Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the core region of the soybean β-conglycinin α' subunit

Yukie Maruyama, Nobuyuki Maruyama, Bunzo Mikami and Shigeru Utsumi*

Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Correspondence e-mail: sutsumi@kais.kyoto-u.ac.jp

Received 16 July 2003
Accepted 28 November 2003

PDB Reference: soybean β-conglycinin α'-subunit core region, 1uik, r1uiksf.

The crystal structure of the core region of the α' subunit (α'_c) of soybean β-conglycinin has been determined at 2.3 Å resolution. α'_c was superimposed on the known crystal structure of the β-conglycinin β subunit with a small root-mean square deviation of 0.77 Å, which is consistent with the high sequence identity of 75.5% between α'_c and the β subunit. It is known that the thermal stability of the β subunit is higher than that of the α' subunit and that their thermal stabilities are conferred by highly homologous core regions. Comparisons of the three-dimensional structures and primary sequences between α'_c and the β subunit suggest that five factors account for this difference between subunits as regards the difference in thermal stability: (i) the total cavity volume is larger in α'_c, (ii) the cluster of charged residues at the intermonomer interface is smaller in α'_c and α'_c lacks the intermonomer salt bridge of the β subunit, (iii) the solvent-accessible surface is more hydrophobic in α'_c, (iv) there are fewer proline residues in α'_c and (v) a loop region between helix 3 and strand J′ in α'_c is more flexible owing to the insertion of five additional residues. Although more hydrogen bonds were found in α'_c, this difference should be more than compensated for by the combined contributions of these other factors.

1. Introduction

Owing to its high nutritional value, soybean protein is an extremely attractive food source (Friedman & Brandon, 2001). In addition, its physicochemical properties, such as its gel-forming and emulsifying abilities, as well as its biological activities, such as its hypocholesterolemic effects (Carroll & Kurowska, 1995; Sirtori *et al.*, 1995), are of great usefulness. These properties depend on the two major components of soybean, β-conglycinin (7S globulin) and glycinin (11S globulin), which account for about 30 and 40% of the total seed proteins, respectively (Utsumi, 1992; Utsumi *et al.*, 1997).

β-Conglycinin is composed of three types of subunits: α, α' and β. Together, these subunits form seven heterotrimers with random combinations, as well as three homotrimers. The core regions are highly homologous between the three subunits. In addition, the α and α' subunits have N-terminal extension regions composed of 125 and 141 amino-acid residues, respectively, which are rich in acidic amino-acid residues. Another difference between the three subunits is that two carbohydrate moieties are attached to the α and α' subunits, whereas only one carbohydrate moiety is attached to the β subunit. The physicochemical properties of recombinant wild-type β-conglycinin homotrimers and of the native α and α' subunits and deletion mutants (α_c and α'_c) without the N-terminal extension regions have previously been investigated (Maruyama *et al.*, 1998, 1999, 2002*a*). These investiga-

tions demonstrated that (i) thermal stability differs between the three subunits in the order β (363.8 K) > α' (355.7 K) > α (351.6 K), (ii) the homologous core regions are responsible for these differences in thermal stability, (iii) the solubility of these subunits depends on the number of N-terminal extensions and carbohydrate moieties and (iv) the surface hydrophobicity is known to be higher in the α and α' subunits than in the β subunit and their surface hydrophobicities are conferred by the core region of each subunit.

We have recently reported the crystal structures of native and recombinant β-conglycinin β homotrimers (Maruyama *et al.*, 2001) and we have demonstrated that the quaternary structure of the β trimer is very similar to those of two other seed 7S globulins, *i.e.* canavalin from jack bean (Ko *et al.*, 1993, 2000, 2001) and phaseolin from kidney bean (Lawrence *et al.*, 1990, 1994). However, the α and α' subunits have never been successfully crystallized, probably because of the high carbohydrate content of the molecule (Morita *et al.*, 1996) and also because of the existence of N-terminal extension regions. Therefore, in the present study we attempted to crystallize the deletion mutant α'_c of the α' subunit, which consists of only the core region, and the crystallization was successful. We then determined the three-dimensional structure of this mutant and discuss the structural factors that account for the observed difference in thermal stability between the α' and β subunits.

2. Materials and methods

2.1. Protein purification and crystallization

The deletion mutant α'_c lacks 141 N-terminal residues from the α' subunit. This mutant was expressed in *Escherichia coli* HMS174(DE3) (Novagen) and was purified as described previously by Maruyama *et al.* (1998). To obtain protein suitable for the production of high-quality crystals, it was necessary to achieve a higher level of purity than that of the previously purified α'_c (Maruyama *et al.*, 1998). The protein fraction was subjected to chromatography with a TSKgel Butyl-650M (Tosoh) column equilibrated with buffer *A* [35 m*M* sodium phosphate buffer pH 7.6 containing 0.5 *M* NaCl, 1 m*M* EDTA, 0.02% NaN_3 and 0.1 m*M* (*p*-amidinophenyl)methanesulfonyl fluoride (*p*-APMSF)] containing 30% ammonium sulfate. The sample was then eluted with a linear gradient of ammonium sulfate (30–0%) in buffer *A*. The purified protein was dialyzed against buffer *B* (10 m*M* HEPES pH 7.5 containing 0.5 *M* NaCl, 1 m*M* EDTA, 0.02% NaN_3 and 0.1 m*M* *p*-APMSF) and concentrated to 10 mg ml^{-1} for crystallization using a Centriprep YM-10 (Millipore). Crystals were prepared by the hanging-drop vapour-diffusion method. Single crystals were grown from 17% PEG 3350 (Hampton Research), 0.8 *M* $MgCl_2$ and 8% glycerol in buffer *B* at 293 K.

2.2. Data collection and refinement

Crystals were frozen and stored in liquid nitrogen. A set of diffraction data was collected at λ = 0.9 Å using an Oxford PX210 CCD detector at SPring-8 beamline BL44XU (Beamline for Macromolecule Assemblies, Institute for Protein Research, Osaka University, Japan). The data were processed with *d*TREK* (Pflugrath, 1999). Molecular replacement using the recombinant β-conglycinin β homotrimer (PDB code 1ipk) as a model gave a clear solution using the program *AMoRe* (Navaza, 1994). The model was replaced by the α' sequence and was refined by simulated annealing with molecular dynamics using a slow-cooling protocol in the program *CNS* (Brünger *et al.*, 1998). Several rounds of positional refinement and individual *B*-factor refinement to 2.3 Å resolution were carried out using *CNS*. 5% of the reflections were randomly selected and excluded from refinement for the calculation of a free *R* factor. After each cycle of refinement, the model was adjusted manually using *TURBO-FRODO* (AFMB-CNRS, France). Isolated electron densities greater than 3σ in the $F_o - F_c$ map and/or 1.2σ in the $2F_o - F_c$ map were assigned as water molecules when the locations were sterically reasonable.

