Acta Cryst. (1996). D**52**, 1161–1168

Atomic Resolution Structure of Concanavalin A at 120 K

SEAN PARKIN,[a]* BERNHARD RUPP[a] AND HÅKON HOPE[b]

[a]*Biology and Biotechnology Research Program, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA, and* [b]*Department of Chemistry, University of California, Davis, CA 95616, USA. E-mail: sp@oedipus.llnl.gov*

(*Received 23 April 1996; accepted 8 July 1996*)

Abstract

The structure of native concanavalin A has been refined to a resolution of 1.2 Å against data collected at 120 K. The space group is *I*222, with $a = 61.954\,(8)$, $b = 86.053\,(11)$, $c = 89.079\,(11)$ Å. The structure was refined by restrained weighted least-squares minimization of $\sum w(F_o^2 - F_c^2)^2$ with *SHELXL*92/3/6. The final model contains all of the atoms from 237 amino acids, two metal ions and 271 water molecules spread over 287 sites. Disorder is modelled over two conformations for 30 amino-acid side chains. The final weighted R index on F^2 (wR_2) on all data was 30.4%. Conventional R indices based on F were 14.2 and 11.8% for all data and for data with $F > 4\sigma(F)$, respectively.

1. Introduction

Concanavalin A (conA), a 237-amino-acid lectin from the jack bean (*Canavalia ensiformis*), has long been exploited in biochemical research because of its specific saccharide-binding properties. In spite of its widespread use, however, the biological function of conA remains unknown.

Crystals of conA form in the space group *I*222, with the molecules in an approximately tetrahedral arrangement. There is one molecule per asymmetric unit. The first structure determinations were at low resolution (Reeke, Becker & Quiocho, 1971; Hardman, Wood, Schiffer, Edmondson & Ainsworth, 1971; Edelman *et al.*, 1972; Hardman & Ainsworth, 1972). These were followed by determinations at higher resolution (Becker, Reeke, Wang, Cunningham & Edelman, 1975; Reeke, Becker & Edelman, 1975), all of which were based on the chemical sequence (Wang, Cunningham, Waxdal & Edelman, 1975; Cunningham, Wang, Waxdal & Edelman, 1975). Unrefined coordinates were deposited in the Protein Data Bank at Brookhaven (Bernstein *et al.*, 1977). The gene sequence of conA by Min, Dunn & Jones (1992) changed the assignment of 13 of the residues.

Native conA has two metal binding sites, a Ca^{2+} site and a transition metal site predominantly occupied by Mn^{2+}. It was shown by Kalb, Yariv, Helliwell & Papiz (1988) that repopulation of the latter site by a single metal ion species improved the diffraction quality of crystalline conA. A cationic size effect was also noted, such that the diffraction limit improved from 1.46 Å for native crystals to 1.19 Å for Cd^{2+}- and Mn^{2+}-substituted crystals. For smaller cations, such as Zn^{2+} and Ni^{2+}, the improvement was less dramatic. Room-temperature structures of conA with the transition-metal site occupied solely by Co^{2+} and by Ni^{2+} were subsequently refined at 1.6 and 2.0 respectively, by Emmerich *et al.* (1994). A third metal binding site was also identified in a room temperature 2.0 Å refinement of Cd^{2+} substituted conA by Naismith *et al.* (1993).

Since native conA crystals were known to diffract to better than 1.5 Å at room temperature, we thought that substantial improvements in the diffraction limit would be seen upon cooling to cryogenic temperatures. From the simple reasoning of Hope (1988), a temperature reduction from around 300 to 100 K should increase the diffraction limit from around 1.5 Å to well beyond 1.2 Å, which is commonly taken as the threshold of true atomic resolution. In fact, we routinely observed flash-cooled crystals of conA to diffract to about 0.9 Å resolution, albeit rather weakly. Since low-temperature refinements at atomic resolution can reveal information that is not available in room-temperature studies, we felt that such an analysis of conA was timely. Unfortunately, because of instrumental limitations (which have recently been overcome), we were unable to collect data beyond $2\theta = 90°$ (~1.1 Å resolution). In this paper we present an atomic resolution (1.2 Å) model of native conA refined using *SHELXL*93 (Sheldrick & Schneider, 1996) against Cu $K\alpha$ data collected at 120 K.

2. Materials and methods

2.1. *Data collection and processing*

Crystals of native conA were prepared (at pH 6.5) following the method described by Kalb *et al.* (1988). The cell dimensions at 120 K were $a = 61.954\,(8)$, $b = 86.053\,(11)$, $c = 89.079\,(11)$ Å, which represents a volume contraction in the range 1–4% compared with the room-temperature cell constants (see for example Weisgerber & Helliwell, 1993; Emmerich *et al.*, 1994). Such contractions are normal for molecular crystals upon cooling (Hartmann *et al.*, 1982; Hope, 1988). Variations in cell volume of up to 3% were

