Acta Cryst. (1993). D**49**, 561–571

Refined Structure of Cadmium-Substituted Concanavalin A at 2.0 Å Resolution

BY JAMES H. NAISMITH,* JARJIS HABASH, STEPHEN HARROP, JOHN R. HELLIWELL,† WILLIAM N. HUNTER, TOMMY C. M. WAN‡ AND SUSANNE WEISGERBER

*Department of Chemistry, University of Manchester, Oxford Road, Manchester M*13 9*PL, England*

A. JOSEPH KALB (GILBOA)

Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

AND JOSEPH YARIV

Laboratoire de Cristallographie, URA 144, *CNRS, Université de Bordeaux* I, 33405 *Talence, France*

(*Received* 12 *March* 1993; *accepted* 22 *June* 1993)

Abstract

The three-dimensional structure of cadmium-substituted concanavalin A has been refined using *X-PLOR*. The *R* factor on all data between 8 and 2 Å is 17.1%. The protein crystallizes in space group *I*222 with cell dimensions $a = 88.7$, $b = 86.5$ and $c = 62.5$ Å and has one protein subunit per asymmetric unit. The final structure contains 237 amino acids, two Cd ions, one Ca ion and 144 water molecules. One Cd ion occupies the transition-metal binding site and the second occupies an additional site, the coordinates of which were first reported by Weinzierl & Kalb [*FEBS Lett.* (1971), **18**, 268–270]. The additional Cd ion is bound with distorted octahedral symmetry and bridges the cleft between the two monomers which form the conventional dimer of concanavalin A. This study provides a detailed analysis of the refined structure of saccharide-free concanavalin A and is the basis for comparison with saccharide complexes reported elsewhere.

1. Introduction

Concanavalin A is a saccharide-binding protein from the jack bean and was isolated in crystalline form by Sumner (1919). Although the biological role of concanavalin A is still unknown, its specific saccharide-binding properties make it an ideal object for the study of protein–saccharide interactions. Together with the structurally related family of legume lectins it has become a popular tool in various branches of research where saccharide specificity is useful (for a review, see Lis & Sharon, 1986).

In solution the protein is a dimer (Kalb & Lustig, 1968). Each concanavalin A monomer consists of 237 amino acids and the protein crystallizes in space group *I*222, with a monomer in the asymmetric unit. The unit-cell parameters of the naturally occuring protein are approximately $a = 88$, $b = 86$ and $c = 62$ Å. Two groups reported the structure of concanavalin A at low resolution (4 Å) (Reeke, Becker & Quiocho, 1971; Hardman, Wood, Schiffer, Edmundson & Ainsworth, 1971; Edelman *et al.*, 1972; Hardman & Ainsworth, 1972). One group subsequently gave an account of a high-resolution (2 Å) structure, including an analysis of dimer and tetramer formation (by crystal symmetry) (Becker, Reeke, Wang, Cunningham & Edelman, 1975; Reeke, Becker & Edelman, 1975). Both groups deposited unrefined coordinates with the Protein Data Bank (Bernstein *et al.*, 1977). These structures were based on the chemical sequence of concanavalin A (Wang, Cunningham, Waxdal & Edelman, 1975; Cunningham, Wang, Waxdal & Edelman, 1975). Both groups refined their structures, one by difference Fourier techniques (Becker, Reeke, Cunningham & Edelman, 1976) and the other using a restrained least-squares method (Hardman, Agarwal & Freiser, 1982). Neither group described differences with respect to the unrefined structures nor deposited their refined coordinates. The least-squares structure did, however, describe in detail the coordination of the two metals at 1.75 Å resolution. The study also confirmed the presence of a *cis* peptide at Ala207-Asp208.

The gene sequence of concanavalin A was determined in 1985 (Carrington, Auffret & Hanke, 1985) and showed 15 amino-acid changes to the chemical sequence. Recently the gene sequence has

* Present address: Research Laboratories, Howard Hughes Medical Institute, University of Texas, 5323 Harry Hines Boulevard, Dallas, TX 75235-9050, USA.

† To whom all correspondence should be addressed.

‡ Present address: Department of Biomolecular Sciences, Kings College London, 26–29 Drury Lane, London WC2 5RL, England.

been redetermined (Min, Dunn & Jones, 1992); this changes the assignment of two residues (151 and 155) back to those determined by the chemical sequence. The amino-acid sequence of concanavalin A is shown in Fig. 1. The protein has an unusual post-translational modification (Bowles, 1990) which involves a ligation at residues 118 and 119. This ligation appears to be only 60% complete so that, as isolated, the protein is a mixture of intact and 'nicked' subunits.

Concanavalin A has two distinct metal-binding sites both of which must be occupied in order for saccharide to bind. These are the transition-metal binding site (*S*1) with affinity for Ni, Co, Zn, Mn, or Cd and a calcium-binding site (*S*2) which is specific for Ca or Cd ions (Kalb & Levitzki, 1968; Shoham, Kalb & Pecht, 1973). As isolated the native protein contains a mixture of metals at the transition-metal site, predominantly manganese. Concanavalin A has been crystallized with several different ions in the transition-metal site with Ca^{2+} occupying the other site (Kalb, Yariv, Helliwell & Papiz, 1988). They concluded that having a mixed population of metals in the transition-metal binding site made the diffraction limit of the crystals worse. They also pointed to a likely correlation between the radius of the metal ion and the resolution limit of the crystals postulating that ions smaller than Mn^{2+} were altering the position of the residues in the transition-metal binding site thus increasing disorder in the crystal lattice.

We present here the refined 2 Å structure of concanavalin A where the trransition-metal binding site has been populated by cadmium. This structure is part of broader study we have undertaken to determine the protein–saccharide interactions of concanavalin A (Derewenda *et al.*, 1989; Harrop, 1992; Naismith, 1992; and papers in preparation). Such a programme of reseach requires the structure of the protein with and without bound saccharide to properly evaluate the determinants of sugar binding at a molecular level. The present studies provide a well refined saccharide-free structure based upon the most reliable amino-acid sequence currently available.

2. Methods

2.1. *Data collection*

Crystals of cadmium-substituted concanavalin A were grown according to a known protocol (Kalb *et al.*, 1988). The crystals contain a mixture of intact and nicked subunits. Data up to 2 Å resolution were

