' resolutions such as 1.25 Å. When the bond-length restraints are removed, the stability of this as the protonation indicator is good to 1.30 Å resolution in both cases (Fig. 7). Between 1.30 and 1.50 Å resolution the two cases behave differently. The *B*-factor estimates (Fig. 8) are stable for both cases up to 1.50 Å resolution. This is certainly a key requirement for the bond lengths investigated in Fig. 7 to be reliable.

2.5. Extension to all aspartic and glutamic acids in concanavalin A

Based on the assessment of the bond-length behaviour in the carboxylic group of Asp28 under the diffraction data truncation tests described in §2.4, the stability of the *B* factors and the role of restraints at a variety of resolutions now allow us to apply the bond length as a key indicator to other aspartic and glutamic acids, firstly in concanavalin A and then in thrombin. In concanavalin A there are 18 aspartic acid residues and seven glutamic acids. From previous structural studies at ultrahigh X-ray resolution, significant bond-length differences were observed in the side-chain groups of Asp28 and Asp82 as well as Glu102, suggesting that their side chains were protonated and therefore a carboxylic group (Deacon, 1996; Price, 1999). Supplementary material *B* Figs. 1–18 show the bond lengths of the carboxylate group and carboxyl group for all the aspartic acids in native (Mn,Ca) concanavalin A studied at 0.92 Å resolution (Deacon *et al.*, 1997; PDB code 1nls), Ni,Ca concanavalin A at 0.94 Å, (Mn,Ca) concanavalin A at 1.20 Å (Parkin *et al.*, 1996) and another (Mn,Ca) concanavalin A study at 0.94 Å (Price, 1999). The bond lengths and/or aspartic acid residues that have been omitted from the graphs are those from residues having a double conformation. Supplementary material *B* Figs. 1–3 and 9–11 show the results when restraints are in place; these largely follow the expectation seen from the interpretations in §2.4. When restraints

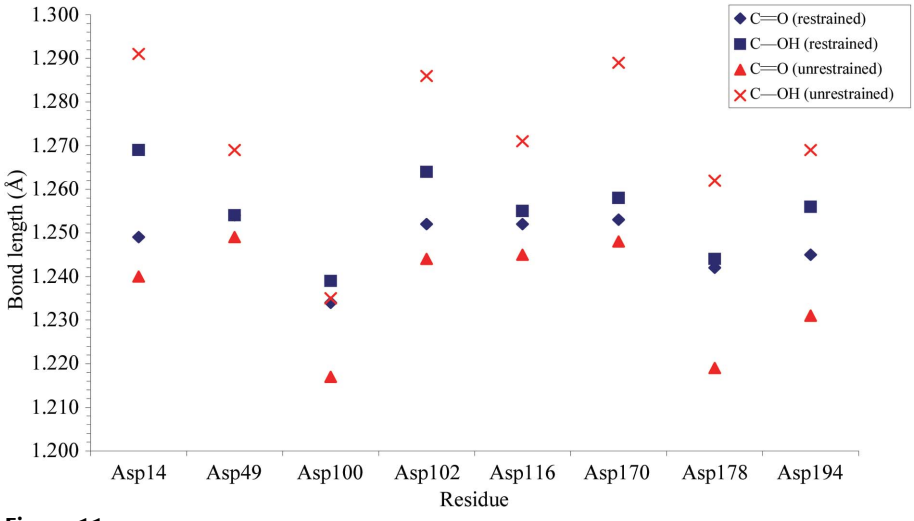

Figure 11
Carboxylate-group bond lengths in aspartates for the thrombin–hirugen–gw420128 complex at 1.39 Å resolution. No indications of protonation are observed, unlike in Figs. 9 and 10 for the 1.26 and 1.32 Å studies.

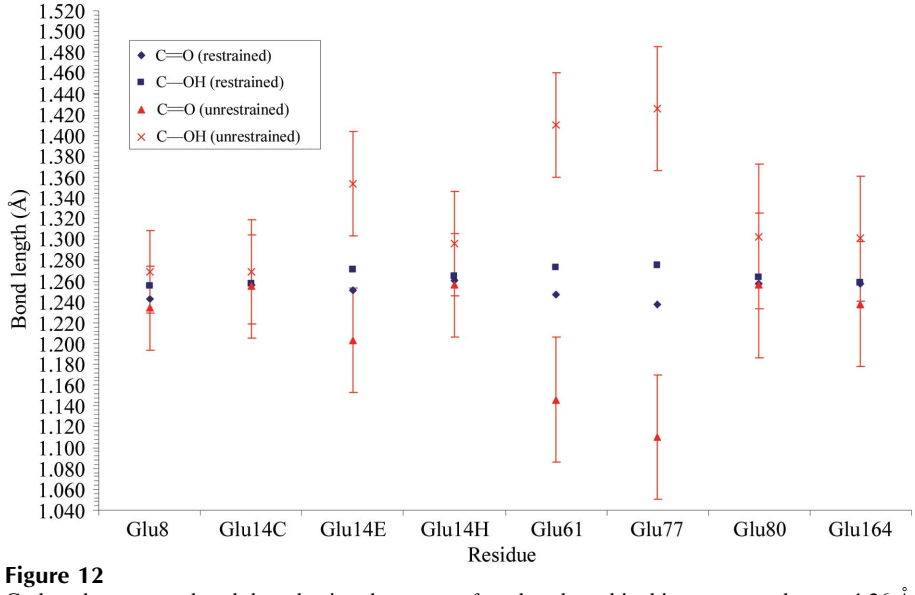

Figure 12
Carboxylate-group bond lengths in glutamates for the thrombin–hirugen complex at 1.26 Å resolution. Glu14E shows indication of protonation, with Glu61 and Glu77 showing possible evidence, but the distances are far from ideal values.

are removed (supplementary material *B* Figs. 4–6 and 12–14), the indication of protonation is clear even at 1.20 Å resolution for Asp28; a key feature of this indication is that the determined values hold close to the Cambridge Structural Database values and that the standard uncertainties are approximately 0.014 Å. The bond distances hold even at 1.50 Å resolution, but there is little to no confidence because the other aspartic acids do not encompass these (and other) values, *i.e.* when we know them to be protonated. Glu8, which makes a hydrogen-bonding interaction with Asp28 and is unprotonated, shows good indication of this up to 1.20 Å resolution, but the other glutamic acids show variable behaviour. Thus, the key structural aspects of Glu8 are whether its carboxyl distances are 1.25 Å and its hydrogen bond with Asp28. In the graphs, the carboxylate groups have been labelled as carboxylic, so that it is easier to differentiate the shorter bond from the longer bond.

