Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Thermus thermophilus* HB8 H-protein of the glycine-cleavage system, resolved by a six-dimensional molecular-replacement method

Tadashi Nakai,[a]* Jun Ishijima,[a] Ryoji Masui,[a,b] Seiki Kuramitsu[a,b] and Nobuo Kamiya[a]

[a]RIKEN Harima Institute/SPring-8, 1-1-1 Kouto, Mikazuki, Sayo-gun, Hyogo 679-5148, Japan, and [b]Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

Correspondence e-mail: nakaix@spring8.or.jp

Received 3 March 2003
Accepted 4 July 2003

PDB Reference: *T. thermophilus* H-protein, 1onl, r1onlsf.

The glycine-cleavage system is a multi-enzyme complex consisting of four different components (the P-, H-, T- and L-proteins). Recombinant H-protein corresponding to that from *Thermus thermophilus* HB8 has been overexpressed, purified and crystallized. Synchrotron radiation from BL44B2 at SPring-8 was used to collect a native data set to 2.5 Å resolution. The crystals belonged to the hexagonal space group $P6_5$ and contained three molecules per asymmetric unit, with a solvent content of 39%. Because of the large number of molecules within a closely packed unit cell, this structure was solved by six-dimensional molecular replacement with the program *EPMR* using the pea H-protein structure as a search model and was refined to an R factor of 0.189 and a free R factor of 0.256. Comparison with the pea H-protein reveals two highly conserved regions surrounding the lipoyl-lysine arm. Both of these regions are negatively charged and each has additional properties that are conserved in H-proteins from many species, suggesting that these regions are involved in intermolecular interactions. One region has previously been proposed to constitute an interaction surface with T-protein, while the other may be involved in an interaction with P-protein. Meanwhile, the lipoyl-lysine arm of the *T. thermophilus* H-protein was found to be more flexible than that of the pea H-protein, supporting the hypothesis that H-protein does not form a stable complex with L-protein during the reaction.

1. Introduction

The glycine-cleavage system (GCS) is a multienzyme complex consisting of four different components (the P-, H-, T- and L-proteins). This complex catalyzes the oxidative cleavage of glycine in a multistep reaction (Fig. 1). First, the P-protein catalyzes the decarboxylation of the glycine molecule, concomitantly with the transfer of the residual methylamine group to the distal S atom on the lipoyl group of the oxidized H-protein (H_{ox}). This generates a methylamine-loaded H-protein (H_{met}) (Fig. 1*a*). Next, the T-protein catalyzes the transfer of a methylene group from H_{met} to tetrahydrofolate, resulting in the release of NH_3 and the generation of reduced H-protein (H_{red}) and methylenetetrahydrofolate (Fig. 1*b*). Finally, the dihydrolipoyl group of H_{red} is oxidized by the L-protein and H_{ox} is regenerated, thereby completing the catalytic cycle (Fig. 1*c*) (Motokawa *et al.*, 1995).

Crystal structures of two of the four components of this system have been determined: H-protein from pea (Pares *et al.*, 1994) and the L-proteins from pea (Faure *et al.*, 2000) and seven other species (Mattevi *et al.*, 1991, 1992, 1993; Mande *et al.*, 1996; Li de la Sierra *et al.*, 1997; Toyoda, Kobayashi *et al.*, 1998; Toyoda, Suzuki *et al.*, 1998). We have recently over-

expressed and purified every component of the GCS from an extremely thermophilic bacterium, *Thermus thermophilus* (*Tth*) HB8 (Nakai *et al.*, unpublished results). We have also crystallized and carried out preliminary X-ray characterization of the *Tth* P-protein (Nakai *et al.*, 2003) as well as the *Tth* H- and L-proteins (Nakai *et al.*, unpublished results), with the goal of determining the structures of each component and their various complexes.

In the GCS, H-protein, a monomeric protein (~14 kDa), plays a pivotal role by interacting successively with the three other component enzymes through a lipoic acid that is covalently bound to Lys63 (Fig. 1*a*), termed the lipoyl-lysine arm. Intermolecular interactions between H-protein and each of the other components have been investigated by various methods: those between H- and P-proteins by biochemical experiments (Neuburger *et al.*, 2000), between H- and T-proteins by small-angle X-ray scattering (Cohen-Addad *et al.*, 1997), cross-linking experiments (Okamura-Ikeda *et al.*, 1999) and NMR spectroscopy (Guilhaudis *et al.*, 2000) and between H- and L-proteins by biochemical experiments (Neuburger *et al.*, 2000), small-angle X-ray scattering and X-ray crystallography (Faure *et al.*, 2000). However, no three-dimensional structure has yet been reported for any complexes formed by these components.

Meanwhile, the crystal structures of H-protein in four different forms [oxidized, intermediate-loaded (methylamine-loaded), reduced and apo forms] have been reported (Pares *et al.*, 1994; Cohen-Addad *et al.*, 1995; Faure *et al.*, 2000; Macherel *et al.*, 1996). Every form of the H-protein that has been structurally analyzed so far was isolated from pea-leaf mitochondria, although the GCS is widely distributed in bacteria and in the mitochondria of plants and mammals (Okamura-Ikeda *et al.*, 1993; Douce *et al.*, 2001; Kikuchi & Hiraga, 1982).

The structure determination of an H-protein from another organism, in particular from a distantly related species such as a bacterium, should provide an important contribution to general understanding of the GCS. Moreover, the subsequent structural comparison with the pea H-protein will provide a new perspective on the interaction between H-protein and the other GCS-component enzymes. Here, we report the crystal structure of a bacterial H-protein from *Tth* at 2.5 Å resolution. We then compare this structure with its mitochondrial counterpart from pea and discuss the regions that are likely to be involved in intermolecular interactions based on their structural conservation.

In the course of this study, we found that this structure could not be solved by conventional molecular-replacement (MR) using the pea H-protein as a search model, even though there were no gaps in the sequence alignment and its sequence identity with the *Tth* H-protein was 44%. This was therefore a very valuable test case to evaluate exactly what it is that causes difficulties in solving this kind of MR problem. In this paper, we include the detailed MR process using the six-dimensional MR program *EPMR* by which this structure was eventually solved efficiently.

