Acta Cryst. (1993). D**49**, 478–489

Determination of Three Crystal Structures of Canavalin by Molecular Replacement

BY TZU-PING KO, JOSEPH D. NG, JOHN DAY, AARON GREENWOOD AND ALEXANDER MCPHERSON

Department of Biochemistry, University of California, Riverside, California 92521, *USA*

(*Received* 12 *November* 1992; *accepted* 21 *April* 1993)

Abstract

Canavalin, the major reserve protein of the jack bean, was obtained in four different crystal forms. From the structure determined by multiple isomorphous replacement in a hexagonal unit cell, the structures of three other crystals were determined by molecular replacement. In two cases, the rhombohedral and cubic crystals, placement was facilitated by coincidence of threefold molecular symmetry with crystallographic operators. In the orthorhombic crystal the canavalin trimer was the asymmetric unit. The rhombohedral, orthorhombic and cubic crystal structures were subsequently refined using a combination of several approaches with resulting R factors of 0.194, 0.185 and 0.211 at resolutions of 2.6, 2.6 and 2.3 Å, respectively. Variation in the conformation of the molecule from crystal to crystal was small with an r.m.s. deviation in Cα positions of 0.89 Å. Packing is quite different among crystal forms but lattice interactions appear to play little role in the conformation of the molecule. Greatest variations in mean position are for those residues that also exhibit the greatest thermal motion. Crystal contacts in all crystals are mediated almost exclusively by hydrophilic side chains, and three to six intermolecular salt bridges per protein subunit are present in each case.

Introduction

Four crystal forms, seen in Fig. 1, of the seed storage protein canavalin from the jack bean, *Canavalia ensiformis*, have been described (McPherson & Rich, 1973; McPherson & Spencer, 1975; Ng, Ko & McPherson, 1993; Smith, Johnson, Andrews & McPherson, 1982; Sumner, 1919; Sumner & Howell, 1936; Sumner, Gralen & Eriksson-Quensal, 1938) and their crystal data are reviewed in Table 1. Three of the crystals have as asymmetric unit a single subunit of canavalin with $M_r \simeq 47000$. The orthorhombic form has the entire trimeric molecule of $M_r = 140000$ as its asymmetric unit. All crystals are obtained under similar crystallization conditions and all are composed of the native plant protein that has been cleaved approximately in half by the action of a protease (Ko, Ng & McPherson, 1993; Ng *et al.*, 1993; Smith *et al.*, 1982).

The molecular structure of canavalin, shown in Fig. 2, was determined by multiple isomorphous replacement (MIR) techniques and refined by crystallographic methods using the $P6_3$ crystal form (Ko *et al.*, 1993). This analysis demonstrated the structure of canavalin to be extremely similar to another reserve protein, phaseolin, the structure of which has also been determined (Lawrence *et al.*, 1990). As shown in Fig. 3, the monomer is composed of two very similar domains. Each domain consists of a 'core' β-barrel and an extended 'loop' which contains three α-helices.

Knowing the molecular structure of canavalin in one crystal form made it possible, using molecular replacement (Rossmann & Blow, 1962; Rossmann, 1990), to determine the structure of the other crystal forms as well. Refinement of a protein structure in different crystal forms allows identification of changes in details of the molecule's structure and these are often of biological significance. Such studies permit delineation of conformationally flexible regions, the binding modes of ligands, and the effects of intermolecular crystal contacts.

Furthermore canavalin is now receiving attention as a tool in the investigation of protein crystallization mechanisms because of its availability, its propensity to crystallize, its known structure and the reproducibility of its crystallization behavior. These investigations include, among others, light scattering (Kadima, McPherson, Dunn & Jurnak, 1990) and time-lapse video microscopy (Koszelak, Martin, Ng & McPherson, 1991). Canavalin is also currently a focus of research carried out on the crystallization of proteins in microgravity aboard the US Space Shuttle (Day & McPherson, 1992; DeLucas *et al.*, 1986; McPherson, Greenwood & Day, 1991). Finally, the genetic engineering of this seed storage protein is now under study for the purpose of enhancing its nutritional properties (Ng *et al.*, 1993). Thus a general understanding of its crystal forms and the molecular details responsible for their appearance is of increasing importance.

Materials and methods

The preparation of canavalin from defatted jack bean meal was described earlier (Smith *et al.*, 1982;

ISSN 0907-4449

Sumner, 1919; Sumner & Howell, 1936). The final step of the preparation involves cleavage of the native protein with trypsin (Smith *et al.*, 1982; Sumner & Howell, 1936), and its crystallization from 1.0%(*w*/*v*) NaCl at pH 6.8. All crystals analyzed here were grown from canavalin that had been recrystallized at least three times after trypsin cleavage.

The rhombohedral crystal form of canavalin is of space group *R*3 and demonstrates a very high degree of pseudo *R*32 symmetry which is readily apparent in its diffraction pattern. The unit-cell dimensions are $a = b = c = 83$ Å with $\gamma = 111.1°$. The equivalent triply centered hexagonal unit cell has dimensions of $a = b = 136.8$ and $c = 75.7$ Å. There is one trimeric molecule in the rhombohedral unit cell so that the asymmetric unit consists of a single subunit of $M_r \simeq 47000$. The pseudo dyad axes present in the diffraction pattern relate the amino- and carboxyl-terminal domains of the subunit to one another and are perpendicular to the molecular and crystallographic threefold axis (Ko *et al.*, 1993; Ng *et al.*, 1993).

The rhombohedral crystals used in the molecular-replacement analysis described here were grown by vapor-diffusion techniques on nine-well glass depression plates in sealed plastic sandwich boxes (McPherson, 1982; 1990). The 25 ml reservoirs consisted of Dulbecco's phosphate buffered saline (Gibco, Grand Island, NY) at pH 6.8. Dulbecco's saline has an NaCl concentration of 1.5%(*w*/*v*) and contains trace amounts of both Mg^{2+} and Ca^{2+}. The protein microdrops were composed of 10 µl of stock protein solution combined with 10 µl of the reservoir. The stock protein solution was 30 mg ml^{-1} of canavalin dissolved in H_2O with minimal 0.1 *M* NH_4OH. Generally the crystals required 1–2 d to grow to full size at room temperature and 3–5 d at 277 K.

As with most of the crystal forms of canavalin, the average intensity of the diffraction pattern of the rhombohedral crystals declines rapidly starting at about 3.2 Å and virtually disappears in the range 2.7–2.3 Å, though weak data is present. In addition,