2.6. Aspartic and glutamic acids in thrombin

The results of the concanavalin A study suggest key indicators of protonation state. A first key indicator is that the bond distances should match the short and long bond values simultaneously. Furthermore, it might be expected that the other aspartic and glutamic acid residues are most likely to be unprotonated and should show values that are equal and of 1.25 Å. These criteria can now be applied to the three thrombin structures as plotted in Figs. 9–14. The full matrix inversion (to obtain s.u.s) did not work for the 1.39 Å thrombin structure; as Professor George Sheldrick pointed out, such calculation instability arising from a lack of diffraction data can occur with *SHELXL*-97 (personal communication). Nonetheless, the Cruickshank DPI values (Cruickshank, 1999) output from *REFMAC*5 for each of the thrombin structures indicate small coordinate errors (see Table 1). It should be noted that *REFMAC*5 outputs the DPI 'coordinate' error, *i.e.* $\sigma(x)$; this is a crucial point and we double-checked this for 1nls native concanavalin A (Deacon *et al.*, 1997), which was used by both Cruickshank (1999) and Blow (2002) as a case study, *i.e.* $\sigma(x)$ for 1nls is 0.02 Å and as output by *REFMAC*5. The DPI values in Table 1 are derived from the *SHELXL* final refinement, *i.e.* including multiple occupancies. In the current context of relating the DPI $\sigma(x)$ to average errors on unrestrained bond lengths, σ(length) is $2^{1/2}\sigma(x)$, *i.e.* $1.414\sigma(x)$ (Cruickshank, 1999), with the assumptions that this is for diffraction data errors alone and that the two atoms concerned show approximately isotropic vibration within their anisotropically treated refinement. For Ni,Ca concanavalin A presented here $\sigma(x)$ = 0.017 Å, making σ(length) = 0.024 Å. To relate this to specific rather than average bond lengths requires adjustment according to the *B* factors of the relevant pair of atoms; low average *B* factors will lead to a more precise value than the average σ(length). Note that the functional form is quadratic with *B* factor (with a nonzero value if extrapolated to *B* = 0) and related to this the

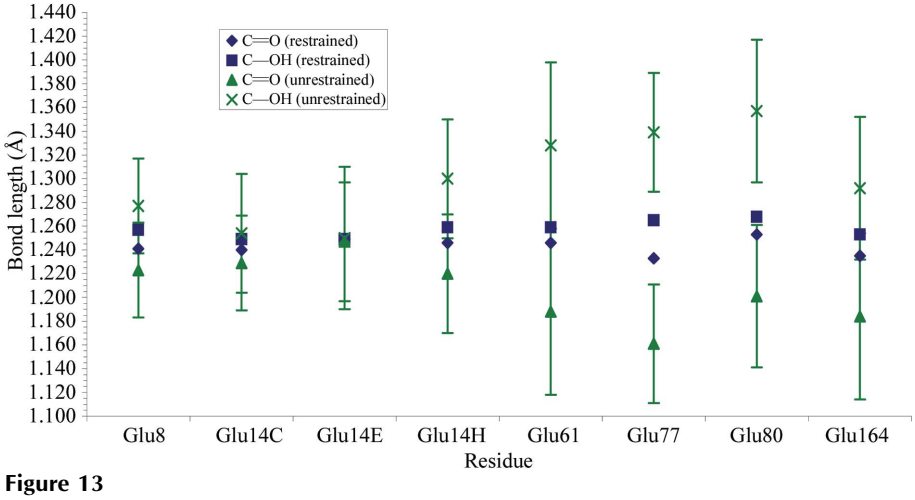

Figure 13
Carboxylate-group bond lengths in glutamates for the thrombin–hirugen–gw473178 ternary complex at 1.32 Å resolution. Glu8 shows evidence of protonation, in agreement with the ternary complex at 1.39 Å resolution (Fig. 14). The other residues that show evidence of protonation are Glu14H, Glu61 and Glu164. Glu77 and Glu80 also show possible evidence of protonation, but the bond lengths are beyond the ideal bond lengths. Glu14E differs from that in the 1.26 Å study (a different crystal effect?).

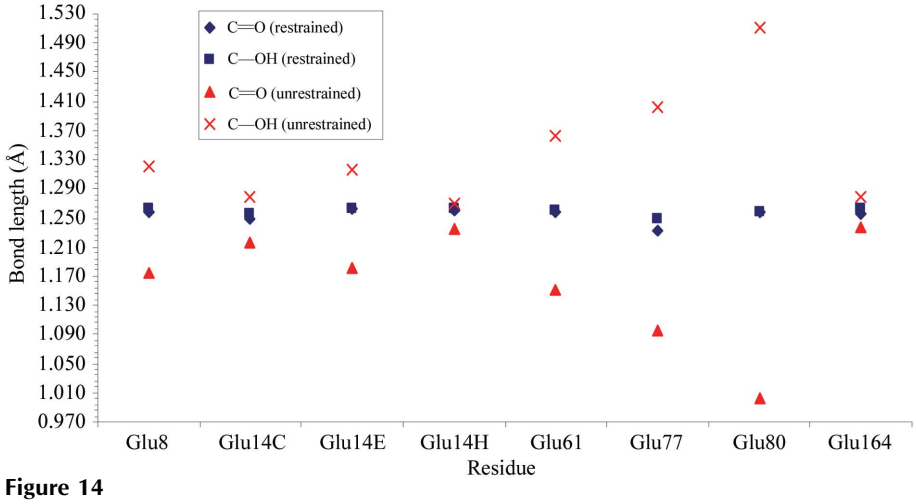

Figure 14
Carboxylate-group bond lengths in glutamates for the thrombin–hirugen–gw420128 complex at 1.39 Å resolution. Glu8 shows evidence of protonation, in agreement with the ternary complex (with inhibitor gw473178) in Fig. 13. Glu14E shows indication of protonation and agrees with the thrombin–hirugen binary complex in Fig. 11. Glu61 and Glu77 again show possible evidence of protonation, but the values for Glu80 are very anomalous. Note the lack of error bars owing to failure of the full matrix inversion, which is awkward (see §3 regarding how to extrapolate from the average Cruickshank DPI).