ISSN 0907-4449

Table 1. *Data coverage and intensity statistics*

Resolution shell (Å)	Unmerged Friedel pairs: All data	No. with $I > 2\sigma(I)$	Coverage (%)	Mean $I\sigma(I)$	Merged Friedel pairs: All data	No. with $I > 2\sigma(I)$	Coverage (%)	Mean $I/\sigma(I)$	R(int) (%)
∞–3.35	6431	3591	96.8	41.35	6379	3564	98.5	42.54	2.31
3.35–2.60	7092	3927	94.0	33.33	7002	3891	98.6	34.46	2.29
2.60–2.25	5700	3854	80.2	19.15	5386	3669	95.8	20.11	3.59
2.25–2.00	6537	4601	70.5	14.64	5227	4232	95.5	15.58	3.80
2.00–1.85	5628	4006	68.8	10.66	4856	3519	94.1	11.04	5.29
1.85–1.70	6818	4957	60.1	6.85	5324	3946	84.4	7.21	7.49
1.70–1.60	7325	4388	72.5	8.47	5829	3595	84.2	9.42	6.92
1.60–1.50	9719	5694	74.9	6.97	7136	4382	85.2	7.76	8.22
1.50–1.40	12237	7320	72.2	5.10	7971	5081	83.9	5.95	10.26
1.40–1.35	7346	4447	70.3	3.87	4156	2765	83.0	4.57	12.64
1.35–1.30	8375	5167	69.1	3.44	4350	2999	83.1	4.10	14.59
1.30–1.25	9474	6022	67.1	2.89	4421	3161	83.2	3.43	15.95
1.25–1.20	10773	7109	64.9	2.38	4377	3268	83.7	2.82	18.19
1.20–1.15	9355	6461	48.1	2.22	3478	2770	65.2	2.62	19.50

also seen between crystals grown and cooled under essentially identical conditions. Therefore, it is possible that a flash-cooled crystal could yield cell dimensions comparable to or even slightly larger than those of a different crystal at room temperature. Such observations are probably indicative of real structural differences, especially with respect to disordered groups and the solvent regions (Hope, 1988). In addition, a strange temperature dependence of the b and c cell dimensions was observed and verified on three different crystals (Parkin & Hope, 1993). Preliminary evidence suggests that this is associated with a slight rearrangement of a loop in the vicinity of Asn162, but a full account will be presented later. A crystal of approximate dimensions $0.90 \times 0.50 \times 0.35$ mm was transferred to a modified mother liquor solution containing ~30% 2-methyl-2,4-pentanediol (MPD) and mounted in a fine glass loop at the end of a copper mounting pin. It was placed in the cold-stream path of a modified Siemens LT-2 low-temperature apparatus attached to a Siemens P4RA diffractometer. A Cu $K\alpha$ data set extending to 1.68 Å (27 201 unique reflections) was collected from this crystal, and used in preliminary least-squares refinements. Later, a second crystal, roughly $0.80 \times 0.45 \times 0.30$ mm in size, was mounted in a similar manner into the cold stream of a modified Siemens LT-2 low-temperature machine attached to a Huber goniometer equipped with dual multiwire area detectors from Area Detector Systems Corporation (Poway, California, USA). A data set extending to a nominal resolution of ~1.1 Å, with an overall $R_{\rm merge}$ of 4.2% was collected using Cu $K\alpha$ X-rays from a Rigaku RU200 generator. Because of geometric limitations the coverage of reciprocal space in this data set was not complete. The merging R value between the two data sets was about 11%, using an anisotropic scaling routine in the *XtalView* package (McRee, 1992). Since the individual $R_{\rm int}$ values were both below 5%, we decided not to merge the scaled diffractometer data with the area detector data set. Owing to a combination of weak counting statistics and poor coverage of reciprocal space in the outermost resolution shell between 1.2 and 1.1 Å, we regard the effective resolution limit of the data to be 1.2 Å. The cell dimensions of the first crystal were obtained from a least-squares fit of $\sin^2\theta$ values for 24 reflections well distributed in reciprocal space. They are more reliable than those obtained from the area detector data ($a = 61.883$, $b = 86.116$, $c = 89.145$ Å; a weighted average of values optimized for each wedge of area detector data) and have the additional advantage of an associated standard uncertainty (hitherto known as estimated standard deviation). For this reason, further refinement of the model using this 1.2 Å resolution data set used the cell dimensions obtained for the first crystal. The difference between cell constants of the first crystal and the second was about 0.1% in each dimension. The 1.2 Å data set consists of 112 810 unique reflections. Coverage and intensity statistics are given in Table 1.