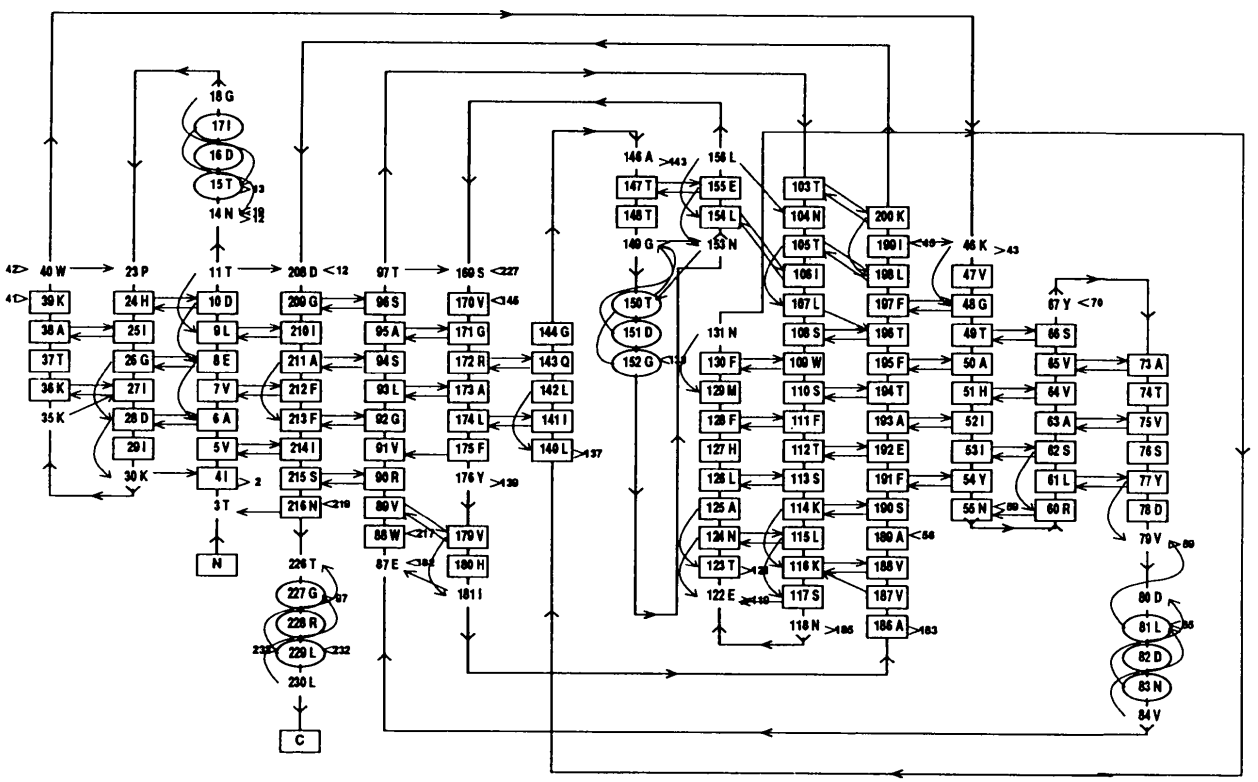

Fig. 1. Schematic representation of the secondary structure in concanavalin A. The two β-sheets are separated and the hydrogen bonds involved in their formation are shown. The diagram was produced using *HERA* (Hutchinson & Thornton, 1991).

collected using photographic film at a wavelength of 1.488 Å on station PX7.2 (Helliwell *et al.*, 1982) at the Daresbury SRS. Four crystals of approximate dimensions 1.0 × 0.4 × 0.4 mm were used to collect the data. The films were digitized using a 50 μm raster in a Joyce–Loebl Scandig 3 microdensitometer and processed using the *MOSFLM* suite of programs (Wonacott, Dockerill & Brick, 1980). Finally the data were merged using the programs *ROTAVATA* and *AGROVATA* (*CCP*4; SERC Daresbury Laboratory, 1979). The average redundancy of the data was 5.1. Missing data were then measured using an AFC-5*R* rotating-anode diffractometer at Cu $K\alpha$. Table 1 summarizes the final data set used for the refinement. The cell dimensions were determined as $a = 88.7$, $b = 86.5$ and $c = 62.5$ Å. The Matthews number (Matthews, 1968) of cadmium-substituted concanavalin A, calculated from the contents of the unit cell, is 2.44 Å^3 Da^{-1} (44% solvent). The density measurement of the crystal by Greer, Kaufman & Kalb (1970) produced nearly the same result of 2.35 Å^3 Da^{-1}.

Table 1. *Data collected on cadmium-substituted concanavalin A*

R_{merge} only applies to data collected at the Daresbury synchrotron, the other statistics are for the final merged data set.

	All data			Data with $F > 1.0\sigma(F)$	
Resolution (Å)	No. of reflections	Complete (%)	R_{merge}* (%)	No. of reflections	Complete (%)
8.0–6.3	289	100	5.0	288	99.7
6.3–4.5	968	100	5.4	963	99.5
4.5–3.7	1193	100	5.6	1184	99.2
3.7–3.2	1439	99.9	5.3	1434	99.6
3.2–2.8	2002	99.9	5.4	1978	98.7
2.8–2.6	1496	99.7	6.6	1467	97.8
2.6–2.4	2011	99.4	6.7	1958	96.7
2.4–2.2	2810	98.5	7.8	2719	95.3
2.2–2.1	1782	98.3	9.2	1712	94.4
2.1–2.0	2188	97.6	11.3	2066	92.1
8.0–2.0	16178	99.1	6.0	15769	96.6

* $R_{merge} = \sum\sum I(h)_i - \langle I(h)\rangle / \sum\sum I(h)_i$ where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

Table 2. *Annealing protocols*

Round	Initial temperature (K)	*R* factor* (%)	R.m.s. deviation from ideality* Bond (Å)	Angle (°)
1	3000	19.8	0.017	3.5
2	1100	20.6	0.014	3.3
	700	18.7	0.015	3.3
3	550	19.2	0.014	3.2

* This refers to the value after positional and temperature refinement which was always run after dynamics simulation.

2.2. *Refinement*

The unrefined coordinates of Reeke *et al.* (1975) from the Protein Data Bank (PDB code 3CNA) were used as a starting model for this study; the *R* factor was 44.6%. Initial positional refinement was carried out with *PROLSQ* (Hendrickson & Konnert, 1980) using data from 8 to 3 Å (eight cycles). The data set was then extended to 2.5 Å for a further 12 cycles, to 2.2 Å for the next six cycles and to 2 Å thereafter. Sequence changes were made according to the gene sequence of Carrington *et al.* (1985) after ten cycles of refinement at 2.0 Å. A further 40 cycles of *PROLSQ* positional and temperature-factor refinement were run. The *R* factor was 20.6%. Further refinement was carried out using *X-PLOR* (version 2.1) (Brünger, 1990) implemented on a Silicon Graphics 4D power series work station. Data in the resolution range 8 to 2 Å with $F > 1\sigma(F)$ were used. No metal–ligand bond-distance restraints were employed and the metal ions were given no charge. We found that charged metal ions led to unrealistically short metal–ligand distances. All molecular dynamics (slow cooling, simulated annealing) employed 25 K drops in temperature to a final temperature of 300 K. At each temperature there were 50 timesteps each of 0.5 fs. Postional refinement was carried out using the *POWELL* minimization option of *X-PLOR*. Electron-density maps were calculated using the *CCP*4 package (SERC Daresbury Laboratory, 1979). Inspection of the model and of the $2F_o - F_c$ and $F_o - F_c$ electron-density maps was carried out using *FRODO* (Jones, 1978) which had been adapted to run on an Evans and Sutherland ESV series work station (P. R. Evans, personal communication). Solvent accessibility was calculated using *DSSP* (Kabsch & Sander, 1984) which employs a 1.4 Å probe.

The refinement proceeded in 11 stages, each stage comprised a round of *X-PLOR* refinement followed by manual intervention using *FRODO*. The first three stages involved simulated-annealing protocols and are summarized in Table 2. The use of low-temperature (< 1500 K) slow-cooling protocols has been reported by Saper, Bjorkman & Wiley (1991). We found that these short dynamic simulations in round two improved the *R* factor of the structure after it had appeared to converge after *POWELL* minimization. However, on the third round of refinement only a marginal improvement was seen. Simulated annealing was followed by *POWELL* minimization and restrained temperature-factor refinement.