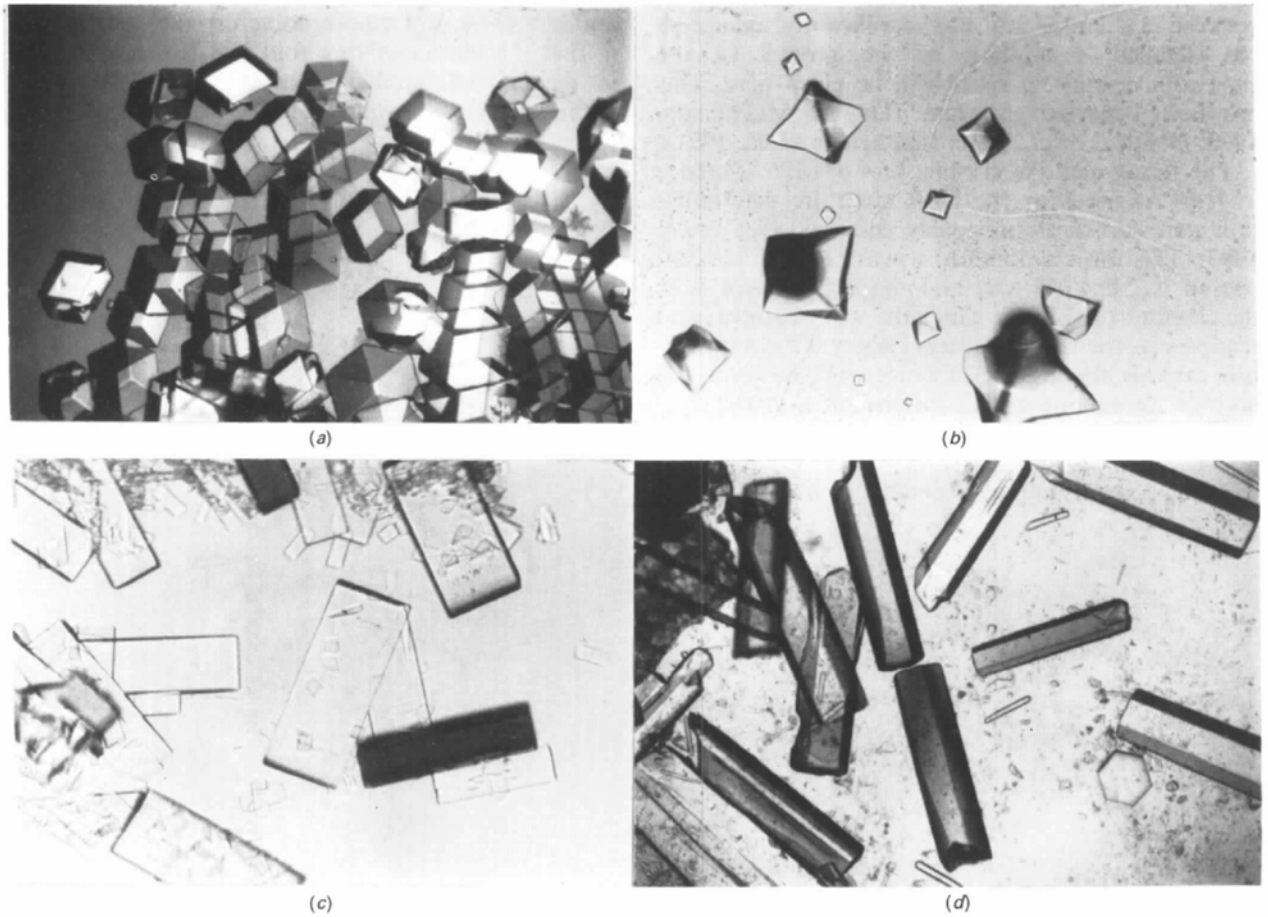


Fig. 1. Photographs of the four known crystal forms of proteolytically modified canavalin, all grown under very similar conditions. (*a*) The *R*3 rhombohedral form, (*b*) cubic crystals of canavalin having space group $P2_13$, (*c*) orthorhombic $C222_1$ crystals and (*d*) hexagonal $P6_3$ crystals.

Table 1. *Crystal data for jack bean canavalin*

Crystal form	Space group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Z	Resolution limit (Å)	V_m	Growth conditions
Hexagonal	$P6_1$	126.35	126.35	51.64	90	90	120	2	2.6	2.49	2.0 3.0% NaCl, pH 6.8, 50 m*M* phosphate buffer, 277 K
Rhombohedral	$R3$	83.0	83.0	83.0	111.1	111.1	111.1	1	2.6	2.78	1.0% NaCl, pH 6.8, 50 m*M* phosphate buffer, 295 K
Hexagonal equivalent		136.8	136.8	75.7	90	90	120	3			
Orthorhombic	$C222_1$	136.5	150.3	133.4	90	90	90	8	2.6	2.33	3.0% NaCl, pH 6.8, 50 m*M* phosphate buffer, 277 or 295 K
Cubic	$P2_13$	106.0	106.0	106.0	90	90	90	4	2.1	2.03	1.0% NaCl, pH 6.8, 50 m*M* phosphate buffer, 277 K

the decline of the diffraction pattern is anisotropic and is more pronounced in the direction of the threefold axis than in directions perpendicular to it. Because this anisotropy appears in crystals with different packing arrangements, this effect likely arises from some inherent property of the molecule.

In an attempt to overcome the problems associated with intensity decline, experiments were carried out on United States space-shuttle missions to grow canavalin crystals in a microgravity environment (McPherson *et al.*, 1991). In the case of the rhombohedral crystals, these efforts met with some limited success and the quality of the diffraction data between 3.2 and 2.3 Å was significantly enhanced. No detailed description of the growth of the canavalin crystals in space will be given here. This has been reported elsewhere (Day & McPherson, 1992; DeLucas *et al.*, 1986; McPherson *et al.*, 1991).

The hexagonal $P6_3$ crystals ($a = b = 126.35$ and $c = 51.64$ Å) used for the MIR structure determination were grown as previously described (Ko *et al.*, 1993). This form is obtained when the NaCl concentration is 2.0–4.0%(*w*/*v*) and the temperature is in the region of 277 K. Crystals will dissolve and reappear as the rhombohedral variety if the temperature exceeds about 283 K. Once grown, however, the crystals appear insensitive to changes in NaCl concentration.

Orthorhombic crystal growth is somewhat unpredictable. Crystals appear generally at 275–283 K and are promoted by salt concentrations of 3–4%(*w*/*v*) NaCl. They frequently appear and coexist in the same samples as rhombohedral or hexagonal crystals and are stable for periods of several years. The various forms of canavalin crystals often interconvert through dissolution and regrowth with orthorhombic crystals generally the final, and presumably most stable, form to appear.

Orthorhombic crystals of canavalin occasionally grow larger than 2 mm along one edge but, nevertheless, do not yield diffraction to a significantly higher resolution than other forms. The space group is $C222_1$ with cell dimensions $a = 136.5$, $b = 150.3$ and $c = 133.4$ Å. Orthorhombic crystals are unique in that the molecular threefold axis is not expressed in the crystallographic symmetry and there is an entire trimeric canavalin molecule of $M_r = 140\,000$ as

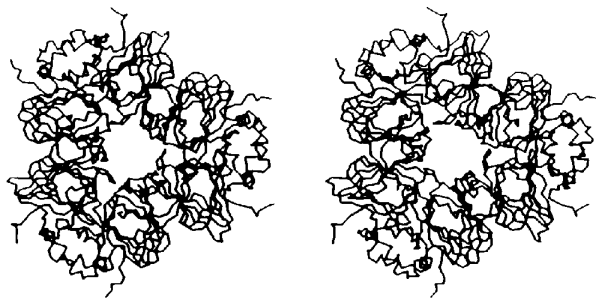


Fig. 2. A stereo diagram of the canavalin trimer based on an MIR solution in the $P6_3$ crystal (Ko *et al.*, 1993). The diameter of the large central hole is about 18 Å, the outside diameter about 86–88 Å and the thickness of the disk-shaped molecule is about 35–40 Å.