2. Materials and methods

2.1. Protein expression

General construction of the plasmids was carried out in *Escherichia coli* strain DH5α. The recombinant plasmid pT7Blue-*gcvH*, which contains the gene for *Tth* H-protein, was supplied by the RIKEN Structural Genomics Initiative (Yokoyama *et al.*, 2000). pT7Blue-*gcvH* was digested with *Nde*I and *Bgl*II and the resulting DNA fragment containing the gene was inserted between *Nde*I and *Bam*HI sites in pET-11a (Novagen). The resulting plasmid pET11a-*gcvH* was used to transform the expression host *E. coli* BL21(DE3). Cultivation and target-gene expression were performed according to the method of Macherel *et al.* (1996). Cells were grown at 310 K in Luria–Bertani medium containing 50 μg ml^{-1} ampicillin. These cultures were induced at an

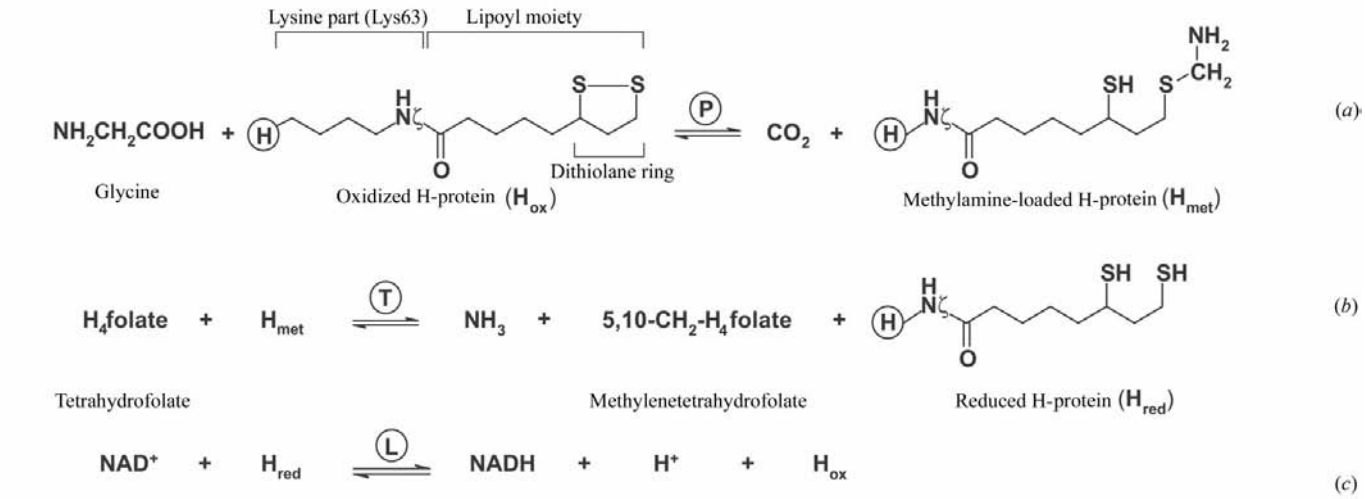

Figure 1
The multistep reaction catalyzed by the glycine-cleavage system. H, P, T and L in the circles represent the respective proteins.

optical density (at 600 nm) of 0.4–0.6 using 1.0 m*M* isopropyl-β-D-thiogalactopyranoside and were grown for 2 h. Chloramphenicol was then added (150 μg ml^{-1} final concentration) to block the *de novo* synthesis of H-protein. After 5 min, lipoic acid (100 μg ml^{-1} final concentration) was added and the culture was incubated for 1 h; the cells were then harvested and stored at 253 K.

2.2. Protein purification

Unless noted otherwise, proteins were purified at room temperature. Frozen cells (10 g from 10 l of culture) were thawed, suspended in 5 m*M* MES–NaOH buffer pH 6.0 containing 5 m*M* 2-mercaptoethanol and 500 m*M* NaCl and then disrupted by sonication. The cell lysate was incubated at 333 K for 10 min, kept on ice for 12 min and then ultracentrifuged (200 000*g*) for 60 min at 277 K and desalted on a HiPrep 26/10 desalting column (Amersham Biosciences) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0. The resulting supernatant was applied to a SuperQ-Toyopearl column (20 ml; Tosoh) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0. Protein was eluted with a linear gradient of 0–1.0 *M* NaCl in the same buffer. Fractions containing the target protein were collected and desalted on a HiPrep 26/10 desalting column (Amersham Biosciences) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0. The sample was applied to a Resource Q column (6 ml; Amersham Biosciences) equilibrated in the same buffer; protein was then eluted with a linear gradient of 0–700 m*M* NaCl. Fractions containing the target protein were then loaded onto a HiLoad 16/60 Superdex 75 prep-grade column (Amersham Biosciences) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 and 150 m*M* NaCl and eluted with the same buffer. After the addition of 1 m*M* dithiothreitol (DTT), the peak fractions were concentrated and stored at 277 K. At each step, the fractions were analyzed by SDS–PAGE with a 15%(*w*/*v*) acrylamide gel.

2.3. Crystallization

Preliminary crystallization conditions for *Tth* H-protein were determined with Crystal Screens 1 and 2 (Hampton Research) using the hanging-drop vapour-diffusion method at 291 K. Several crystal forms were obtained and the most promising set of crystallization conditions, corresponding to Crystal Screen 1 solution No. 3, was optimized. A droplet containing 2 μl of protein solution [10 mg ml^{-1} protein, 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl and 1 m*M* DTT] was mixed with an equal volume of reservoir solution [0.4 *M* ammonium phosphate, 24%(*v*/*v*) glycerol pH 4.3] and equilibrated against 400 μl of reservoir solution to give crystals of *Tth* H-protein. The crystals appeared within a few days and grew to maximum dimensions of 0.1 × 0.1 × 0.3 mm. Mass-spectrometric analysis of washed and dissolved crystals was performed using a Voyager DE-STR matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (PE Biosystems).

Table 1
Data-collection and refinement statistics.

Values in parentheses correspond to reflections observed in the highest resolution shell.

Diffraction data	
Resolution range (Å)	20.0–2.50 (2.58–2.50)
No. of measured reflections	135290
No. of unique reflections	11730
Redundancy	11.5 (5.4)
Completeness (%)	99.6 (98.5)
$R_{\rm merge}$† (%)	10.9 (35.5)
Average $I/\sigma(I)$	13.7 (2.5)
Refinement	
Resolution range (Å)	20.0–2.50 (2.66–2.50)
R factor	0.189 (0.243)
Free *R* factor	0.256 (0.320)
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.006
Bond angles (°)	1.2
Mean *B* factors	
Main-chain atoms (Å^2)	26.0
Side-chain atoms (Å^2)	27.4
Water atoms (Å^2)	30.5

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_i I_{hkl,i}$, where I is the observed intensity and $\langle I\rangle$ is the averaged intensity for multiple measurements.