2.3. Structure comparison

To compare the α'_c and β structures, the coordinates of the recombinant β homotrimer were used as a β model, since the resolution of the recombinant β was slightly higher than that of the native trimer. No significant differences were observed between the native and recombinant trimers (Maruyama *et al.*, 2001). The accessible surface area (ASA) was estimated by *NACCESS* (Hubbard & Thornton, 1993) with a probe radius of 1.4 Å. Hydrogen bonds were assigned using *CONTACT*, part of the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994), in which the definition of the cutoff distance between the acceptor and donor atoms was 3.3 Å, the minimum calculated angle O—H—N was 120° and the minimum angle donor atom—O—C was 90°. The cavity estimate and volume refinement were carried out using the program *VOIDOO* (Kleywegt & Jones, 1994). The hydropathy (Kyte & Doolitle, 1982) was taken as the average score for five continuous residues and the surface hydropathy was calculated as the sum of the hydropathies of the surface residues (relative ASA > 7%). Since residues that are present in one protein and absent in the other can affect the results of the structural comparison (*e.g.* solvent-accessible areas and cavity volumes), these residues were omitted from all monomers of both protein models, *i.e.* from the N-terminus (142) to position 147, 313–325, 432–444 and 535 to the C-terminus in the α'_c numbering. Thus, some of the values used for the structural comparison were slightly different from those given in the previous report (Maruyama *et al.*, 2001).

2.4. Differential scanning calorimetry (DSC) measurement

DSC experiments were carried out as described previously (Maruyama *et al.*, 1998). All DSC measurements were performed with a protein concentration of 0.5 mg ml^{-1} in 35 m*M* sodium phosphate pH 7.6, 1 m*M* EDTA, 0.02% NaN_3 and 0.1 m*M* *p*-APMSF. The ionic strength was adjusted to 0.5 or 1.0 by the addition of NaCl. The DSC scan rate was 1 K min^{-1} for all experiments.

Table 1
Overview of the statistics of data collection and refinement.

Values for the outer resolution shell are given in parentheses.

Data statistics	
Space group	$P3_121$
Observed reflections	386694 (38357)
Unique reflections	76689 (7604)
R_{sym} (%)	10.7 (19.4)
Completeness (%)	91.4 (93.8)
Unit-cell parameters (Å)	$a = b = 164.96$, $c = 110.05$
Resolution (Å)	30–2.3 (2.38–2.30)
Refinement statistics	
Final model	
Protein atoms	8759
Water molecules	584
Magnesium ions	3
Resolution range (Å)	6–2.3 (2.38–2.30)
R_{cryst} (%)	22.1 (28.4)
R_{free} (%)	26.9 (31.0)
R.m.s.d. on bond angles (°)	1.289
R.m.s.d. on bond lengths (Å)	0.0063
Average B values (Å^2)	
Protein atoms	32.7
Water molecules	37.6
Magnesium ions	43.8

2.5. Construction, expression and purification of the β mutants

Two single mutations and one double mutation were introduced into the β coding region by PCR using the expression plasmid pECβ (Maruyama *et al.*, 1998) and single-mutated pECβ as templates, respectively. The sequences of the mutant coding regions were confirmed by DNA-sequencing analysis. The resultant expression plasmids were transformed to *E. coli* BL21(DE3). The mutants were expressed and purified in a manner similar to that used for the wild-type β (Maruyama *et al.*, 1998), with the exception that a Resource Q anion-exchange column (Amersham Biosciences) was used in place of the Q-Sepharose column.

3. Results

3.1. Final model quality

The crystal structure of α'_c was determined at 2.3 Å resolution. The data-collection and refinement statistics are summarized in Table 1. 99.7% of the amino-acid residues were within the allowed region of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) as calculated using *PROCHECK* (Laskowski *et al.*, 1993). All monomers contained a single residue, Tyr531, in the disallowed region. The corresponding residues in all of the models of 7S globulins of known crystal structure also had φ–ψ angles similar to those of Tyr531 in α'_c (Lawrence *et al.*, 1994; Ko *et al.*, 2000; Maruyama *et al.*, 2001). An isolated electron density greater than that of a water molecule was seen in the molecular-packing regions of each monomer. This dense region was coordinated by six water molecules with a distance of 2.24 ± 0.22 Å. This region may be a magnesium ion arising from the addition of $MgCl_2$ during crystallization. In addition to the N- and C-terminal regions, two regions, 314–325 and 432–443, were omitted from the final model because the electron density was diffuse at the main-chain level. In the β trimer, all the corresponding residues also had diffuse density, with the exception of residues 175–182 (corresponding to 314–325 in α'_c) in one of three β monomers forming a trimer. These findings, together with the fact that the resolution of α'_c was significantly higher than that of the β trimer, indicated that these regions were disordered.

3.2. Overall structure

There is a trimer in the asymmetric unit of the α'_c crystal. Figs. 1 and 2 show trimeric and monomeric models of α'_c, respectively. Each monomer can be divided into two very similar modules, the N- and C-terminal modules, related by a pseudo-dyad axis. Each module contains a core β-barrel (jelly-roll) domain and an extended loop domain containing several helices. The secondary-structure elements were assigned using *DSSP* (Kabsch & Sander, 1983) and were named according to the strands in the β subunit (Maruyama *et al.*, 2001). Each β-barrel consists of a larger β-sheet A′ABIDGZ and a smaller β-sheet J′JCHEF. Three monomers were superimposed onto each other with an r.m.s. deviation of less than 0.2 Å for all main-chain atoms.