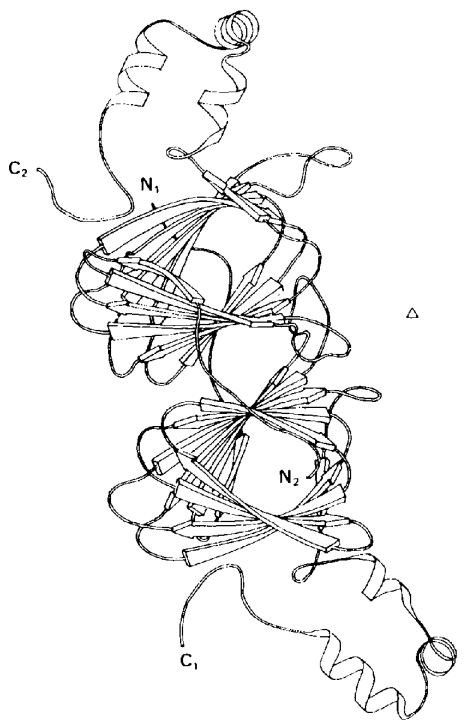
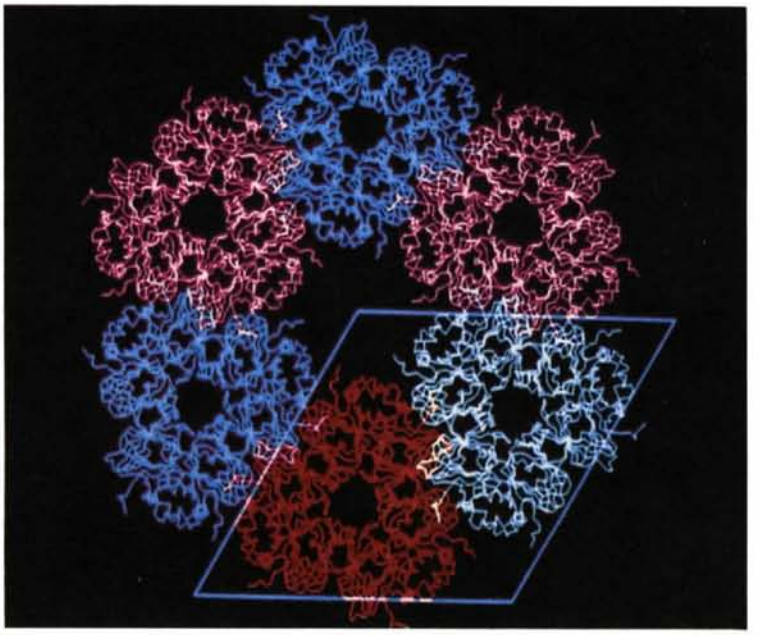


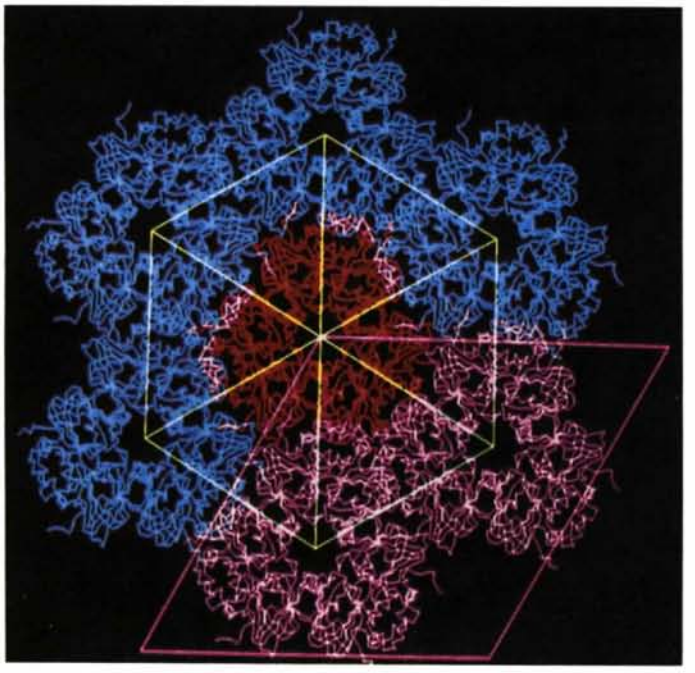
Fig. 3. A schematic drawing of the canavalin subunit. The polypeptide termini are labeled as N_1, C_1, N_2 and C_2. The triangle indicates the location of the triad axis around which three monomers are arranged.

the asymmetric unit which implies that the three canavalin subunits of $M_r = 47000$ need not be structurally identical or symmetry related.

The last crystal form is the most elusive and no reproducible crystal-growth conditions can be given with confidence. The crystals have an octahedral habit, grow to about 0.5 mm in their longest dimension, and are of cubic space group $P2_13$. The unit-cell dimensions are $a = b = c = 106$ Å and there are four canavalin trimers of $M_r = 140000$ in the unit cell or



(*a*)



(*b*)

Fig. 4. (*a*) The $P6_3$ unit cell is shown with its two canavalin trimers centered on threefold crystallographic axes. (*b*) The $R3$ unit cell of canavalin with its single oligomer centered on a threefold axis of the crystal. The pseudo-twofold axes of the molecule, which lend high pseudo-$R32$ symmetry to the diffraction pattern from this crystal, lie very close to the *a* and *b* crystallographic axes.

one subunit of $M_r = 47000$ as the asymmetric unit. The threefold axis of each molecule must coincide with a crystallographic threefold axis along a body diagonal of the unit cell. Cubic crystals are of interest because their diffraction patterns appear to extend to higher resolution than those of other forms, and we have obtained good diffraction data from these crystals to about 2.1 Å resolution. The appearance of cubic crystals seems to be very much dependent on the particular protein preparation used to grow the crystals, and when such a preparation is encountered all of the resultant crystals are cubic. The cubic crystals are not seen to coexist with or transform into others.

For data collection crystals were mounted by conventional means in quartz capillaries and intensities collected at 291 K using a two-panel multiwire area-detector system manufactured by San Diego Multiwire Systems equipped with a 2θ table (Hamlin *et al.*, 1981; Xuong, Nielson, Hamlin & Anderson, 1985). The X-ray source was a Rigaku RU-200 rotating anode operated at 45 kV and 145 mA and fitted with a Supper Co. graphite-crystal monochromator to yield Cu $K\alpha$ radiation. Crystals were rotated through 60° sectors of ω at various combinations of χ and φ angles. Frame size was 0.12 or 0.14° and counting times varied with the properties of the crystals from 60 to 150 s per frame. Data from several crystals of each kind were collected to assure high redundancy of equivalent reflections, though in the cubic case only a single crystal was employed.

Molecular-replacement procedures were those contained in the program *MERLOT* (Fitzgerald, 1988) including the rotation function of Crowther (Crowther, 1972; Crowther & Blow, 1967). Fourier calculation used the *FFT* of Ten Eyck (1985) and computer graphics were carried out with *FRODO* (Jones, 1982, 1985).

Initial refinement was with the constrained–restrained least-squares method of *CORELS* (Sussman, Holbrook, Church & Kim, 1977; Sussman, 1985) followed by the restrained least-squares approach in *TNT* (Ten Eyck, Weaver & Matthews, 1976; Tronrud, Ten Eyck & Matthews, 1987). Simulated annealing and subsequent restrained crystallographic least squares used *X-PLOR* (Brünger, Kuriyan & Karplus, 1987; Brünger, 1988, 1991). In the final stages, the structure was further refined in *TNT*. During refinement $2F_o - F_c$ maps were examined with *FRODO* to monitor progress and guide rebuilding, though this was seldom necessary.