2.4. Data collection

A crystal of *Tth* H-protein was transferred to a nylon cryoloop (Hampton Research) and flash-cooled in a nitrogen-gas stream at 100 K. X-ray diffraction data were collected at 100 K on a MAR CCD detector using synchrotron radiation with a wavelength of 1.00 Å from BL44B2 at SPring-8 (Adachi *et al.*, 2001). The crystal belonged to the hexagonal system, space group $P6_1$ or $P6_5$, with unit-cell parameters $a = b = 55.8$, $c = 191.2$ Å. The presence of three monomeric molecules per asymmetric unit gives a crystal volume per protein mass ($V_{\rm M}$) of 2.03 Å^3 Da^{-1} and a solvent content of 39% by volume. The data processing was completed using the program *HKL*2000 (Otwinowski & Minor, 1997) (Table 1).

2.5. Structure determination

The MR program *EPMR* (Kissinger *et al.*, 1999, 2001) was used to solve the structure of the *Tth* H-protein using the pea H-protein (PDB code 1hpc; Pares *et al.*, 1994) as a search model, in which all non-glycine residues were truncated to alanine. All data in the resolution range 15–4 Å (2791 reflections) were used. Each evolutionary search was carried out over 50 generations using a population size of 300, as described by Kissinger *et al.* (1999). 200 runs were used in each search procedure.

2.5.1. The first MR solution. As a first step in MR, the correct enantiomeric space group of the crystal was determined according to the procedure used in the case of cytochrome c' (Kissinger *et al.*, 1999). In the published application, one molecule was contained in the asymmetric unit and the space group was $P6_522$. In contrast, in our case three molecules were contained in the asymmetric unit and the space group was $P6_5$. As shown in Fig. 2(*a*), the highest correlation coefficients (CCs) obtained for $P6_1$ and $P6_5$ were 0.187 and

0.213, respectively, indicating the correct space group to be $P6_5$. This result confirmed that *EPMR* was superior in identifying the correct space group, even in the case of an asymmetric unit containing three molecules.

Having established the correct space group, the search procedure was performed and resulted in two different sets of correct solutions having CCs of 0.213 (R factor = 0.580) and 0.201 (R factor = 0.587). Defining the rotation and translation operator as op$X$ = (θ_1, θ_2, θ_3, x, y, z), where X is an operator identifier, θ_1, θ_2 and θ_3 are rotation parameters specified in Eulerian angles (°) and x, y and z are translation parameters in orthogonal coordinates (Å), the operator for the former set of solutions (op$A$) was (59.17, 73.65, 49.70, 0.66, 21.70, 0.00) and that for the latter (op$B$) was (58.31, 58.16, 198.33, 33.35, 31.51, 0.00). The highest CC obtained for an incorrect solution was 0.187, indicating that the correct solutions could clearly be identified above the background. It should be noted that the highest CC obtained in space group $P6_1$ was also 0.187.

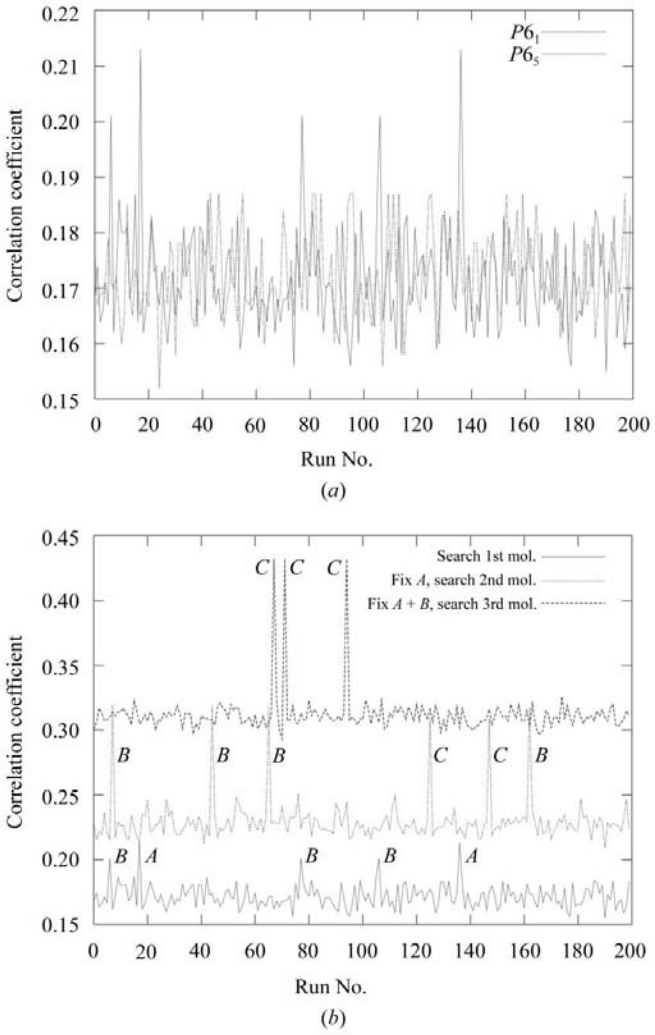

Figure 2
Results of six-dimensional MR searches carried out with the program *EPMR*. (*a*) Search for the first molecule in the two different enantiomeric space groups and (*b*) sequential searches for three molecules in the asymmetric unit.

2.5.2. The second and third MR solutions. The search for the second molecule was performed according to the procedure described for FK506-binding protein (Kissinger *et al.*, 1999), in which two molecules were contained in an asymmetric unit. Meanwhile, to our knowledge this is the first report demonstrating that *EPMR* can be applied successfully to MR problems in which the target crystal structure contains more than two molecules per asymmetric unit.

The partial contribution of one of the first solutions (op$A$) was included as part of the structure factors during the search for the second molecule (Fig. 2*b*). This search procedure gave two different sets of second solutions, with CCs for the combined solutions of 0.319 (R factor = 0.554) and 0.309 (R factor = 0.553). The first operator (opB') was (57.95, 60.05, 199.23, 33.35, 31.24, 144.13). This corresponded to the op$B$ operator in which the value of z differed from that in opB' as a result of choosing one common origin in the z coordinate. The second operator (op$C$) was (78.26, 56.43, 266.51, 18.39, 37.21, 55.37).

The partial contributions of the first and the second solutions (op$A$ and opB') were part of the structure factors during the search for the third molecule. The search procedure gave a set of equivalent third solutions, with a CC for the combined solutions of 0.432 (R factor = 0.512). The orientation and position of this solution corresponded to op$C$. The three molecules to which rotation and translation operators op$A$, opB' and op$C$ have been applied are defined as molecules A, B and C, respectively.

The total run time for the three search procedures was about 1536 min of CPU time on a PC with two AMD Athlon 1.6 GHz CPUs, 512 MB memory and running RedHat Linux 7.2. Only one CPU was used at a time, although the PC had two CPUs.