Fig. 3 shows a structure-based sequence alignment between the two subunits, with a classification of the secondary structure and solvent-accessible and solvent-inaccessible residues. The sequences of the α'_c and β subunits are highly homologous, with an identity of 75.5%. Although the sequence identities of the buried and exposed residues are not significantly different between the α'_c and β subunits, most of the significant amino-acid replacements altering the side-chain volume and/or

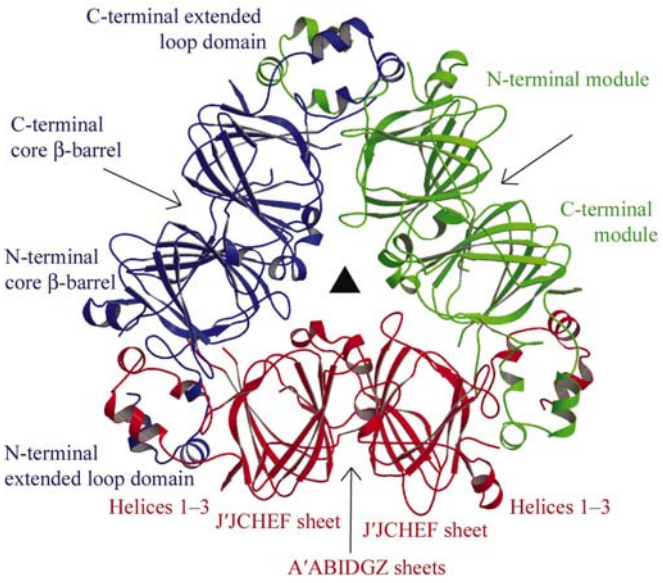

Figure 1
Ribbon diagram of the α'_c trimer as seen along a molecular threefold axis (black triangle). The three monomers are shown in green, blue and red. Each monomer contains two similar modules related by a pseudo-dyad axis (arrow) perpendicular to the threefold axis. This figure was prepared using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994).

Table 2
Residues around the cavities inside β-barrels.

Residues that differ between the α'_c and β sequences are shown in bold.

Cavity	Volume† (Å^3)	Residues														
N-module																
α'_c	54.3	L160	F161	**V168**	**L190**	E191	F192	L200	H203	**Y209**	Y247	**M260**	**T262**	F273	**S275**	F277
β	23.8	L22	F23	I30	**V52**	**E53**	F54	L62	H65	**F71**	Y109	I122	**K124**	Y135	**D137**	F139
C-module																
α'_c	81.6	**L376**	F393	S395	V397	**F405**	H408	N410	V415	L417	I462	**V468**	V470	F478	A480	R490
β	20.8	**F233**	F250	S252	**V254**	**L262**	H265	N267	V272	L274	I319	**F325**	**V327**	F335	A337	R347

† Total cavity volume of three monomers detected with the program *VOIDOO* (Kleywegt & Jones, 1994) using a probe radius of 1.4 Å.

polarity occur on the molecular surface. Helix 5, located in a loop region between the strand J′ of the N-terminal module and strand Z of the C-terminal module, has not previously been observed in the β subunit. In contrast, helix 0′ of the N-terminal module seen in the β subunit is absent from α'_c. The sequences in helices 0′ and 5 are identical in the α'_c and β subunits, and the conformations of these helices are not very different from those of the corresponding regions in the respective subunits. Whether or not the secondary structures can be determined may depend on small deviations in the main-chain torsion angles.

In addition to the crystal structures of the α'_c and β subunits of β-conglycinin, the crystal structures of four other seed-storage globulins have been reported: two 7S globulins, jack bean canavalin (Ko *et al.*, 2001) and kidney bean phaseolin (Lawrence *et al.*, 1994), and two 11S globulins, soybean proglycinin A1aB1b homotrimer (Adachi *et al.*, 2001) and mature-glycinin A3B4 homohexamer (Adachi *et al.*, 2003). The sequence alignment of α'_c with other seed-storage proteins is shown in Fig. 4. All β-strands that form the two core β-barrels are common among these proteins except for the Z strand in the N-terminal module. Similarly, all proteins have α-helices 2 and 3. Two of the 11S globulins lack the Z strand and helix 1 of the N-terminal module.

All C^{α} atoms of the α'_c monomer could be superimposed on the corresponding atoms of the β monomer with a small r.m.s. deviation of 0.77 Å; this is consistent with the high sequence identity of these subunits. The plot of the r.m.s. deviation of C^{α} atoms between the α'_c and β subunits against the amino-acid sequence (Fig. 5) shows that the deviations of the β-barrels are relatively small. Large deviations (>1.5 Å) are seen around a deletion site (between residue 154 and 155 in the α'_c), two disordered regions, two helices 3, two loop regions in the N-terminal module (between strand A and helix 0 and between strand J and helix 1) and two loop regions in the C-terminal module (between helices 4 and 5 and between strands Z and A′). Most residues responsible for the large deviations in Fig. 5 are located on the molecular surface and are thought to be flexible.

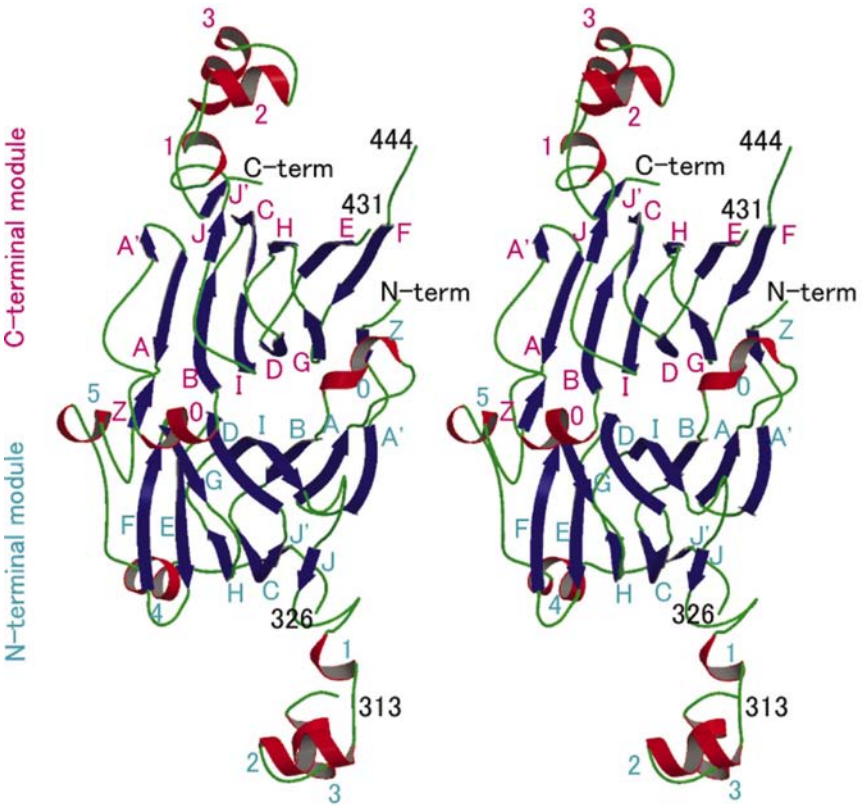

Figure 2
A stereoscopic view of the α'_c monomer. The β-strands and the α- and 3_{10}-helices are shown in blue and red, respectively. This figure was prepared using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994).

