Acta Cryst. (1994). D**50**, 813–825

Structures of Three Crystal Forms of the Sweet Protein Thaumatin

BY TZU-PING KO, JOHN DAY, AARON GREENWOOD AND ALEXANDER MCPHERSON

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

(*Received* 14 *July* 1993; *accepted* 19 *May* 1994)

Abstract

Three crystal forms of the sweet-tasting protein thaumatin from the African berry *Thaumatococcus daniellii* have been grown. These include two naturally occurring isoforms, *A* and *B*, that differ by a single amino acid, and a recombinant form of isoform *B* expressed in yeast. The crystals are of space groups *C*2 with $a = 117.7$, $b = 44.9$, $c = 38.0$ Å, and $\beta = 94.0°$, $P2_12_12_1$ with $a = 44.3$, $b = 63.7$ and $c = 72.7$ Å, and a tetragonal form $P4_12_12$ with $a = b = 58.6$ and $c = 151.8$ Å. The structures of all three crystals have been solved by molecular replacement and subsequently refined to *R* factors of 0.184 for the monoclinic at 2.6 Å, 0.165 for the orthorhombic at 1.75 Å, and 0.181 for the tetragonal, also at 1.75 Å resolution. No solvent was included in the monoclinic crystal while 123 and 105 water molecules were included in the higher resolution orthorhombic and tetragonal structures, respectively. A bound tartrate molecule was also clearly visible in the tetragonal structure. The r.m.s. deviations between molecular structures in the three crystals range from 0.6 to 0.7 Å for Cα atoms, and 1.1 to 1.3 Å for all atoms. This is comparable to the r.m.s. deviation between the three structures and the starting model. Nevertheless, several peptide loops show particularly large variations from the initial model.

Introduction

Thaumatin is a protein from the arils of the African shrub *Thaumatococcus daniellii* that produces an intense sweet sensation when it interacts with appropriate receptors on the tongue. The degree of sweet taste response is several thousandfold that produced by an equivalent weight of sucrose, making it a good candidate for non-calorific sugar substitutes. The molecule has a molecular weight of 22000, contains 207 amino-acid residues and eight disulfide bonds.

In the plant, the native protein exists as at least two gene products, referred to here as thaumatins *A* and *B*, that differ from one another only by the change of an asparagine at position 46 (isoform *A*) to a lysine (isoform *B*). Thaumatin *B* has been cloned into yeast (Lee *et al.*, 1988) and is also available in the recombinant form. It should be noted that some disagreement exists in the precise amino acid at position 113 of thaumatin *B* and, therefore, likely to be in thaumatin *A* as well. The original sequence of Iyengar *et al.* (1979) identifies residue 113 as asparagine while the sequence of the protein cloned into yeast was found to have an aspartic acid at this position (Lee *et al.*, 1988). This would, perhaps, not be a significant point except that Lee *et al.* (1988) reported that when aspartic acid is substituted by asparagine at that position, the protein is no longer sweet.

Both native isoforms *A* and *B* were crystallized by vapor diffusion from polyethylene glycol (PEG)-containing solutions (McPherson & Weickmann, 1990) and more recently, the recombinant isoform *B* was crystallized by vapor diffusion from sodium potassium tartrate solutions. The molecular structure of thaumatin *A* was solved, initially at 3.1 Å resolution, using multiple isomorphous replacement (MIR) methods (de Vos *et al.*, 1985), and subsequently refined to 1.65 Å (Ogata, Gordon, de Vos & Kim, 1992). This structure provided the search model for our own molecular-replacement efforts and the starting model for our refinement in all of the three crystal forms.

Methods

Crystal growth

The purification and characterization of the two isoforms of thaumatin has been described previously, as has the growth of the orthorhombic crystals (isoform *A*, $P2_12_12_1$, $a = 44.3$, $b = 63.7$, $c = 72.7$ Å) and the monoclinic crystals (isoform *B*, *C*2, $a = 117.7$, $b = 44.9$, $c = 38.0$ Å, $\beta = 94.0°$). Both were obtained using the vapor-diffusion approach with PEG 3350 as the precipitant (McPherson, 1982, 1990). Both crystal forms diffract to well beyond 2.0 Å resolution. The monoclinic form, however, is very frequently twinned and is somewhat difficult to work with.

A third crystal form of recombinant thaumatin, expressed in yeast (Lee *et al.*, 1988) was grown by vapor diffusion using Cryschem plastic plates at room temperature over 24–48 h. The microdrops contained 5 μl of 15 mg ml^{-1} thaumatin plus 5 μl of

Table 1. *Statistics for X-ray data collected from thaumatin crystals*

	d_{min} (Å)	Average $I/\sigma(I)$	No. obs.	No. of reflections	R factor	% Complete
$P2_12_12_1$	2.74	59.90	120875	10546	0.0544	100.0
	2.17	14.62	55349	10456	0.1172	98.8
	1.90	7.40	40745	10016	0.1738	97.3
	1.72	3.54	25781	8821	0.2332	80.9
	1.60	1.56	11337	6977	0.2201	68.2
	Total	19.24	254087	46816	0.0718	89.3
$C2$	4.46	36.99	9629	9629	0.0950	88.1
	3.54	23.50	7074	2036	0.1009	86.1
	3.09	13.41	6043	1989	0.1249	83.5
	2.81	7.57	3136	1350	0.1435	57.6
	2.61	4.84	1854	892	0.1679	38.1
	Total	19.86	27736	8350	0.1062	70.8
$P4_12_12$* (1)	4.29	64.00	15903	1846	0.0344	100.0
	3.40	39.28	9072	1781	0.0584	98.3
	2.97	25.27	7466	1765	0.0839	97.8
	2.70	16.47	6222	1702	0.1039	95.0
	2.51	11.90	4119	1393	0.1237	79.0
	Total	32.68	42782	8487	0.0544	94.6
$P4_12_12$* (2)	2.35	40.27	35403	4060	0.0569	73.8
	2.05	28.30	46149	5608	0.0700	100.0
	1.87	16.80	41033	5548	0.0957	99.5
	1.73	8.72	23803	4335	0.1329	76.0
	Total	23.18	146388	19551	0.0728	89.3

* Data sets from the tetragonal crystals were collected for low-resolution part (1) and high-resolution part (2). These data do not contain anomalous scattering information. They were both scaled to the initial F_{calc} from molecular replacement and then merged with $R = 0.17$ for 1635 equivalents.

the reservoir solution. The reservoir solution was 1 *M* sodium potassium tartrate ($KNaC_4H_4O_6.4H_2O$) containing 0.1 *M* ADA (sodium-*N*-2-acetamido-iminodiacetic acid) at pH 6.5. This crystal, similar to that grown previously from ammonium sulfate solutions (van der Wel, van Soest & Royers, 1975), is of space group $P4_12_12$ with $a = b = 58.6$ and $c = 151.8$ Å. These crystals grow as tetragonal bipyramids whose dimensions commonly exceed 1 mm and they produce observable intensities to at least 1.5 Å resolution with virtually no intensity decline with X-ray exposure over 48 h. As for the monoclinic and orthorhombic crystals, the tetragonal form has a single thaumatin molecule as the asymmetric unit. However, its solvent content is significantly higher than that of the other two forms.

Data collection

X-ray diffraction intensities were collected at 290 K from all three forms of the thaumatin crystals, mounted by conventional methods in quartz capillaries (McPherson, 1982), using a San Diego Multiwire Systems two-detector multiwire area-detector system (Hamlin *et al.*, 1981; Xuong, Nielson, Hamlin & Anderson, 1985) with a 2θ table. Frame sizes were 0.12–0.14° with count times of 60–100 s. The X-ray source was a Rigaku RU-200 rotating-anode generator fitted with a Supper graphite crystal monochromator and operated at 45 mA and 150 kV. The radiation was Cu $K\alpha$ and a Huber three-circle goniostat was used for orientation of the crystals.

Generally several specimens were used for data collection for each crystal form to yield data of high redundancy. X-ray data from an individual crystal could be recorded for 48 to 60 h with little apparent decay. Data-collection statistics are presented in Table 1. The maximum resolution to which data were collected was 1.6 Å for the orthorhombic and 1.75 Å for the tetragonal crystals, but only 2.6 Å for the monoclinic crystals.