2.2. *Solution and refinement*

The structure was refined on a variety of computers, including a DEC Vaxstation 3100 M38 and Silicon Graphics Indigo, Indy and Power Challenge machines using various chronological incarnations of *SHELXL* [β and γ test versions of *SHELXL*92, *SHELXL*93 (Sheldrick & Schneider, 1996) and a β test version of *SHELXL*96]. An initial model was derived from the 2.4 Å resolution Protein Data Bank (Bernstein *et al.*, 1977) entry 3CNA (Hardman & Ainsworth, 1976). Initial model rebuilding was carried out by a combination of refinement with specific amino acids removed from the model followed by analysis of electron-density peak lists output by *SHELXL*92. By the time the full 1.2 Å data set became available, the model was complete apart from two poorly defined loop regions around residues His120 and Asn162. This model contained 205 water molecules, and had an R value of 16.8% [based on F for $F > 4\sigma(F)$]. In the later

Table 2. *Summary of model building and refinement progress*

Round	wR_2 (%)	R_1 (%)	R_{free} (%)	ND	NR	NP	Notes
0*	71.50	33.52	—	27364	7162	7492	Model constrained, 66 waters, isotropic U's diffractometer data
1*	48.52	20.43	—	27364	7367	8038	205 waters, further refinement with diffractometer data
2†	42.96	15.29	18.60	101133	7558	8449	Coordinate shake (r.m.s. = 0.18 Å), refinement against 1.2 Å data, absorption correction with *XABS2*§
3†	36.42	12.95	16.24	101133	35455	19102	Anisotropic thermal parameters, H atoms and disorder modelled
4†	36.18	12.83	16.23	101133	35455	19102	Further refinement
5†	35.01	12.49	16.11	101133	35437	19102	Disorder fine tuning, water editing, further refinement
6†	32.46	12.41	17.25¶	105289	35437	19102	Coordinate shake (r.m.s. = 0.18 Å), 50 cycles least squares against all data available data from ∞ to 1.15 Å
7‡	30.67	11.87	16.71	105289	35167	19482	Water structure, disorder rebuilt
8‡	30.36	11.75	—	112810	35167	19482	Coordinate shake, final refinement on all data

*Refinement with *SHELXL92*. †Refinement with *SHELXL93*. ‡Refinement with *SHELXL96*. §Parkin, Moezzi & Hope (1995). ¶ *SHELXL93* creates the data subset used for R_{free} by skipping every nth reflection. When all data were included in the refinement, this data subset was not the same as before. The value of 'n' was also changed from 10 to 15 at this point to allow more reflections to contribute to the refinement. The continuity of the trend in R_{free} is thus broken at this point, and the higher value is not an indicator of overfitting.

stages of refinement, model rebuilding and analysis also made use of the program *Xfit* from the *XtalView* package (McRee, 1992). The various stages in the refinement, and the imposition of constraints and restraints using the 'standard' values of Engh & Huber (1991) were similar to those outlined in detail for a refinement of bovine pancreatic trypsin inhibitor (BPTI) by Parkin, Rupp & Hope (1996). This scheme involves increasing the complexity of the model while monitoring the progress of the R_{free} statistic (Brünger, 1992, 1993). With the geometric integrity of the polypeptide chain maintained by bond-length and angle restraints, the model was refined to convergence using a conjugate-gradient algorithm to minimize $\sum w(F_o^2 - F_c^2)^2$ with individual isotropic thermal parameters. Water molecules were added from difference map information and disorder was modelled when evidence from 'omit' electron-density maps calculated with a variety of coefficients including $F_o - F_c$ and $2F_o - F_c$ suggested multiple conformations. Further refinement cycles were followed by the addition of more water molecules, fine tuning of disordered side chains, calculation of anisotropic thermal parameters, addition of H atoms and correction of absorption effects using the program *XABS2* (Parkin, Moezzi & Hope, 1995). Further sophistication of the model was halted when the downward trend of R_{free} ceased. As a final step, the model was subjected to a coordinate shake (r.m.s = 0.18 Å) to minimize 'memory' bias, and refined to convergence against the full data set. The conventional R factor at this point was 11.8% [based on F for $F > 4\sigma(F)$]. A similar approach is described in more detail by Sheldrick & Schneider (1996) and by Frazão *et al.* (1995). Table 2 gives an overview of the refinement progress.

3. Results and discussion

3.1. *Overall features*

The asymmetric unit of conA contains a single molecule with approximate dimensions 42 × 40 × 40 Å. Two molecules related about a crystallographic twofold parallel to the c axis form pseudo-ellipsoidal dimers. These dimers are related about a second crystallographic twofold to form tetramers. The secondary structure is dominated by two large antiparallel β-sheets of 64 and 57 amino acids each that span the majority of the surface of the molecule. The rest of the protein is made up of loops and turns joining the β-strands. The transition-metal site is six coordinate, involving interactions with Oε2 Glu8, Oδ2 Asp10, Oδ2 Asp19, Nε2 His24 and two water molecules. The seven-coordinate Ca^{2+} is bound to both O atoms in the carboxyl group of Asp10, and is also connected to O Tyr12, Oδ1 Asn14, Oδ1 Asp19 and to two water molecules. Metal–ligand distances are compared with room-temperature values in Table 3. The general characteristics of the molecule itself are largely unchanged relative to room-temperature structures, and have been described in detail by Naismith *et al.* (1993). The major differences between the model obtained in this refinement and that in other studies are largely a consequence of the low temperature (all other studies have been at room temperature) and the higher limiting resolution (1.2 Å compared with 1.6 Å for the refinement of Co^{2+}-substituted conA). The higher resolution and low temperature, however, have led to a much more extensive water structure and clear definition of some previously poorly defined loops. We reserve detailed description here to structural features in which the present model has significant changes or additional information compared with previous room-temperature models.