2.6. Structure refinement

The molecular-replacement electron-density map was of good quality and the model of the molecule was gradually built into the map through several cycles of model building using the program *Xfit* within the software package *XtalView* (McRee, 1992). As three molecules were contained in the asymmetric unit, early cycles of refinement employed strict non-crystallographic symmetry (NCS) constraints with only a single-molecule model. In the initial model, the side chains that were identical between pea and *Tth* H-protein were included, while a polyalanine model was used for molecular replacement. Structure refinement was carried out by simulated annealing and energy minimization with the program *CNS* (Brünger *et al.*, 1998), using X-ray data from 20.0 to 2.5 Å resolution, to give an R factor of 0.363 and a free R factor of 0.381. At this stage, the strict NCS constraints were removed and after several cycles of refinement and manual rebuilding the R factor and free R factor were reduced to 0.206 and 0.260, respectively. Water molecules were picked up in the difference map based on peak heights and distance criteria. Further model-building and refinement cycles resulted in an R factor of 0.189 and a free R factor of 0.256, using 11 137 reflections

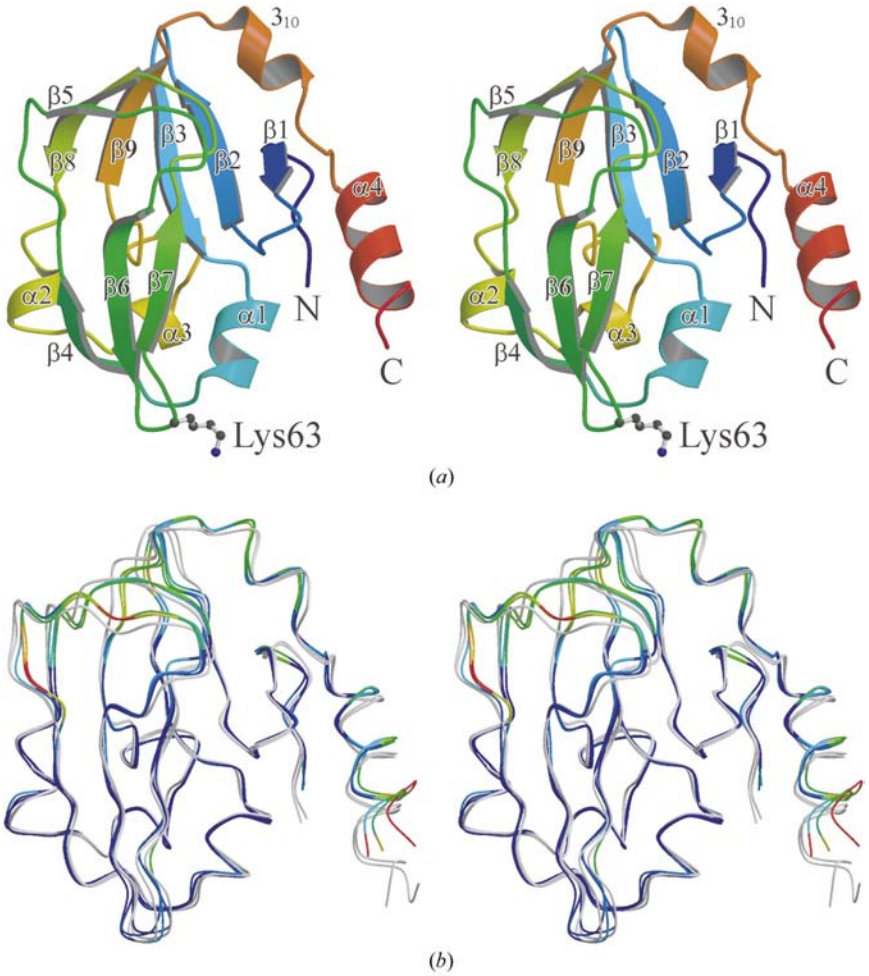

Figure 3
Stereoview of the structures of H-protein. (*a*) The structure of the *Tth* H-protein with secondary-structure assignment. The structure is coloured from the N-terminus (blue) to the C-terminus (red). (*b*) Superimposition of the *Tth* (coloured) and pea (grey) H-protein structures by least-squares fitting of main-chain atoms. Displacements for equivalent C^{α} atoms between *Tth* and pea H-proteins are represented by a colour gradient from blue (<0.56 Å) to red (>2.24 Å). Since the model of pea H-protein in the oxidized form (PDB code 1hpc) contains two molecules per asymmetric unit, two sets of displacements were calculated between each of the molecules *A*–*C* of the *Tth* H-protein and the two molecules of the pea H-protein. The shorter value of the displacement for each set of equivalent C^{α} atoms was selected for colour mapping.

$[F_o > 2\sigma(F_o)]$ observed in the 20.0–2.5 Å resolution range (Table 1). Fig. 3(*a*) displays a C^{α} backbone tracing of the *Tth* H-protein in ribbon-model drawings.

2.7. Programs for data analysis and graphical representation

Secondary structures were assigned using the program *DSSP* (Kabsch & Sander, 1983). Stereochemistry was assessed using the program *PROCHECK* (Laskowski *et al.*, 1993). Ribbon drawings, electron densities and electrostatic potential surfaces were generated using *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1997) and *GRASP* (Nicholls *et al.*, 1991), respectively, and were rendered with *Raster3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Structure determination

Structure determinations using MR methods were originally attempted using the programs *AMoRe* (Navaza, 1994) and *X-PLOR* (Brünger, 1992) with Patterson correlation refinement (Brünger, 1990). In attempting to apply these classical methods to the data described here, no reasonable solution was found using the cross-rotation search. This was true even when another search was carried out subsequent to the structure determination with *EPMR*, using the refined model of the *Tth* H-protein as the search model. This can probably be attributed to the large number of symmetry elements and molecules in the unit cell and the relatively low (∼39%) solvent content.

In the conventional MR method, the rotational and translational parameters of the search model are treated separately and an independent determination of their values is attempted. This normally results in a reduction of the signal-to-noise ratio, which is remarkable in cases where the unit cell has a large number of molecules and/or has a low solvent content. In contrast, novel MR algorithms such as that used in *EPMR* determine the rotational and translational parameters simultaneously, resulting in an improved signal-to-noise ratio (Kissinger *et al.*, 1999).