Results

Table 2 contains statistics for data collected from the three crystal forms. Rhombohedral crystal intensities tended to fall rapidly beyond 3 Å resolution and those from orthorhombic crystals were little better, though somewhat improved in quality. The more rarely observed cubic crystals yielded reflections with quality superior to the others, and data were collected to nearly 2 Å resolution.*

Table 2. *Statistics for X-ray data collected from canavalin crystals*

	d_{min} (Å)	Average $I/\sigma(I)$	No. observed	No. of reflections	R factor
$R3$	4.27	44.54	8275	2164	0.0450
	3.39	20.43	7747	2511	0.0786
	2.96	6.11	3801	2365	0.0019
	2.69	2.89	3235	2273	0.1392
	2.50	1.80	2983	2212	0.1507
	Total	14.98	26041	11525	0.0644
$C222_1$	4.45	49.58	48791	8715	0.0432
	3.53	26.90	30734	7697	0.0505
	3.08	12.97	22411	7317	0.0741
	2.80	6.59	19874	7151	0.1209
	2.60	3.84	18236	7058	0.1744
	Total	21.31	140046	37938	0.0548
$P2_13$*	3.46	41.25	29326	10223	0.0289
	2.75	15.74	13590	7332	0.0458
	2.40	6.48	10498	6179	0.0817
	2.18	4.39	8893	5659	0.0986
	2.02	2.91	5608	3963	0.1102
	Total	18.39	67915	33356	0.0359

* Data set from the cubic crystal was collected for space group $P222$ and then merged into 21297 unique reflections with $R = 0.0308$ for 19322 equivalents.

Determinations of the orientations and positions of the canavalin model in the various unit cells were initiated well before an accurate refined structure had been derived by MIR techniques (Ko *et al.*, 1993). Thus molecular-replacement solutions were deduced using considerably less than optimal models for the protein, though every molecular-replacement solution was ultimately confirmed using the refined structure. For rotation-function searches in all crystal forms, a polyalanine model corresponding to the canavalin polypeptide was used to obviate the early uncertainty in side-chain positions. Of additional value were cross-rotation function searches using observed structure amplitudes from the hexagonal crystal, in which the orientation and position of the canavalin trimer (seen in Fig. 4*a*) was known, and those observed from a second crystal characterized by unknown disposition. While the former search often yielded several possible solutions the latter usually resolved uncertainties. Packing relationships and lattice contacts were also useful in identifying correct solutions.

* Atomic coordinates and structure factors of the $P6_3$ crystal (Reference: 1CAV, R1CAVSF), the $R3$ crystal (Reference: 1CAW, R1CAWSF) and the $C222_1$ crystal (Reference: 1CAX, R1CAXSF) have been deposited with the Protein Data Bank, Brookhaven National Laboratory. Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37084). At the request of the authors, the atomic coordinates and the structure factors will remain privileged until 1 July 1994. A list of deposited data is given at the end of this issue.

In the $R3$ crystal the molecular triad of the oligomer coincides with the crystallographic threefold axis; therefore, solution required the 'top–bottom' orientation of the molecule and its relative rotation about the polar threefold axis. A complication was that the molecule itself has pseudo 32 point-group symmetry, thus it appears structurally similar along the triad from top to bottom, or *vice versa*. In addition, the disk-shaped molecule seen on edge appears similar at numerous angles of rotation about the threefold axis.

Fig. 5(*a*) is the cross-rotation function using observed structure amplitudes from the hexagonal crystal and those of the rhombohedral crystal for $\beta = 0$ and 180°. These values for β correspond to the 'top-up' and 'bottom-up' orientation of the disk-shaped oligomer and the symmetry of the plot reflects the pseudo-molecular symmetry. A peak, common to all resolution ranges, for $\beta = 180$ and $\alpha = 37.5°$ was the maximum of the function. By symmetry this peak repeats at $\alpha = 97.5°$. Searches with a polyalanine model of the canavalin trimer, however, only showed a maximum at $\alpha = 97.5°$, with none at $\alpha = 37.5°$.

The refined model from the $P6_3$ crystal analysis, when placed in the $R3$ unit cell and subjected to rigid-body refinement using *CORELS*, in which each domain was a constrained group, yielded a residual of $R = 0.40$ and a correlation coefficient of 0.63 for all data with $F/\sigma > 3$ in the resolution range 20.0–2.6 Å. Fig. 4(*b*) illustrates the packing of canavalin trimers in the $R3$ unit cell with alignment of molecular pseudo-dyad axes nearly along the a and b crystallographic axes.

Solution of the $C222_1$ crystal was more involved because the threefold axis of canavalin (assuming it was perfectly or approximately maintained) did not correspond to crystallographic symmetry and the entire trimer of $M_r = 140000$ composed the asymmetric unit. Rotation searches were carried out both with structure amplitudes calculated from a polyalanine model of the canavalin trimer and, independently, those observed from the $P6_3$ crystal. A representative plot as a function of β angle is seen in Fig. 5(*b*) for the latter search. For both approaches a bifurcated central peak about $\beta = 90°$ appeared for all possible values of α and γ. In particular, the maximum near $\alpha = 0$ indicated the molecular triad to be close, but not parallel, to the a axis.

A self-rotation function with $\kappa = 120$ and 240° yielded results consistent with those above. Maxima appeared as four pairs of peaks about spherical angles $\varphi = 0$ and $\psi = 90°$. The relative rotation of the molecule about the oligomeric triad, γ, was more uncertain, particularly when early models were used as probes. Cross-rotation functions with the

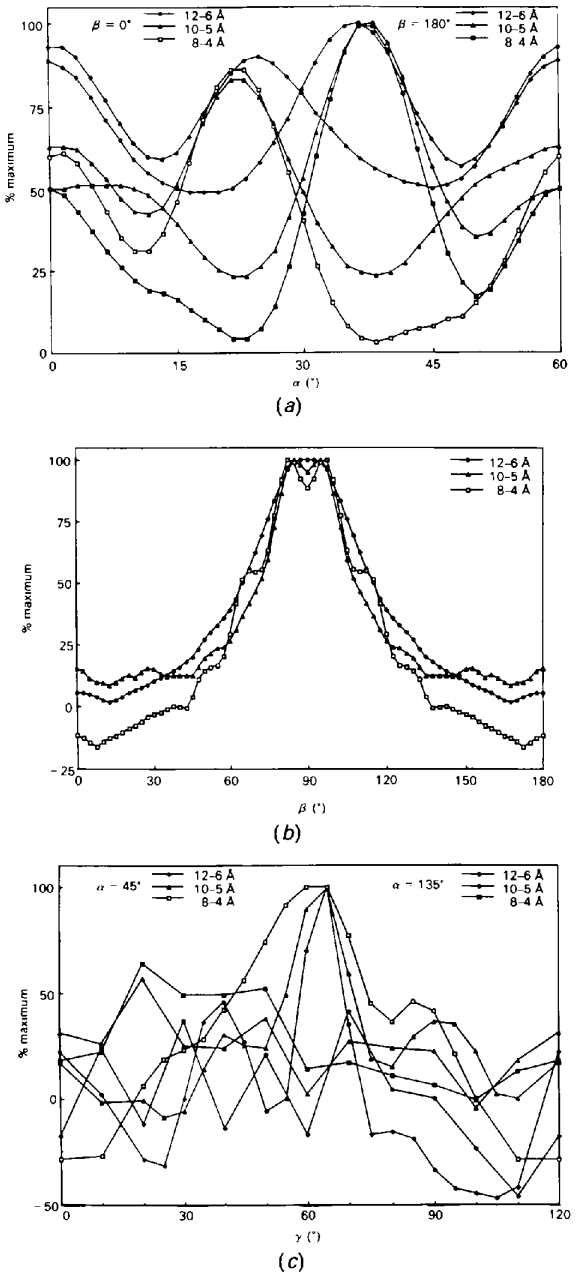


Fig. 5. (*a*) The cross-rotation function between the unknown $R3$ crystal and the $P6_3$ crystal containing the known structure and disposition of the canavalin molecule. α represents the only permissible rotation angle since the trimeric molecule must lie exactly on a crystallographic threefold axis in both crystal forms. (*b*) Shown here is the profile of a cross-rotation function for the $C222_1$ crystal using as search model a polyalanine backbone corresponding to the canavalin polypeptide. All resolution ranges and values for α and γ yield essentially the same results near $\beta = 90°$. (*c*) Cross-rotation function for the cubic $P2_13$ crystal form using as search model a polyalanine backbone corresponding to the canavalin molecule. γ is the only permissible rotation angle since the threefold axis of canavalin must lie along a body diagonal of the unit cell. The peak at 62.5° is the maximum for data from three resolution ranges for which the rotation function was computed.

