Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 4850–4853, May 1994
Biochemistry

X-ray structure determination at 2.6-Å resolution of a lipoate-containing protein: The H-protein of the glycine decarboxylase complex from pea leaves

SERGE PARES*, CLAUDINE COHEN-ADDAD*†, LARRY SIEKER*‡, MICHEL NEUBURGER§, AND ROLAND DOUCE§

*Institut de Biologie Structurale, Centre National de la Recherche Scientifique et Commissariat à l'Energie Atomique, 41 av. des Martyrs, 38027 Grenoble Cedex, France; ‡Department of Biological Structure, SM 20, University of Washington, Seattle, WA 98195; and §Physiologie Cellulaire Végétale, Unité de Recherche Associée Centre National de la Recherche Scientifique, n°576, Departement de Biologie Moléculaire et Structurale/Commissariat à l'Energie Atomique, F38054, Grenoble Cedex 9, France

Communicated by Pierre Joliot, February 1, 1994 (received for review November 2, 1993)

ABSTRACT **H-protein, a lipoic acid-containing protein of the glycine decarboxylase (EC 1.4.4.2) complex from pea (*Pisum sativum*) was crystallized from ammonium sulfate solution at pH 5.2 in space group $P3_121$. The x-ray crystal structure was determined to 2.6-Å resolution by multiple isomorphous replacement techniques. The structure was refined to an *R* value of 23% for reflections between 15- and 2.6-Å resolution ($F > 2\sigma$), including the lipoate moiety and 50 water molecules, for the two protein molecules of the asymmetric unit. The 131-amino acid residues form seven β-strands arranged into two antiparallel β-sheets forming a "sandwich" structure. One α-helix is observed at the C-terminal end. The lipoate cofactor attached to Lys-63 is located in the loop of a hairpin configuration. The lipoate moiety points toward the residues His-34 and Asp-128 and is situated at the surface of the H-protein. This allows the flexibility of the lipoate arm. This is the first x-ray determination of a lipoic acid-containing protein, and the present results are in agreement with previous theoretical predictions and NMR studies of the catalytic domains of lipoic acid- and biotin-containing proteins.**

During the course of the C_2 cycle (photorespiration) in leaf cells, glycine molecules formed in the peroxisomes are immediately cleaved by a complex of proteins (glycine decarboxylase multienzyme complex known also as the glycine cleavage system) localized in the mitochondrial matrix (1, 2). Glycine decarboxylase [glycine: lipoylprotein oxidoreductase (decarboxylating and acceptor-aminomethylating), EC 1.4.4.2] has been purified from pea leaf mitochondria (3) and consists of four protein components, which have been tentatively named "P-protein" (a homodimer of 114-kDa peptides containing pyridoxal phosphate), "H-protein" (a 14.1-kDa lipoamide-containing protein), "T-protein" [a monomer of 41 kDa catalyzing the tetrahydrofolate-polyglutamate (H_4FGlu_n)-dependent step of the reaction], and "L-protein" (lipoamide dehydrogenase; a homodimer of 50-kDa peptides containing FAD). The glycine decarboxylase complex from plant leaf mitochondria is closely related to similar enzyme complexes found in bacteria, including *Ptetococcus glycinophilus* (4) and *Athrobacter globiformis* (5), and in the mitochondria of animal tissues (6). In humans, its absence due to genetic deficiency leads to a dramatic accumulation of glycine in the blood and to severe neurological diseases (7). The glycine cleavage system, which is linked to serine hydroxymethyltransferase, comprises about half of the soluble proteins in the mitochondria from fully expanded green leaves (2, 8). This is in contrast to the situation observed in mammalian mitochondria, where glycine decarboxylase is only a small fraction of the total matrix protein (6).

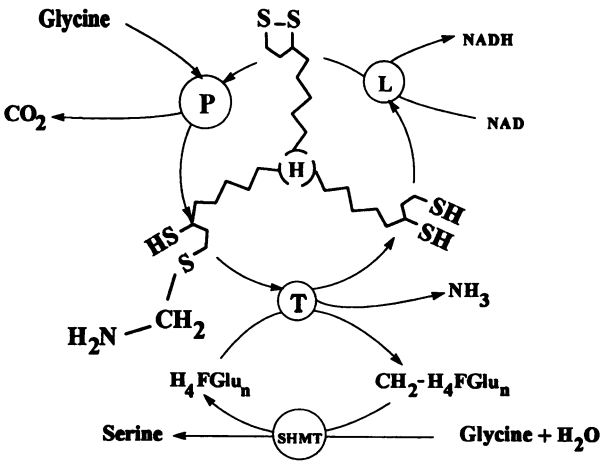

FIG. 1. Scheme outlining the reactions involved in oxidative decarboxylation and deamination of glycine in plant mitochondrial matrix. P, H, T, and L are the protein components of the glycine decarboxylase system. The serine hydroxymethyltransferase (SHMT) is involved in the recycling of methylenetetrahydrofolate-polyglutamate ($CH_2H_4FGlu_n$; *n* = numbers of glutamate residues) into H_4FGlu_n.

The H-protein provides the attachment site for the lipoic acid via an amide linkage to the ε-amino group of a lysine residue [lysine-63 on the pea H-protein (9)]. The lipoate moiety very likely interacts in a flexible manner with each of the other three proteins (P-, T-, and L-proteins) in the multienzyme complex (Fig. 1). The structural and mechanistic heart of the glycine cleavage complex is provided by H-protein. We have isolated and characterized cDNA clones encoding the H-protein of pea leaf mitochondria (10). The deduced primary structure revealed that the 131-amino acid polypeptide is cytoplasmically synthesized with a 34-amino acid mitochondrial targeting peptide. This protein shows a strong sequence homology with the H-protein isolated from liver mitochondria (11). So far it has proved to be impossible to crystallize any intact lipoate-containing protein. In the current study, we have determined the three-dimensional structure of the lipoate-containing H-protein at 2.6 Å¶ to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: H_4FGlu_n, tetrahydrofolate-polyglutamate; MIR, multiple isomorphous replacement; FAST, fast-scanning area-sensitive television detector.