From a technical viewpoint, the results of these successive searches with *EPMR* have confirmed the importance of carrying out a number of runs, as described by Kissinger *et al.* (2001). As shown in Fig. 2(*b*), the *EPMR* search led to successful solutions for all three molecules in the asymmetric unit. The correct solutions for the first, second and third molecules were obtained after only five, six and three out of 200 runs each of the search procedure, respectively. In examining the data in Fig. 2(*b*), we note that if only 100 runs of a third molecule search had been carried out and these runs had corresponded to the latter half of the 200 runs that were actually carried out for the third molecule search, we would not have obtained any solution. This is becuase of the nature of stochastic search algorithms, as explained by Kissinger *et al.* (2001). However, by increasing the total number of runs, the likelihood of obtaining correct solutions is markedly increased, even in cases such as the difficult MR problem described here.

Upon completing the search for the first molecule the solution corresponding to op*C* was not obtained, whereas solutions corresponding to op*A* and op*B*′ were obtained (Fig. 2*b*). The difference Fourier calculation with *EPMR* (the

−g0 flag was used in order to bypass the MR search), in which only the molecule positioned by op*C* was used, yielded a CC of 0.170. This value was not only lower than the CCs corresponding to op*A* and op*B*′, but it was also lower than the highest CC obtained for an incorrect solution. In order to analyze what had caused such a low value to be calculated for a correct solution, we compared the structures of molecules *A*–*C* and the local packing of these molecules. Although the structures and average thermal factors of the molecules are almost the same as described in the next section, the local packing of the three molecules is quantitatively different. The number of neighbouring molecules for molecules *A*, *B* and *C* calculated using van der Waals contacts are eight, nine and 11, respectively. Consequently, the numbers of atoms participating in these intermolecular contacts are 220, 251 and 265, respectively. These values indicate that the CC value corresponding to a particular molecule decreases as the packing density around the molecule increases. If an MR search is carried out in a tightly packed crystal form, intermolecular Patterson vectors (cross-vectors) and intramolecular Patterson vectors (self-vectors) have a tendency to overlap, creating difficulties in resolving these MR problems. Therefore, localized tight packing around a particular molecule may cause difficulties in searching the molecule.

Table 2
R.m.s. deviations (upper right; Å) and maximum displacements (lower left; Å) for equivalent main-chain atoms in pea and *Tth* H-proteins.

Letters in parentheses denote the chain names of the independent molecules in the asymmetric unit.

	Pea (*A*)	Pea (*B*)	*Tth* (*A*)	*Tth* (*B*)	*Tth* (*C*)
Pea (*A*)	—	0.33	0.85	0.97	1.06
Pea (*B*)	2.24	—	0.86	1.02	1.10
Tth (*A*)	2.43	3.04	—	0.55	0.60
Tth (*B*)	4.10	4.14	4.29	—	0.68
Tth (*C*)	5.86	6.33	4.41	6.67	—

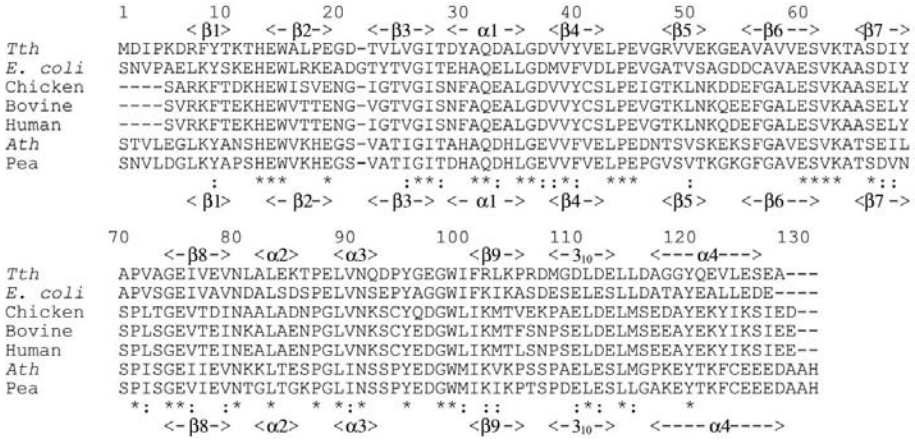

Figure 4
Amino-acid sequence alignments of seven H-proteins with secondary-structure designations for the pea and *Tth* H-proteins. The sequence data were obtained from Swiss-Prot. Identical and homologous residues are marked with asterisks and colons, respectively.

3.2. Structure of *Tth* H-protein

Tth H-protein is a monomeric protein having a molecular weight of 14 083 Da, with 128 amino-acid residues. The final model in the asymmetric unit contains 3 × 127 amino-acid residues and 115 water molecules, with an *R* factor of 0.189 at 2.5 Å resolution. The model lacks three Met1 residues for the three independent molecules (*A*–*C*) owing to disorder of each amino-terminal residue. The structures of molecules *A*–*C* are quite similar; least-squares fitting of main-chain atoms in the three molecules gave root-mean-square (r.m.s.) deviations of 0.55–0.68 Å (Table 2). As the average error of the structure is estimated from a Luzzati plot to be 0.27 Å (Luzzati, 1952), the three molecules have essentially the same three-dimensional structure. The average thermal factors of the main-chain atoms in molecules *A*–*C* are 26.3, 24.9 and 26.9 Å^2, respectively. The model is of good quality, with all the residues falling in the most favourable (83.8%) and additionally allowed (16.2%) regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968).

The overall structure of the *Tth* H-protein is shown in Fig. 3(*a*). The core of this structure consists of two antiparallel β-sheets forming a hybrid barrel–sandwich structure, one having six antiparallel strands (β1, β2, β3, β9, β8 and β5) and the other containing a strand (β4) and two adjacent antiparallel strands (β6 and β7) joined by a loop (β-hairpin motif). At the tip of this loop, the side chain of the critical lysine residue (Lys63) is lipoylated (Fig. 1*a*), although this model lacks the lipoyl moiety owing to poor electron density at that site. There are also four α-helices (α1, α2, α3 and α4) and a 3_{10}-helix (3_{10}).

3.3. Structural comparison between *Tth* and pea H-proteins

A sequence alignment of the *Tth* H-protein and H-proteins from other organisms is shown in Fig. 4. The sequence identities of the *Tth* H-protein with *E. coli*, chicken, bovine, human, *Arabidopsis thaliana* (*Ath*) and pea H-proteins are 55, 38, 41, 40, 41 and 44%, respectively, showing that the sequence of the *Tth* H-protein is most similar to that of the H-protein from *E. coli*, which was the other prokaryotic protein in the alignment. In comparison with the pea H-protein, the only other protein for which a crystal structure has been determined so far, the *Tth* H-protein is seen to have neither deletions nor insertions, except for the shortening of the three C-terminal residues. Thus, the numbering of the residues in the *Tth* H-protein corresponds to that of the pea H-protein. The secondary structures of the *Tth* and pea H-proteins are almost the same except for two residues in the C-terminal α-helix (α4) (Fig. 4).