Table 3. *Representative peak list of the rotation function for the $C222_1$ canavalin crystal*

Only the first two peaks in the maps are shown; symmetry-related peaks are omitted.

Resolution ranges (Å)	Model rotation Eulerian angles (°) α	β	γ	Percent maximum	Cross rotation with $P6_3$ crystal Eulerian angles (°) α	β	γ	Percent maximum
12–6	2.5	85.0	10.0	100.0	0.0	90.0	12.5	100.0
	0.0	82.5	105.0	71.7	0.0	90.0	44.2	71.1
10–5	3.3	85.0	10.0	100.0	2.5	85.0	10.0	100.0
	0.0	85.0	45.0	43.9	0.0	90.0	40.0	48.2
8–4	6.7	82.5	10.0	100.0	2.5	82.5	10.0	100.0
	5.0	80.0	40.0	67.7	5.0	85.0	59.2	76.6

Table 4. *Representative peak list of the translation function for the $C222_1$ canavalin crystal*

Translation vector* From	To	Peak position T_a	T_b	T_c	Percent maximum	Predicted center x	y	z
Molecule 2	Molecule 1	0.00	0.48	0.36	100.0	—	0.24	0.18
		0.00	0.70	0.90	51.2			
Molecule 3	Molecule 1	0.72	0.48	0.50	100.0	0.36	0.24	—
		0.20	0.98	0.50	100.0			
		0.44	0.96	0.50	50.5			
		0.94	0.46	0.50	50.5			
Molecule 4	Molecule 1	0.72	0.00	0.86	100.0	0.36	—	0.18
		0.42	0.00	0.84	49.6			

* Molecules are numbered according to the order of appearance in *International Tables for X-ray Crystallography* (1969, Vol. I).

Table 5. *Representative peak list of the translation function for the cubic canavalin crystal*

Translation vector From	To	Peak position T_a	T_b	T_c	Percent maximum	Predicted center x	y	z
Molecule 2	Molecule 1	0.54	0.06	0.50	100.0	0.02	0.03	—
		0.22	0.58	0.50	63.9			
Molecule 3	Molecule 1	0.50	0.52	0.02	100.0	—	0.01	0.01
		0.50	0.24	0.08	69.1			
Molecule 4	Molecule 1	0.04	0.50	0.56	100.0	0.02	—	0.03
		0.56	0.50	0.12	64.7			

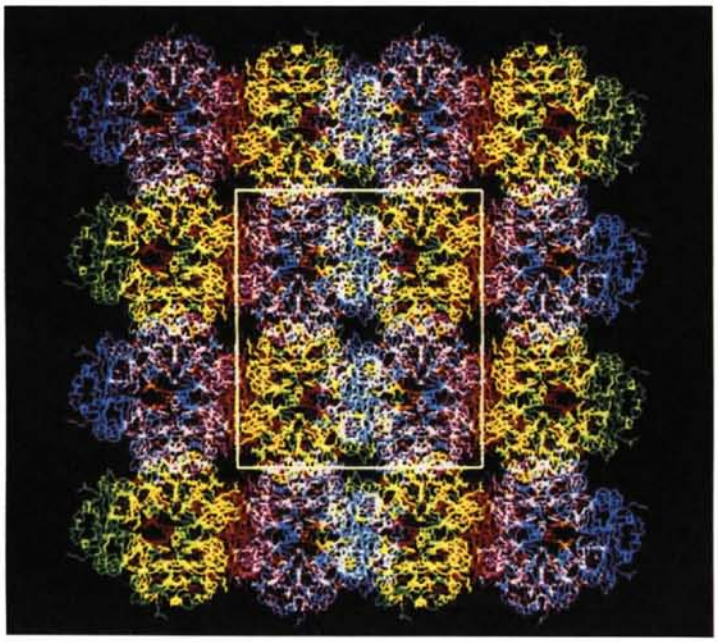
hexagonal crystal structure amplitudes, however, consistently indicated $\gamma = 10°$ to be the correct solution. As the quality of the probe model improved, this was confirmed. Table 3 shows some comparative maxima for these searches over several resolution ranges. In most cases peaks for secondary solutions were only 50–80% that of the solution at $\alpha = 4$, $\beta = 84$ and $\gamma = 10°$.

Translation functions were calculated for several resolution ranges and all were interpretable, though the range 14–6 Å was best. The solution predicted the center of the molecule to be (0.36, 0.24, 0.18) and crystallographically equivalent positions. Table 4 lists peaks for the three Harker sections of the translation-function maps showing the maxima to be self consistent. Fig. 6(*a*) shows the packing of the canavalin molecules in the $C222_1$ unit cell, the orientations of threefold axes, and the absence of unacceptable contacts. With imposition of threefold symmetry, rigid-body refinement yielded $R = 0.41$ and a correlation coefficient of 0.62 in the resolution range 20–2.6 Å, using all F_{obs} greater than 3σ.

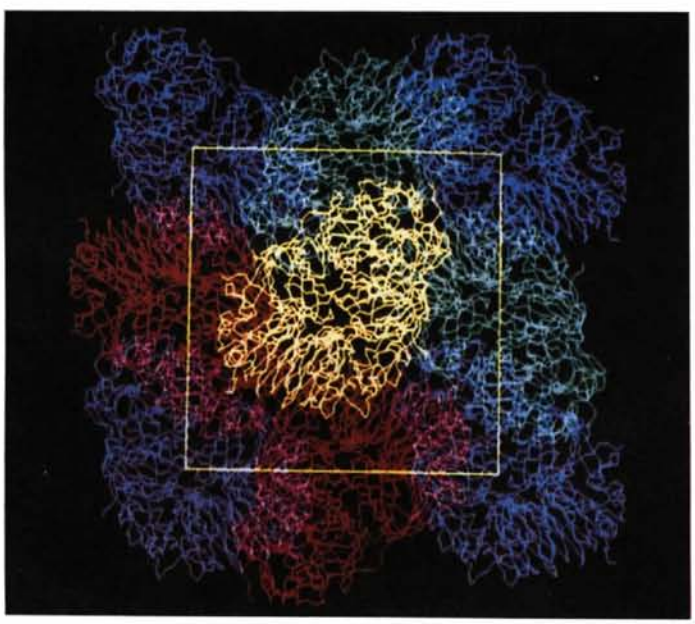
Molecular threefold axes were coincident with the body diagonals of the cubic unit cell but the relative rotation required determination. In addition, the canavalin molecule resided at an indeterminate position along the body diagonal. A molecular-replacement search in this crystal, therefore, had two degrees of freedom. Rotation functions were again calculated using a polyalanine model as well as observed structure amplitudes from the hexagonal crystal. The search was carried out as a function of γ, with $\beta = 54.74$ and $\alpha = 45$ or 135°. As shown in Fig. 5(*c*) a search with the polyalanine model yielded a single unambiguous maximum at $\gamma = 62.5$ and $\alpha = 45°$. The result with the hexagonal structure amplitudes, which exhibited maxima at γ between 0 and 5 and $\alpha = 45°$, was fully consistent with that solution.