Molecular replacement

Molecular-replacement procedures were those incorporated in the program system *MERLOT* (Fitzgerald, 1988) which included the rotation function of Crowther (Crowther, 1972; Crowther & Blow, 1967). Fourier calculations used the *FFT* of Ten Eyck (1985) and images were created and analyzed using the programs *FRODO* (Jones, 1982, 1985) and *RIBBONS* (Carson & Bugg, 1986) running on Evans and Sutherland PS390 graphics systems and an SGI 320 VGX workstation, respectively. All other calculations were performed on the UCR campus VAX network and the SGI 320 system.

At the time that we carried out our molecular-replacement attempts the refined thaumatin structure (Ogata *et al.*, 1992) had not yet appeared, and only C_α coordinates for the 3.1 Å MIR structure (de Vos *et al.*, 1985) were available from the Brookhaven Protein Data Bank (Abola, Bernstein, Bryant, Koetzle & Weng, 1987; Bernstein *et al.*, 1977). We were able, however, to construct a polyalanine chain using the C_α positions as guide points that closely resembled the thaumatin structure. This polyalanine model, as our subsequent results demonstrated, proved adequate in providing correct and unambiguous molecular-replacement solutions for all of the three crystal forms.

Based on the orientations and positions obtained from the polyalanine model, the refined thaumatin structure, when it became available, was subsequently placed in the three different unit cells. In every case rigid-body refinement of this latter structure in each of the three crystals yielded R values in the range 0.28–0.32, thereby confirming the earlier molecular-replacement results.

Crystallographic refinement

Initial refinement of the polyalanine thaumatin model was carried out in each unit cell using the program *CORELS* (Sussman, Holbrook, Church & Kim, 1977; Sussman, 1985). Subsequent refinement of the complete model employed simulated annealing by *X-PLOR* (Brünger, Kuriyan & Karplus, 1987;

Brünger, 1991) and the restrained least-squares approach in *TNT* (Ten Eyck, Weaver & Matthews, 1976; Tronrud, Ten Eyck & Matthews, 1987). $2F_o - F_c$ and $F_o - F_c$ Fourier syntheses were used to improve the models, monitor changes, and identify solvent molecules, again using *FRODO*.

Results

The polyalanine model of thaumatin comprising 1035 non-H atoms in 207 residues was placed with its center of mass at the origin of an orthogonal $P1$ unit cell having $a = b = c = 100$ Å and structure factors were calculated to 3.7 Å resolution. Spherical harmonics were calculated in *MERLOT* for resolution ranges 12–6, 10–5 and 8–4 Å and these used for rotation-function searches for all three crystal forms.*

Figs. 1(*a*)–1(*c*) are plots of the rotation-function value against the search angle β for the maximum in each section. Unambiguous solutions appeared in the rotation function for the orthorhombic crystal ($\alpha = 100$, $\beta = 65$, $\gamma = 10°$) and the monoclinic crystal ($\alpha = 147$, $\beta = 99$, $\gamma = 102°$). A few possible solutions were also found for the tetragonal crystal but they were considerably less certain than for the other two crystals.

Translation-function searches using *MERLOT* based on the rotation-function solutions for the orthorhombic and monoclinic crystals, as shown in Tables 2 and 3, produced unambiguous solutions, with the function maxima appearing as self consistent sets on the Harker sections. The translation function using the best tetragonal rotation function, however, was unconvincing.

Translation-function solutions for the $P2_12_12_1$ and the $C2$ crystals were verified by calculating R-value maps for the 500 strongest reflections. In Fig. 2(*a*) are shown the minimal values in the map sections for the orthorhombic crystals and in Fig. 2(*b*) part of the map section for the monoclinic crystal. Further rigid-body refinement at 6 Å with *CORELS* using the polyalanine model yielded R values and correlation coefficients (CC) of $R = 0.46$, CC = 0.52 for the $P2_12_12_1$ crystals, and $R = 0.47$, CC = 0.49 for the $C2$ crystals. While maintaining strong restraints on the peptide geometry, the polyalanine model was refined by *CORELS* to $R = 0.40$, CC = 0.51, and $R = 0.36$, CC = 0.54, in the orthorhombic and the monoclinic unit cells, respectively, both at 3.2 Å resolution.

In earlier molecular-replacement studies of other proteins (Ko, 1992; Ko, Ng, Day, Greenwood & McPherson, 1993) we observed that cross-rotation functions between two different native crystal forms of the same protein molecules using the observed diffraction intensities could be very useful in determining the relative orientations of the molecules

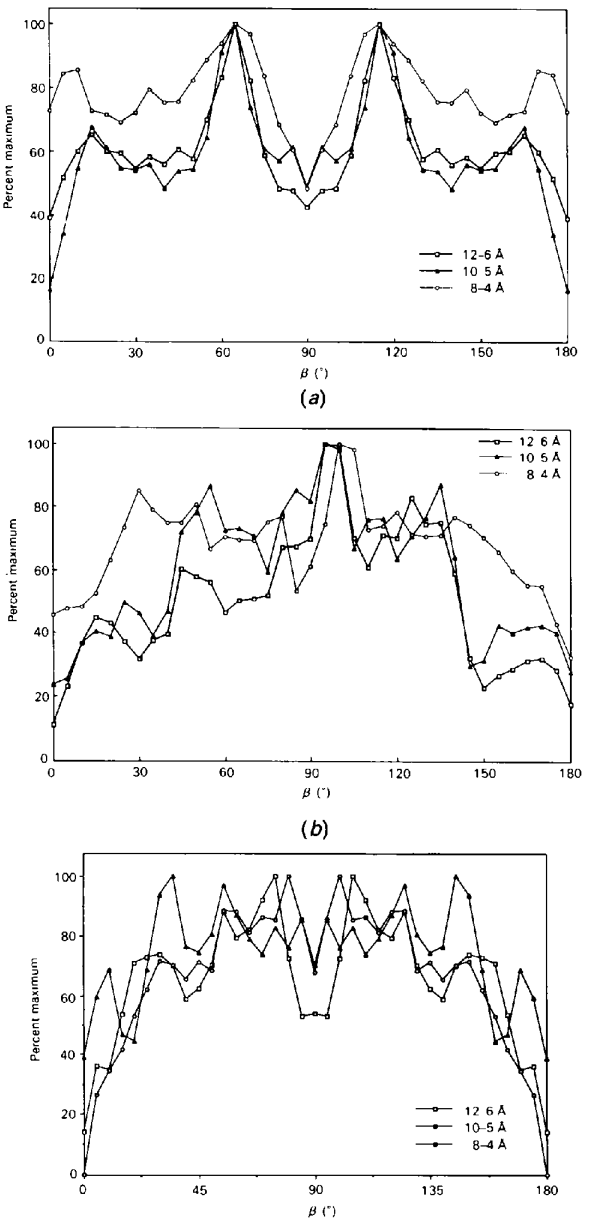


Fig. 1. Profiles of cross-rotation functions of the thaumatin polyalanine model with the three native crystals (*a*) $P2_12_12_1$, (*b*) $C2$, (*c*) $P4_12_12$. In (*a*) and (*b*), different resolution ranges yielded maxima with consistent β-angles. In (*c*) the plot exhibits multiple possibilities for a global maximum.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1THU, 1THV, 1THW, R1THUSF, R1THVSF and R1THWSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. (Reference: GR0271). At the request of the authors the atomic coordinates and structure factors will remain privileged until 1 July 1995.