After the first round of refinement, inspection of the electron-density maps located 100 water molecules. Water molecules were included in the model if they made reasonable hydrogen-bonding contacts and were in $> 3\sigma$ $F_o - F_c$ density; they were deleted if they failed to reappear in 1σ $2F_o - F_c$ density. Many adjustments to side-chain positions were made on the basis of the electron-density maps. The three N-terminal residues and the loops at 119–123, 149–154 and 159–164 were not in good $2F_o - F_c$ density and the $F_o - F_c$ maps in these regions were

unclear. In view of the many changes in amino-acid conformation made during manual intervention all temperature factors were set to 15 Å^2. After the second round of refinement the electron density was much clearer and it was possible to begin to model the problem loops and the N-terminus. A further 20 water molecules were added during this inspection.

One water molecule which had been included after the first stage of refinement was now at a peak of 20σ $F_o - F_c$ electron density. The temperature factor of this 'water' was 3 Å^2 (the minimum allowed). The residues around this water strongly suggested an octahedral metal-coordination site. The coordinates of this water were identical to the coordinates of a cadmium ion which had been located by a Patterson analysis of cadmium-substituted concanavalin A (Wenzierl & Kalb, 1971). The water was changed to a cadmium ion. This site had also been identified by Reeke *et al.* (1975) as the binding site of several of the heavy metal ions originally used to solve the protein structure. The metal and its symmetry equivalent are located in the cleft between the monomers which form the conventional dimer. The $F_o - F_c$ density at Arg155 and the neighbouring peptide bond between residues 150 and 151 was unclear. A further six stages of refinement were carried out; these comprised *POWELL* minimization (typically 20 cycles), temperature-factor refinement and minor adjustments to the model (addition/deletion of water molecules and/or changes in side-chain orientations). The problem with the model at residues 150 and 155 remained unresolved. The sequence of Min *et al.* (1992) became available at this stage. The change to Glu from Arg at 155 fitted the density and also allowed the neighbouring peptide bond at residues 150 and 151 to rotate by 180° giving a better fit to the density. The position of the carboxylate side chain of Asp151 was still uncertain; however, after further refinement of the structure it was fitted to $F_o - F_c$ density.

2.3. *Confirmation of the cis peptide*

The presence or absence of the *cis* peptide between Ala207 and Asp208 in the holo protein was resolved by Hardman *et al.* (1982). Whether or not it is present in demetallized concanavalin A is unclear given that there are two contradictory results (Shoham *et al.*, 1979; Reeke, Becker & Edelman, 1978). To demonstrate the presence of the *cis* peptide in this structure, residues 205 to 210 (inclusive) were omitted, 40 cycles of *POWELL* minimization were run and an $F_o - F_c$ difference electron-density map was calculated. The model is in excellent agreement with the difference density. This energetically unfavourable conformation is thought to be stabilized by an interaction with water 50 which interacts with the Oδ2 and O of Asp208 (Hardman *et al.*, 1982). Water 50 is ligated to the calcium ion and hence the continuing interest in whether the *cis* peptide is present in the demetallized structure.

2.4. *Modelling of the metal-ion sites*

The final model contains two cadmium ions and one calcium ion. The assignment of the third metal as cadmium was confirmed by an omit-map calculation. All three metals were omitted from the structure and 40 cycles of *POWELL* minimization were run to reduce model bias. The three highest peaks in the resulting $F_o - F_c$ map are shown in Table 3 and corresponded to the three sites. These results show that the *S*3 site has more difference electron density than the *S*2 site (27 *versus* 19σ). Since, the *S*2 site is known to contain a well ordered calcium ion, the *S*3 site must, therefore, contain an ion which is more electron dense than calcium. The *S*3 metal may be modelled as part calcium ion and part cadmium ion, or as a sole cadmium ion with partial occupancy or as a full occupancy cadmium ion with a high *B* factor. At this resolution it is not possible to distinguish unambiguously between these possibilities. For the well ordered *S*1 and *S*2 sites the metal temperature factors are close to the average of their ligands (Tables 7 and 8). The same is true for a full occupancy cadmium (Table 9) modelled at *S*3, suggesting that this is the best model.

2.5. *Definition of hydrogen bonds*

Hydrogen bonds were identified by applying the following criteria: a maximum distance between donor and acceptor atom of 3.5 Å and a minimum distance of 2.5 Å, a maximum deviation from linearity of *D*—H···*A* of 100°, and a maximum deviation from linearity of H···*A*—*AA* of 120°. Baker & Hubbard (1984) suggest a cutoff deviation value of 60° for *D*—H···*A* for about 90% of hydrogen bonds, but give a lower limit of about 90° deviation for weak bonds. These criteria allowed the identification of physically reasonable hydrogen bonds.

2.6. *Quality assessment of the final model*

The concanavalin A monomer is folded to form an ellipsoidal dome of approximate height 42 Å with the greatest cross section approximately 40 by 30 Å. The base of the dome is smaller with a cross section approximately 40 by 25 Å. In the final model all atoms were in at least 0.5σ $2F_o - F_c$ density. Main-chain breaks at the 1σ level of the $2F_o - F_c$ electron-density map occurred at residues 119, 120 and 160. No further water molecules could be satisfactorily identified. The highest peak in the $F_o - F_c$ map (8σ)

Table 3. *Omit-map peak heights for the three metals*

Metal	Site	σ in $F_o - F_c$ omit map	B factor (Å^2)
Cadmium	*S*1	50	14.4
Calcium	*S*2	19	11.2
Cadmium	*S*3	27	38.3

occurred near the side chain of Asp80 and would appear to be a partially occupied conformation of the side chain. Multiple side-chain conformations were also suspected for Ser21, His51 and Arg172. The multiple conformations were not modelled. The final structure contains 1809 protein atoms (237 amino acids), one calcium and two cadmium ions and 144 water molecules (*W*4 to *W*147), one of which occurs at a special position (along the twofold *c* axis). This water (*W*147) was not included in the refinement but included at half occupancy in map calculations. The final *R* factor for all data from 8 to 2 Å was 17.1% [$F > \sigma(F)$ 16.8%]. In the final structure the r.m.s. deviations from ideality were: bond lengths 0.011 Å, bond angles 2.9°, dihedral angles 27.7° and improper angles 1.2°. From a Luzzati plot (not shown) the error in the position of any one atom is estimated as ranging from 0.15 to 0.25 Å (Luzzati, 1952).