Table 3. *Metal–ligand distances at 120 K and room temperature (RT)*

Room-temperature values are taken from Emmerich *et al.* (1994).

(*a*) Transition-metal-to-ligand distances, thermal parameters [B (Å^2)] in parentheses

	Native 120 K	Native RT	Co^{2+} RT	Ni^{2+} RT
Oε2 Glu8	2.15 (7.8)	2.23 (14)	2.14 (8)	2.21 (16)
Oδ1 Asp10	2.11 (8.0)	2.13 (15)	2.12 (7)	2.13 (12)
Oδ1 Asp19	2.17 (8.8)	2.26 (13)	2.19 (9)	2.17 (18)
Nε2 His24	2.24 (8.4)	2.25 (9)	2.20 (10)	2.13 (10)
OW A	2.18 (8.9)	2.26 (14)	2.19 (10)	2.26 (12)
OW B	2.28 (8.4)	2.13 (9)	2.24 (8)	2.28 (14)
Average	2.19 (8.4)	2.21 (13)	2.18 (9)	2.20 (14)

(*b*) Ca^{2+}-to-ligand distances, thermal parameters [B (Å^2)] in parentheses

	Native 120 K	Native RT	Co^{2+} RT	Ni^{2+} RT
Oδ1 Asp10	2.53 (8.0)	2.45 (15)	2.47 (7)	2.51 (12)
Oδ2 Asp10	2.45 (8.2)	2.28 (10)	2.39 (9)	2.36 (13)
O Tyr12	2.37 (8.4)	2.31 (10)	2.32 (9)	2.39 (13)
Oδ1 Asn14	2.35 (8.4)	2.55 (15)	2.37 (10)	2.57 (16)
Oδ2 Asp19	2.41 (8.6)	2.30 (11)	2.32 (10)	2.40 (13)
OW C	2.44 (8.6)	2.37 (13)	2.40 (10)	2.45 (12)
OW D	2.34 (8.0)	2.32 (10)	2.42 (7)	2.36 (15)
Average	2.41 (8.3)	2.37 (12)	2.38 (9)	2.43 (13)

3.2. *Secondary structure*

The two large β-sheets are changed little at 120 K, but there are a number of notable differences in residues situated in the loops. The root-mean-square (r.m.s.) deviation for the Cα atoms between this model and the 1.6 Å Co^{2+}-substituted conA of Emmerich *et al.* (1994) is 0.34 Å; it is 0.32 Å for all backbone atoms (unless stated otherwise, r.m.s. deviations given are for the Co^{2+}-substituted conA, as it is the highest resolution model available for comparison). This is comparable to the r.m.s. deviations found between room and low-temperature structures of BPTI (Parkin *et al.*, 1996), but is considerably higher than the r.m.s. deviations between the various well refined native and substituted conA's (0.10–0.17 Å, Emmerich *et al.*, 1994). For the β-sheet regions alone, the r.m.s. deviation for Cα atoms is 0.18 (*c.f.* 0.08–0.13 Å for different room-temperature models).

The loop and turn stretches in conA contain the major differences between this low-temperature model and the room-temperature models, and all are associated with regions having higher than average thermal parameters (see §3.3 below). The overall r.m.s. deviation for side-chain atoms is 1.33 Å but within the β-sheets it is lower, at 1.05 Å. There are significant deviations at the N-terminal alanine, at a loop in the vicinity of Thr15 and at the loop between Val159 and Pro165. The largest difference is at a loop between residues Asn118 and Thr123 (Fig. 1*a*) at one extreme of the molecule. The Nε2 atom of His121 is close (≃ 3.0 Å) to its equivalent atom on a twofold-related molecule. There was a lot of weak electron density in this region, although no firm evidence of other multiple conformers could be found. Fig. 1(*b*) is an electron-density map (coefficients $2F_o - F_c$) of this loop contoured at the 0.67σ level. Other than the side chains of residues Lys46, Lys101 and Arg158, the only other significant differences between this model and Co^{2+}-substituted conA involve those residues for which multiple conformations were modelled (see §3.4). Hydrogen bonds between backbone atoms forming the β-sheets average 2.9 Å, while those for backbone to side chain and side-chain to side-chain interactions both average 2.8 Å.

3.3. *Correction of sequence errors and side-chain orientations*

Determination of the gene sequence of conA by Carrington, Auffret & Hanke (1985) changed the assignment of 15 amino acids from the chemical sequence of Wang *et al.* (1975). Two of these were subsequently changed back after a redetermination by Min *et al.* (1992). It was in fact clear from the early stages in the refinement that anomalies existed in the starting model sequence assignment. Of the misassigned residues, some were of obviously different geometry (Ser72, Ala73, Thr74, Ala189, Ser190, Thr196 were changed from Ala, Thr, Ser, Ser, Ala, Ala, respectively), and stood out clearly in electron-density maps. Other changes required the substitution of residues with similar geometries (*i.e.* Asn and Asp, Gln and Glu, Thr and Val). A combination of hydrogen-bond and thermal-parameter information was used to orient the side chains correctly.