A translation function yielded the results seen in Table 5 which were both unambiguous and self consistent. The molecular center was at (0.02, 0.02, 0.02). The canavalin trimer in the indicated disposition was examined for packing deficiencies using *FRODO* and virtually no unacceptable contacts were observed. The model was subjected to rigid-body refinement using *TNT* and this gave an R of 0.41 and a correlation coefficient of 0.60 for all data from 20 to 2.3 Å resolution with $F/\sigma > 3$. The cubic unit cell containing its complement of four canavalin trimers is seen in Fig. 6(*b*).

The crystal structures of canavalin were refined by successive application of the procedures in *CORELS*, *TNT*, *X-PLOR* and then again *TNT* as described above. The starting model in every case was that deduced from the $P6_3$ crystal by MIR techniques (Ko *et al.*, 1993), after appropriate placement and rigid-body refinement. $2F_o - F_c$ maps were calculated for all three models prior to further refinements. No significant discrepancies between the models and corresponding maps were observed, however, and little modification of the model was necessary. An exception was deletion of the N-terminal residue of the



(*a*)



(*b*)

Fig. 6. (*a*) The $C222_1$ orthorhombic unit cell is shown with the canavalin trimers that comprise its asymmetric units. The molecular threefold axes lie almost parallel with the crystallographic *a* axis, which is perpendicular to the plane. (*b*) The packing of canavalin oligomers is shown here for the $P2_13$ unit cell. The threefold axes of the trimers correspond to crystallographic triad axes lying along body diagonals. There are four oligomers in one unit cell.

Table 6. *TNT refinement of the rhombohedral R3 canavalin crystal structure*

Unit cell (hexagonal): $a = b = 136.8$, $c = 75.7$ Å, $\alpha = \beta = 90$, $\gamma = 120°$, resolution 8.0–2.6 Å, $I/\sigma \geq 3.0$, No. of reflections = 10127, native $K = 3.361$, overall $B = 0.0$, $K_{solvent} = 0.80$, $B_{solvent} = 311.6$, overall $R = 0.194$, correlation coefficient = 0.8916.

Resolution breakdown										
d_{min} (Å)	5.12	4.26	3.79	3.47	3.24	3.06	2.91	1.79	2.69	2.60
No. of F_{obs}	1205	1091	969	968	1013	1017	1029	1025	1038	772
% complete	74.0	67.0	59.5	59.4	62.2	62.4	63.2	62.9	63.7	47.4
R_{shell}	0.201	0.154	0.165	0.175	0.190	0.196	0.219	0.233	0.235	0.254
R_{sphere}	0.201	0.178	0.174	0.175	0.177	0.179	0.183	0.187	0.190	0.194

Thermal parameter distribution (total No. of atoms = 2924, average $B_{iso} = 20.88$ Å^2)							
B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	60.0	> 60.0
No. of atoms	683	894	680	344	214	65	44

Stereochemical deviation							
Category	Bond length	Bond angle	Torsional angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	2978	4012	1793	101	419	326	377
R.m.s. deviation	0.020 Å	2.911°	24.892°	0.019 Å	0.020 Å	0.098 Å	0

Table 7. *TNT refinement of the orthorhombic $C222_1$ canavalin crystal structure*

Unit cell: $a = 136.5$, $b = 150.3$, $c = 133.4$ Å, $\alpha = \beta = \gamma = 90°$, resolution 8.0–2.6 Å, $I/\sigma \geq 3.0$, No. of reflections = 27332, native $K = 7.045$, overall $B = 0.0$, $K_{solvent} = 1.0$, $B_{solvent} = 336.2$, overall $R = 0.185$, correlation coefficient = 0.9069.

Resolution breakdown										
d_{min} (Å)	5.12	4.26	3.79	3.47	3.24	3.06	2.91	2.79	2.69	2.60
No. of F_{obs}	4118	3854	3401	3177	3005	2683	2321	1987	1752	1034
% complete	100.0	93.5	82.5	77.1	72.9	65.1	56.3	48.2	42.5	25.1
R_{shell}	0.213	0.153	0.166	0.174	0.184	0.198	0.205	0.212	0.216	0.226
R_{sphere}	0.213	0.181	0.177	0.177	0.178	0.179	0.181	0.183	0.184	0.185

Thermal parameter distribution (total No. of atoms = 8745, average $B_{iso} = 21.03$ Å^2)							
B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	60.0	> 60.0
No. of atoms	2034	2665	2002	1116	571	237	120

Stereochemical deviation							
Category	Bond length	Bond angle	Torsional angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	8907	12000	5361	300	1254	850	1128
R.m.s. deviation	0.020 Å	2.987°	24.604°	0.019 Å	0.020 Å	0.015 Å	0

Table 8. *TNT refinement of the cubic $P2_13$ canavalin crystal structure*

Unit cell: $a = b = c = 106.0$ Å, $\alpha = \beta = \gamma = 90°$, resolution 8.0–2.3 Å, $I/\sigma \geq 3.0$, No. of reflections = 11968, native $K = 1.230$, overall $B = 0.0$, $K_{solvent} = 1.0$, $B_{solvent} = 237.9$, overall $R = 0.211$, correlation coefficient = 0.8823.

Resolution breakdown										
d_{min} (Å)	4.65	3.82	3.37	3.09	2.88	2.71	2.58	2.47	2.38	2.30
No. of F_{obs}	1736	1703	1636	1355	1179	1101	941	851	752	714
% complete	99.5	97.6	93.8	77.7	67.6	63.1	53.9	48.8	43.1	40.9
R_{shell}	0.232	0.171	0.189	0.210	0.223	0.235	0.233	0.227	0.236	0.252
R_{sphere}	0.232	0.201	0.198	0.200	0.203	0.205	0.207	0.208	0.209	0.211

Thermal parameter distribution (total No. of atoms = 2924, average $B_{iso} = 21.96$ Å^2)							
B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	60.0	> 60.0
No. of atoms	521	888	819	429	182	71	14

Stereochemical deviation							
Category	Bond length	Bond angle	Torsional angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	2978	4012	1793	101	419	242	377
R.m.s. deviation	0.019 Å	3.015°	24.427°	0.018 Å	0.020 Å	0.096 Å	0

second domain (Gln 245) which was observed to overlap with symmetry-related atoms.

A complication occurs with the $C222_1$ crystal where the asymmetric unit consists of an entire trimer. The molecule can be refined as identical subunits related by a perfect threefold axis, as three correlated subunits lacking exact threefold symmetry, or each of the three subunits could be treated as independent entities. The structure of canavalin in the $C222_1$ unit cell was refined according to the first (most ideal) and last (least ideal) sets of assumptions.

Table 6 presents refinement statistics for the $R3$ unit cell and Table 7 the case of least symmetry in the $C222_1$ crystal form. Table 8 contains the statistics for canavalin refined in the cubic unit cell. The refinement of all three crystal forms was quite acceptable with the models held to relatively stringent geometrical standards. Fig. 7 is a

Ramachandran plot which includes dihedral angles from all of the refined models. Most of the φ, ψ angles are within the acceptable regions (Morris, MacArthur, Hutchinson & Thornton, 1992). The few violations are located either within the amino- or carboxyl-terminal strands or in the connecting loops between β-strands and α-helices, especially the large loops between strands *A* and *B* of the first domain, and between strands *E* and *F* of the second domain (Ko *et al.*, 1993).

The *R* factor for the *R*3 crystal was 0.194 with a correlation coefficient of 0.892, comparable to that for the $P6_3$ crystal. No water molecules were included in the model and a resolution range of 8.0–2.6 Å was employed. For the $C222_1$ crystal the final *R* was 0.31 and the correlation coefficient was 0.74 to a resolution of 3.0 Å with maintenance of strict threefold symmetry in the molecule and $R = 0.185$ with a correlation coefficient of 0.907 when that constraint was relaxed. X-ray diffraction data for the cubic unit cell extended to the highest resolution and provided the most precise result. The *R* factor was 0.211 with a correlation coefficient of 0.882 at 2.3 Å. Again, no water molecules were included in the model. The Luzzati plot (Luzzati, 1952) in Fig. 8 indicates the estimated errors in the atomic coordinates were about 0.25 Å for all models of the three crystals.