3.3. Solvent atoms

Many water molecules interact with each other *via* weak hydrogen-bonding networks in the central channel of the trimer and at intermolecule interfaces. It is noteworthy that the water molecules inside the β-barrels and water-mediated β-sheets found in the canavalin and phaseolin structures (Ko *et al.*, 2000) are also present in α'_c and that they occupy positions very similar to those in canavalin structures (the differences are within 0.9 Å). These positions include Wat28/27/54, 17/3/6, 50/5/65, 120/10/52 and

19/358/15 inside each barrel of three monomers (corresponding to Wat504, 505, 506, 508 and 515 in the cubic crystal of canavalin) and Wat21/129/71, 69/11/31 and 122/116/94 (Wat507, 540 and 550 of canavalin) mediating the hydrogen bonds of the β-strands.

3.4. Packing of modules

26 cavities were detected in the α'_c trimer and the total volume was 172 Å^3, compared with 20 and 79 Å^3 in the β trimer. Large cavities are found inside the β-barrels of both modules, which are important for enzymes belonging to the cupin superfamily (Dunwell *et al.*, 2000) to bind the metal ions and/or substrates. Seed-storage proteins are also members of the cupin superfamily but have lost the metal-binding activity, probably owing to the substitution of conserved histidine residues (Dunwell *et al.*, 2000). The deviations of the core β-barrels of the α'_c and β subunits are small, as shown in Fig. 5. Therefore, the differences in these cavity volumes between the α'_c and β subunits are derived from residue substitutions rather than from their structural diversity. Among the seven substitutions observed around the cavity of the N-terminal β-barrel (Table 2), two substitutions, Thr262 and Ser275 in α'_c for Lys124 and Asp137 in the β subunit, largely contribute to the difference in the cavity volume. Leu376, Phe405 and Val468 in the C-terminal β-barrel of α'_c are replaced by Phe233, Leu262 and Phe325 in the β subunit. Because the effects on the cavity volume of the substitutions of Leu376 and Phe405 in α'_c by Phe233 and Leu262, respectively, in the β subunit may compensate for each other, the difference in the cavity volume inside the β-barrel of the C-terminal module is likely to arise from the replacement of Val468 in α'_c by Phe325 in the β subunit.

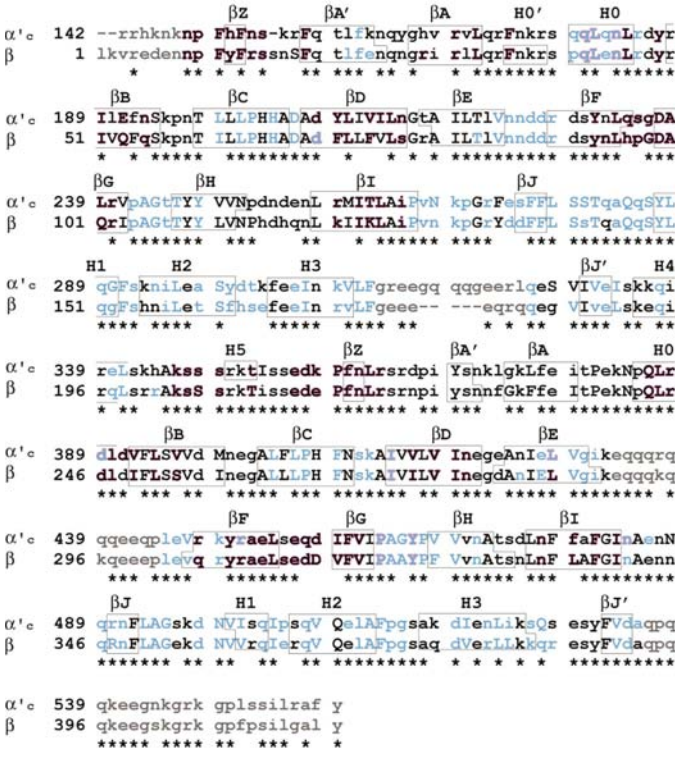

Figure 3
Structure-based sequence alignment of the α'_c and β subunits. Lower case and upper case letters represent solvent-accessible (relative ASA > 7%) and solvent-inaccessible (relative ASA < 7%) residues, respectively. The blue characters show residues at the intermonomer interface and pink overlaid characters represent those at the intermodule interface. Residues omitted from the final models are shown in gray. '*' indicates the position of residues in common between the two. Helices 1 and 5 are 3_{10}-helices and the other helices in α'_c are α-helices.

The number of intramonomer hydrogen bonds comprising intramodule and intermodule interactions (Table 3) was 442 in α'_c compared with 396 in the β subunit. An intramodule salt bridge strictly preserved among 7S globulins is also found between Glu378 (strand A) and Arg490 (strand J), which is located inside a β-barrel of the C-terminal module in α'_c. Other intramodule salt bridges (Arg169–Glu191, Asp208–Arg240 and Lys303–Glu305 in the N-terminal module, and Arg363–Asp398 in the C-terminal module) are present on the molecular surface. Two of these salt bridges, Asp208–Arg240 and Arg363–Asp398, correspond to two salt bridges in the β subunit, namely Asp70–Arg102 and Arg220–Asp314, respectively.

3.5. Intermodule interface

Within a monomer, two modules are tightly associated through their larger β-sheet (strands A′ABIDGZ) and helix 0. To estimate the area of the intermodule interface, the ASA values of individual modules and monomers were calculated (Table 4). The buried areas between the two modules were found to be 1541 and 1543 Å^2 in the α'_c and β monomer, respectively. In both proteins, around 70% of the buried areas are occupied by non-polar atoms that stabilize the monomer form *via* hydrophobic interactions. The hydrophobic residues participating in the hydrophobic interactions (C—C distance <4 Å) at the α'_c intermodule interface are Pro149, Phe150, Phe152, Phe157, Leu171, Phe174, Leu181, Leu184, Tyr187, Ile189, Leu210, Val212, Leu214, Tyr231, Leu233, Ala238, Leu239, Ile261, Leu263 and Ile265, giving a total of 20 residues in the N-terminal module, and Pro359, Phe360, Leu387, Leu390, Val392, Leu394, Ile414, Val416, Val418, Leu428, Tyr450, Ile459, Phe460, Val461, Pro463, Phe479, Phe481 and Ile483, giving a total of 18 residues in the C-terminal module. These values are less than the corresponding totals of 21 residues in the N-terminal and 22 residues in the C-terminal module of the β subunit. Many hydrogen bonds are found in the interface (Table 3), including the main-chain interactions of strands Z and G forming the β-sheet (Fig. 2). Two intermodule salt bridges between Arg240 and Asp389 and between Arg259 and Glu456 were observed in α'_c. The latter salt bridge is also found in the β subunit (between Lys121 and Glu313).