O atoms are more precisely determined in position than N atoms than C atoms (see Cruickshank, 1999). There remains the question of which $B_{\rm average}$ value to use; for example, excluding the bound water B values in this $B_{\rm average}$ seems reasonable.

The acidic residues that were selected for computational evaluation had the lowest B factors (<35 Å^2). Thus, Asp49, Asp116, Asp170, Asp178 and Asp194 show indications of protonation in the 1.26 Å study. Most interestingly, the 1.26 and 1.32 Å studies both indicate that Asp102 from the His-Ser-Asp catalytic triad is not protonated in thrombin (Figs. 9, 10 and 11). This is consistent with the neutron protein crystallography work performed by Kossiakoff & Spencer (1980) on trypsin, who identified that the His57 in trypsin was protonated (as deuterium) rather than Asp102.

For the glutamic acids (Figs. 12, 13 and 14), only Glu14E shows protonation at 1.26 Å, which is corroborated at 1.39 Å but surprisingly not at 1.32 Å. (Note the conventional numbering scheme for thrombin is to preserve the relationship with respect to chymotrypsinogen amino-acid numbering but which requires addition of a suffix; Stubbs & Bode, 1993). There are possible grounds for protonation for Glu61 and Glu77, which are also corroborated by the 1.32 and 1.39 Å studies.

2.7. Determination of the protonation states using neutron crystallography

The acidic residues (Asp and Glu) and the histidines in (Mn,Ca) concanavalin A have been thoroughly analyzed at various resolutions to see at what resolution level the peaks for the D atoms remain visible. Figs. 15–18 show the neutron density maps at various resolutions.

2.7.1. Aspartic acid protonation-state determination. For the aspartic acid and glutamic acid residues, only Asp82 was clearly seen to be protonated. However, Asp28 could possibly also be protonated: for this residue the $F_{\rm o} - F_{\rm c}$ peaks are weaker (<1.5σ) and not completely in the correct geometry. As can be seen in Fig. 15, the D atom of Asp82 remains visible up to 2.75 Å resolution.

2.7.2. Histidine protonation-state determination. Strong peaks (≥2.5σ) were seen for the D atoms of His24, His51 and

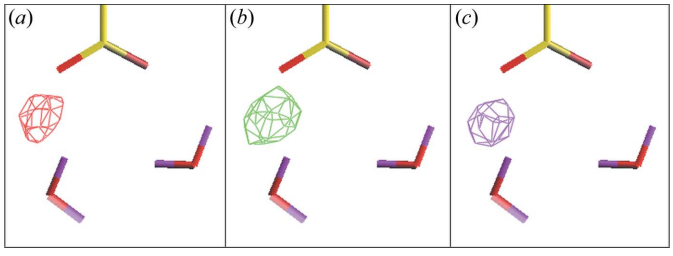

Figure 15
Positive neutron $F_{\rm o} - F_{\rm c}$ maps for Asp82 (all at 2.5σ) at various resolutions (calculated from the model with no D atoms included). (*a*) 2.2 Å resolution; (*b*) 2.5 Å resolution; (*c*) 2.75 Å resolution.

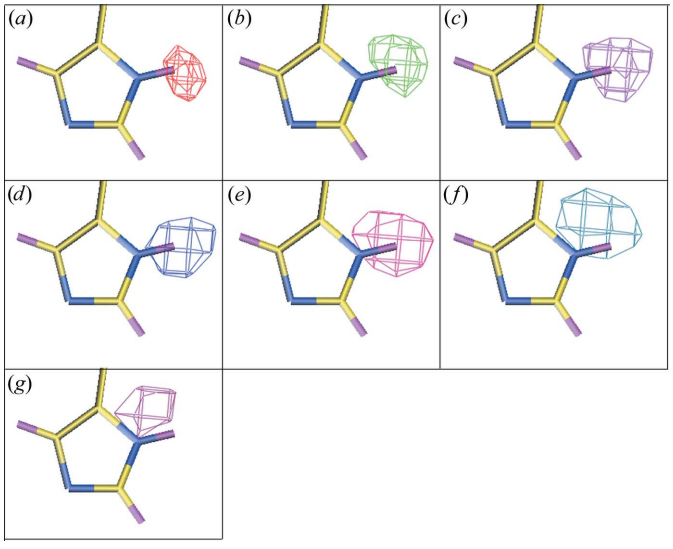

Figure 16
Positive neutron $F_{\rm o} - F_{\rm c}$ maps for His24 (all at 2.5σ) at various resolutions (calculated from the model with no D atoms included). (*a*) 2.2 Å resolution; (*b*) 2.5 Å resolution; (*c*) 2.75 Å resolution; (*d*) 3.0 Å resolution; (*e*) 3.25 Å resolution; (*f*) 3.5 Å resolution; (*g*) 3.75 Å resolution.