Local differences in C$^\alpha$ positions between the *Tth* and pea H-proteins are

shown in Fig. 3(*b*). The overall structure of the *Tth* H-protein is similar to that of the pea H-protein; least-squares fitting of main-chain atoms in those proteins gave r.m.s. deviations of 0.85–1.11 Å, with maximum displacements of 2.43–6.33 Å (Table 2). Sequential fitting cycles, in which the most deviant residue was removed from the list of residues in each cycle, provided information on structurally conserved regions in the proteins. The main-chain atoms of residues 4, 7–17, 23–43, 47, 53–61, 64, 66–71, 79–101, 111, 114–118 and 120 of the proteins were fitted, with r.m.s. deviations of 0.55–0.59 Å and maximum displacements of 1.24–1.62 Å. These r.m.s. deviations are comparable with those calculated between the independent molecules in the asymmetric unit (Table 2). These data indicate that the region surrounding the lipoyl-lysine residue, comprised primarily of one β-sheet (β4, β6 and β7) and three α-helices (α1, α2 and α3) around this β-sheet, are particularly conserved in the three-dimensional structures of the *Tth* and pea H-proteins.

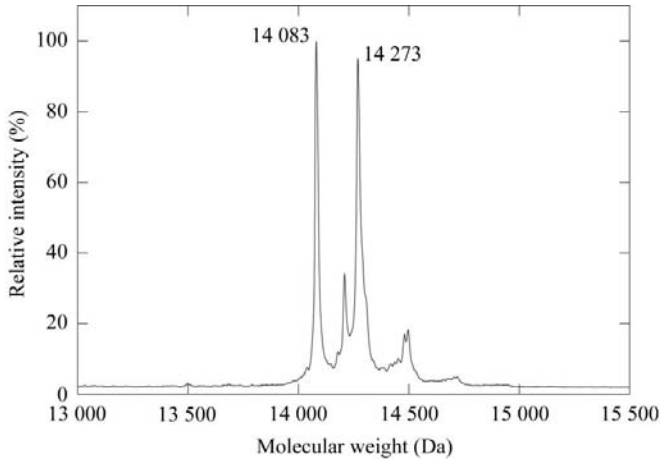

Figure 5
Mass-spectrometric analysis of washed and dissolved crystals. The 14 083 Da and 14 273 Da peaks correspond to unlipoylated and lipoylated proteins, respectively.

3.4. Structure of the region around the lipoyl-lysine

Although mass-spectrometric analysis of thoroughly washed crystals confirmed the lipoylation of the *Tth* H-protein in the crystallized material, the proportion of lipoylated protein was only about 50%, *i.e.* the other half of the protein is unlipoylated (Fig. 5). This indicates that precise comparison of the lipoyl moieties between the *Tth* and pea H-proteins must await further studies of fully lipoylated *Tth* H-protein. Nevertheless, we describe the structure of the lipoyl-lysine arm here because the electron density for it was partial and was significantly visible above the background at the 1σ level. We also use the 'half-lipoylated' *Tth* protein for comparison with the lipoylated pea H-protein because the bound lipoic acid should not play any role in the protein structure, as in the case of the pea H-protein (Macherel *et al.*, 1996).

The electron density for the lipoyl moiety of the lipoyl-lysine attached to the *Tth* H-protein was relatively poor (Fig. 6), which could be attributed in part to the low proportion of lipoylated protein. In particular, for molecules *B* and *C*, no electron densities were observed for the side chains of Lys63 or for the lipoyl moieties. However, in molecule *A* electron density was clearly observed for the side chain of Lys63 and for the amide group (N^{ζ} and C═O in Fig. 1*a*) of the lipoyl-lysine, while density for the dithiolane-ring moiety of the lipoyl-lysine was absent. The electron density for the amide group corresponds to the extended electron density shown in Fig. 6. In contrast to the *Tth* H-protein, the entire lipoyl-lysine is ordered in all crystal forms of the pea H-protein and the lysine residue does not move significantly. The lipoyl moiety of the pea H-protein interacts through van der Waals contacts with the side chain of His34, which is replaced by Ala34 in the *Tth* H-protein; thus, this interaction is absent in the *Tth* H-protein. Therefore, the lipoyl moiety of the *Tth* H-protein is probably more flexible than that of the pea H-protein.

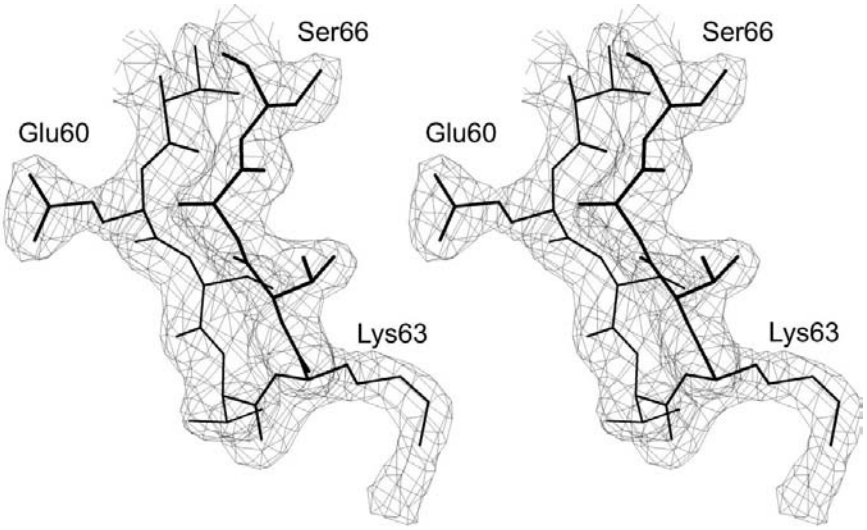

Figure 6
Stereoview of a $2F_o - F_c$ electron-density map contoured at the 1.0σ level around the lipoyl-lysine of molecule *A* (see text). Extended electron density at the distal amino group of Lys63 corresponds to part of the lipoyl moiety which is covalently bound to Lys63 through an amide linkage. The orientation is the same as in Fig. 3.