After substantial model rebuilding and refinement, all residues were situated in recognisable electron density. The ends of some of the Gln and Asn side chains as well as His51, however, were misoriented. Correct assignment of O- and N-atom positions in Gln and Asn was made by analysis of isotropic thermal parameters and electron-density maps. The correct orientation of His51 was clear from analysis of electron-density maps (Fig. 2) and was confirmed by the hydrogen-bonding environment around the N atoms.

3.4. *Thermal parameters*

On average, backbone isotropic U values are lower than their corresponding room temperature values (compared with the Co^{2+}-substituted conA of Emmerich *et al.*, see Fig. 3*a*,*b*). This is particularly apparent in the less well defined regions of the molecule such as the loops around His121 and Asn162 and at the N terminus. The reduction is much less dramatic than that observed for BPTI at 125 K, however (Parkin *et al.*, 1996). One possible explanation for this is that at the lower resolution limits of the various room-temperature conA structure refinements (typically 1.6–2.0 Å), much less of the variation in scattering factor *versus* $\sin\theta/\lambda$ is present in the data, making the refinement of individual thermal parameters less reliable. This reasoning ties in well with the observation that there are a large number of models present in the Protein Data Bank

with clearly nonsensical thermal parameters (*e.g.* those driven down to some programmed lower limit by the least-squares process), particularly at poorer resolution. Thermal parameters in the backbone atoms are also generally lower at 120 K. The number of side chains with thermal parameters that are vastly higher than their neighboring atoms is greatly reduced at low temperature. This may be partly due to the 'SIMU' restraint option of *SHELXL*, which weakly restrains U values on adjacent atoms to be similar. Another probable partial explanation is the larger number of side chains that were modelled in multiple conformations in this work. The trace of thermal parameters with amino-acid sequence conforms quite closely to the secondary structure, with the smallest vibrational amplitudes associated with the more structurally well defined β-sheets.

Calculation of restrained anisotropic thermal parameters was associated with a drop in $R_{\rm free}$ of 2.36% (Table 2), and so likely represents a real improvement in the model, though part of this may be attributable to the addition of H atoms and modelling of disorder. In contrast, refinement of anisotropic U values for BPTI (Parkin *et al.*, 1996) gave no significant drop in $R_{\rm free}$. The difference is most likely to result from data quality; the higher resolution shells of BPTI data were known to be rather weak.