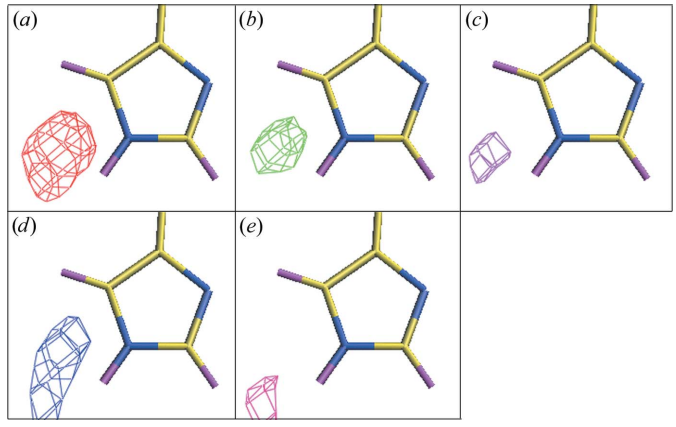

Figure 17
Positive neutron $F_{\rm o} - F_{\rm c}$ maps for His51 (all at 2.5σ) at various resolutions (calculated from the model with no D atoms included). (*a*) 2.2 Å resolution; (*b*) 2.5 Å resolution; (*c*) 2.75 Å resolution; (*d*) 3.0 Å resolution; (*e*) 3.25 Å resolution.

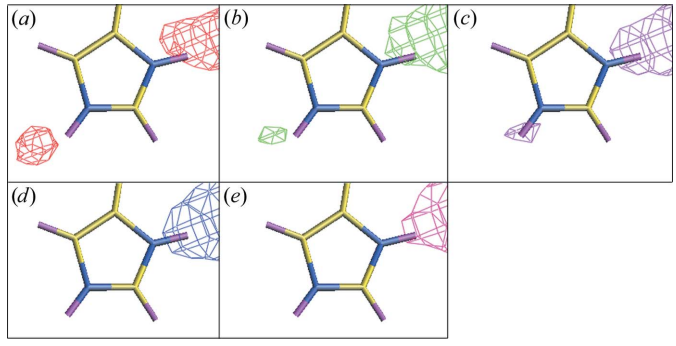

Figure 18
Positive neutron $F_{\rm o} - F_{\rm c}$ maps for His180 (all at 2.5σ) at various resolutions (calculated from the model with no D atoms included). (*a*) 2.2 Å resolution; (*b*) 2.5 Å resolution; (*c*) 2.75 Å resolution; (*d*) 3.0 Å resolution; (*e*) 3.25 Å resolution.

His180, while for the other histidine residues (His121, His127 and His205) the peaks were not visible at reasonable σ levels ($<1.5\sigma$). His24 (Fig. 16) was seen to be singly protonated at the $N^{\delta1}$ position, while His51 was seen to be singly protonated at the $N^{\varepsilon2}$ position. His180 was seen to be doubly protonated, *i.e.* at both the $N^{\delta1}$ and $N^{\varepsilon2}$ positions. As can be seen, the peak for the D atom (DD1) of His24 is still visible (at 2.5σ) up to 3.75 Å resolution. For His51 (Fig. 17) the peak for the D atom (DE2) is still visible (at 2.5σ) up to 3.25 Å resolution. For His180 (Fig. 18) the peaks for the D atoms (DD1 and DE2) are still visible (at 2.5σ) up to 2.75 Å resolution.

3. Discussion

3.1. Remarks on the confirmation of single bonds (C—OH) and double bonds (C═O) of COOH groups in concanavalin A

The significant bond-length differences seen in the various data-truncation and refinement studies confirm that some C—O bonds did deviate from an ionized bond-length distance of 1.249 Å, from which it can be inferred that the carboxylate group is protonated. Our studies at truncated resolutions with the restraints left on also showed the expected behaviour that the bond lengths (C—O in COO^-) were forced to converge towards the ionized bond dictionary length of 1.249 Å. However, as seen with Asp28 (Fig. 6) and Glu102 (supplementary material *B* Fig. 15), it is still possible to distinguish a single bond from a double bond up to 1.25 Å resolution. In Ni,Ca concanavalin A the standard uncertainty of the observed difference between the bond lengths C═O and C—OH for Asp28 and Asp82 at 0.94 Å resolution are 3.9σ and 3.2σ, respectively. For Glu102 the standard uncertainty of the observed difference between the bond lengths is 7σ. These values confirm that the bond-length differences are significant and therefore protonated.

When the restraints are removed at ultrahigh resolution with a high X-ray data-to-parameter ratio (6:1), the bond lengths are naturally slightly closer to the ideal values than with the restraints on. The bond lengths that do deviate outside the ideal bond length usually have high atomic *B* factors and are therefore more disordered. Daopin *et al.* (1994) demonstrated that atomic precision in proteins strongly depends on the *B* factor. In our study, the carboxylic groups in Asp28 and Glu102 have *B* factors less than 10 $Å^2$ and this is why the bond lengths have acceptable values (without restraints), even after resolution truncation to 1.30 Å (Figs. 19, 20, 21 and 22). What of the $2F_o - F_c$ electron-density map evidence? Figs. 19–22 show the $2F_o - F_c$ electron-density maps of Asp28 and Glu102. The $O^{\delta1}$ and $O^{\delta2}$ of Asp28 (Fig. 19) and $O^{\varepsilon1}$ and $O^{\varepsilon2}$ of Glu102 (Fig. 20) can be differentiated by their densities and different bond lengths, which are consistent with one atom being predominantly double-bonded (continuous electron density) and the other being predominantly single-bonded (broken electron density). This latter indicator was observed by Minasov *et al.* (2002) and we have also adopted it here as shown in Figs. 19–22 (with unrestrained bond lengths from CGLS refinement and standard uncertainties from full-matrix inversion).