3.6. Intermonomer interface

The intermonomer interface is primarily formed by the extended-loop domain and the smaller J′JCHEF β-sheet. Table 4 shows the decrease in ASA caused by trimerization. 7254 and 7367 Å^2 are buried upon trimerization of the α'_c and β subunits, respectively. Non-polar atoms occupy around 67%

Table 3
Number of hydrogen bonds, including salt bridges.

The numbers in parentheses are the numbers of salt bridges.

	N-module	C-module	Intermodule	Intermonomer
α'_c	216 (3)	204 (2)	22 (2)	26 (0)
β	188 (1)	187 (2)	21 (2)	27 (1)

Table 4
Accessible and buried surface areas.

	Non-polar atoms		Polar atoms		Total	
	α'_c	β	α'_c	β	α'_c	β
Accessible surface area (Å^2)						
N-module	5791	5701	5290	5439	11080	11140
C-module	5825	5348	4400	4609	10225	9957
Monomer	9505	8889	8719	9123	18225	18012
Trimer	18665	16951	21500	22358	40165	39309
Buried surface area (Å^2)						
Intermodule†	1056	1080	486	463	1541	1543
Intermonomer‡	4925	4858	2329	2506	7254	7363

of the intermonomer interfaces of the α'_c and β trimers. The hydrophobic residues participating in the hydrophobic interactions at the α'_c intermonomer interface are Phe161, Leu199, Leu201, Pro202, Val223, Tyr248, Val267, Phe276, Leu278, Tyr287, Leu288, Phe291, Ile295, Ala298, Tyr300, Ile307, Val310, Leu311, Phe312, Ile333 and Leu341, giving a total of 21 residues in the N-terminal module of one monomer, and Leu404, Pro407, Ile414, Val429, Ile431, Leu445, Val447, Ala464, Tyr466, Pro467, Val468, Val469, Leu493, Ala494, Val500, Ile501, Ile504, Pro505, Val508, Leu511, Ala512, Phe513, Pro514, Ile520, Leu523, Ile524 and Ala535, giving a total of 27 residues in the C-terminal module of the other monomer. These values are lower and higher than the corresponding totals of 23 residues in the N-terminal and 25 residues in the C-terminal module of the β subunit, respectively. Along with a number of hydrophobic interactions, many hydrogen bonds were found, as shown in Table 3. In addition, water-mediated hydrogen bonds were observed in the intermonomer interfaces, as also reported for the canavalin structures, despite the non-preserved positions.

Many charged residues form a cluster on the molecular surface around the intermonomer interface of the β subunit (Fig. 6*a*). The corresponding cluster in α'_c (Fig. 6*b*) is smaller owing to the substitution of seven residues, Glu353, Arg359, Glu362, Arg363, Arg379, Lys383 and Arg385 in the β subunit by Ser496, Ser502, Pro505, Ser506, Asn522, Ser526 and Ser528 in α'_c, respectively. One of these, the replacement of Lys383 by Ser in α'_c, breaks an intermonomer salt bridge, Glu168–Lys383, formed in the β subunit.

The α, α' and β subunits of β-conglycinin are able to form heterotrimers as well as each homotrimer. Interactions between monomers in each trimer are thought to be strong because the isolated heterotrimers are stable (Maruyama *et al.*, 2002*b*). The sequence identity of the intermonomer interfaces of the α'_c and β subunits (82%) is higher than that of the overall sequence identity (75.5%). The intermonomer interface, however, contains some highly deviating regions, as illustrated in Fig. 5. Although subunit interfaces are important for specific interactions in oligomer proteins, the conformation of the intermonomer interface of β-conglycinin trimers with many hydrophobic residues and water molecules is relatively flexible. This is probably the reason why these three subunits of β-conglycinin can form stable heterotrimers with random combinations.

3.7. Molecular surface

The ASA values of the entire molecules shown in Table 4 indicate that the α'_c trimer has a slightly larger surface area and higher non-polar atom ratio (46%) than the β trimer (43%). The higher surface hydrophobicity of α'_c was also indicated by the sum of the hydropathies of surface residues, −179 for the α'_c monomer and −196 for the β monomer. These observations may be consistent with the previous experimental results obtained from ANS analysis (Maruyama *et al.*, 1999) and hydrophobic column chromatography (Maruyama *et al.*, 2002*b*). However, the patterns of distribution of charged and hydrophobic residues on the molecular surface of the α'_c and β trimers were very similar. The most significant differences were seen on the molecular surface around the intermonomer interface shown in Fig. 6, where the β subunit is rich in charged residues.

3.8. Thermal stability

It has been reported that thermal stability differs between the three homotrimers of β-conglycinin in the order β (363.8 K) > α' (355.76 K) > α (351.6 K) at an ionic strength of 0.5 and pH 7.6 (Maruyama *et al.*, 1998), where α'_c associates into a hexamer (a dimer of trimers). However, the salt concentration (I > 2.9) of the crystallization condition of α'_c was too high for α'_c to associate into a hexamer (Maruyama *et al.*, 1998). We therefore examined the thermal stability of the β, α'_c and α_c subunits by DSC at an ionic strength of 1.0 and pH 7.6, where α'_c gave a single peak corresponding to a trimer in the gel-filtration experiment (Maruyama *et al.*, 1998). Although all trimers were more stable under the present conditions than under the conditions employed previously (Maruyama *et al.*, 1998), the order of the T_m values was consistent: β (371.1 K) > α'_c (366.7 K) > α_c (362.7 K).

In order to investigate the effect of differences in the charge distribution between the α'_c and β subunits (Fig. 6) on their thermal stabilities, three mutants of β (R379N, K383S and R373N/K383S) were prepared and subjected to DSC analysis. The T_m values of these mutants at pH 7.6 and an ionic strength of 0.5 were 363.9, 361.9 and 361.9 K, respectively, compared with 364.0 for the wild-type β subunit.