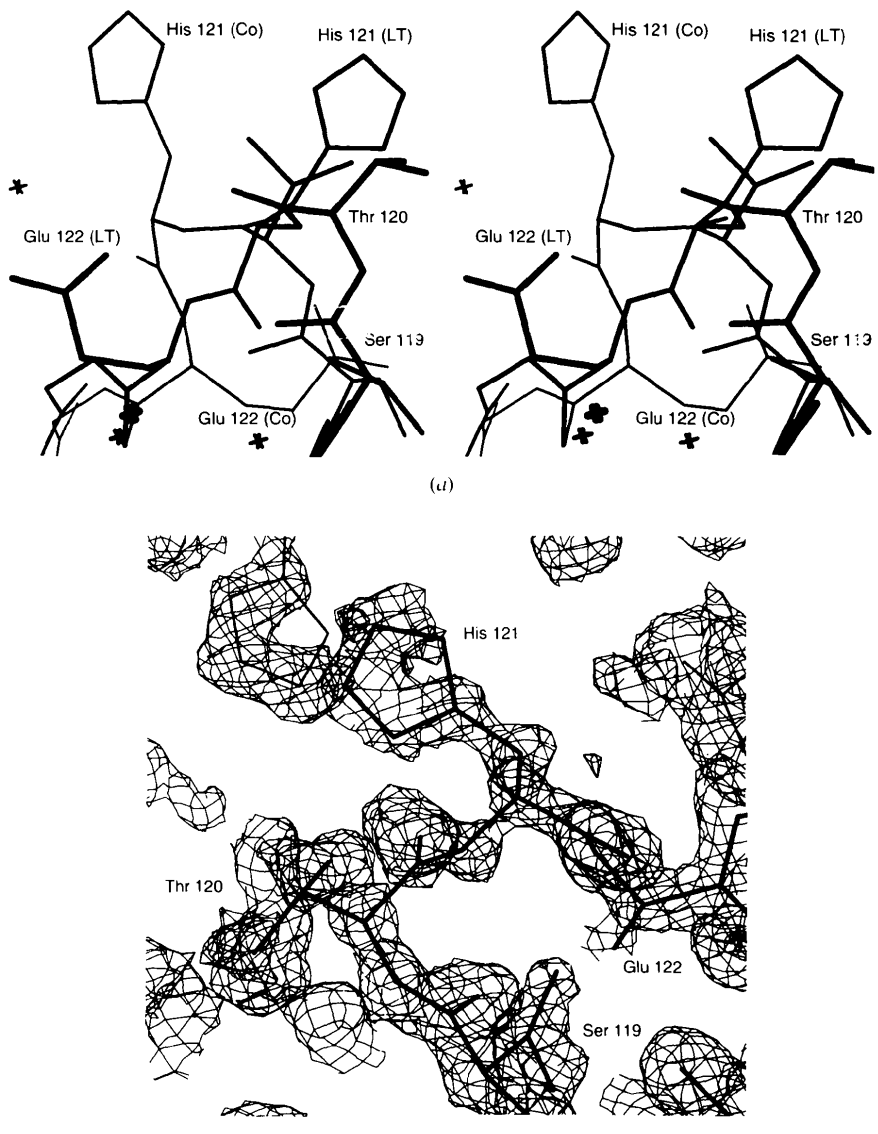

Fig. 1. (*a*) Stereoview (cross-eye) comparison of the loop region centred around residue His121 for this low-temperature model (thick lines) and the room-temperature model of Co^{2+}-substituted conA. Water-molecule positions are indicated by crosses. (*b*) Electron-density plot (0.67σ level, coefficients $2F_o - F_c$) of the poorly defined loop region around residue His121.

3.5. *Disorder*

In the final model of conA from this refinement, 30 residues have disorder modelled explicitly over two conformations. Some of the atoms involved still had rather high thermal parameters (in comparison with neighbouring atoms), and those side chains could possibly be disordered over more than two sites. In particular, one of the two major conformations of Met42 (disordered at Cγ onwards) appears to have two positions for atom Cε, but it was not modelled as such.

The disordered side chains of Leu9, Ile25 and Val65, buried inside the protein, point toward each other, but only the disorder of Ile25 and Val65 appear correlated. On the protein surface, Thr15 and Ser21 are adjacent in space but their disorder does not appear to be a direct result of interference. Side-chain disorder of the proximal residues Ser160, Ser161, Asn162 and Gln166, on the other hand, is correlated. Mutual disorder exists between the side chains of Arg60 and Asp78 and the same residues of a twofold-related molecule. Asp218 is also disordered and is bridged by water *via* a twofold to Asp218 of a related molecule. Disorder of some side chains is correlated with disorder on other molecules, notably Gln132 with a twofold-related Ser185 and both Lys114 and Lys116 (on one molecule) with a twofold-related Glu192. The majority of the disordered side chains (Ile27, Ser113, Ser119, Met129, Ser134, Lys135, Asp151, Val159, Ser164, Thr196, Ser215, Ser223), however, do not appear to be interfering with other side chains.

Of the residues identified as potential disorder candidates at room temperature (Emmerich *et al.*, 1994), only Ile25 and Val65 were found to be disordered in this study. There was no evidence for a second conformation of His51, Ser72 or Thr74. A comparison of atomic resolution models of BPTI (Parkin *et al.*, 1996), also found large differences in disorder between room and low temperature. Although it is possible that these differences may represent slightly different crystal growth conditions, it is also possible that they reflect structural rearrangements occurring during the cooling process.

3.6. *Solvent structure*

A total of 271 water molecules spread over 287 sites were included in the final model. In addition, a means of modelling the contribution of diffuse water by Babinet's principle (Langridge *et al.*, 1960; Driessen *et al.*, 1989) was included using the 'SWAT' option in *SHELXL*93 (the β-test version of *SHELXL*96 used in the later stages of refinement used a different formulation, described by Moews & Kretsinger, 1975). One water molecule is located on a twofold axis (Fig. 4), as was reported to be the case at room temperature in all the well refined models.

There are four distinct regions where water molecules penetrate the surface of the conA molecule. A single water is enclosed by hydrogen bonds to backbone O atoms of Phe175 and Ala177 and to the N atom of Val91.

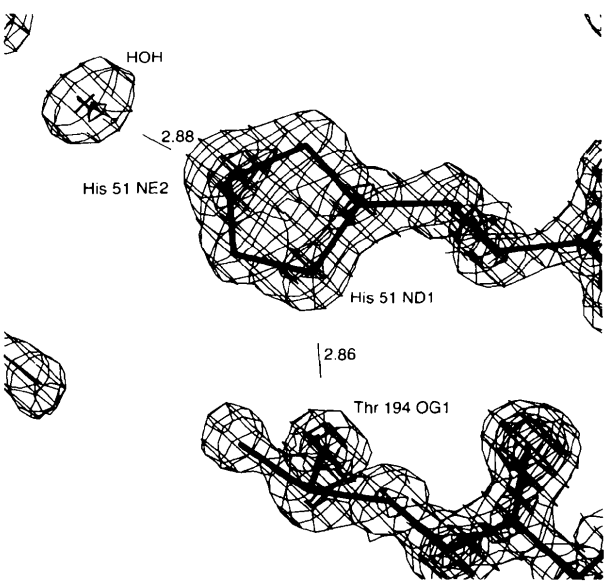

Fig. 2. Electron-density plot (1σ, 3σ level, coefficients $2F_o - F_c$) of His51, showing increased electron density associated with N atoms and the hydrogen bonds that were used to correctly orient the side chain.