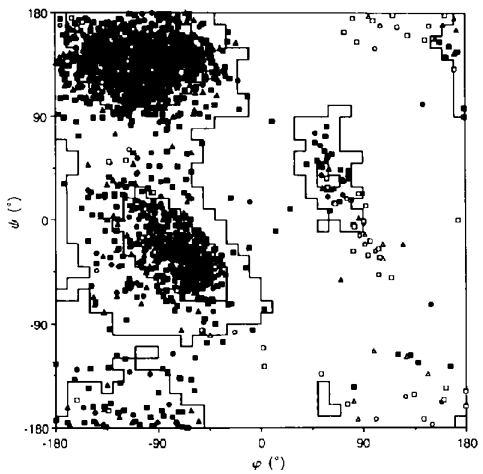


Fig. 7. A Ramachandran plot showing most φ,ψ angles to be within allowed regions according to Morris *et al.* (1992). Symbols: (▲) *R*3, non-glycine residues; (△) *R*3, glycines; (■) $C222_1$, non-glycine residues; (□) $C222_1$, glycines; (●) $P2_13$, non-glycine residues; (○) $P2_13$, glycines.

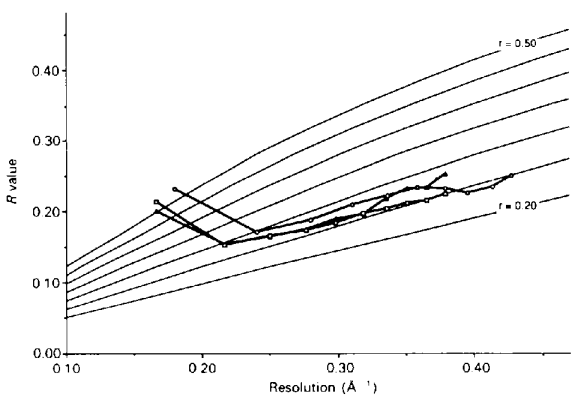


Fig. 8. Luzzati plots for the refined models. Symbols: (△) *R*3; (□) $C222_1$; (○) $P2_13$.

Table 9. *R.m.s. differences between model coordinates (Å) refined in different crystals*

Models	C$^\alpha$	Backbone*	All atoms
$P6_3$–*R*3	1.017	1.193	1.525
$P6_3$–$C222_1$†	0.424	0.519	0.637
$P6_3$–$P2_13$	0.813	0.953	1.314
*R*3–$C222_1$	0.984	1.164	1.506
*R*3–$P2_13$	1.133	1.309	1.701
$C222_1$ $P2_13$	0.747	0.879	1.245
$C222_1$ M_1‡	0.952	1.102	1.453
$C222_1$–M_2	0.872	1.019	1.386
$C222_1$–M_3	0.867	1.009	1.388
$C222_1$–combined	0.898	1.044	1.409

* Including C$^\alpha$, C$^\beta$, N, C and O, *i.e.* for polyalanine model.
† Model was refined with strict local threefold constraint.
‡ Models were refined for each of the individual subunits, the combined deviations of the three subunits are listed in the following line.

Discussion

Fig. 9 presents the variation about the mean position for main-chain atoms for the models in the different crystal forms. Overall deviations between model coordinates are described in Table 9. For the orthorhombic case, having a trimer as asymmetric unit,

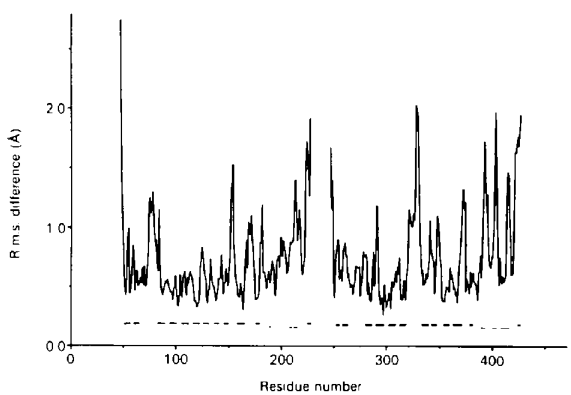


Fig. 9. Deviation between main-chain atoms of canavalin models refined in four different unit cells. The differences are presented as root-mean squares of distances between N, C$^\alpha$ and C atoms in each amino-acid residue for all possible pairs of the models. Near the bottom of the plot, heavy dashes represent ranges for the β-strands, and light dashes the α-helices.

Table 10. *Possible salt bridges involved in different crystal contacts between trimers*

		$C222_1$*		
$P6_3$	$R3$	(*a*)	(*b*)	$P2_13$
Arg 52–Asp 225	Arg 52–Glu 404	Asp 1255–Arg 1388	Arg 1052–Glu 2205	Arg 52–Glu 408
Lys 55–Asp 152	Lys 61–Asp 225	Asp 1246–Arg 3252	Arg 1182–Glu 3314	Lys 55 Glu 404
Lys 61–Asp 133	Arg 312–Glu 411	Glu 1271–Lys 3247	Glu 1214–Arg 3052	Glu 114–Lys 131
Lys 131–Glu 342		Lys 1272–Asp 3125	Glu 1329–Lys 2224	Glu 214–Arg 334/335
Lys 224–Asp 255		Arg 1277 Asp 3246	Arg 1335–Glu 2214	Asp 246–Arg 312
		Asp 2246–Arg 3376	Arg 2182–Asp 3204	Arg 252–Glu 310
		Glu 2271–Arg 3254	Glu 2314–Lys 3061	Arg 254–Glu 290
		Lys 2272–Asp 3255	Arg 2334–Glu 3342	Asp 255–Lys 264

* The numbers 1000, 2000 and 3000 were added for different subunits in a trimer. (*a*) and (*b*) represent two distinct crystal contact areas between the molecules, see text for detail.

the average of the three chains was used. The polypeptides are virtually identical with only very small local deviations. The maximum variations in position are confined principally to the four amino- and carboxyl-terminal strands and to connecting loops and extended regions.

The strands of the β-barrels essentially superimpose and are the most invariant components of the structure (Ko *et al.*, 1993) and have an r.m.s. deviation for 84 corresponding C^α coordinates of 0.78 Å. Extended α-helix-containing elements of the subunits forming interfaces within the trimer are somewhat more variable but still consistent. We would conclude that crystal-packing interactions play an almost insignificant role in determining the conformation of the molecule. The variation in polypeptide conformation throughout the molecule is closely correlated with the local temperature factors, with regions of greatest variability almost the same as those exhibiting the highest thermal parameters. Again, these are connecting loops between β-strands and elements of the extended 'wings' of the subunits.

In three of the four crystal forms (hexagonal, rhombohedral and orthorhombic) the packing of the canavalin trimers takes advantage of the flat, almost plate-like shape of the molecule. The molecules are stacked like disks and form layers through the crystals. The cubic arrangement, however, is very different and the molecules are arranged not only with large disk surfaces in contact, but with intricate edge to disk face interactions as well. The contents of the cubic unit cell are, therefore, a complex interdigitation of molecules involving far more and varied contacts than for any of the other crystal forms. As an apparent consequence of this involved packing, the cubic crystals contain less solvent than any of the other forms (see Table 1), and they diffract to a higher resolution.