Table 2. *Representative peak list of the translation function for the polyalanine thaumatin model in the $P2_12_12_1$ unit cell*

Translation vector		Peak position			Percent	Predicted center		
From	To	*Ta*	*Tb*	*Tc*	maximum	*x*	*y*	*z*
Molecule 2	Molecule 1	0.52	0.16	0.50	100.0	0.01	0.08	—
		0.92	0.64	0.50	86.5			
		0.86	0.92	0.50	83.3			
Molecule 3	Molecule 1	0.50	0.66	0.60	100.0	—	0.08	0.30
		0.50	0.76	0.70	72.6			
Molecule 4	Molecule 1	0.02	0.50	0.12	100.0	0.01	—	0.31
		1.00	0.50	0.12	94.1			
		0.50	0.50	0.10	78.8			

Table 3. *Representative peak list of the translation function for the polyalanine thaumatin model in the C2 unit cell*

Translation vector		Peak position			Percent	Predicted center		
From	To	*Ta*	*Tb*	*Tc*	maximum	*x*	*y*	*z*
Molecule 2	Molecule 1	0.70	0.00	0.64	100.0	0.35	—	0.32
		0.20	0.50	0.64	100.0			
		0.14	0.00	0.50	70.4			
		0.64	0.50	0.50	70.3			

in the two unit cells. This is presumably true because the procedure obviates the effects of errors in a search model. Accordingly, we calculated rotation-function maps between the orthorhombic crystal, in which the dispositions of the thaumatin molecules

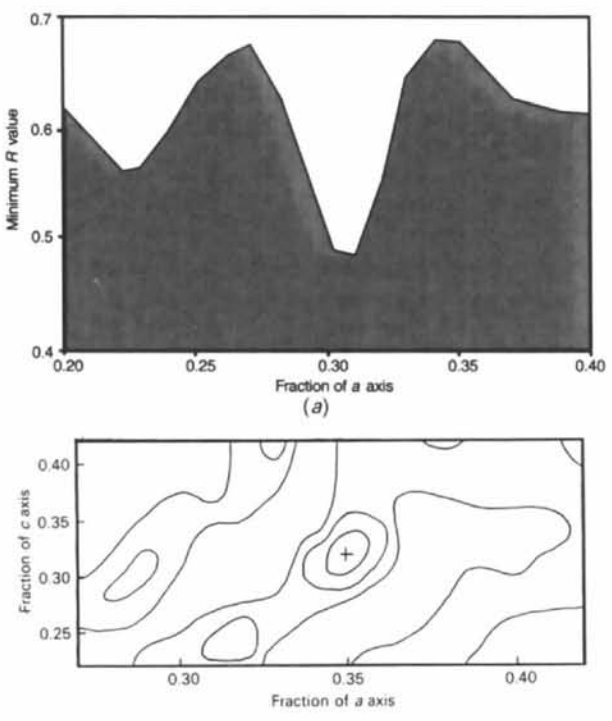


Fig. 2. (*a*) *R*-value profile for the orthorhombic thaumatin crystal showing the minimum value in each of the *z* sections. (*b*) Portion of the *R*-value map section for the monoclinic crystal. The global minimum is marked by a cross. Maps were calculated for the 500 strongest reflections at 6 Å resolution using a polyalanine search model. The grid spacing was 0.01 in each axial direction of the unit cell.

Table 4. *Representative peak list of the cross-rotation function of the orthorhombic and the tetragonal thaumatin crystals*

Resolution	Eulerian angles (°)			Percent
ranges (Å)	α	β	γ	maximum
12–6	45.0	30.0	0.0	100.0
	52.5	80.0	17.5	91.6
	55.0	82.5	17.5	90.2
	62.5	87.5	42.4	79.3
10–5	25.0	87.5	45.0	100.0
	65.0	90.0	45.0	92.1
	42.5	52.5	20.0	72.4
	43.7	67.5	35.0	67.4
8–4	22.5	85.0	45.0	100.0
	66.3	70.0	35.0	78.3
	47.5	70.0	7.5	66.2
	70.0	90.0	45.0	63.7

were known, and the tetragonal crystal under investigation. Table 4 is a list of cross-rotation function maxima for the three resolution ranges. A peak at $\alpha = 24$, $\beta = 86$ and $\gamma = 45°$ was present in the two searches using higher resolution data but did not appear strong in the 12–6 Å range.

Matrices for the Eularian angles derived from the cross-rotation function (M_2), and for all probable solutions obtained from the rotation-function search using the polyalanine model (M_3) were calculated and analyzed for agreement by applying the known operation [M_1 ($\alpha = 100$, $\beta = 65$, $\gamma = 10°$)] required to place the polyalanine model in the orthorhombic unit cell. That is, they were tested for the condition that $M_3 = M_2 \times M_1$. The results were quite clear in this instance; the peak at $\alpha = 24$, $\beta = 86$ and $\gamma = 45°$ in Table 4 (after symmetry transformation to $\alpha = 66$, $\beta = 86$ and $\gamma = 135°$) corresponds to a matrix M_2 that converts M_1 into a matrix corresponding to angles $\alpha = 1$, $\beta = 56$ and $\gamma = 270°$, which was present as a secondary peak in the original rotation-function search using the polyalanine model.

The translation-function search using this solution again failed, however, to yield a convincing solution. This procedure was complicated as well by the space-group ambiguity for the tetragonal crystal which could have been either $P4_12_12$ or $P4_32_12$. R-value maps were computed using the 472 strongest reflections between 12 and 6 Å resolution for both space groups for all possible translation-function solutions, *i.e.* for 1/8 unit-cell volume. The results are summarized in Fig. 3. A global minimum of $R = 0.50$ was found for $P4_12_12$ at (0.21, 0.41, 0.22) which upon *CORELS* rigid-body refinement of the polyalanine model yielded $R = 0.46$ and a correlation coefficient of 0.50, at 6 Å resolution.

Again, with strong geometry restraints, the crude model was refined by *CORELS* in the tetrgonal unit cell to an $R = 0.42$ and CC = 0.47 at 3.2 Å resolution. Difference maps allowed us to identify a few of the side chains. Gradual improvement was attempted on the model by the inclusion of side chains at probable positions, but difficult situations were encountered, especially in the large loop domain between residues 130 and 180. The R values at this stage were about 0.32 at 2.6 Å resolution. Similar results were obtained from the other two crystal forms.

When the entire thaumatin model, including both alternative conformers and all solvent molecules, was placed according to the orientations and positions deduced as above and again subjected to only rigid-body refinement, the resultant R value for the monoclinic crystal was $R = 0.304$ at 2.6 Å, for the orthorhombic $R = 0.287$ at 1.75 Å, and for the tetragonal unit cell $R = 0.317$, also at 1.75 Å. The correlation coefficients were 0.738, 0.840 and 0.822, respectively. Although the data were collected to 1.6 Å for the orthorhombic crystal form, the number of reflections

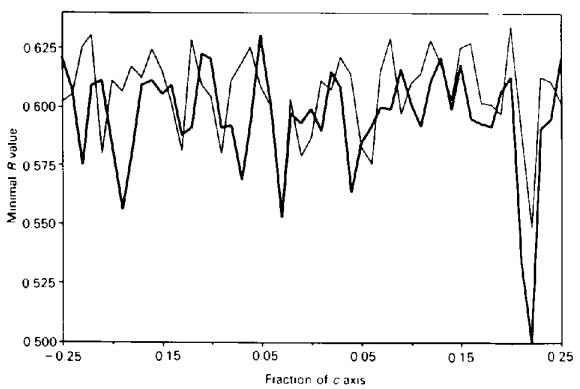


Fig. 3. R-value profiles for the tetragonal thaumatin crystal. Maps were calculated for the 472 strongest reflections between 12 and 6 Å, using a polyalanine model, in the $P4_12_12$ (thick line) and $P4_32_12$ (thin line) unit cell. The grid spacing for this calculation was 0.01 Å in each axial direction. The global minimum confirmed the space group of this crystal form as $P4_12_12$.

declined rapidly beyond 1.8 Å. Therefore, only 1.75 Å data were used in the refinement. The model was then checked thoroughly against $2F_o - F_c$, $F_o - F_c$ difference maps and omit maps with systematic deletions of up to ten consecutive residues.

For the $C2$ crystal, because the difference Fourier maps were produced only to 2.6 Å resolution, limited detail regarding side-chain conformations could be discerned. However, a significant difference between the original model and the map was observed for the segment containing residues 116–120. As shown in Fig. 4(*a*), the altered conformation could be inferred directly from the map. In addition, shifts in the loop region including residues 151–152 and the C terminus were also clear. The conformation of Cys159 and Cys164 was nearer the Alternative-1 conformer of the original model. Although the density was weak for the C terminus, a possible rotation, and change in the dihedral angles of Thr206 could be inferred. Instead of forming salt bridges with Arg8 and Arg125, the terminal carboxyl group is now interacting with two lysines from neighboring molecule (see also *Discussion* below). These two arginines interact with a nearby Asn113, which may possibly be an Asp. Another change in the sequence is a lysine replacing the asparagine at residue 46. However, this was not very well defined in the difference map.