3.2. The protonation state of acidic residues in binary and ternary thrombin–hirugen complexes

In Fig. 9, the carboxylate-group bond lengths in some aspartates are shown for the thrombin–hirugen binary complex at 1.26 Å resolution. The restrained refinement bond-length values all restrained to the target of 1.25 Å. After unrestrained refinement, Asp49, Asp116, Asp170, Asp178, Asp194 show evidence of protonation. The 1.32 Å ternary complex (with inhibitor gw473178) shown in Fig. 10 illustrates that Asp116 and Asp178 show indication of protonation, in

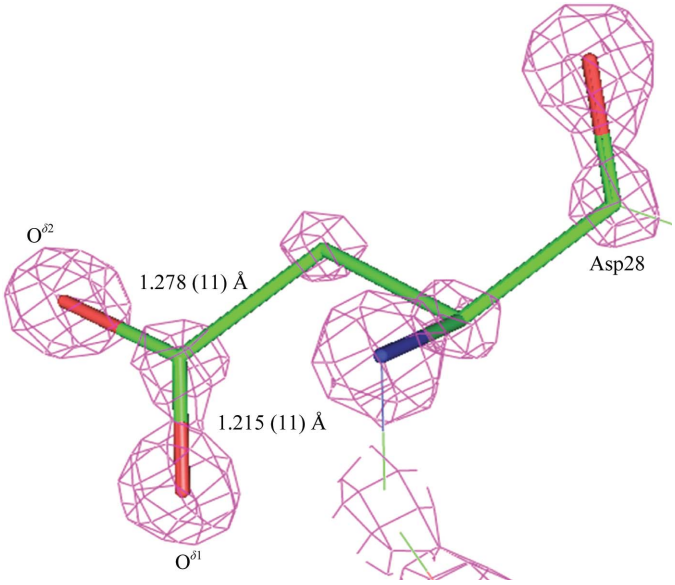

Figure 19
Asp28 at 0.94 Å resolution; $2F_o - F_c$ electron-density map contoured at 5.0 r.m.s.

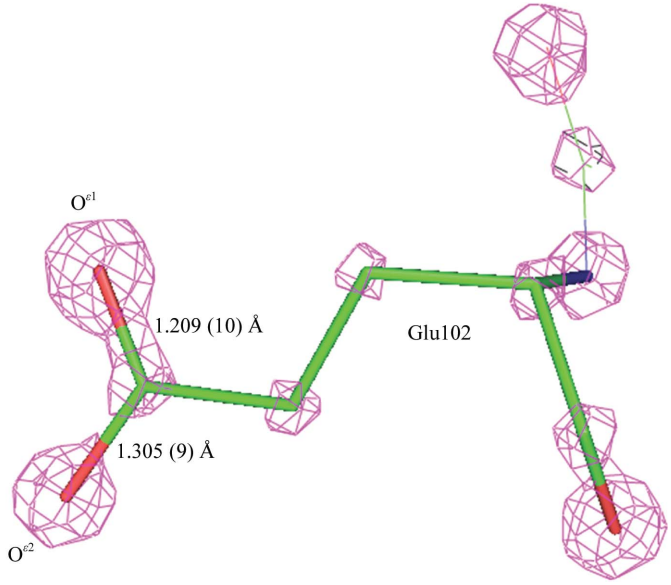

Figure 20
Glu102 at 0.94 Å resolution; $2F_o - F_c$ electron-density map contoured at 5.0 r.m.s.

agreement with the binary complex at 1.26 Å (Fig. 9). For the ternary complex with inhibitor gw420128 at 1.39 Å (Fig. 11), there is no evidence of protonation on any residue.

Most interestingly, the 1.26 and 1.32 Å studies both indicate that Asp102 from the His-Ser-Asp catalytic triad is not protonated in thrombin (Figs. 9 and 10). This is consistent with the neutron protein crystallography work peformed by Kossiakoff & Spencer (1980) on trypsin, who identified that the His57 in trypsin was protonated (as deuterium).

Fig. 12 shows the carboxylate-group bond lengths in glutamates for the thrombin–hirugen binary complex at 1.26 Å resolution. Only Glu14E shows indication of protonation after unrestrained refinement; Glu61 and Glu77 show impossibly short and long bond lengths. For the thrombin–hirugen–gw473178 ternary complex at 1.32 Å resolution (Fig. 13) Glu8 shows evidence of protonation, in agreement with the ternary complex at 1.39 Å resolution (Fig. 14). The other residues that show evidence of protonation are Glu14H, Glu61 and Glu164. Glu77 and Glu80 show possible evidence of protonation, but the bond lengths are beyond the ideal bond lengths. In the thrombin–hirugen–gw420128 ternary complex at 1.39 Å resolution (Fig. 14) Glu14E shows indication of protonation after unrestrained refinement; this agrees with the binary complex.

The standard uncertainties in Figs. 9–13 were ∼0.04–0.05 Å for thrombin compared with 0.014 Å for the 1.2 Å truncated concanavalin A. Thus, even by comparison with dictionary values one is assessing the observed bond distance at a '2σ' level of treatment of agreement, whereas the truncated concanavalin A is a '5σ' treatment. In any given study where the standard uncertainties are not at a '3σ' level (0.025 Å) then a repeat experiment under more favourable experimental conditions or with an improved order crystal should be sought.

3.3. The protonation state of acidic residues in native (Mn,Ca) concanavalin A using neutron crystallography

The neutron structure reported here showed that Asp82 was protonated, with possible protonation of Asp28. As the resolution is truncated, the $F_o - F_c$ peak does not diminish until 2.75 Å resolution. Asp82 $O^{\delta2}$ was protonated in the neutron structure, whereas in the X-ray structure $O^{\delta1}$ was protonated. This could be a consequence of using two different crystals, in one of which one residue has undergone a rotation around the C^{γ} atom. The ultrahigh-resolution X-ray structure of Ni concanavalin A revealed the protonation states of Asp28, Asp82 and Glu102 from bond-length differences. Clearly, the strength of neutron and X-ray evidence is that Asp82 is protonated, with the possibility of protonation in Asp28. The other residues could show differences arising from the different crystals used.