The most striking feature of concanavalin A is the two large β-sheets in the structure comprising 111 of the 237 amino acids in concanavalin A. There is no α-helix in the structure and the remaining amino acids form a series of loops and turns. The secondary structure of the protein was assigned using the algorithm of Kabsch & Sander (1984). Using the *HERA* program (Hutchinson & Thornton, 1990) a schematic representation of the secondary structure is shown in Fig. 1. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the monomer shown in Fig. 2 indicates that all non-glycine residues lie within or on the edge of the energetically allowed regions. Asp69 ($\varphi = 63°$, $\psi = 13°$) is only slightly outside this region and is well modelled as judged by electron-density maps. χ_1 is the dihedral angle about Cα and Cβ. In this structure 17% of residues have a χ_1 value of 60° (g^-), 24% of the residues have a value of 180° (t) and the remainder have the value 300° (g^+). (This analysis excludes proline, alanine and glycine residues.) The number of g^- conformations is slightly higher than predicted from previous analysis (Bhat, Sasisekharan & Vijayan, 1979; Janin, Wodak, Levitt & Maigret, 1978; McGregor, Islam & Sternberg, 1987). Lys46, 59, 135, Gln137, 166 and Glu155 have non-*trans* dihedral conformations at the bond between Cβ and Cγl (χ_2). Gln137 and Glu155 are in good ($> 2\sigma$ $F_o - F_c$) density and their environment dictates their conformation. The three lysine side chains and Gln166 are the 'best fits' to weak electron density.

3. Structural results

3.1. β-sheet structure

The average hydrogen-bond distance in the β-sheet structure is 2.8 Å (*cf.* 2.9 Å in Baker & Hubbard, 1984). The hydrogen-bonding pattern of the residues forming the β-sheets can clearly be seen in the secondary-structure diagram, Fig. 1. The larger of the sheets occurs on a face of the molecule and it is this sheet which forms the twelve-stranded anti-parallel sheet in the dimer. The sheet in the monomer is six stranded and is made up of 64 residues. The smaller β-sheet is seven stranded and consists of 47 residues and is located inside the molecule. The hydrophobic pocket sandwiched between the two sheets has been found to bind non-polar substituents (Hardman & Ainsworth, 1973). Another smaller hydrophobic pocket occurs below the smaller β-sheet behind the Ca^{2+} binding site.

3.2. *Bends, turns and loops*

These form the remainder of the secondary structure of concanavalin A and encompass 116 amino acids. A residue by residue assignment of the Reeke *et al.* (1975) structure using the algorithm of Kabsch and Sander has been reported (Kabsch & Sander, 1984). Leszczynski & Rose (1986) defined a new class of secondary structure, the omega loop (Ω) and analyzed an unrefined structure of concanavalin A (Hardman & Ainsworth, 1972). A further classifi-

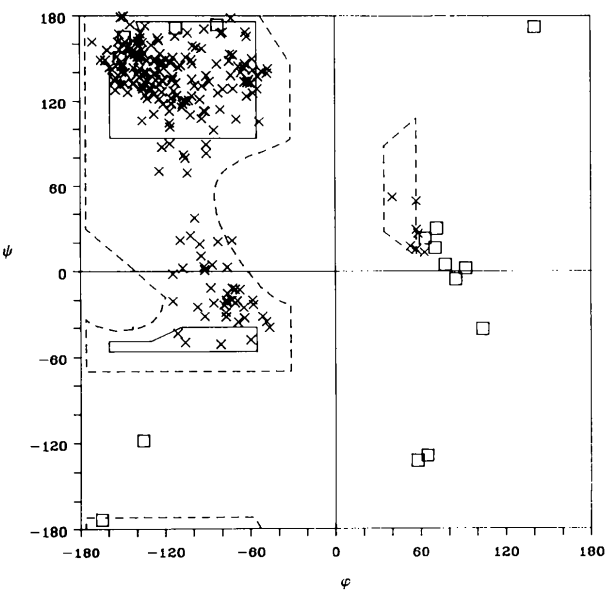

Fig. 2. A Ramachandran plot of the monomer. Glycine residues are represented by □ and non-glycine residues by × . Asp69 is the only non-glycine residue outside the allowed region.

Table 4. *Assignment of loop structure in concanavalin A*

The assignment is compared to a previous study denoted L&R (Leszcynski & Rose, 1986).

Residues	Classification	Comment
13–22	Ω	Same as L&R
40–46	Strap	New
81–86	Zeta	New
97–102	Ω	Originally 97–104 in L&R
118–122	Ω	No longer a strict L&R loop
131–135	Strap	New
135–139	Strap	New
147–156	Ω	Same as L&R
159–165	Ω	Originally 160–165 in L&R
176–185	Strap	New
201–207	Ω	No longer a strict L&R loop
225–231*	Zeta	Previously Ω 222–235 in L&R
229–237*	Ω	Same as L&R

* These form a compound loop.

cation of loops has been made by Ring, Kneller, Langridge & Cohen (1992), adding two classifications namely strap and zeta loops. Concanavalin A was not included in their analysis of known structures. Table 4 shows a classification of loops according to the criteria of Ring *et al.* (1992) in the refined structure; differences to Leszczynski & Rose (1986) are noted.

3.3. *Main-chain to side-chain hydrogen bonds*

There are 105 such interactions within the monomer and these have an average distance of 3.0 Å. One Asn/Gln side-chain to main-chain link occurs, it is formed by Asn44 Nδ2···O Ser201 and Gly45 N···Oδ1 Asn44. Such links are of special interest as these were proposed by Baker & Hubbard (1984) to provide important cross links within the protein.

3.4. *Side-chain to side-chain hydrogen bonds*

There are 34 such interactions within the monomer and these have an average distance of 3.0 Å. Several salt bridges occur within the monomer. These are Arg60 with Asp78, Lys101 with Asp203, Arg172 with Asp145, and Arg228 with Asp16.

3.5. *Interactions involving water molecules*

There are 296 water/protein hydrogen bonds with an average length of 3.0 Å. There are 104 water/water hydrogen bonds with an average length of 3.0 Å.

3.6. *Temperature factors*

The average temperature factor of main-chain atoms is 16.8 Å^2, of side-chain atoms 21.5 Å^2 and of all the atoms 19.0 Å^2. The variation of the main-chain temperature factors along the entire length of the protein is shown in Fig. 3. As expected the temperature factors are greatest for the loop regions connecting the β-strands. The loops around residues 120, 150 and 160 all have temperature factors above 30 Å^2. These loops have weak density and are all on the surface of the protein. The loop at 120 is involved in crystal contracts but may exist in other discrete conformations. Furthermore, as concanavalin A is a mixture of both unligated and ligated forms, this loop may be present only as a fraction of the protein in the crystal. As this loop is involved in crystal contacts its presence or absence may affect the crystallization and diffraction limit of concanavalin A. The temperature factors of the water molecules range from 12 to 56 Å^2 with an average of 35 Å^2.

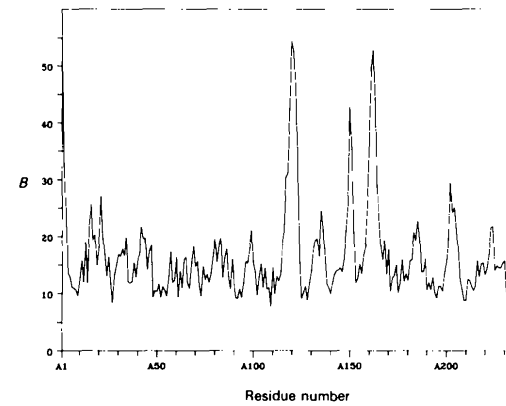

Fig. 3. Plot of the backbone temperature factor of each residue. The peaks at regions 120, 150 and 160 occur at loops on the surface of the protein.