4. Discussion

Thermal stability, which is related to functions such as heat-induced gelation, is an important characteristic of soybean proteins. The α, α' and β homotrimers of these proteins exhibit different thermal stabilities in the order β > α' > α; these

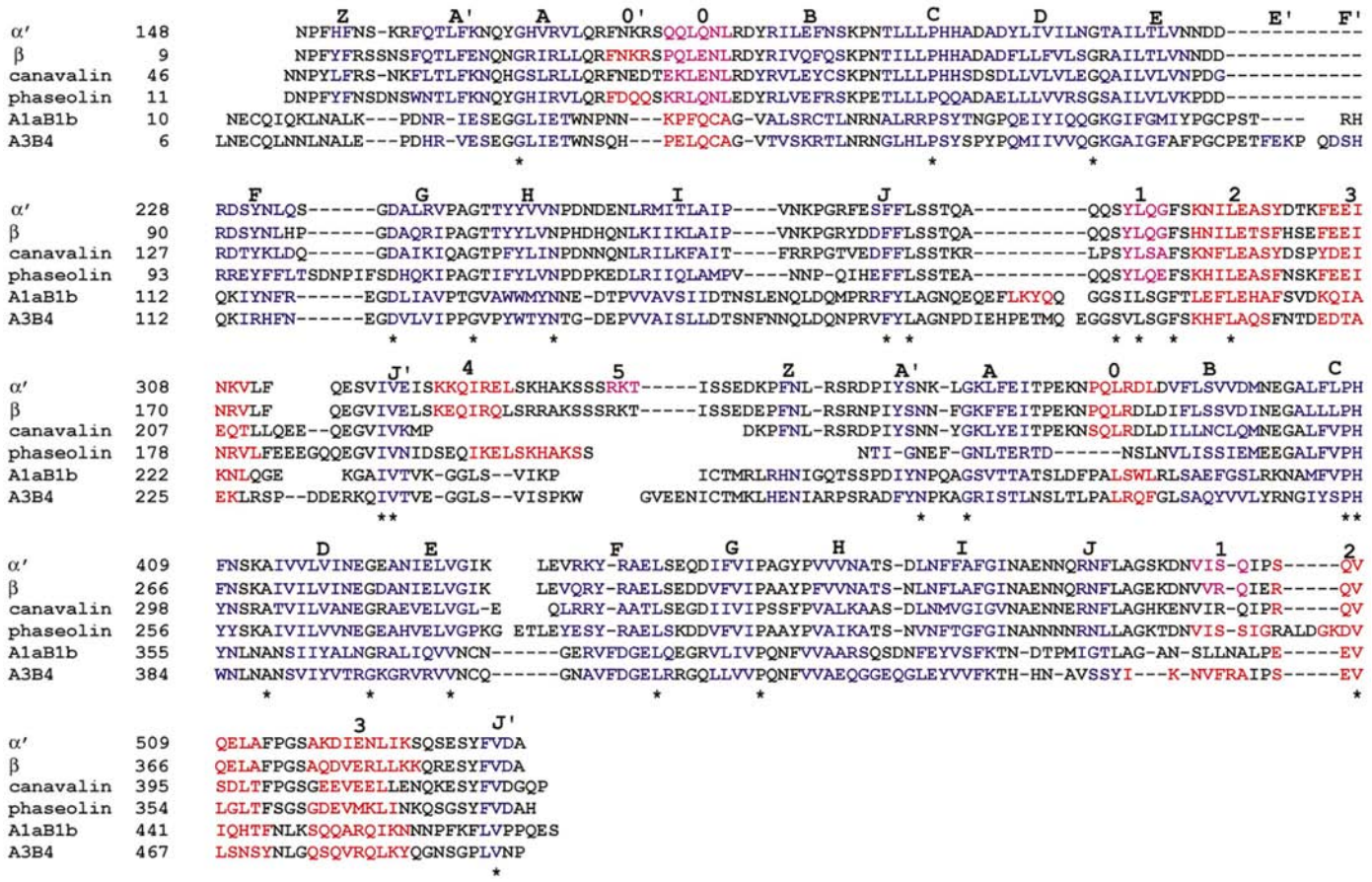

Figure 4
Sequence alignment of six seed-storage globulins for which crystal structures have been described. Blue, red and pink represent β-strands, α-helices and 3_{10}-helices, respectively. The symbols '-' and '*' represent gaps and conserved residues, respectively. A space represents a disordered region omitted from this figure. Pairwise structure alignments were originally performed using *DALI* (Holm & Sander, 1994) and all sequences were then aligned manually. Secondary-structure elements were assigned using *DSSP* (Kabsch & Sander, 1983). With regard to the molecules with three monomers in the asymmetric unit, secondary-structure elements that were observed in at least two monomers are shown. PDB references are as follows: 1ipk for β, 1dgw for canavalin, 2phl for phaseolin, 1fxz for A1aB1b and 1od5 for A3B4.

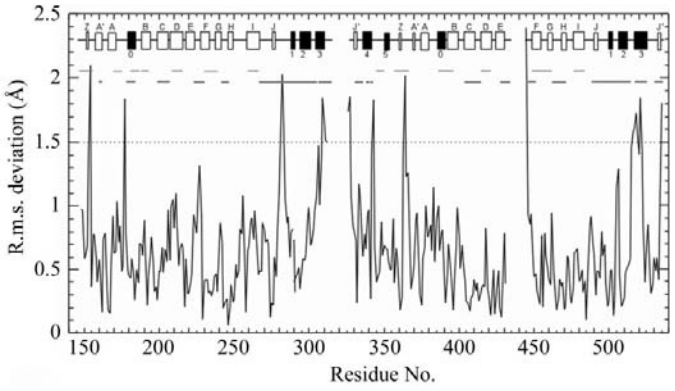

Figure 5
The deviation of C^{α} atoms between α'_c and the β subunit *versus* residue number. Secondary-structure elements (boxed) and the intermodule (thin grey line) and intermonomer (thick grey line) interfaces are also displayed. The residue number and helix and strand names are based on those of α'_c. The dashes indicate a deviation of 1.5 Å, which is approximately twice the length of the r.m.s. deviation (0.77 Å) of all C^{α} atoms of the two proteins.

thermal stabilities are conferred by the core regions, which are highly homologous between these subunits (Maruyama *et al.*, 1998). Although the overall structures of the α'_c and β homotrimers are very similar, a detailed comparison of the two structures revealed five structural factors that may account for the difference observed in thermal stability.