In the methylamine-loaded form of the H-protein (H_{met}) from pea, the methylamine-loaded lipoyl-lysine arm binds into the cleft formed by the main chains of Ser12 and Asp67 and the side chains of Glu14, Ile27, Ala31, Leu35 and Ala64 (Cohen-Addad *et al.*, 1995). These seven residues are conserved in the *Tth* H-protein, except for Ala64 (Thr64 in *Tth*); the main-chain atoms of the *Tth* and pea H-proteins superimpose with r.m.s. deviations of 0.28–0.61 Å and with maximum displacements of 0.50–1.18 Å. Furthermore, the arrangements of the side chains are also well conserved between them. Therefore, the lipoamide-methylamine arm of *Tth* H_{met} is likely to bind into the cleft, as in the case of

the pea H_{met}, even if the lipoyl-lysine arm of the *Tth* H-protein in its oxidized form can swing freely.

3.5. Implication for regions that may interact with other component enzymes

Since H-protein interacts successively with the three other component enzymes of the GCS and with lipoyltransferase, which catalyzes the lipoylation of the specific lysine residue of H-apoprotein, certain regions in H-protein must be recognized by these enzymes. Meanwhile, a structural comparison between the *Tth* and pea H-proteins reveals structurally conserved regions, suggesting that these regions play important roles in the reaction mechanism, which involves intermolecular interactions. From this viewpoint, we have examined candidates for these kinds of regions in detail.

As shown in Fig. 7, electrostatic potential surfaces were calculated for the *Tth* and pea H-proteins, revealing two negative surface regions, designated I and II, that are highly conserved between the two species. Region I is situated around β4, β6 and β7 (Figs. 3*a*, 7*a* and 7*b*), in which the negative charges are attributed to Glu42, Glu60 and Asp67 in both the *Tth* and pea H-proteins. Of these residues, Glu42 in the *Tth* and pea proteins corresponds to Asp43 in the *E. coli* protein, which has been shown by cross-linking experiments to participate in the interaction with its cognate T-protein (Okamura-Ikeda *et al.*, 1999). The negative charges of Glu60 and Asp67 are completely conserved among the seven species (Fig. 4), suggesting that these residues play an important role in the GCS. These two negative residues correspond to Glu56 and Glu63 of bovine H-protein; the latter residues are necessary for lipoylation catalyzed by lipoyltransferase (Fujiwara *et al.*, 1991, 1996). Recently, using NMR spectroscopy, Guilhaudis *et al.* (2000) have shown the residues involved in intermolecular interactions between H- and T-proteins and these interactions include Glu60 and Asp67. Thus, the observation that the three-dimensional structures and electrostatic surfaces of region I are conserved between the two H-proteins are additional indications of interacting regions between H-protein and the relevant enzymes, at least in the case of T-protein and lipoyltransferase.

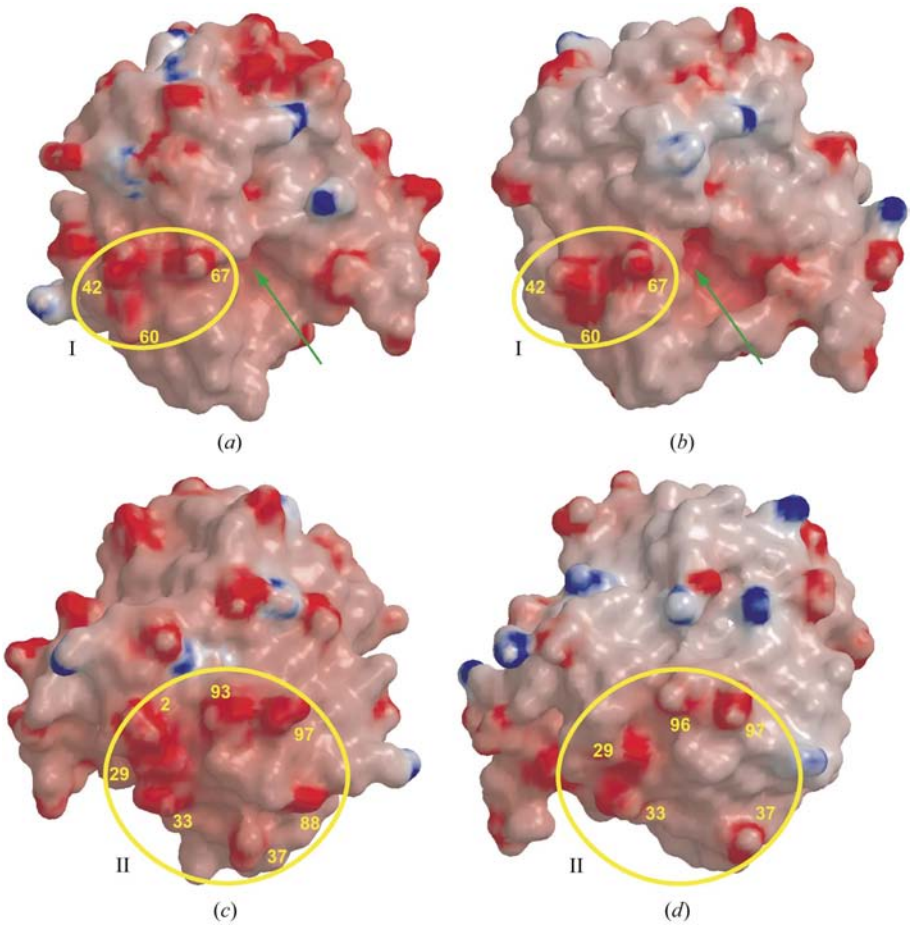

Figure 7
Electrostatic potential surfaces for the *Tth* H-protein (*a* and *c*) and for the pea H-protein (*b* and *d*). The negative (red), positive (blue) and neutral (white) charges are mapped onto the van der Waals surfaces in the range $-20k_BT$ (red) to $+20k_BT$ (blue), where k_B is Boltzmann's constant and T is the absolute temperature. The orientations in (*a*) and (*b*) are the same as in Fig. 3 and those in (*c*) and (*d*) are rotated by 180° around a vertical axis in the plane of the figure. Yellow ellipses indicate the electrostatically conserved regions, denoted I and II. Residue numbers of the negative residues situated in these regions are shown in yellow. Green arrows indicate the clefts into which the methylamine-loaded lipoyl-lysine arm of H_{met} binds.