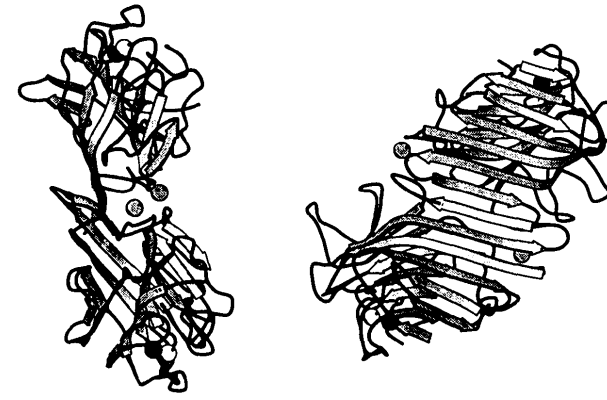

Fig. 4. Two views of the conventional dimer of concanavalin A. The large anti-parallel β-sheet is clearly visible. The *S*1 site (Cd^{2+}) is black, the *S*2 site (Ca^{2+}) is white and the *S*3 site (Cd^{2+}) is grey. The *S*3 site bridges between the monomers in the dimer. This diagram and Figs. 5–7 were produced with *MOLSCRIPT* (Kraulis, 1991).

Table 5. *Contacts which form the dimer*

Monomer 2 is generated from monomer 1 (the asymmetric unit) by applying the symmetry operation $1-x, 1-y, z$.

Monomer 1	Monomer 2	Distance (Å)
Contacts which form extended β-sheets		
N Ala125	O Met129	2.8
N His127	O His127	2.8
N Met129	O Ala125	2.9
Remaining protein/protein hydrogen-bond interactions		
Nε1 Trp88	O Asp136	3.2
Nη2 Arg90	OH Tyr176	3.4
Nδ2* Asn124	O Gln132	2.8
N* Asp131	Oδ1 Asn124	2.8
Nδ2 Asn131	O Thr123	2.8
N* Gln132	Oδ1 Asn124	3.2
Nε2 Gln132	Oγ Ser117	2.7
Nε2 Gln132	O Ser185	3.4

Hydrogen bonds through a single bridging water

Monomer 1	Water	Monomer 2	Distance (Å)
Oε1 Glu122	68	O Asn131	3.4, 3.0
O Ala125	66	O Met129	3.0, 3.1
O Asn131	126	Oε1 Glu122	3.0, 2.6
Nε2 Gln132	33	O Glu122	3.0, 3.1
Nε2 Gln137	82	Nδ1 His180	3.2, 2.8
Oδ2 Asp139	63	O Pro178	2.7, 2.7

van der Waals contacts (atoms less than 4.0 Å apart)

Monomer 1	Monomer 2	No. of contacts
Glu87	Asp136	3
Trp88	Asp136, Gln137, Lys138, Asp139	33
Arg90	Tyr176	2
Ser117	Gln132	2
His121	Asn131	3
Glu122	Asn131, Gln132	12
Thr123	Met129, Asn131	9
Asn124	Met129, Phe130, Asn131, Gln132	19
Ala125	Phe128, Met129	9
Leu126	His127, Phe175	3
His127	His127	3 + $\frac{1}{2}$
Gln132	Ser185	2
Ser134	His180, Glu183	2
Asp136	Glu183	1
Lys138	Pro178	2
Asp139	Pro178	1
Tyr176	Tyr176, Ala177, Pro178	7
Total number of unique van der Waals contacts		**114 + $\frac{1}{2}$**

* Important side-chain–main-chain interaction of the type highlighted by Baker & Hubbard (1984).

3.7. *Dimer formation*

The crystallographic twofold axis (parallel to the *c* axis) generates another monomer (monomer 2) which together with the asymmetric unit form the conventional dimer of concanavalin A. The conventional dimer contains a twelve-stranded anti-parallel β-sheet. The dimer is shown in Fig. 4. The extended β-sheet is clearly visible. The six metal ions which are bound to the dimer are also shown. There are extensive contacts involved in dimer formation and the unique contacts are listed in Table 5 (each of the contacts listed occurs twice, once for each monomer in the dimer). Hydrogen bonds were identified as described previously, van der Waals interactions are listed if the protein atoms are closer than 4.0 Å. The dimer interactions are composed of 22 protein/protein hydrogen bonds (six of which link the two β-sheets), 12 hydrogen bonds through a single bridging water, ten hydrogen bonds through multiple bridging waters and 229 van der Waals contacts (<4.0 Å). Formation of the dimer buries a total of 2600 Å^2 (or 1300 Å^2 per monomer) of surface-accessible area. The dimer has two bridging cadmium metal sites (one per monomer) each using as ligands two amino acids from one monomer and one amino acid from the other.

3.8. *Tetramer formation*

The dimer is then paired across a further twofold axis (parallel to either *a* or *b* axes) generating another dimer (monomers 3 and 4) to form a tetramer of concanavalin A. The resulting tetramer has 222 symmetry and is shown in Fig. 5. The contacts which stabilize the tetramer occur between monomer 1 and monomer 3 as well as between monomers 2 and 4. There are no direct contacts between monomer 1 and monomer 4 or between monomers 2 and 3. Monomer 3 is generated by the symmetry operator $(1-x, y, -z)$; monomer 4 by the operation $(x, 1-y, -z)$. This analysis uses the tetramer numbering scheme of Reeke *et al.* (1975). Table 6 lists the contacts involved in tetramer formation. Thus the '1–3 dimer' interaction is composed of 18 protein/protein hydrogen bonds, 20 hydrogen bonds through a single bridging water, four hydrogen bonds through multiple bridging waters and 114 van der Waals contacts. These numbers are doubled to give the total number of interactions involved in tetramer formation. Tetramer formation buries a total of 10 400 Å^2 (or 2600 Å^2 per monomer) of surface-accessible area. The extended β-sheet is partially shielded by tetramer formation.

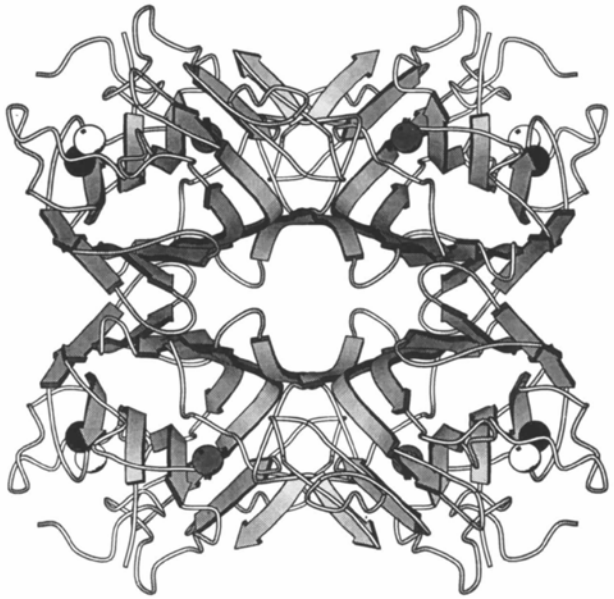

Fig. 5. The tetramer of concanavalin A viewed down a molecular twofold axis.