Firstly, although the protein is tightly packed, cavities are often observed inside the molecule and these cavities decrease the thermal stability (Pace, 1992). The total cavity volume in α'_c is much larger than that in the β subunit. Although most of the notable substitutions between the α'_c and β occur at solvent-exposed sites, as shown in Fig. 3, several amino acids are substituted inside the molecule. One of the most significant differences between the two proteins is Thr262, which lines the cavity of the N-terminal β-barrel of α'_c and which is substituted by Lys124 in the β subunit (Table 2). The side chain of this lysine residue is completely buried in the molecule, but does not form a salt bridge. Thus, the presence of this lysine residue inside the β molecule is thought to destabilize the molecule. However, the partially buried and non-ion-paired Asp137 is

also present around the cavity of the N-terminal β-barrel at a distance of about 6 Å from Lys124, a distance that may be short enough to allow interaction between the two residues. Asp137 is replaced by a smaller non-charged residue (Ser275) in α'_c, which also contributes to the difference in the cavity volume. Lys124 buries the cavity with its long side chain and forms a hydrogen bond with the OH group of Tyr109; therefore, Lys124 may actually stabilize the structure of β. Owing to the compensating substitutions of Leu376 and Phe405 in α'_c by Phe233 and Leu262, respectively, in the β subunit, the substitution of Val468 in the α'_c by Phe in the β subunit mainly contributes to the difference in the cavity volume of the C-terminal β-barrel. It is of note that a double-mutated β subunit substituting Ile122 and Lys124 for the α'_c sequence (Met and Thr, respectively) was slightly destabilized (Maruyama *et al.*, 2003). The simulated model of mutant β (I122M/K124T) showed a larger cavity volume: 173 Å^3 for the total volume and 31.1 Å^3 for the N-terminal β-barrel cavity. Maruyama *et al.* (2002*b*) similarly concluded that the cavity volume was the reason for the difference in thermal stability between the native β prepared from a mutant soybean cultivar and the recombinant β subunit.

Secondly, as long ion-pair networks have been found in several enzymes from a hyperthermophilic bacterium (Yip *et al.*, 1995; Lim *et al.*, 1997), they are thought to be a common mechanism of stabilizing proteins in hyperthermophiles (Karshikoff & Ladenstein, 2001). As described above, we found a different charge distribution between the α'_c and β subunits at the molecular surface around the intermonomer interface (Fig. 6). This difference is derived from substitutions of seven residues. One of these, the replacement of Lys383 by Ser in α'_c, breaks an intermonomer salt bridge, Glu168–Lys383, formed in the β subunit. We constructed three mutants (R379N, K383S and R379N/K383S) of the β subunit and analyzed their thermal stabilities by DSC. The T_m value of K383S lacking an intermonomer salt bridge Glu168–Lys383 as in α'_c was 361.9 K, which was 2 K lower than that of the wild-type β subunit. On the other hand, the T_m values of R379N (363.9 K) lacking a free charged residue near Lys383 and the double mutant R379N/K383S (361.9 K) were identical to that of the wild-type β subunit and K383S, respectively. These results suggest that the intermonomer salt bridge Glu168–Lys383 contributes somewhat to the higher thermal stability of the β subunit, but other free charged residues on the molecular surface have little effect on thermal stability under the experimental high ionic strength conditions.

Thirdly, hydration of non-polar groups apparently destabilizes proteins (Spassov *et al.*, 1995). The ratio of non-polar atoms and the sum of the surface hydropathies revealed that the solvent-accessible surface of α'_c was more hydrophobic than that of the β subunit, which was consistent with the experimental results. This was thus considered to be another factor contributing to the observed difference in thermal stability.

Fourthly, the number of proline residues, which stabilize the protein structure by decreasing the entropy of the denatured structure, is 57 in the α'_c trimer, compared with 63 in the β trimer. According to Matthews *et al.* (1987), this difference contributes 20 kJ mol^{-1} in ΔG. There are three prolines, Pro41, Pro97 and Pro409, in the β monomer that are replaced by other residues in the α'_c monomer. One of these, Pro409, is located in the C-terminal region and its coordinates are unknown. The polypeptide φ/ψ angles of Pro41 and Pro97 are $-53 \pm 6/-33 \pm 7^\circ$ and $-52 \pm 5/130 \pm 4^\circ$, respectively. These values are reasonable for a proline and therefore Pro41 and Pro97 do not decrease the enthalpy of the native structure.

Finally, shorter loops are yet another stabilizing factor (Vogt *et al.*, 1997; Chakravarty & Varadarajan, 2002). A previous study has shown that insertion of polyglycines into a loop region of the rop protein results in a significant length-dependent destabilization (Nagi & Regan, 1997). The loop between helix 3 and strand J′ of the N-terminal module is five residues shorter in the β subunit than in α'_c (Fig. 3). This region is located on the molecular surface and is disordered in the crystal. The electron densities of this loop region were broken in all monomers of the α'_c trimer and in two of the monomers of the β trimer, but appeared in one monomer of the β trimer which gave the most clear density map in three monomers, despite the lower resolution of the β subunit. Therefore, a relatively rigid conformation of this loop region in the structure of the β subunit was suggested.

Water molecules are another factor that affects the structural stability of proteins (Takano *et al.*, 1997). In the case of 7S seed proteins, several equivalent water molecules inside the β-barrel and water-mediated β-sheets were found at common positions, as

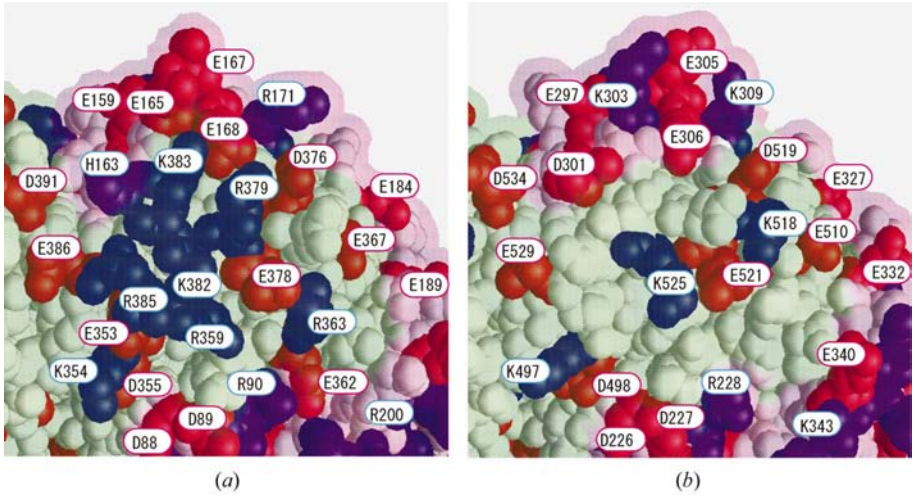

Figure 6
Comparison of the charge distribution at the molecular surfaces of (*a*) β and (*b*) α'_c. This region is near the interface between two interacting monomers shown in pink and green. Positively (Lys, Arg and His) and negatively (Glu and Asp) charged residues are shown in blue and red, respectively. This figure was prepared using *GRASP* (Nicholls *et al.*, 1991).

described above, and these particular waters may play an important role in protein stability. Furthermore, water-mediated interactions were observed at the intermonomer interface of α'_c as well as in canavalin structures, despite their different positions. However, the influence of these water molecules on thermal stability of these proteins cannot be addressed here, because the resolution of the β subunit structure was too low to introduce water molecules into the final model. To discuss the contribution of water molecules to the difference in thermal stability between the α'_c and β subunit, higher resolution data will be needed.