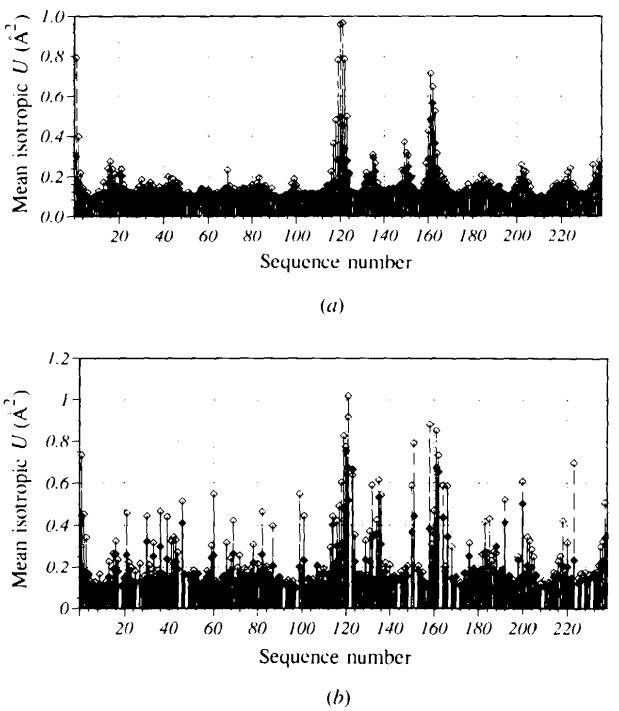

Fig. 3. Thermal parameters *versus* residue number for conA crystals at 120 K (solid, this work) and for room temperature Co^{2+}-substituted conA (outline, Emmerich *et al.*, 1994). (*a*) Backbone atoms; (*b*) side-chain atoms.

Two water molecules, adjacent to each other, are held within the protein molecule by residues Glu102, Thr103, Asn104, Thr105, Glu155, Leu156, Thr157 and Arg158. A chain of four hydrogen-bonded water molecules is held in a cleft at one end of the molecule by Tyr54, Ser56, Leu81, Asp82, Ser113, His180, Ile181, Ala189 and Phe191, but they are not cut off from the outside of the protein. The rest of the internal waters are located around the metal binding sites. Two of them are bound to the Ca^{2+} ion, while two (different ones) are bound to the transition-metal site (modelled as Mn^{2+} in this refinement). The internal waters tend to have low U values, but not significantly different than well ordered waters directly hydrogen bonded to the protein molecule.

The remainder of the waters surround the protein. Around hydrophobic side chains, there are hydrogen-bonded water ring structures similar to those observed in crambin (Teeter, 1984) and BPTI (Parkin *et al.*, 1996). Hydrogen bonds between water and backbone atoms average 2.9 Å, while side-chain to water hydrogen bonds average 2.8 Å. The average water to water hydrogen-bond distance is 2.8 Å. The level of detail in the solvent regions of this study is somewhat higher than that at room temperature. The 2.0 Å structures of Cd^{2+}-(Naismith *et al.*, 1993) and Ni^{2+}- (Emmerich *et al.*, 1994) substituted conA each had 144 waters, while the 1.6 Å model of Co^{2+}-substituted conA (Emmerich *et al.*, 1994) had 149 water molecules. It is clear that the increased clarity of electron-density maps obtained from low-temperature data can greatly facilitate the modelling and rationalization of solvent structure in protein crystals.

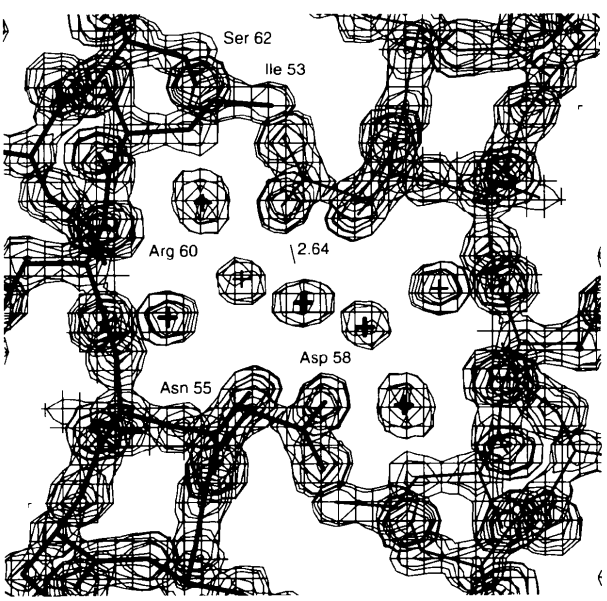

Fig. 4. Electron-density plot (1σ, 3σ, 5σ level, coefficients $2F_o - F_c$) showing the water atom situated on a twofold axis running perpendicular to the page. This water has a distorted tetrahedral hydrogen-bonding geometry.