For the $P2_12_12_1$ crystals, the higher resolution difference maps at 1.75 Å revealed more detail. Residue 46 is definitely an Asn, as in the original model. Based on the electron density alone, however, we can not identify residue 113 as asparagine or aspartic acid. For the loop region 116–120, seen in Fig. 4(*b*), similar changes as for the $C2$ structure were observed. Density near residues 160–163 was weak and indicated shifts in the polypeptide chain. The disulfide of Cys159–Cys164 had a similar conformation to that of Alternative-1 in the original model. In addition, several changes of side-chain conformations were clearly indicated by the difference maps. For other residues with alternative conformers, the density usually strongly favored one over the other. In some cases (*e.g.* Arg29 and Glu168) density was pronounced and unambiguous, but resembled neither of the original alternatives. Solvent molecules were added conservatively, only for the strongest densities. There were ultimately 123 waters in our model, compared with 236 in the original. Some sites are common while many others are different.

The 1.75 Å map of the $P4_12_12$ crystal provided additional structural detail of thaumatin B. Residue 46 was confirmed to be lysine, consistent with the DNA sequence. Again, we could not discriminate residue 113 as Asn or Asp, but according to the cloned gene sequence it was assumed to be Asp. The difference map, shown in Fig. 4(*c*), also indicated

changes in the loop 116–120, consistent with those indicated by the other two crystals. There were some adjustments in main-chain conformation for residues 156 and 157, as well as the C terminus. The peptide linking Pro205 and Thr206 was inverted and the dihedral angles of Thr206 were also modified. In this crystal form, the terminal carboxyl group still forms a salt bridge with Arg8, but not with Arg125. Cys159 and Cys164 assume conformations similar to Alternative-2 in the original model. Other changes are confined principally to side-chain rotations. There are 105 water molecules and an L-tartrate which we could identify with certainty. Coordinates for the tartrate molecule were obtained from the mirror image of the stereoisomer D-tartaric acid as determined by Okaya, Stemple & Kay (1966). It was fitted directly to the density, shown in Fig. 5, without any further manipulation.

Following application of the above changes in the models using *FRODO*, as guided by difference Fourier maps, the polypeptide geometry was regularized with an overall deviation of less than 5% from ideal values. Several cycles of refinement were then carried out using *TNT*, additional difference Fourier maps calculated for each crystal form, and the model examined again using *FRODO*. Only minor additional modifications were necessary. These were made in regions around loop 60–61 and to the three consecutive glycines 142–144. Both regions are likely to be very flexible.

Subsequent refinement employed the simulated-annealing procedure of *X-PLOR* with a starting temperature of 3000 K. This yielded R values of 0.241, 0.255 and 0.270 for the $C2$, $P2_12_12_1$ and $P4_12_12$ crystals, respectively. It was followed by several cycles of conjugate-gradient refinement and

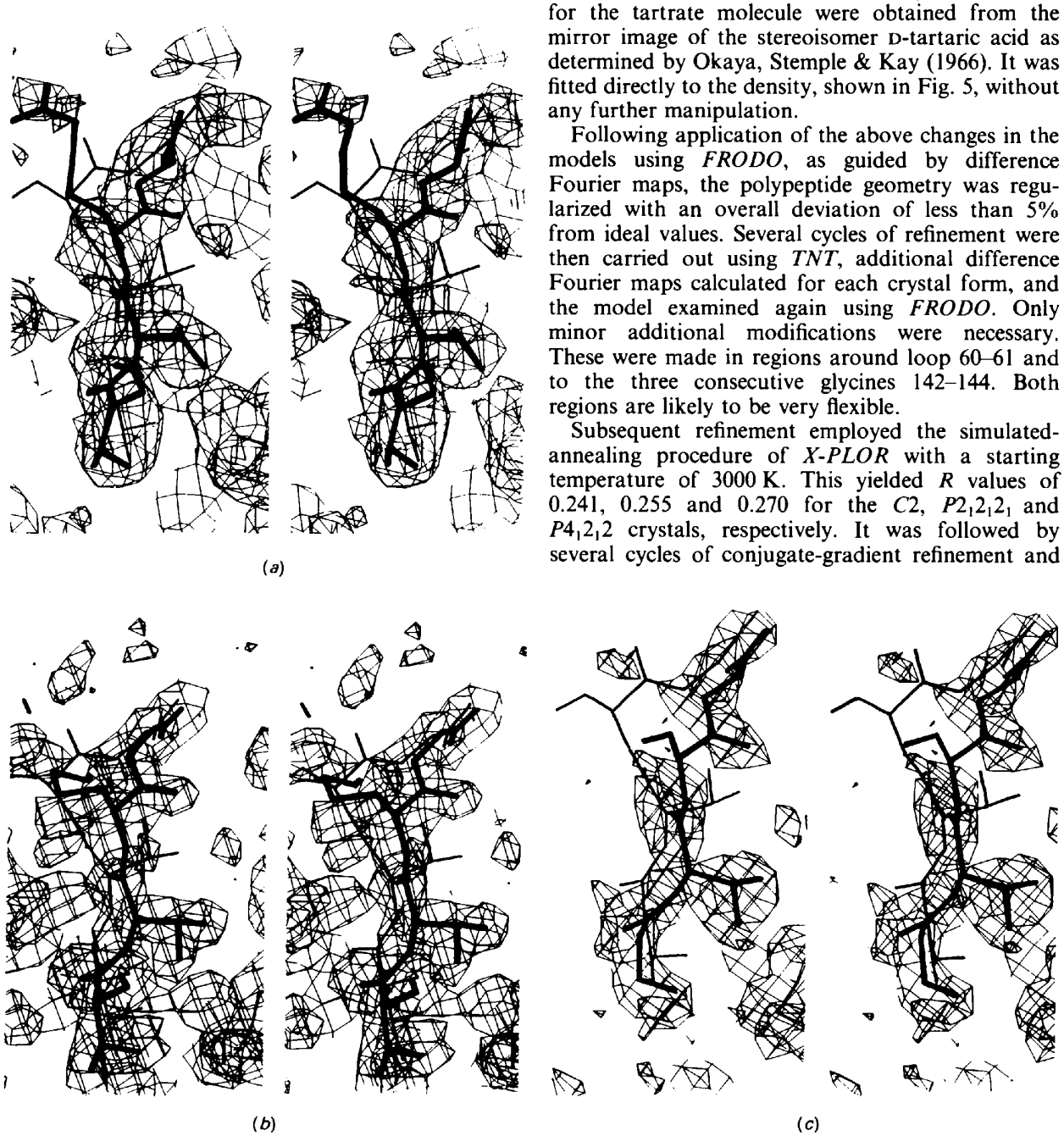


Fig. 4. Stereo diagrams showing the altered structure of the loop containing residues 116–120 in the (*a*) $C2$, (*b*) $P2_12_12_1$ and (*c*) $P4_12_12$ thaumatin crystals. The corresponding $2F_o - F_c$ electron-density maps calculated using phases from initial rigid-body refinement of the original model (thin lines) indicate the changes in the structures reported here (heavy lines). The ranges of data used were 20–2.6 Å for $C2$, 20–1.75 Å for $P2_12_12_1$ and $P4_12_12$. R values were 0.304, 0.287 and 0.317, respectively. The omit maps for the same region were almost identical.

Table 5. *TNT refinement of the monoclinic C2 thaumatin crystal structure*

Unit cell $a = 117.7$, $b = 44.9$, $c = 38.0$ Å, $\alpha = 90.0$, $\beta = 94.0$, $\gamma = 90.0°$; resolution 8.0–2.6 Å; $I/\sigma \geq 3.0$; No. of reflections = 4622; native $K = 1.94$; overall $B = 0.0$ Å^2; $K_{solvent} = 0.884$; $B_{solvent} = 143.7$ Å^2; overall $R = 0.184$; correlation coefficient = 0.873.

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}	564	554	558	552	557	544	400	367	338	188
% Complete	92.5	90.9	91.5	90.5	91.4	89.2	65.6	60.2	55.4	30.8
R_{shell}	0.211	0.150	0.169	0.168	0.179	0.192	0.190	0.218	0.226	0.218
R_{sphere}	0.211	0.178	0.175	0.174	0.175	0.177	0.178	0.181	0.183	0.184

Thermal-parameter distribution

Total No. of atoms 1552 Average $B_{iso} = 17.76$ Å^2

B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	> 50.0
No. of atoms	173	851	438	84	6	0

Stereochemical deviation

Category	Bond length	Bond angle	Torsion angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	1596	2155	947	32	240	93	211
R.m.s. deviation (Å,)	0.020	2.954	21.81	0.017	0.020	0.104	0

finally concluded with *TNT*, in which the individual temperature factors of atoms were also refined.