Table 6. *Contacts which form the tetramer*

All of these occur between monomer 1 and monomer 3. Monomer 3 is generated from monomer 1 by $1-x, y, -z$.

Protein/protein hydrogen-bond interactions

Monomer 1	Monomer 2	Distance (Å)
Nε1* Arg60	Oδ1 Asp58	3.5
Nη2 Arg60	O Asp58	3.4
Oγ Ser62	Oδ1 Asp58	2.7
N Asn69	Oδ1 Asn118	3.2
Nζ* Lys114	Oε1 Glu192	3.3
Nζ* Lys114	Oε2 Glu192	3.0
Nζ* Lys116	Oε1 Glu192	3.1
Nε2 His121	Oγ Ser108	2.8
Nε2 His121	Oδ1 Asn131	3.4

Hydrogen bonds through a single bridging water

Monomer 1	Water	Monomer 2	Distance (Å)
Nδ1 His51	64	Nζ Lys116	3.1, 2.7
Nδ1 Asn55	147†	Oδ2 Asp58	3.3, 2.5
Oδ2 Asp58	77	O Arg60	2.6, 3.1
N Asn69	87	Oγ Ser109	2.8, 3.4
O Ser117	18	Oγ Ser66	3.0, 3.1
O Asn118	104	Oγ Ser66	2.6, 3.0
Oγ1 Thr120	135	Oγ1 Thr196	3.4, 3.0
Nε2 His121	135	Oδ1 Asn131	3.0, 3.0
Nε2 His121	35	Oγ1 Thr196	3.2, 2.9
Oγ1 Thr194	64	Nζ Lys116	3.4, 2.7

van der Waals contacts (<4.0 Å)

Monomer 1	Monomer 3	No. of contacts
Thr49	Thr120	2
His51	Lys116, Val187	2
Ile53	Asn55	3
Val57	Ser62, Val64, Thr74, Ser76	1
Asp58	Arg60, Ser62, Ser76	8
Arg60	Arg60	6
Cγ1 Val64	Val187	1
Tyr67	Asn118	4
Pro68	Asn118	5
Asn69	Asn118	1
Ala70	Asn118	2
Ser72	Val187	2
Ser108	His121	4
Lys114	Glu192	1
Lys116	Glu192	5
Thr120	Thr196	1
His121	Asn131, Thr196	5
Total number of unique van der Waals contacts		**57**

* Salt bridge.

† Special position water; coordinates 44.35, 23.25, 0.00 (lies on crystallographic twofold axis).

3.9. *Additional crystal-packing interactions*

There are 28 protein/protein hydrogen bonds, 34 interactions through a single bridging water molecule and a total of 190 van der Waals contacts. These contacts affect residues in three main areas: 12–16, 182–184 and 203–204. The conformation of Trp182 with $\chi_1 = 67$, $\chi_2 = -101°$ is different from that normally found $\chi_1 = 180$, $\chi_2 = 90°$ (McGregor *et al.*, 1987). It appears to be stabilized by a π-stacking interaction with His205 of a symmetry-related molecule.

3.10. *Metal sites*

The cadmium (*S*1) site is shown in Fig. 6. It is six coordinate and octahedral. The ligand distances are shown in Table 7. Numbers in parentheses denote the values for manganese reported for the 1.75 Å structure in the paper by Hardman *et al.* (1982). The average temperature factor of the ligands is 14 Å^2 which is identical to that of the metal. No restraints were imposed on the metal temperature factors or metal–ligand bond distances. The waters which ligate the metal also make hydrogen bonds with the protein itself. *WB* is bonded to *W*9 (bonded to O Ile17) with a distance of 2.9 Å, and to Oγ Ser34 with a distance of 2.7 Å. *WA* is hydrogen bonded to O Val32 at a distance of 2.9 Å as well as to Oε1 Glu8 at a distance of 2.7 Å and to *WD* (calcium ligand) at a distance of 3.2 Å.

The calcium (*S*2) site is seven coordinate and is shown in Fig. 6. It has pseudo-octahedral geometry with Asp10 binding in a bidentate manner capping the sixth vertex of an octahedron. Table 8 lists the bond distances at the calcium with distances reported for the 1.75 Å structure shown in parentheses. The average temperature factor of the ligands is 13 Å^2, this is close to the value for the calcium ion (11 Å^2). The difference in the average distances for the two structures is again small although there are differences for individual ligands. An extensive study of calcium carboxylate interactions in small molecules has been carried out (Einspahr & Bugg, 1981). The average distance for bidentate coordination of calcium by a carboxylate was reported as 2.53 Å; however, a range of values from 2.4 to 2.8 Å was cited. Such coordination was usually symmetrical but not exclusively so. The authors reported an average distance of 2.38 Å for unidentate coordination with a range from 2.2 to 2.6 Å. The minimum distance for the second non-bonded O atom was given as 2.94 Å.

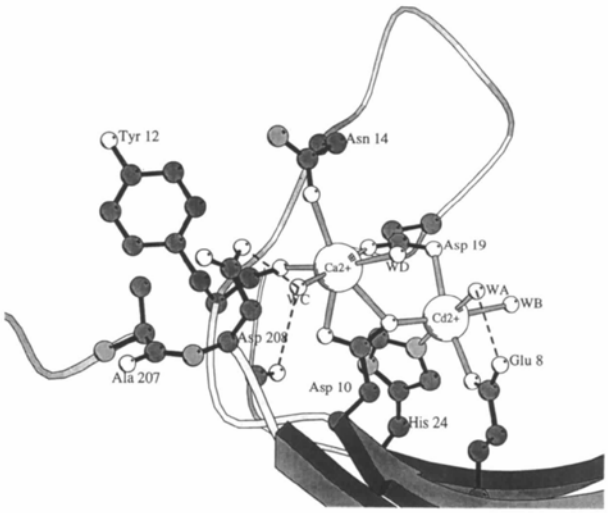

Fig. 6. The *S*1 (Ca^{2+}) and *S*2 (Cd^{2+}) metal sites. Both sites have octahedral-like geometry with Asp10 capping an axial position of the Ca^{2+} ion. Asp19 and Asp10 are bound to both metals. Asp208 is shown and it makes hydrogen bonds with *WC* which are thought to stabilize the *cis* peptide. The two metal ions are separated by 4.18 Å.

Table 7. *Metal–ligand distances at S1 (cadmium)*

The values for Mn^{2+} from Hardman *et al.* (1982) are shown in parentheses. *WA* is *W5* and *WB* is *W6*.

Residue	Atom	Distance (Å)	*B* factor (Å^2)
Glu8	Oε2	2.33 (2.31)	16
Asp10	Oδ1	2.14 (2.28)	16
Asp19	Oδ1	2.40 (2.28)	17
His24	Nε2	2.34 (2.29)	13
WA	O	2.13 (2.32)	13
WB	O	2.50 (2.20)	16
Average	—	**2.31 (2.28)**	**14**

Table 8. *Metal–ligand distances at S2 (calcium)*

The values from Hardman *et al.* (1982) are shown in parentheses. *WC* is *W50* and *WD* is *W98*.