The packing of canavalin in the four unit cells was analyzed to identify interactions responsible for crystallization, anticipating that common motifs might be generally observed. This was not the case. Packing of molecules is quite different for each crystal form with few contacts or interactions shared among them.

Canavalin subunits in $P6_3$ crystals are in contact with others principally about the 2_1 screw axis of the unit cell, and along the threefold axis, that is 'bottom of the disk' to 'top of the disk' contacts. There are 33 individual amino-acid residues involved, which are predominantly hydrophilic. There are five probable intermolecular salt bridges in the unit cell, per canavalin subunit, and these are indicated in Table 10. In addition, the amino-terminal strands of the carboxyl domains of each subunit interact with the *E–F* loop of the same domain of another subunit (Ko *et al.*, 1993) related by translation along *z*.

Canavalin trimers in the rhombohedral crystal assume a very stable packing arrangement where virtually all of the intermolecular contacts are at the faces of the rhombohedral unit cell. There are 26 interactions that we can identify with confidence. Again they are predominantly hydrophilic in character. There are three likely salt bridges for each canavalin subunit; these are also shown in Table 10.

Because the trimer comprises the asymmetric unit in the orthorhombic crystals, the variety of lattice contacts is substantially greater than for other crystal forms. The packing of oligomers, seen in Fig. 6(*a*), can be described as layers of trimers arranged in planes normal to the crystallographic *a* axis. Within a single layer the trimers are rather well separated and the only close association is with another molecule related by the crystallographic twofold axis. There are symmetrically related pairs of salt bridges involving Lys 224 and Asp 361, and a constellation of hydrogen bonds produced by Glu 114, Gln 115 and Asp 154. The remaining packing interactions in the $C222_1$ crystals involve contacts with molecules in the planes immediately above and below the layer in which a molecule resides. These are numerous and, again, the contacts are almost exclusively hydrophilic. Table 10 identifies 16 salt bridges participating in contacts between layers of molecules, eight for one side and eight for the other.

Although the cubic unit cell has only a canavalin subunit as its asymmetric unit, the packing is very involved and difficult to describe. Virtually all exterior surfaces of the subunits contribute to lattice interactions. There are 45 residues which we can

identify with some confidence, but others may be present as well. At least eight salt bridges participate in the crystal contacts and they are also included in Table 10.

A feature common to three crystal forms, but not the cubic, is the substantial amount of solvent in the crystals. This is about 60% by volume and may be a consequence of the hydrophilic surface of the canavalin molecules. Salt bridges have no common pattern shared by the three crystal forms, nor were invariant specific interactions between the molecules noted. In other words, there seems to be no especially preferred mechanism for interfacing one molecule with another, and the polymorphism and transformation of canavalin between different unit cells may be a result of shuffling between complementary hydrophilic residues on the surfaces.

In the hexagonal crystals there are, in addition to the 18 Å diameter channels along the threefold axes, vast channels 50 Å in diameter continuous along the 6_3 screw axes. The crystals are mechanically very sturdy, but this extensive water presence suggests an explanation for the limited resolution of the diffraction patterns. In the $R3$ and $C222_1$ crystals there are again broad channels and extensive networks of intersticies filled with solvent. These were highly visible in electron micrographs of negatively stained microcrystals (McPherson & Spencer, 1975) and are prominent features of the crystal structures.

As for the cubic crystal, judging from its very involved pattern of contacts between the trimers in the unit cell, we suspect that there is a unique combination of the interactions between surface residues which favors the compact arrangement of the protein molecules and leads to low solvent content in this peculiar crystal form. These stabilizing interactions will not be clarified until a thorough refinement of the model to a higher resolution is completed, for which further work is underway.

This research was supported by grants from NSF, the NIH and NASA. The authors gratefully acknowledge the assistance of Ms Marie Greene and Debora Felix.

Note added in proof: The model of canavalin in the cubic crystal has been further refined and contains some changes in the regions 217–220, 273–280 and 390–395. The current model, submitted to the Protein Data Bank, contains residues 44–224 for the N-terminal domain and 241–424 for the C-terminal domain. The final R factor at 2.3 Å for the $P2_13$ crystal was 0.193.

References

BRÜNGER, A. T. (1988). In *Crystallographic Computing* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & W. R. TAYLOR. Oxford: Clarendon Press.

BRÜNGER, A. T. (1991). *Annu. Rev. Phys. Chem.* **42**, 197–223.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

DAY, J. & MCPHERSON, A. (1992). *Protein Sci.* **1**, 1254–1268.

DELUCAS, L. J., SUDDATH, F. L., SNYDER, R., NAUMANN, R., BROOM, M. B., PUSEY, M., YOST, V., HERREN, B., CARTER, D., NELSON, B., MEEHAN, E. J., MCPHERSON, A. & BUGG, C. E. (1986). *J. Cryst. Growth*, **76**, 681–693.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

HAMLIN, R., CORK, C., HOWARD, A., NIELSON, C., VERNON, W., MATTHEWS, D. & XUONG, N.-H. (1981). *J. Appl. Cryst.* **14**, 85–89.

JONES, T. A. (1982). *Computational Crystallography*, edited by D. SAYRE, pp. 307–317. New York: Oxford Univ. Press.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

KADIMA, W., MCPHERSON, A., DUNN, M. F. & JURNAK, F. A. (1990). *Biophys. J.* **57**, 125–132.

KO, T.-P., NG, J. D. & MCPHERSON, A. (1993). *Plant Physiol.* **101**, 729–744.

KOSZELAK, S., MARTIN, D., NG, J. & MCPHERSON, A. (1991). *J. Cryst. Growth*, **110**, 177–181.

LAWRENCE, M. C., SUZUKI, E., VARGHESE, J. N., DAVIS, P. C., VAN DONKELAAR, A., TULLOCH, P. A. & COLMAN, P. M. (1990). *EMBO J.* **9**, 9–15.

LUZZATI, P. V. (1952). *Acta Cryst.* **5**, 802–810.

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*, ch. 4, pp. 82–159. New York: John Wiley.

MCPHERSON, A. (1990). *Eur. J. Biochem.* **198**, 1–23.

MCPHERSON, A., GREENWOOD, A. & DAY, J. (1991). *Adv. Space Res.* **11**, 343–356.

MCPHERSON, A. & RICH, A. (1973). *J. Biochem. (Tokyo)*, **74**, 155–160.

MCPHERSON, A. & SPENCER, R. (1975). *Arch. Biochem. Biophys.* **169**, 650–661.

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992). *Proteins*, **12**, 345–364.

NG, J. D., KO, T.-P. & MCPHERSON, A. (1993). *Plant Physiol.* **101**, 713–728.

ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

SMITH, S. C., JOHNSON, S., ANDREWS, J. & MCPHERSON, A. (1982). *Plant Physiol.* **70**, 1199–1209.

SUMNER, J. B. (1919). *J. Biol. Chem.* **37**, 137–142.

SUMNER, J. B., GRALEN, N. & ERIKSSON-QUENSAL, I. (1938). *J. Biol. Chem.* **125**, 45–48.

SUMNER, J. B. & HOWELL, S. F. (1936). *J. Biol. Chem.* **113**, 607–610.

SUSSMAN, J. L. (1985). *Methods Enzymol.* **115**, 271–303.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800–804.

TEN EYCK, L. F. (1985). *Methods Enzymol.* **115**, 324–337.

TEN EYCK, L. F., WEAVER, L. A. & MATTHEWS, B. W. (1976). *Acta Cryst.* A**32**, 349–350.

TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

XUONG, N.-H., NIELSON, C., HAMLIN, R. & ANDERSON, D. (1985). *J. Appl. Cryst.* **18**, 342–360.