As described above, the difference in thermal stability of the α'_c and β homotrimers is likely to have resulted from a combination of many factors, as in a previous report analyzing numerous mutants (Funahashi *et al.*, 2002). As depicted in Table 3, more hydrogen bonds were observed in each module of the α'_c than in the β. This suggests more stable packing of α'_c, which is not consistent with the experimental data. This difference should be more than compensated for by the accumulation of the small effects of many factors accounting for the relatively lower thermal stability of α'_c, such as the larger cavity volumes, the lack of an intermonomer salt bridge, the higher surface hydrophobicity, the smaller number of proline residues and the longer loop. The combined effect of these contributions would then result in the β subunit being more stable than α'_c.

This work was supported in part by grants from the Japan Society for the Promotion of Science (to SU and YM), the Program for the Promotion of Basic Research Activities for Innovative Biosciences (to SU and BM) and the Protein 3000 project of the Ministry of Education, Culture, Sports, Science and Technology (to BM and SU).

References

Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura K., Mikami, B. & Utsumi, S. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 7395–7400.

Adachi, M., Takenaka, Y., Gidamis, A. B., Mikami, B. & Utsumi, S. (2001). *J. Mol. Biol.* **305**, 291–305.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–821.

Carroll, K. K. & Kurowska, E. L. (1995). *J. Nutr.* **125**, 594S–597S.

Chakravarty, S. & Varadarajan, R. (2002). *Biochemistry*, **41**, 8152–8161.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dunwell, J. M., Khuri, S. & Gane, P. J. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 153–179.

Friedman, M. & Brandon, D. L. (2001). *J. Agric. Food Chem.* **49**, 1070–1086.

Funahashi, J., Takano, K., Yamagata, Y. & Yutani, K. (2002). *J. Biol. Chem.* **277**, 21792–21800.

Holm, L. & Sander, C. (1994). *Proteins*, **19**, 165–173.

Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS Computer Program.* Department of Biochemistry and Molecular Biology, University College, London.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Karshikoff, A. & Ladenstein, R. (2001). *Trends Biochem. Sci.* **26**, 550–556.

Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.

Ko, T.-P., Day, J. & McPherson, A. (2000). *Acta Cryst.* D**56**, 411–420.

Ko, T.-P., Kuznetsov, Y. G., Malkin, A. J., Day, J. & McPherson, A. (2001). *Acta Cryst.* D**57**, 829–839.

Ko, T.-P., Ng, J. D. & McPherson, A. (1993). *Plant Physiol.* **101**, 729–744.

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kyte, J. & Doolitle, R. F. (1982). *J. Mol. Biol.* **157**, 105–132.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lawrence, M. C., Izard, T., Beuchat, M., Blagrove, R. J. & Colman, P. M. (1994). *J. Mol. Biol.* **238**, 748–776.

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.

Lim, J.-H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B.-Y., Kim, S.-H. & Cho, Y. (1997). *J. Mol. Biol.* **270**, 259–274.

Maruyama, N., Adachi, M., Takahashi, K., Yagasaki, K., Kohno, M., Takenaka, Y., Okuda, E., Nakagawa, S., Mikami, B. & Utsumi, S. (2001). *Eur. J. Biochem.* **268**, 3595–3604.

Maruyama, N., Katsube, T., Wada, Y., Oh, M. H., Barba de la Rosa, A. P., Okuda, E., Nakagawa, S. & Utsumi, S. (1998). *Eur. J. Biochem.* **258**, 854–862.

Maruyama, N., Maruyama, Y., Tsuruki, T., Okuda, E., Yoshikawa, M. & Utsumi, S. (2003). *Biochem. Biophys. Acta*, **1648**, 99–104.

Maruyama, N., Mohamad Ramlan, M. S., Takahashi, K., Yagasaki, K., Goto, H., Hontani, N., Nakagawa, S. & Utsumi, S. (2002*a*). *J. Am. Oil Chem. Soc.* **79**, 139–144.

Maruyama, N., Mohamad Ramlan, M. S., Takahashi, K., Yagasaki, K., Goto, H., Hontani, N., Nakagawa, S. & Utsumi, S. (2002*b*). *J. Agric. Food Chem.* **50**, 4323–4326.

Maruyama, N., Sato, R., Wada, Y., Matsumura, Y., Goto, H., Okuda, E., Nakagawa, S. & Utsumi, S. (1999). *J. Agric. Food Chem.* **47**, 5278–5284.

Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.

Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.

Morita, S., Fukase, M., Yamaguchi, M., Fukuda, Y. & Morita, Y. (1996). *Biosci. Biotech. Biochem.* **60**, 866–873.

Nagi, A. D. & Regan, L. (1997). *Fold. Des.* **2**, 67–75.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.

Pace, C. N. (1992). *J. Mol. Biol.* **226**, 29–35.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

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

Sirtori, C. R., Lovati, M. R., Manzoni, C., Monetti, M., Pazzucconi, F. & Gatti, E. (1995). *J. Nutr.* **125**, 598S-605S.

Spassov, V. Z., Karshikoff, A. D. & Ladenstein, R. (1995). *Protein Sci.* **4**, 1516–1527.

Takano, K., Hunahashi, J., Yamagata, Y., Fujii, S. & Yutani, K. (1997). *J. Mol. Biol.* **274**, 132–142.

Utsumi, S. (1992). *Adv. Food Nutr. Res.* **36**, 89–208.

Utsumi, S., Matsumura, Y. & Mori, T. (1997). *Food Proteins and Their Applications*, edited by S. Damodaran & A. Paraf, pp. 257–291. New York: Marcel Dekker.

Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.

Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R. & Consalvi, V. (1995). *Structure*, **3**, 1147–1158.