3.4. The protonation state of histidines in native (Mn,Ca) concanavalin A using neutron crystallography

The crystal used for neutron data collection had been soaked in D_2O for several months prior to data collection. Therefore, those H atoms attached to O or N will generally have exchanged for deuterium. This was verified by the neutron maps, which clearly show peaks for the exchangeable H atoms that are visible in the positive $F_o - F_c$ maps, indicating that they are indeed D atoms and not H atoms (which would be seen in the negative $F_o - F_c$ maps). In the 2.20 Å resolution neutron structure of native (Mn,Ca) concanavalin A reported here, the H atoms on the N atoms of the imidazole rings were replaced by deuterium and this was visible in the positive $F_o - F_c$ neutron maps. As the resolution was truncated from 2.20 to 3.00 Å, the neutron density maps did not diminish and this is obviously a consequence of the deuterium scattering as well as that of nitrogen and carbon. In the 0.94 Å resolution X-ray structure of Ni concanavalin A, it was not

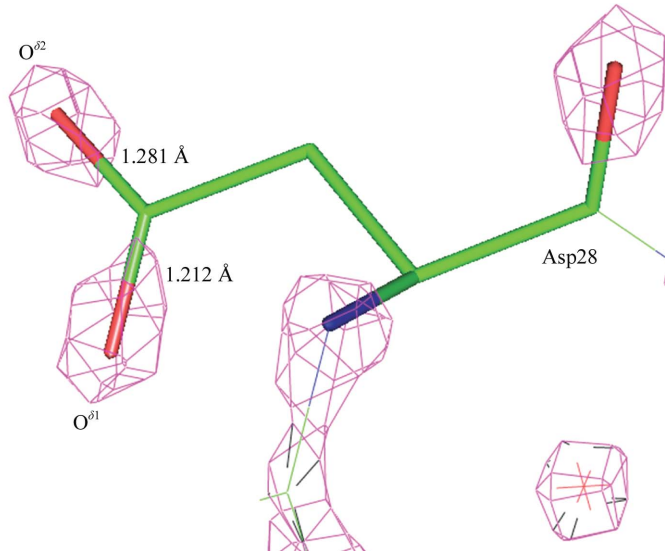

Figure 21
Asp28 at 1.30 Å resolution; $2F_o - F_c$ electron-density map contoured at 5.0 r.m.s.

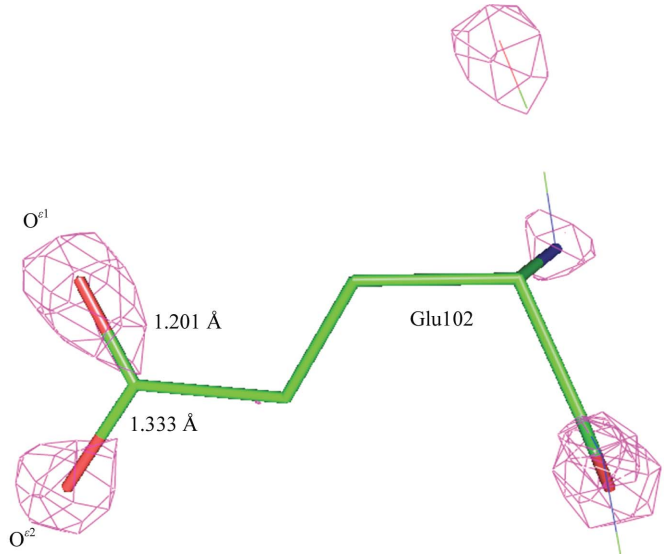

Figure 22
Glu102 at 1.30 Å resolution; $2F_o - F_c$ electron-density map contoured at 5.0 r.m.s.

possible to distinguish a double bond from a single bond in the histidine imidazole rings owing to the small differences in their expected bond lengths; therefore difference electron density alone was used. Thus, we use neutron crystallography to finally confirm the protonation states of such N atoms beyond doubt.

4. Conclusions regarding the determination of protonation states

To determine the protonation states, the following factors are listed in the order of their importance.

(i) A residue needs to be single occupancy. This is a key requirement and is illustrated by the case of Asp82 in concanavalin A (see supplementary material *B*, p. 6).

(ii) Anisotropic refinement at a minimum resolution of 1.20 Å (or an X-ray data-to-parameter ratio of ∼3.5:1). The resolution truncation to 1.20 Å of the X-ray structures of Ni concanavalin A and Mn concanavalin A revealed the protonation states for Asp28 and Glu102 (with and without restraints). At 1.20 Å, the observed difference between the bond lengths (C═O *versus* C—OH) in Asp28 in native (Mn,Ca) concanavalin A with restraints is 2.8σ and for Ni concanavalin A without restraints the observed difference between the bond lengths is 3.1σ (Fig. 7). Obviously, truncation of a 0.94 Å resolution data set to 1.2 Å is a favourable situation [for example, $\langle I/\sigma(I)\rangle \geq 9.2$ in this resolution shell, still yielding an X-ray data-to-parameter ratio of 3.57]. At resolutions poorer than 1.20 Å it was difficult to discriminate the bond lengths of C═O and C—OH. Fig. 7 also shows that even at 1.30 Å resolution bond-length discrimination is possible in favourable cases.

For thrombin, the full-matrix inversion worked when adding in $R_{\rm free}$ reflections. The 'statistically proper' discrimination of long, short and mixed (*i.e.* equal length) bond distances for a carboxyl/carboxylate side chain with errors in bond distances of ∼0.04 Å is not certain and this leads to the next requirement, namely agreement with dictionary values. In any case, an improved experiment or improved order crystal should be sought. In this context, this means trying to improve on X-ray data-to-parameter ratios of 3.4, 2.7 and 2.5 for the thrombin–hirugen, thrombin–hirugen–gw473178 and thrombin–hirugen–gw420128 cases at 1.26, 1.32 and 1.39 Å, respectively.

(iii) Agreement with expectation values from the dictionary values. It is very important that both the bond lengths of the carboxylic group remain within the ideal definitive experimental bond distances for a double bond (C═O), a single bond (C—OH) or a delocalized bond (C···O). The chemical microenvironment effects discussed in §1.2 illustrate the need for a spread of target values for C—OH and C═O in the dictionary. Also, with the increase of fully deuterated proteins in neutron protein crystallography work, X-ray characterization will be performed on this protein form. There is the possibility that 'dictionary distances' and their microenvironment spread of values will need some further amendment, *i.e.* variability to be allowed for deuteration.