Residue	Atom	Distance (Å)	*B* factor (Å^2)
Asp10	Oδ1	2.52 (2.46)	16
Asp10	Oδ2	2.29 (2.50)	13
Tyr12	O	2.32 (2.42)	13
Asn14	Oδ1	2.60 (2.60)	17
Asp19	Oδ2	2.30 (2.45)	12
WC	O	2.35 (2.45)	13
WD	O	2.50 (2.27)	11
Average	—	**2.41 (2.45)**	**13**

The values quoted below are all comfortably within these ranges with the exception of Asp10 which has one distance (2.29 Å) smaller than usual for a bidentate system but within the range of unidentate coordination. However, the errors in these values at this resolution preclude detailed conclusions about the interaction of Asp10 with calcium. It does seem clear, however, that in our structure the coordination of calcium by Asp10 is more asymmetric than that reported previously (Hardman *et al.*, 1982). *WC* is hydrogen bonded to Oδ2 Asp208 at a distance of 2.8 Å and to O Asp208, at a distance of 2.9 Å. *WD* is hydrogen bonded to *WA* (cadmium ligand) at a distance of 3.2 Å and to O Arg228, at a distance of 2.8 Å. The calcium (*S2*) site is 8.7 Å distant from the saccharide-binding site whose structure at 2.9 Å resolution was described by Derewenda *et al.* (1989) and recently extended to 2.0 Å (Naismith, 1992, and papers in preparation).

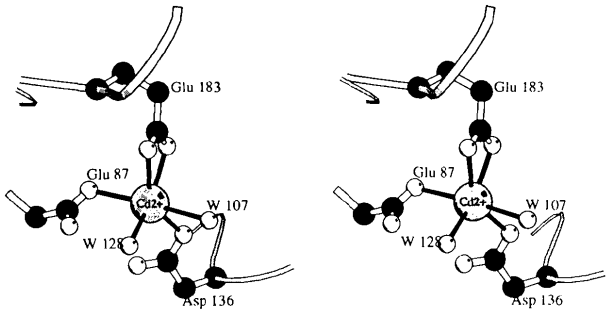

Fig. 7. The *S3* metal-binding site. The site has distorted octahedral-like geometry.

Table 9. *Metal–ligand distances at S3 (cadmium)*

Residue	Atom	Distance (Å)	*B* factor (Å^2)
Glu87	Oε1	2.49	20
Asp136*	Oδ2	2.36	44
Glu183	Oδ1	2.37	24
Glu183	Oδ2	2.49	31
*W*107	O	2.80	34
*W*128	O	2.71	30
Average	—	**2.54**	**31**
Asp136*	Oδ1	2.91	47

* Residue from a symmetry-related subunit.

The cadmium (*S3*) site shown in Fig. 7 is six coordinate and is in a distorted octahedral environment. The metal–ligand distances are shown in Table 9. The metal–ligand distances are a little longer than would be expected. However, the site is quite flexible (average temperature factor 31 Å^2). The temperature factor of the cadmium (38 Å^2) is close to the average for the ligand atoms. *W*128 is hydrogen bonded to Oδ1 Glu87, at a distance of 2.7 Å. *W*107 is hydrogen bonded to *W*88 at a distance of 2.8; *W*140 at a distance of 3.2 Å and *W*142 at a distance of 2.9 Å. Oδ1 Asp136, although not coordinated in this model, is included for completeness, as it is within the error of coordination distances.

4. Discussion

4.1. *Comparison to native concanavalin A structures*

We now make the comparison of our structure with the two concanavalin A structures already deposited in the Protein Data Bank [identification 2CNA (Hardman & Ainsworth, 1972) and 3CNA (Reeke *et al.*, 1975)]. The average main-chain r.m.s. deviation of these two structures from the cadmium-substituted concanavalin A described here is 0.9 Å (the average main-chain r.m.s. deviation between 2CNA and 3CNA is 1.1 Å). The deviation for the β-sheet residues is approximately half these values. There are too many differences in side-chain orientations to list usefully here, whereas some key differences can be highlighted. In our structure, Asp16 forms a salt link with Arg228; however, no such interaction is found in either of the other two structures. In 3CNA the side chain of Leu99 protrudes into the saccharide-binding pocket, whereas it did not in either our structure or 2CNA. The C terminus was different in the structure from our study as compared to either of the other two structures. This region was not well modelled in at least one of the unrefined structures (Reeke *et al.*, 1975) and it would appear that the side-chain and main-chain density were confused. In both 2CNA and 3CNA native concanavalin A structures, the amino acids which form the third metal-binding site had different orientations than were found in this structure of Cd concanavalin

A, precluding formation of the third metal-binding site. A native structure of concanavalin A has recently been refined at 2 Å resolution using a combination of Laue and monochromatic data (Weisgerber & Helliwell, 1993). This native concanavalin A structure is almost identical (within error) to the Cd concanavalin A structure we report here. However, the two structures are different at Asp136 and Glu183 which is where the native structure lacks the third metal site and the side chains adopt different conformations.

4.2. *Comparison to other legume lectins*

As the overall topology of our structure is unchanged, the comparison of concanavalin A with pea lectin and favin is still valid (Reeke & Becker, 1988). Their study showed that the overall topology of the two β-sheets was remarkably well conserved as was the *cis* peptide and the metal-binding sites. Since that report, lectins from *Lathyrus ochrus* (Bourne *et al.*, 1990), *Erythina corallodendron* (Shaanan, Lis & Sharon, 1991) and *Griffonia simplicifolia* (Delbaere *et al.*, 1989, 1990) have been determined. These proteins have the same overall fold as concanavalin A consisting of two β-sheets. They also contain a *cis* peptide linked to a calcium ion by a water and a transition-metal binding site. These structurally homologous lectins share considerable gene sequence homology (Reeke & Becker, 1988, and references therein). Interestingly the lectins from *E. corallodendron* and *G. simplicifolia* form the unconventional '1–3' dimer; this was attributed to glycosylation preventing the '1–2' dimer formation (*E. corallodendron*) or stabilizing the '1–3' dimer (*G. simplicifolia*).

4.3. *Metal substitution at the third site*

In the two unrefined structures from the database and the native structure of Weisgerber & Helliwell (1993) the third metal-site ligands adopt different conformations and no site is formed. This site was first found by Patterson analysis of the cadmium-substituted structure (Weinzierl & Kalb, 1971). The site can also be occupied by heavy-metal ions (Pb^{2+}, Sm^{3+}) soaked into the crystal. This was used to solve the structure (Becker *et al.*, 1975). This site has also been shown to bind trivalent lanthanide ions (Gd^{3+}, Tb^{3+}, Eu^{3+}) (Barber, Fuhr & Carver, 1975). The radii of the ions (identified so far) bound at this site vary from 1.06 (Tb^{3+}) to 1.33 Å (Pb^{2+}). This third site is non-specific in its metal requirement.