On the other hand, region II is situated around α1, α3 and β4 (Figs. 3*a*, 7*c* and 7*d*), in which the negative charges are attributed to Asp2 in *Tth*, Asp29, Asp33, Asp37 (Glu37 in pea), Glu88 in *Tth*, Asp93 in *Tth*, Glu96 in pea and Glu97 (Asp97 in pea). In region II, the negative charges of residues Asp33 (or Glu) and Asp37 (or Glu) are completely conserved among the seven species (Fig. 4), suggesting that these residues may also play an important role in the GCS. In contrast to the situation with region I, there is no further information on this region concerning intermolecular interactions. Nevertheless, we expect that region II may be involved in interactions with other component enzymes because the tendency for the negative residues in region II to form a cluster in the *Tth* H-protein is similar to that in the pea H-protein and a similar cluster forms in region I in both the *Tth* and pea H-proteins. In the case of region II, the most likely candidate for an interaction partner would be P-protein for three reasons: firstly, molecular recognition and interaction between H- and P-proteins has been suggested by biochemical experiments, in which H-apoprotein was found to behave as a competitive inhibitor relative to H-protein in the glycine decarboxylation reaction catalyzed by P-protein (Neuburger *et al.*, 2000); secondly,

region II is not likely to be involved in the interaction between H- and T-proteins because upon complex formation between H- and T-proteins, significant chemical shift variations were not observed in this region using NMR spectroscopy (Guilhaudis *et al.*, 2000); thirdly, H-protein is unlikely to form a stable complex with L-protein, for reasons described in the next paragraph.

In contrast to the theories regarding interactions with P- or T-protein, Faure *et al.* (2000) have strongly suggested that L-protein does not form a stable complex with H-protein during the reaction. As evidence for the absence of complex formation, they demonstrated that the apparent molecular mass of mixtures of L- and H-proteins did not increase to the expected value for the complex. They further showed that the structure of the lipoyl-lysine arm of pea H_{red} found in the crystal could not penetrate into the active site of the L-protein unless the arm moves. On the other hand, the lipoyl-lysine arm of *Tth* H_{red} is likely to be more flexible than those of pea H_{red} and H_{ox}, because His34 of pea H-protein is replaced by Ala34 in *Tth*, as described above. The highly flexible lipoyl-lysine arm of *Tth* H-protein would allow it to swing into the active site of L-protein and this may be more appropriate for the reaction than a rigid arm. This also suggests that there would not be a strong interaction between *Tth* H- and L-proteins, for the same reason given for the pea proteins: owing to the shape and size of the L-protein active site, the arm is required to be oriented in specific directions during the reaction, in which case, as described by Faure *et al.* (2000), there would be no strong interaction between the two proteins. We have recently analyzed the three-dimensional structure of the *Tth* L-protein (data not shown) and have confirmed that the structure of its active site is essentially identical to that of the pea L-protein. Thus, the highly flexible arm observed in the *Tth* H-protein is consistent with the absence of a stable complex being formed between the H- and L-proteins.

We thank Dr Takaaki Hikima and Mr Taiji Matsu of the Division of Bio-Crystallography Technology, RIKEN Harima Institute for their help with data collection at SPring-8. This work was supported in part by the Special Postdoctoral Researchers Program at RIKEN to TN and by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (13780495) to TN.

References

Adachi, S., Oguchi, T., Tanida, H., Park, S. Y., Shimizu, H., Miyatake, H., Kamiya, N., Shiro, Y., Inoue, Y., Ueki, T. & Iizuka, T. (2001). *Nucl. Instrum. Methods Phys. Res. A*, **467**, 711–714.

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T. (1992). *X-PLOR. Version* 3.1. *A System for X-ray Crystallography and NMR*. Yale University, Connecticut, USA.

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**, 905–921.

Cohen-Addad, C., Faure, M., Neuburger, M., Ober, R., Sieker, L., Bourguignon, J., Macherel, D. & Douce, R. (1997). *Biochimie*, **79**, 637–643.

Cohen-Addad, C., Pares, S., Sieker, L., Neuburger, M. & Douce, R. (1995). *Nature Struct. Biol.* **2**, 63–68.

Douce, R., Bourguignon, J., Neuburger, M. & Rébeillé, F. (2001). *Trends Plant Sci.* **6**, 167–176.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Faure, M., Bourguignon, J., Neuburger, M., MacHerel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C. & Douce, R. (2000). *Eur. J. Biochem.* **267**, 2890–2898.

Fujiwara, K., Okamura-Ikeda, K. & Motokawa, Y. (1991). *FEBS Lett.* **293**, 115–118.

Fujiwara, K., Okamura-Ikeda, K. & Motokawa, Y. (1996). *J. Biol. Chem.* **271**, 12932–12936.

Guilhaudis, L., Simorre, J. P., Blackledge, M., Marion, D., Gans, P., Neuburger, M. & Douce, R. (2000). *Biochemistry*, **39**, 4259–4266.

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

Kikuchi, G. & Hiraga, K. (1982). *Mol. Cell Biochem.* **45**, 137–149.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.

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

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

Li de la Sierra, I., Pernot, L., Prange, T., Saludjian, P., Schiltz, M., Fourme, R. & Padron, G. (1997). *J. Mol. Biol.* **269**, 129–141.

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

Macherel, D., Bourguignon, J., Forest, E., Faure, M., Cohen-Addad, C. & Douce, R. (1996). *Eur. J. Biochem.* **236**, 27–33.

McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.

Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N. & Hol, W. G. J. (1996). *Structure*, **4**, 277–286.

Mattevi, A., Obmolova, G., Kalk, K. H., van Berkel, W. J. & Hol, W. G. (1993). *J. Mol. Biol.* **230**, 1200–1215.

Mattevi, A., Obmolova, G., Sokatch, J. R., Betzel, C. & Hol, W. G. (1992). *Proteins*, **13**, 336–351.

Mattevi, A., Schierbeck, A. J. & Hol, W. G. (1991). *J. Mol. Biol.* **220**, 975–994.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Motokawa, Y., Fujiwara, K. & Okamura-Ikeda, K. (1995). *Bithiols in Health and Disease*, edited by L. Packer & E. Cadenas, pp. 389–407. New York: Marcel Dekker.

Nakai, T., Nakagawa, N., Maoka, N., Masui, R., Kuramitsu, S. & Kamiya, N. (2003). *Acta Cryst.* D**59**, 554–557.

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

Neuburger, M., Polidori, A. M., Pietre, E., Faure, M., Jourdain, A., Bourguignon, J., Pucci, B. & Douce, R. (2000). *Eur. J. Biochem.* **267**, 2882–2889.

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

Okamura-Ikeda, K., Fujiwara, K. & Motokawa, Y. (1999). *Eur. J. Biochem.* **264**, 446–453.

Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K. & Motokawa, Y. (1993). *Eur. J. Biochem.* **216**, 539–548.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. & Douce, R. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4850–4853.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

Toyoda, T., Kobayashi, R., Sekiguchi, T., Koike, K., Koike, M. & Takenaka, A. (1998). *Acta Cryst.* D**54**, 982–785.

Toyoda, T., Suzuki, K., Sekiguchi, T., Reed, L. J. & Takenaka, A. (1998). *J. Biochem.* **123**, 668–674.

Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.