(iv) $F_{\rm o} - F_{\rm c}$ map $\geq 2.5\sigma$ and in the correct geometry and the importance of $2F_{\rm o} - F_{\rm c}$ maps. The role of the $F_{\rm o} - F_{\rm c}$ electron-density maps has already been discussed in §1.1. We have illustrated here how the $F_{\rm o} - F_{\rm c}$ map was used to determine the protonation state of His24 in Ni concanavalin A (Fig. 1), which agreed with the neutron (Mn,Ca) concanavalin A structure (see §2.7). However, when the X-ray resolution is in the 1.20 Å resolution range, utility of the $F_{\rm o} - F_{\rm c}$ map to determine the protonation state is suspect. However, the $2F_{\rm o} - F_{\rm c}$ electron-density map we show can be used to compare the electron-density coverage of C═O *versus* C—OH bonds. This analysis provided further evidence in determining the protonation state. *i.e.* a C═O shows continuous $2F_{\rm o} - F_{\rm c}$ electron density, whereas a C—OH bond shows discontinuous density (see §3.1). However, the drawback is that such a $2F_{\rm o} - F_{\rm c}$ map features only appear to be distinguishable when the atomic *B* factors are less than 20 Å^2.

(v) Multiple experiments, *e.g.* multiple X-ray and neutron crystallography experiments. We have refined the crystal structures of Ni and Mn concanavalin A using X-ray data and of (Mn,Ca) concanavalin A using neutron data as well. All results showed clear agreement of the protonation state of Asp82 (protonated), Glu8 (unprotonated) and His24 (singly protonated). The case of Asp28 in this neutron study showed evidence, albeit weaker, for protonation. The neutron crystal structure of Mn concanavalin A has also allowed us to correct the orientation of Asn104 in the X-ray structure of Ni concanavalin A. This was evident after refinement, when the bond-length differences of C═O and C—NH_2 were clearly distinguishable; the $2F_{\rm o} - F_{\rm c}$ map showed continuous electron density for C═O and discontinuous electron density for C—NH_2, with the atomic *B* factors now lower. Thus, by combining more than one technique we gain much greater confidence in determining the protonation states and the remodelling of incorrect conformations, both of which can be crucial in the functional study of proteins.

In the case of thrombin without an inhibitor we have achieved an X-ray diffraction resolution of 1.26 Å, the best thus far to our knowledge, although there is a 1.25 Å resolution thrombin ternary complex (Neumann *et al.*, 2006). The related protein trypsin has been studied with neutron protein crystallography by Kossiakoff & Spencer (1980), who identified that His57 in trypsin was protonated (as deuterium) rather than Asp102. The functionally equivalent (*i.e.* from the His-Ser-Asp catalytic triad) Asp102 in our thrombin X-ray studies is indicated to be unprotonated in both the 1.26 and 1.32 Å studies. However, Kuhn *et al.* (1998) found in a 0.78 Å X-ray structure of a serine protease (*Bacillus lentus* subtilisin; see, for example their Fig. 3*c*) that the critical hydrogen was positioned between His64 and Asp32, in contrast to the findings of Kossiakoff & Spencer (1980) and ours reported here. Kuhn *et al.* (1998) suggested that the different results on the protonation state of each residue was a consequence of deuterium, which was used in the neutron experiment, having a different zero-point energy and thus being confined to one position; a H atom in contrast, they argued, could be shared. Most recently, we note the results of Fuhrmann *et al.* (2006) in their 0.8 Å study of α-lytic protease, who definitively show using X-rays that the H atom is on His57.

Another factor that may become important is that if excessive X-ray irradiation is used in the pursuit of the highest possible diffraction resolution, then radiation damage to the protein crystal could occur even if held at cryotemperature in a glassy state and where even decarboxylation of a side chain may happen; this was not seen in our studies, but for a discussion in this context see Erskine *et al.* (2003). Such cases add further to the interest in neutron protein crystallography studies since neutrons do not cause radiation damage.

In the neutron crystal structure of native (Mn,Ca) concanavalin A, we have shown that the protonation state of histidines can be determined to a modest resolution of 2.75 Å or possibly 3.0 Å. The correct orientations of residues such as glutamine and asparagine have also been determined, therefore improving our structural knowledge of concanavalin A. The general implication of this result is important in that it suggests that higher molecular-weight proteins can be studied but with a lower expectation of the diffraction resolution required; such issues are important when judging whether a project is technically viable either by a researcher or in a neutron beamtime approval committee situation.

In our thrombin crystal structures the protonation states of aspartic acids 49, 116, 170, 178 and 194 and Glu14E showed evidence of protonation when the carboxylate-group bond-length restraints were not applied, but once they were applied (§2.7, Fig. 9) the indication was lost. The functionally important Asp102 was not protonated. The diffraction resolution of 1.26 Å (thrombin–hirugen binary complex) with an X-ray data-to-parameter ratio of 3.5:1 proved sufficient to allow the determination of protonation states at the '2σ' level, but an improved X-ray analysis through the use of a more intense source was not obtained. The Blow (2002) formula includes a $(\text{resolution})^{5/2}$ factor. If we compare the Ni,Ca concanavalin A DPI $\sigma(x)$ at the two resolutions 0.94 and 1.2 Å, which are 0.017 and 0.027 Å, respectively, then we provide a practical comparison as a cross-check of the Blow formula, *i.e.* $0.017 \times (1.2/0.94)^{5/2} = 0.031$ Å, which is closely similar to 0.027 Å. The Blow reformulation (Blow, 2002) of the Cruickshank DPI formula (Cruickshank, 1999) then provides a very useful insight into the impact of diffraction resolution. Thus, if we can just discriminate something at 1.2 Å with a sigma of 1.0 units, then at 1.0 Å we would expect to achieve a sigma of $(1.0/1.2)^{5/2} = 0.63$ units, whereas at 1.4 Å one would have a sigma of $(1.4/1.2)^{5/2} = 1.47$ units. Clearly resolution is worth striving after, as is well known of course, but in this way is neatly quantified. The Blow (2002) and Cruickshank (1999) treatments also clarify explicitly that improved precision comes (linearly) from improved data completeness and lower R factors. If these improvements cannot be accrued using X-rays, as we found with thrombin, then additional collection of neutron data on either the binary or the ternary complexes would still be beneficial.