5. Concluding remarks

We have described the refined structure of saccharide-free Cd-substituted concanavalin A at 2 Å resolution. This is the first report detailing the coordination and geometry of the third metal site. The coordinates of cadmium-substituted concanavalin A have been deposited with the Protein Data Bank.*

The Carnegie Trust for the Universities of Scotland is thanked for the provision of a scholarship to JHN. The use of the *CSSR* and *SEQNET* databases (SERC, Daresbury) is gratefully acknowledged. Synchrotron radiation facilities and support were kindly provided by SERC, Daresbury Laboratory. The University of Manchester, the Hasselblab foundation and the Wellcome Trust are thanked for their support. JH and SH were supported by the SERC under the Molecular Recognition Initiative. JY was an SERC visiting fellow during the period of this work. SW was supported by the British Council and the EC Science Stimulation Plan. We thank Professor Dave Garner for his encouragement, Drs Gordon Leonard and Sue Bailey for helpful discussions, and Dr Jim Raftery for computing support.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CON, R1CONSF). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37091). A list of deposited data is given at the end of this issue.

References

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

BARBER, B. H., FUHR, B. & CARVER, J. P. (1975). *Biochemistry*, **14**, 4075–4082.

BECKER, J. W., REEKE, G. N. JR, CUNNINGHAM, B. A. & EDELMAN, G. M. (1976). *Nature (London)*, **259**, 406–409.

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., MYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BHAT, T. N., SASISEKHARAN, V. & VIJAYAN, M. (1979). *Int. J. Pept. Protein Res.* **13**, 170–184.

BOURNE, Y., ABERGEL, C., CAMBILLAU, C., FREY, M., ROUGE, P. & FONTECILLA-CAMPS, J.-C. (1990). *J. Mol. Biol.* **214**, 571–584.

BOWLES, D. J. (1990). *Annu. Rev. Biochem.* **59**, 873–907.

BRÜNGER, A. T. (1990). *X-PLOR Manual.* Version 2.1. Howard Hughes Medical Institute, Yale Univ., New Haven, USA.

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.

DELBAERE, L. T. J., VANDONSELAAR, M., PRASAD, L., QUIAL, J. W., NIKRAD, P. V., PEARLSTONE, J. R., CARPENTER, M. R., SMILLIE, L. B., SPHOR, U. & LEMIEUX, R. U. (1989). *Trans. Am. Crystallogr. Assoc.* **25**, 65–76.

DELBAERE, L. T. J., VANDONSELAAR, M., PRASAD, L., QUIAL, J. W., PEARLSTONE, J. R., CARPENTER, M. R., SMILLIE, L. B., NIKRAD, P. V., SPOHR, U. & LEMIEUX, R. U. (1990). *Can. J. Chem.* **68**, 1116–1121.

DEREWENDA, Z., YARIV, J., HELLIWELL, J. R., KALB (GILBOA), A. J., DODSON, E. J., PAPIZ, M. Z., WAN, T. & CAMPBELL, J. W. (1989). *EMBO J.* **8**, 2189–2193.
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.
EINSPAHR, H. & BUGG, C. A. (1981). *Acta Cryst.* B**37**, 1044–1052.
GREER, J., KAUFMAN, H. W. & KALB, A. J. (1970). *J. Mol. Biol.* **48**, 365–366.
HARDMAN, K. D., AGARWAL, R. C. & FREISER, M. J. (1982). *J. Mol. Biol.* **157**, 69–86.
HARDMAN, K. D. & AINSWORTH, C. F. (1972). *Biochemistry*, **11**, 4910–4919.
HARDMAN, K. D. & AINSWORTH, C. F. (1973). *Biochemistry*, **12**, 4442–4447.
HARDMAN, K. D., WOOD, M. K., SCHIFFER, M., EDMUNDSON, A. B. & AINSWORTH, C. F. (1971). *Cold Spring Harbour Symp. Quant. Biol.* **36**, 271–276.
HARROP, S. J. (1992). PhD thesis, Univ. of Manchester, England.
HELLIWELL, J. R., GREENHOUGH, T. J., CARR, P. D., RULE, S. A., MOORE, P. R., THOMPSON, A. W. & WORGAN, J. S. (1982). *J. Phys. E*, **5**, 1363–1372.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
HUTCHINSON, E. G. & THORNTON, J. M. (1990). *Proteins*, **8**, 203–212.
JANIN, J., WODAK, S., LEVITT, M. & MAIGRET, B. (1978). *J. Mol. Biol.* **125**, 357–386.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KABSCH, W. & SANDER, C. (1984). *Biopolymers*, **22**, 2577–2637.
KALB, A. J. & LEVITZKI, A. (1968). *Biochem. J.* **109**, 669–672.
KALB, A. J. & LUSTIG, S. (1968). *Biochim. Biophys. Acta*, **168**, 366–367.
KALB, A. J., YARIV, J., HELLIWELL, J. R. & PAPIZ, M. Z. (1988). *J. Cryst. Growth*, **88**, 537–540.
KRAULIS, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
LESZCZYNSKI, J. F. & ROSE, G. D. (1986). *Science*, **234**, 849–855.
LIS, H. & SHARON, N. (1986). *Annu. Rev. Biochem.* **53**, 35–67.
LUZZATI, P. V. (1952). *Acta Cryst.* **5**, 802–810.
MCGREGOR, M. J., ISLAM, S. & STERNBERG, M. J. E. (1987). *J. Mol. Biol.* **162**, 419–444.
MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
MIN, W., DUNN, A. J. & JONES, D. H. (1992). *EMBO J.* **11**, 1303–1307.
NAISMITH, J. H. (1992). PhD thesis, Univ. of Manchester, England.
RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.
REEKE, G. N. JR & BECKER, J. W. (1988). *Curr. Top. Microbiol. Immunol.* **139**, 35–58.
REEKE, G. N. JR, BECKER, J. W. & EDELMAN, G. M. (1975). *J. Biol. Chem.* **250**, 1525–1547.
REEKE, G. N. JR, BECKER, J. W. & EDELMAN, G. M. (1978). *Proc. Natl Acad. Sci. USA*, **78**, 2286–2290.
REEKE, G. N. JR, BECKER, J. W. & QUICHO, F. A. (1971). *Cold Spring Harbour Symp. Quant. Biol.* **36**, 277–284.
RING, C. S., KNELLER, D. G., LANGRIDGE, R. & COHEN, F. E. (1992). *J. Mol. Biol.* **224**, 685–699.
SAPER, M. A., BJORKMAN, P. J. & WILEY, D. C. (1991). *J. Mol. Biol.* **219**, 277–319.
SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, Cheshire, England.
SHAANAN, B., LIS, H. & SHARON, N. (1991). *Science*, **254**, 862–866.
SHOHAM, M., KALB, A. J. & PECHT, I. (1973). *Biochemistry*, **12**, 1914–1917.
SHOHAM, M., YONATH, A., SUSSMAN, J. L., MOULT, J., TRAUB, W. & KALB (GILBOA), A. J. (1979). *J. Mol. Biol.* **131**, 137–155.
SUMNER, J. B. (1919). *J. Biol. Chem.* **37**, 137–142.
WANG, J. L., CUNNINGHAM, B. A., WAXDAL, M. J. & EDELMAN, G. M. (1975). *J. Biol. Chem.* **250**, 1490–1502.
WEINZIERL, J. & KALB, A. J. (1971). *FEBS Lett.* **18**, 268–270.
WEISGERBER, S. & HELLIWELL, J. R. (1993). *J. Chem. Soc. Faraday Trans.* **89**(15), 2667–2676.
WONACOTT, A. J., DOCKERILL, S. & BRICK, P. (1980). *MOSFLM* program. Unpublished notes.