Table 5 presents refinement statistics for the *C*2 unit cell, Tables 6 and 7 those for the $P2_12_12_1$ and $P4_12_12$ unit cells, respectively. The final geometry of each of the models was quite acceptable reflecting the stringent restraints imposed during refinement. Fig. 6 is a Ramachandran plot which shows dihedral angles from all of the refined models. For the orthorhombic and the tetragonal crystals, nearly all of the φ, ψ angles are contained in the allowed regions (Ramakrishnan & Ramachandran, 1965). The violating and marginal residues are located in the loops. These include Asp25 with φ, ψ of about 60 and $-140°$. For the monoclinic crystal, most of the residues are within the allowed region, but quite a few outlying residues were also present, probably reflecting the result of the limited diffraction data set and, therefore, reduced constraint on geometrical parameters.

The *R* factor for the *C*2 crystal was 0.184 with a correlation coefficient of 0.873. No water molecules were included in the model and a resolution range 8.0–2.6 Å was employed. For the $P2_12_12_1$ and the

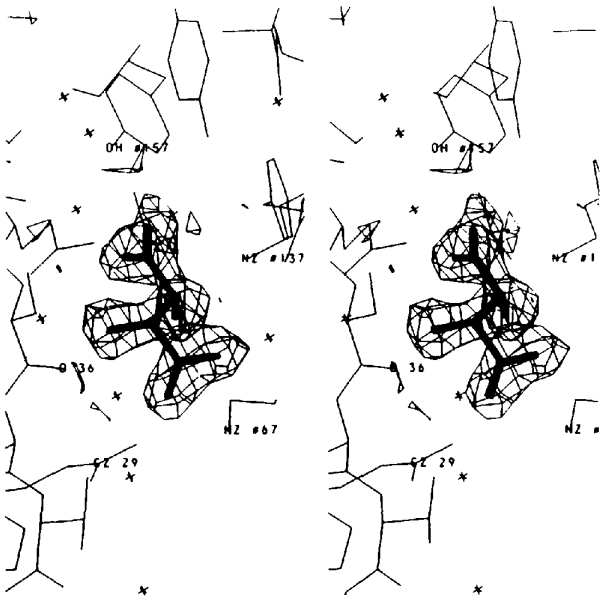


Fig. 5. $2F_o - F_c$ electron-density map for the bound tartrate in the tetragonal crystal. The map was calculated as in Fig. 4. For clarity only a box of density sufficient to encompass the tartrate molecule is shown. Similar density was also observed in the $F_o - F_c$ map based on the original model. Also shown are nearby protein atoms. Seen labeled are the interacting groups from Lys137, Tyr157, Arg29′, Ser36′ and Lys67″.

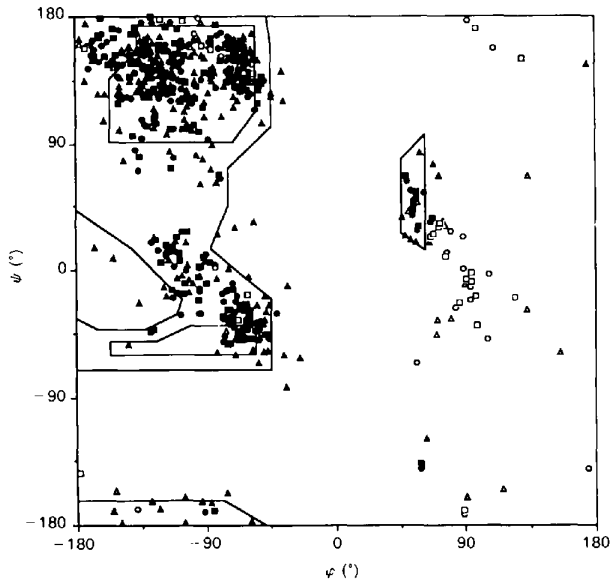


Fig. 6. A Ramachandran plot showing that the φ, ψ angles of all but a few residues in the models are contained within allowed regions. The symbols (▲), (●) and (■) indicate non-glycine residues; (△), (○) and (□) indicate glycine residues for the *C*2, $P2_12_12_1$ and $P4_12_12$ crystals, respectively.

Table 6. *TNT refinement of the orthorhombic $P2_12_12_1$ thaumatin crystal structure*

Unit cell $a = 44.3$, $b = 63.7$, $c = 72.7$ Å, $\alpha = \beta = \gamma = 90.0°$; resolution 12.0–1.75 Å; $I/\sigma \geq 3.0$; No. of reflections = 19968; native $K = 0.869$; overall $B = 0.0$ Å^2; $K_{solvent} = 1.0$; $B_{solvent} = 336.4$ Å^2; overall $R = 0.165$; correlation coefficient = 0.945.

Resolution breakdown

d_{min} (Å)	4.660	3.736	3.275	2.980	2.769	2.608	2.478	2.371	2.281	2.203
No. of F_{obs}	1161	1095	1101	1074	1059	1072	1061	1032	1030	1013
% Complete	99.0	99.5	99.8	99.7	100.0	99.2	100.0	97.0	96.8	96.2
R_{shell}	0.193	0.119	0.133	0.147	0.158	0.166	0.172	0.163	0.164	0.161
R_{sphere}	0.193	0.155	0.148	0.148	0.150	0.151	0.153	0.154	0.155	0.155
d_{min} (Å)	2.134	2.073	2.019	1.970	1.925	1.885	1.847	1.812	1.780	1.750
No. of F_{obs}	1016	1026	969	1006	978	939	915	962	795	664
% Complete	96.5	95.7	93.7	93.6	94.0	89.2	89.0	89.4	76.2	62.9
R_{shell}	0.167	0.177	0.170	0.182	0.186	0.199	0.197	0.208	0.211	0.215
R_{sphere}	0.156	0.157	0.158	0.159	0.160	0.161	0.162	0.163	0.165	0.165

Thermal-parameter distribution

B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	60.0	>60.0	Total	Average (Å^2)
Protein	481	768	150	63	34	31	24	1551	16.08
Solvent	3	27	36	31	14	7	5	123	30.60
Total	484	795	186	94	48	38	29	1674	17.15

Stereochemical deviation

Category	Bond length	Bond angle	Torsion angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	1595	2155	945	33	240	25	211
R.m.s. deviation (Å, °)	0.019	2.925	18.34	0.014	0.019	0.060	0

Table 7. *TNT refinement of the tetragonal $P4_12_12$ thaumatin crystal structure*

Unit cell $a = b = 58.6$, $c = 151.8$ Å, $\alpha = \beta = \gamma = 90.0°$; resolution 12.0–1.75 Å; $I/\sigma \geq 3.0$; No. of reflections = 25864; native $K = 6.07$; overall $B = 0.0$ Å^2; $K_{solvent} = 0.641$; $B_{solvent} = 228.5$ Å^2; overall $R = 0.181$; correlation coefficient = 0.942.

Resolution breakdown

d_{min} (Å)	4.660	3.736	3.275	2.980	2.769	2.608	2.478	2.371	2.281	2.203
No. of F_{obs}	1363	1329	1285	1297	1241	1341	1378	1345	1348	1341
% Complete	98.9	96.5	93.3	94.1	90.1	97.3	100.0	97.6	97.8	97.3
R_{shell}	0.207	0.135	0.143	0.159	0.176	0.178	0.185	0.183	0.188	0.190
R_{sphere}	0.207	0.169	0.161	0.161	0.163	0.165	0.166	0.168	0.169	0.170
d_{min} (Å)	2.134	2.073	2.019	1.970	1.925	1.885	1.847	1.812	1.780	1.750
No. of F_{obs}	1345	1351	1307	1347	1344	1282	1325	1329	1243	723
% Complete	97.6	98.1	94.9	97.8	97.6	93.1	96.2	96.5	90.2	52.5
R_{shell}	0.188	0.188	0.192	0.200	0.200	0.212	0.212	0.230	0.233	0.220
R_{sphere}	0.171	0.172	0.173	0.174	0.175	0.177	0.178	0.179	0.181	0.181

Thermal-parameter distribution

B_{max} (Å^2)	10.0	20.0	30.0	40.0	50.0	60.0	>60.0	Total	Average (Å^2)
Protein	257	917	243	69	33	20	13	1552	17.24
Tartrate	0	7	3	0	0	0	0	10	17.76
Solvent	2	26	32	29	12	4	0	105	27.88
Total	259	950	278	98	45	24	13	1667	17.92

Stereochemical deviation

Category	Bond length	Bond angle	Torsion angle	Trigonal atom	Planar group	Bad contact	Chiral center
No. of restraints	1605	2167	947	34	240	11	213
R.m.s. deviation (Å, °)	0.019	2.932	18.08	0.015	0.019	0.093	0

$P4_12_12$ crystals the final R values were 0.165 and 0.181, the correlation coefficients were 0.945 and 0.942, respectively, both for a resolution range of 12–1.75 Å. The Luzzati plot (Luzzati, 1952) in Fig. 7 indicates the estimated errors in the atomic coordinates for the model of the monoclinic crystal form to be about 0.25 Å and those for the other two models between 0.15 and 0.20 Å. The tartrate bound to the surface of thaumatin in the tetragonal crystal had excellent geometry and differed from the original tartrate structure by an r.m.s. of only 0.193 Å for the four C and six O atoms.