JRH thanks the BBSRC and GlaxoSmithKline (GSK) for the PhD CASE studentship grant to HUA. JRH would like to thank SRS Daresbury laboratory, Dr. A. Joachimiak and colleagues at APS Chicago and the Institut Laue–Langevin for synchrotron X-ray and neutron beamtime awards and facilities, respectively. He is also grateful to Professor George Sheldrick for his helpful and prompt comments to enquiries made regarding the use of *SHELXL*-97, and Professor J. Sutherland at The University of Manchester for discussions on the nature of the carboxyl group. Drs J. Raftery and P. J. Rizkallah are also thanked for discussions and two anonymous referees are thanked for their helpful comments and criticisms. Dr M. Spano of the EMBL Outstation Grenoble is thanked for collaborative work aiming towards larger thrombin crystals for future neutron work. We thank Dr Lisa Keefe and colleagues, for beamtime support at IMCA-CAT, APS.

References

Ahmed, F. R. & Cruickshank, D. W. J. (1953). *Acta Cryst.* **6**, 385–392.
Arzt, S., Campbell, J. W., Harding, M. M., Hao, Q. & Helliwell, J. R. (1999). *J. Appl. Cryst.* **32**, 554–562.
Blakeley, M., Kalb (Gilboa), A. J., Helliwell, J. R. & Myles, D. A. A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 16405–16410.
Blow, D. M. (2002). *Acta Cryst.* D**58**, 792–797.
Borthwick, P. W. (1980). *Acta Cryst.* B**36**, 628–632.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Campbell, J. W. (1995). *J. Appl. Cryst.* **28**, 228–236.
Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. (1998). *J. Appl. Cryst.* **31**, 496–502.
Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules*, p. 49. New York: W. H. Freeman & Co.
Cianci, M. *et al.* (2005). *J. Synchrotron Rad.* **12**, 455–466.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Daopin, S., Davies, D. R., Schlunegger, M. P. & Grüter, M. G. (1994). *Acta Cryst.* D**50**, 85–92.
Dauter, Z. (2003). *Methods Enzymol.* **368**, 288–337.
Deacon, A. M. (1996). PhD thesis. The University of Manchester, UK.
Deacon, A. M., Gleichmann, T., Kalb (Gilboa), A. J., Price, H. J., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4305–4312.
Donohue, J. & Marsh, R. E. (1962). *Acta Cryst.* **15**, 941–945.
Edington, P. & Harding, M. M. (1974). *Acta Cryst.* B**30**, 204–206.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2003). *Protein Sci.* **12**, 1741–1749.
Evans, P. R. (1997). *Jnt CCP4/ESRF–EACBM Newsl.* **33**, 22–24.
Fuhrmann, C. N., Daugherty, M. D. & Agard, D. A. (2006). *J. Am. Chem. Soc.* **128**, 9086–9102.
Greer, J., Kaufmann, H. W. & Kalb, A. J. (1970). *J. Mol. Biol.* **48**, 365–366.
Habash, J., Raftery, J., Nuttal, R., Price, H. J., Wilkinson, C., Kalb (Gilboa), A. J. & Helliwell, J. R. (2000). *Acta Cryst.* D**56**, 541–550.
Habash, J., Raftery, J., Weisgerber, S., Cassetta, A., Lehman, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. W. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4313–4317.
Harris, T. K. & Turner, G. J. (2002). *IUBMB Life*, **53**, 85–98.
Helliwell, J. R., Habash, J., Cruickshank, D. W. J., Harding, M. M., Greenhough, T. J., Campbell, J. W., Clifton, I. J., Elder, M., Machin,

P. A., Papiz, M. Z. & Zurek, S. (1989). *J. Appl. Cryst.* **22**, 483–497.
Howard, E. I., Sanishvili, R., Cachau, E. R., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Zandt, V. M., Sibley, E., Bon, C., Moras, D., Schneider, R. T., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.
Jones, A. T., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kalb (Gilboa), A. J. & Helliwell, J. R. (2001). *Handbook of Metalloproteins*, edited by A. Messerschmidt, R. Huber, T. Poulos & K. Wieghardt, pp. 963–972. New York: John Wiley.
Kalb (Gilboa), A. J., Yariv, J., Helliwell, J. R. & Papiz, M. Z. (1988). *J. Cryst. Growth*, **88**, 537–540.
Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, ch. 17.17. Dordrecht: Kluwer Academic Publishers.
Kossiakoff, A. A. & Spencer, A. (1980). *Nature (London)*, **27**, 414–416.
Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Leslie, A. W. G. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Minasov, G., Wang, X. & Shoichet, B. K. (2002). *J. Am. Chem. Soc.* **124**, 5333–5340.
Nahlovska, Z., Nahlovsky, B. & Strand, T. G. (1970). *Acta Chem. Scand.* **24**, 2617–2628.
Neumann, T., Junker, H.-D., Keil, O., Burkert, K., Ottleben, H., Gamer, J., Sekul, R., Deppe, H., Feurer, A., Tomandl, D. & Metz, G. (2006). *Lett. Drug Des. Discov.* **2**, 590–595.
Ngai, P. K. & Chang, J.-Y. (1991). *Biochem. J.* **280**, 805–808.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* D**52**, 1161–1168.
Price, H. J. (1999). PhD thesis. The University of Manchester, UK.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1997). *SHELXL-97 Program for Crystal Structure Refinement.* University of Göttingen, Göttingen, Germany.
Skrzypczak-Jankun, E., Carperos, V. E., Ravichandran, K. G. & Tulinsky, A. (1991). *J. Mol. Biol.* **221**, 1379–1393.
Stubbs, M. T. & Bode, W. (1993). *Thromb. Res.* **69**, 1–58.
Tsyba, I. & Bau, R. (2002). *Chemtracts Inorg. Chem.* **15**, 233–257.
Würtele, M., Hahn, M., Hilpert, K. & Höhne, W. (2000). *Acta Cryst.* D**56**, 520–523.
Zobel, D., Luger, P., Dreissig, W. & Koritsanszky, T. (1992). *Acta Cryst.* B**48**, 837–848.