4. Concluding remarks

A proper assessment of accuracy and precision is made difficult by the absence of both an independent low-temperature refinement and the absence of standard uncertainties. An estimate of the upper bound coordinate error from a Luzzati plot (Luzzati, 1952) is about 0.12 Å. This must be treated with caution, though, because the Luzzati plot assumes that coordinate errors are the root cause of all variations in the R factor (Guss, Bartunik & Freeman, 1992; Parkin, 1993). The actual deviations depend strongly on the local environment, being somewhat smaller in the β-sheets and larger in the loops and turns, particularly those suffering from disorder. Attempts to quantify accuracy by direct comparison of this low-temperature model with any or all of the room-temperature models is not possible because real differences exist between them. Whether these are a result of crystal growth conditions or cooling remains to be seen. The final model was checked using the *PROCHECK* program (Laskowski, MacArthur, Moss & Thornton, 1993). R.m.s. deviations of bond lengths and angles between this model and standard values (Engh & Huber, 1991) are 0.012 Å and 2.2°. In a Ramachandran plot (not shown), all non-glycine residues were in the 'most favoured' (>95%) or the 'allowed' regions (<5%). All other common indicators of model quality convey favourable statistics for the refinement. R factors for the final model were 11.8% [based on F for $F > 4\sigma(F)$] and 14.2% on all data, while wR_2 (all data) was 30.4%.*

In the final stages of manuscript preparation we became aware of a second independent low-temperature refinement of conA using synchrotron data to even higher resolution (0.94 Å) from Professor Helliwell's group. A comparison of these two models should reveal some important structural differences that can exist between crystals of the same protein.

We are grateful to A. Joseph Kalb (Gilboa) for the kind donation of a sample of conA. SP would like to thank the Regents of the University of California for the award of a Regents Fellowship during the initial stages of this study. Part of this work was performed under the auspices of the US Department of Energy under contract number W-7405-ENG-48 at Lawrence Livermore National Laboratory.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1JBC, R 1JBCSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: am0048).

References

Becker, J. W., Reeke, G. N. Jr, Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1513–1524.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Carrington, D. M., Auffret, A. & Hanke, D. E. (1985). *Nature (London)*, **313**, 64–67.
Cunningham, B. A., Wang, J. L., Waxdal, M. J. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1503–1512.
Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
Edelman, G. M., Cunningham, B. A., Reeke, G. N., Becker, J. W., Waxdal, M. J. & Wang, J. L. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 2580–2584.
Emmerich, C., Helliwell, J. R., Redshaw, M., Naismith, J. H., Harrop, S. J., Raftery, J., Kalb (Gilboa), A. J., Yariv, J., Dauter, Z. & Wilson, K. S. (1994). *Acta Cryst.* D**50**, 749–756.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Frazão, C., Soares, C. M., Carrondo, M. A., Pohl, E., Dauter, Z., Wilson, K. S., Hervs, M., Navarro, J. A., De La Rosa, M. A. & Sheldrick, G. M. (1995). *Structure*, **3**(11), 1159–1169
Guss, J. M., Bartunik, H. D. & Freeman, H. (1992). *Acta Cryst.* B**48**, 790–811.
Hardman, K. D. & Ainsworth, C. F. (1972). *Biochemistry*, **11**, 4910–4919.
Hardman, K. D. & Ainsworth, C. F. (1976). Protein Data Bank entry 3CNA, Brookhaven National Laboratory, Upton, NY 11973, USA.
Hardman, K. D., Wood, M. K., Schiffer, M., Edmondson, A. B. & Ainsworth, C. F. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 271–276.
Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ringe Ponzi, D. & Frauenfelder, H. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4967–4971.
Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.
Kalb, A. J., Yariv, J., Helliwell, J. R. & Papiz, M. Z. (1988). *J. Cryst. Growth*, **88**, 537–540.
Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
McRee, D. E. (1992). *J. Mol. Graphics*, **10**, 44–46
Min, W., Dunn, A. J. & Jones, D. H. (1992). *EMBO J.* **11**, 1303–1307.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
Naismith, J. H., Habash, J., Harrop, S., Helliwell, J. R., Hunter, W. N., Wan, T. C. M., Weisgerber, S., Kalb (Gilboa), A. J. & Yariv, J. (1993). *Acta Cryst.* D**49**, 561–571.
Parkin, S. R. (1993). PhD thesis, University of California at Davis, USA.
Parkin, S. R. & Hope, H. (1993). Am. Crystallogr. Assoc. Meet. Abstract PI20, Albuquerque, NM, USA.
Parkin, S. R., Moezzi, B. & Hope, H. (1995). *J. Appl. Cryst.* **28**, 53–56.
Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* D**52**, 18–29.
Reeke, G. N. Jr, Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525–1547.
Reeke, G. N. Jr, Becker, J. W. & Quiocho, F. A. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 277–284.
Sheldrick, G. M. & Schneider, T. R. (1996). *Methods Enzymol.* In the press.
Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.
Wang, J. L., Cunningham, B. A., Waxdal, M. J. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1490–1502.
Weisgerber, S. & Helliwell, J. R. (1993). *J. Chem. Soc. Faraday Trans.* **89**, 2667–2675.