Fig. 8 shows plots of the all-atom and backbone thermal parameters as a function of residue number. The average temperature factors of protein atoms were in the range 16–18 Å^2, while those of the waters were about 30 Å^2. The tartrate in the tetragonal crystal had an average B value of 17.8 Å^2. The r.m.s.

differences between B values of bonded atoms were 2.8, 2.6 and 2.8 Å^2 for the $C2$, $P2_12_12_1$ and $P4_12_12$ crystals, respectively. For the $C2$ crystal, because the resolution of diffraction data was not very high, an additional restriction was employed during the early stages of temperature-factor refinement, by limiting the B values in the range 10–30 Å^2. This resulted in the more even distribution seen in Fig. 8(*a*).

As shown in Figs. 8(*b*) and 8(*c*), residues with high temperature factors are located at regions near 140, 160 and 190 in the $P2_12_12_1$ crystal, and near 60, 120, 140, 160 and 190 in the $P4_12_12$ crystal. All of these regions are at the surface of the protein molecule. Especially conspicuous is the loop between 150 and 170, which protrudes into the bulk solvent in the orthorhombic crystal and is very flexible. In the tetragonal crystal, this loop is in slight contact with neighboring protein molecules, thus its motion is somewhat restricted. Similarly, high temperature factors were not observed for the loops near 60 and 120 in the orthorhombic crystal because of intermolecular contacts.

Discussion

Fig. 9 is a schematic *RIBBONS* drawing of the polypeptide chain and Fig. 10 shows all non-H atoms in the three models arising from these studies when superimposed on the original model of Ogata *et al.* (1992). This was produced using the *RIGID* procedure in *FRODO* and based on the C_α coordinates in the strands of the β-sandwich. Table 8 lists overall r.m.s. deviations between the models and Figs. 11(*a*)–11(*c*) plot combined deviations in coordinates against residue numbers. The vast majority of differences were less than 0.5 Å, while in only a few regions were they greater than 1 Å. These latter regions involved loops, such as residues 60–61 and 142–144. The disulfide-rich domain II exhibited greatest variation among different models. These differences are strongly correlated to the higher thermal parameters for these regions, indicating their flexibility.

The only major difference was the loop containing residues 116–120. Although the loop can undoubtedly assume a range of conformations, the variation between our three refined models was significantly less than between any of these models and the original. Fig. 12 shows a detail of this region. Aside from this difference, the polypeptide backbone and the disulfide bridges remained virtually unaltered

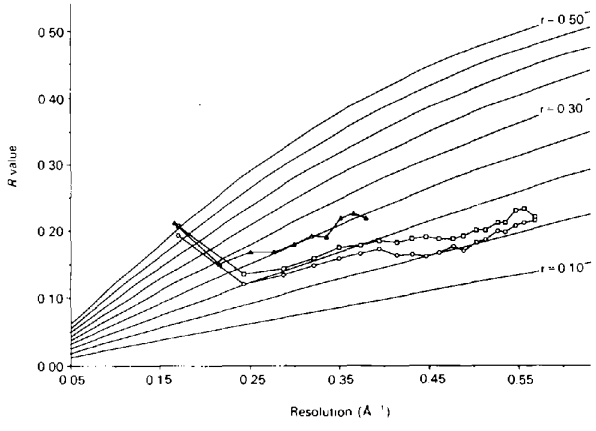


Fig. 7. Luzzati plots for the refined structures of the $C2$ (△), $P2_12_12_1$ (○) and $P4_12_12$ (□) crystals.

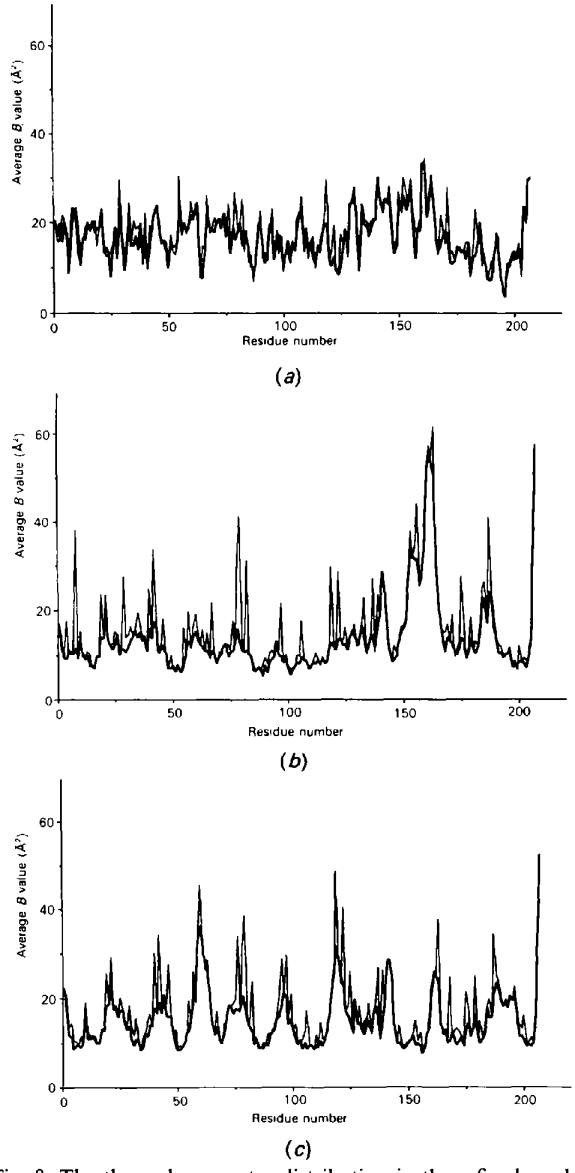


Fig. 8. The thermal-parameter distribution in the refined models of thaumatin in the (*a*) $C2$, (*b*) $P2_12_12_1$ and (*c*) $P4_12_12$ unit cells. Heavy lines represent the main-chain atoms; light lines all atoms.

Table 8. *R.m.s. differences between model coordinates* (Å) *refined in different crystals*

Models*	C_α	Backbone†	All atoms‡
Kim–*C*2	0.741	0.831	1.234
Kim–$P2_12_12_1$	0.485	0.521	1.102
Kim–$P4_12_12$	0.603	0.678	1.060
*C*2–$P2_12_12_1$	0.685	0.772	1.280
*C*2–$P4_12_12$	0.627	0.684	1.206
$P2_12_12_1$–$P4_12_12$	0.593	0.640	1.150

* For alternative conformers in Kim's model, the one with less deviation was employed.

† Including C_α, N, C and O, *i.e.* for polyglycine model.

‡ For residues 46 and 113 only the atoms in common were compared.

with the other exception of Cys164 in the *C*2 crystal which was quite different from the original. Again, however, this may result from the lower resolution of the refinement as well as the greater flexibility of domain II.

Figs. 13(*a*)–13(*c*) are the packing diagrams for the three crystal forms. There are four protein molecules in each of the monoclinic and the orthorhombic unit cells and eight in the tetragonal unit cell. In the *C*2 crystal, each thaumatin molecule is in contact with five other molecules in the lattice. At least 40 residues are involved. All are hydrophilic or polar, except for Ile65, whose side chain is associated with a symmetry equivalent from a neighboring molecule by the crystallographic twofold axis. Intermolecular contacts are extensive, although there are but five interacting residues (including Ile65) near the twofold axis. Thus, the packing resembles bimolecular layers separated by the grid of crystallographic dyads. Lattice interactions include seven possible salt bridges. They are Asp25–Lys97/Lys163;* Glu35–Arg171; Glu42–Arg122; Glu156–Arg76; and carboxyl207–Lys46/Lys187.* Residue 46, the lysine that distinguishes thaumatin B from A, forms a salt bridge with the carboxyl group of the C terminus. This could explain why thaumatin B grew in the *C*2 unit cell. The frequent twinning observed for the monoclinic crystals may be explained as well by the packing of bimolecular layers. During crystal growth, deposition of a nascent molecular layer on the crystal surface could be imprecise due to the weak interactions between layers, across the twofold axes. Accumulation of molecules on the faulty layer would lead to an inverted crystal lattice with consequent twinning.

For the $P2_12_12_1$ crystal, about ten contact regions are more or less evenly distributed over the protein surface. Approximately 28 residues are involved, all are polar, yet only one salt bridge, Glu168–Arg76, is likely. For the $P4_12_12$ crystal, interactions between protein molecules can be divided roughly into eight contact regions. There are about 35 residues involved. Three of them are non-polar: Ile65, Val151 and Phe152. Ile65 interacts with the disulfide of Cys149–Cys158 and Thr160. Val151 and Phe152 are in contact with Pro141. Three salt bridges were identified between the molecules: Glu42–Arg175, Asp60–Arg119 and Glu156–Arg8.

The asymmetric unit of the tetragonal crystal contains a bound tartrate molecule. As seen in Fig. 5 one carboxyl group hydrogen bonds to the phenolic O atom of Tyr157 of one thaumatin molecule while the other salt bridges to Arg29 of a second molecule. The tartrate also interacts with Lys137 of the first molecule and possibly with Lys67 from a third. One hydroxyl group hydrogen bonds to the carbonyl O atom of Ser36 of the second protein molecule while the other appears engaged with the bulk solvent.

Although the resolution of the model was quite high, uncertainty about residue 113, whether it is an asparagine or an aspartate, remained. The side chain of this residue could have potential interactions with the guanidinium groups of Arg8, Arg125 and Arg200. As noted, two of these arginines can also interact with the terminal carboxyl group. This electrostatic attraction, which appears to be an important tertiary interaction, may be a key determinant for the folding process. Loss of sweet taste in the first recombinant thaumatin expressed in yeast, which had an Asn at residue 113, may be a result of its inability to fold into an active conformation.

Another residue of interest is Asp25. In the four known crystal structures this residue always assumes a conformation consistent with a glycine-like type II′ turn. We attempted during refinement to move the φ,ψ angles into one of the 'core' regions by inverting the peptide unit, but found that it moved back to the original conformation upon subsequent refinement. As pointed out by Unger & Moult (1993), deviating main-chain dihedrals might be indicative of functionally important regions in a protein. This aspartate residue is located in the loop insertion between β-strands *B* and *C*. Other homologous plant proteins (Hejgaard, Jacobsen & Svendsen, 1992; Huynh, Borgmeyer & Zobel, 1992; Rebmann, Mauch & Dudler, 1991) lack this six-residue insertion entirely. These thaumatin-like (TL) pathogenesis-related (PR) proteins are widely distributed and some of these have demonstrated antifungal activity. None of them, however, except thaumatin was shown to be sweet. The six-residue *B*–*C* loop is unique to thaumatin. It is on the surface of the protein and, thus, accessible for interaction with other molecules such as receptors. We feel it could be a candidate as a sweetness determinant.

The three-dimensional structure of another sweet protein, monellin, was also determined by Ogata,

* These are residues from other symmetry-related molecules in the crystal.

Hatada, Tomlinson, Shin & Kim (1987). The overall fold of the polypeptide chain is completely different from that of thaumatin. Weak homologous regions were observed only for short segments (Kim, de Vos & Ogata, 1988). It appears that the sweetness of these proteins may not be a consequence of the overall topology but due to some local regions that may act as sweetness determinants. The six-residue insertion between β-strands *B* and *C* of thaumatin seems a good candidate with an appropriately folded polypeptide essential for straining the loop into an active conformation. Further comparison of the structures of sweetness determinants with the molecular configurations of sugars and other sweet compounds may reveal a more general motif.

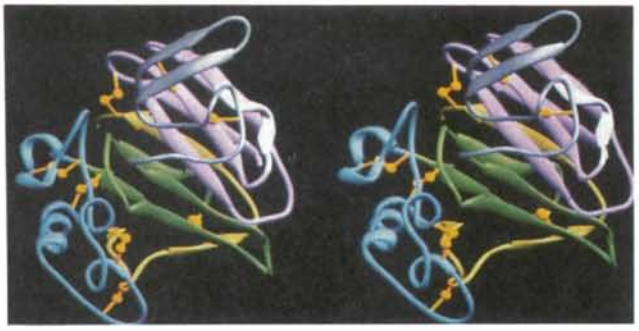


Fig. 9. Stereo *RIBBONS* drawing of the thaumatin molecule. Coordinates from the $P2_12_12_1$ crystal were employed. N-terminal strands (*A–E*), the small disulfide-rich domain (Dom III), the second part of the β-sandwich (strands *F–I*), the large disulfide-rich domain (Dom II) and the C-terminal strands (*J*, *K*) are colored pink, grey, green, blue and yellow, respectively. The disulfides are seen as gold spheres. Highlighted in white is the six-residue insertion which includes Asp25.

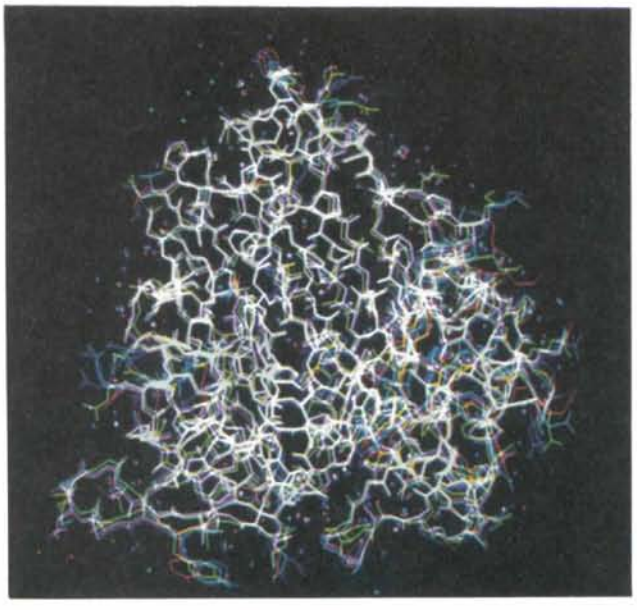


Fig. 10. Superposition of all-atom models after refinement in the *C*2, $P2_12_12_1$ and $P4_12_12$ crystals, colored green, blue and red, respectively, on the original model, colored purple.

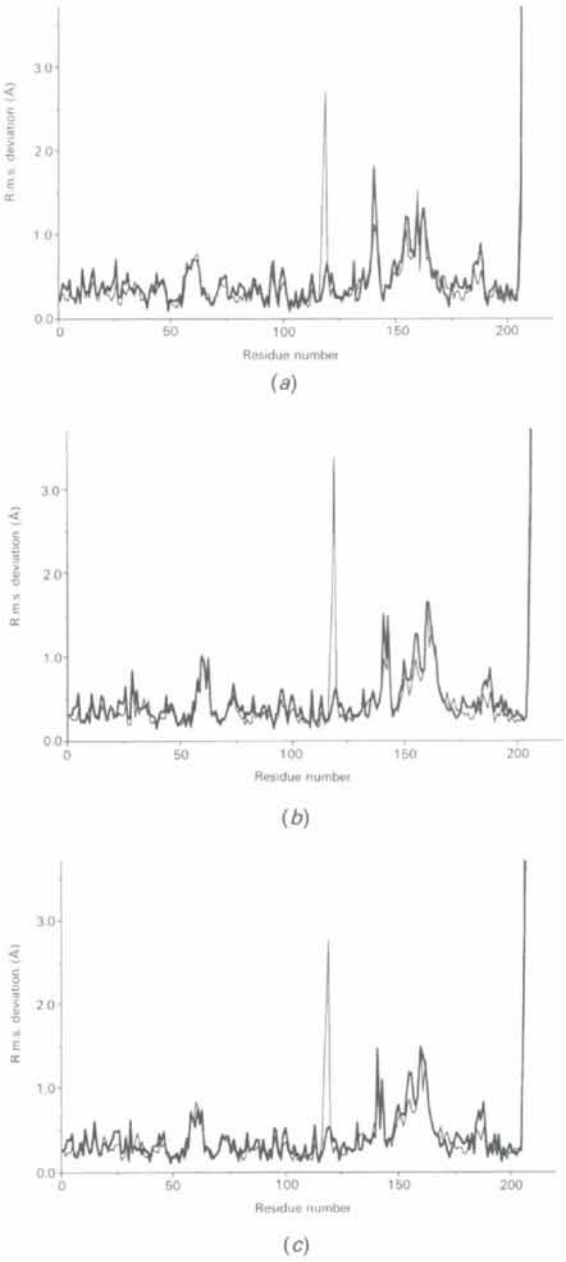


Fig. 11. Deviation between coordinates of (*a*) C_α atoms, (*b*) backbone atoms and (*c*) all atoms, in the thaumatin models for the three new crystal forms (heavy lines) as compared with deviation of these models from the original (light lines).

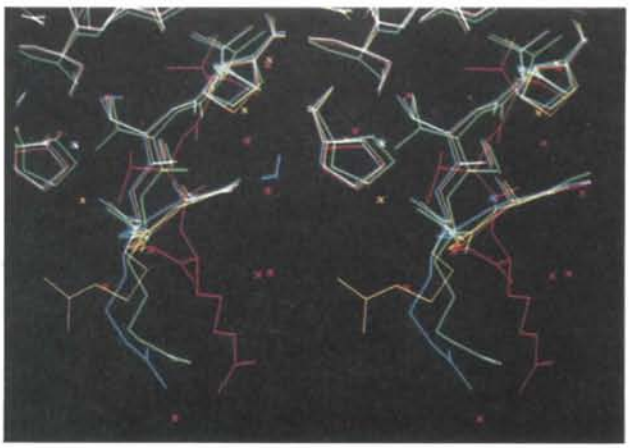
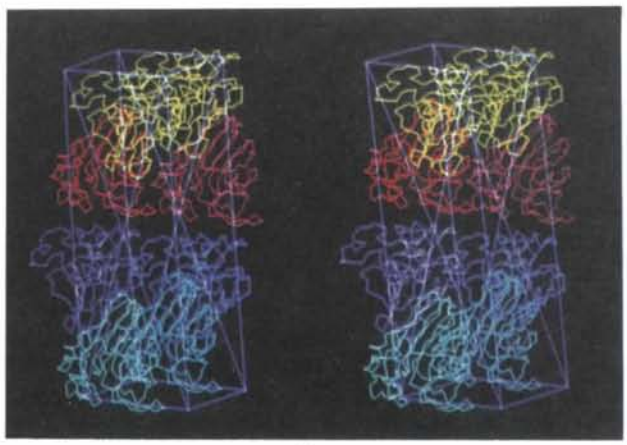
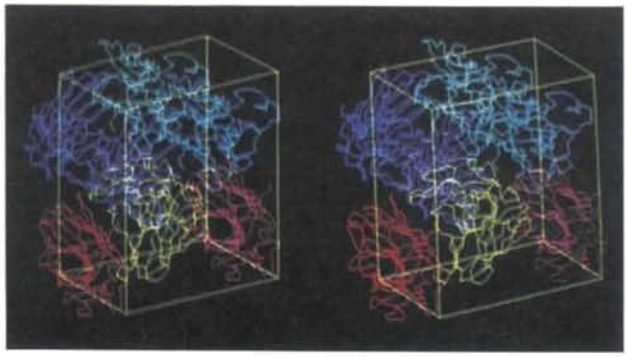


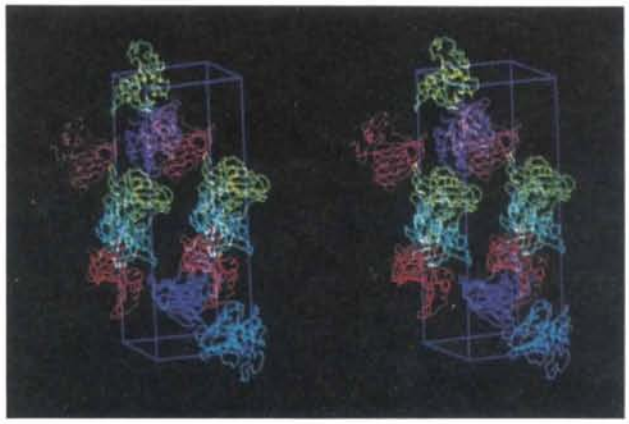
Fig. 12. Stereo photograph of the loop region 116–120 of the refined model in the $C2$ (green), $P2_12_12_1$ (yellow) and $P4_12_12$ (blue) crystals superimposed on the initial model (red). With the exception of the flexible side chain of Arg119, the three new models are entirely consistent with one another, but significantly different from the original.



(a)



(b)



(c)

Fig. 13. Stereo diagrams illustrating the packing of thaumatin molecules in (*a*) $C2$, (*b*) $P2_12_12_1$ and (*c*) $P4_12_12$ crystals. There are four protein molecules in the monoclinic and the orthorhombic unit cells, and eight in the tetragonal. Some neighboring molecules are included for clarity.

References

ABOLA, E. E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WENG, J. (1987). *Crystallographic Database Information Content, Software Systems, Scientific Applications*, edited by F. H. ALLEN, G. BERGERHOFF & R. SIEVERS, pp. 107–132. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

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

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

CARSON, M. & BUGG, C. E. (1986). *J. Mol. Graphics*, **4**, 121–122.

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.

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.

HEJGAARD, J., JACOBSEN, S. & SVENDSEN, I. (1992). *FEBS Lett.* **291**, 127–131.

HUYNH, Q. K., BORGMEYER, J. R. & ZOBEL, J. F. (1992). *Biochem. Biophys. Res. Commun.* **182**, 1–5.

IYENGAR, R. B., SMITS, P. S., VAN DER OUDERAA, F., VAN DER WEL, H., VAN BROUWERSHAVEN, J., RAVESTEIN, P., RICHTERS, G. & VAN WASENAAR, P. D. (1979). *Eur. J. Biochem.* **96**, 193–204.

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.

KIM, S.-H., DE VOS, A. & OGATA, C. (1988). *Trends Biochem. Sci.* **13**, 13–15.

KO, T.-P. (1992). PhD thesis, Univ. of California, Riverside, California, USA.

KO, T.-P., NG, J. D., DAY, J., GREENWOOD, A. & MCPHERSON, A. (1993). *Acta Cryst.* D**49**, 478–489.

LEE, J.-H., WEICKMANN, J. L., KODURI, R. K., GHOSH-DASTIDAR, P., SAITO, K., BLAIR, I. C., DATE, T., LAI, J., HOLLENBERG, S. M. & KENDALL, R. L. (1988). *Biochemistry*, **27**, 5101–5107.

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

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: John Wiley.

MCPHERSON, A. (1990). *Eur. J. Biochem.* **189**, 1–23.

MCPHERSON, A. & WEICKMANN, J. (1990). *J. Biomol. Struct. Dynam.* **7**, 1053–1060.

OGATA, C. M., GORDON, P. F., DE VOS, A. M. & KIM, S.-H. (1992). *J. Mol. Biol.* **228**, 893–908.

OGATA, C., HATADA, M., TOMLINSON, G., SHIN, W.-C. & KIM, S.-H. (1987). *Nature (London)*, **328**, 739–742.

OKAYA, Y., STEMPLE, N. R. & KAY, M. I. (1966). *Acta Cryst.* **21**, 237–243.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

REBMANN, G., MAUCH, F. & DUDLER, R. (1991). *Plant Mol. Biol.* **17**, 283–285.

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., 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.

UNGER, R. & MOULT, J. (1993). *Bull. Math. Biol.* **55**(6), 1183–1198.

VOS, A. M. DE, HATADA, M., VAN DER WEL, H., KRABBENDAM, H., PEERDEMAN, A. F. & KIM, S.-H. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1406–1409.

WEL, H. VAN DER, VAN SOEST, H. & ROYERS, T. (1975). *FEBS Lett.* **56**, 316–317.

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