†To whom reprint requests should be addressed.

¶The C_α atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1 H.P.C.). This information is embargoed for 1 year (coordinates) from the date of publication.

Table 1. Crystallographic, MIR, and phasing analysis

Data set	Native	$Eu(NO_3)_3$* (a)	$Gd(NO_3)_3$* (b)	EtHgCl (c)	$Eu(NO_3)_3$ + EtHgCl (d)
Detector	FAST	IP	IP	IP	IP
Number of measurements	22,630	26,396	29,753	23,977	17,872
Unique reflections	7,773	4,887	5,046	6,507	4,759
Completeness, %	87	90	93	90	88
Resolution limit, Å	2.6	3.1	3.1	2.8	3.1
R_{sym}	0.067	0.078	0.077	0.073	0.065
R_{iso}		0.204	0.130	0.173	0.200
Anomalous scattering used		Yes	Yes	No	No
Number of heavy atom sites		1	1	4	5
R_{Cullis} (20 to 3.2 Å) (number of reflections)		0.635 (645)	0.772 (798)	0.699 (729)	0.685 (609)
Phasing power (20 to 3.2 Å)		1.76	1.01	1.36	1.27
Mean figure of merit (20 to 3.2 Å)		0.84			

Derivatives were obtained by soaking native crystals in a medium containing 2.5 M ammonium sulfate, 0.1 M Tris maleate buffer, and, respectively, 25 mM $Eu(NO_3)_3 \cdot 6\,H_2O$ for 40 hr (column a), 20 mM $Gd(NO_3)_3 \cdot 6H_2O$ for 2 hr (column b), 0.04 mM EtHgCl for 20 hr (column c), or a mixture of $Eu(NO_3)_3$ and EtHgCl (column d). The CCP4 system (Daresbury Laboratory, U.K.) was used for analysis. The enantiomorph group was determined by using the anomalous scattering data. The program PHARE was used for phase determination. A 3.2-Å MIR map was improved by using solvent-flattening procedure (15). $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$, $R_{iso} = \Sigma|F_{PH} - F_P|/\Sigma F_P$, and $R_{Cullis} = \Sigma\,||F_{PH} \pm F_P| - F_H|/\Sigma|F_{PH} - F_P|$ for centric reflections. F_P, F_{PH}, and F_H are native derivative and heavy-atom structure factors, respectively. Phasing power is $f_{H/E}$, where f_H is the r.m.s. heavy-atom structure factor amplitude and E is the r.m.s. lack of closure error. IP, image plate.

*Even though Eu and Gd shared the same heavy atom sites, using the two derivatives separately improved the precision of the calculated phases (the noise was not the same from the two derivatives).

further our understanding of the self-recognition of proteins in the glycine cleavage complex.

MATERIALS AND METHODS

Enzyme Preparation and Crystallization. Pea (*Pisum sativum* L. var. Douce Provence) plants were grown in vermiculite at 28°C (12-hr light period; Philips lamps model TLD 58W83) and at 22°C (12-hr dark period) in a growth cabinet.

Mitochondria were isolated and purified from pea leaves as described by Douce *et al.* (12) by using self-generating Percoll gradients and a linear gradient of 0–10% (wt/vol) polyvinylpyrrolidone 25 (top to bottom). Proteins of high molecular mass from pea mitochondrial matrix retained on an XM-300 Diaflo membrane ("matrix extract") were prepared as described by Bourguignon *et al.* (3). The H-protein of the glycine cleavage system has been purified to apparent homogeneity from the matrix extract by using gel filtration, ion-exchange, and phenyl-Superose fast protein liquid chromatography. The identity of H-protein was determined by measuring the amount of [^{14}C]bicarbonate fixed to the carboxyl-group carbon atom of glycine during the exchange reaction (in the presence of saturating amounts of P-protein) (3).

The H-protein crystals were grown by vapor diffusion at 8°C from a solution containing 15 mg of protein per ml buffered at pH 5.2 with 0.1 M Tris maleate equilibrated against 2 M $(NH_4)_2SO_4$ with the same buffer solution. Seeding techniques led to 0.3 × 0.2 × 0.1 mm^3 bipyramidal crystals diffracting up to 2.4 Å. The space group is $P3_121$. The cell parameters are $a = b = 57.14$ Å and $c = 137.11$ Å, with $V = 384{,}860$ Å^3. The number of molecules per unit cell is 12 (two independent molecules in the asymmetric unit). The crystal solvent content is about 45%. Details on the preliminary crystallographic data have been reported (13).

Data Collection and Multiple Isomorphous Replacement (MIR). X-ray diffraction data were collected from single crystals by using either a FAST/Enraf Nonius area detector (Enraf Nonius FR 571 rotating anode generator, copper radiation, graphite monochromator) or a MARRESEARCH imaging plate (Rigaku RU-200 HB rotating anode, copper radiation, and double-mirror focusing optics). The fast-scanning area-sensitive television detector (FAST) data were processed with the program MADNES (14), and the MARRESEARCH imaging plate data were processed with the program MOSFLM (Daresbury Laboratory, Warrington, U.K.). The structure of the H-protein from pea leaves has been solved by using MIR followed by a solvent-flattening procedure. Analysis of the data collection and of the MIR phase determination is given in Table 1.

RESULTS AND DISCUSSION

The residues were fitted in a 3.2-Å solvent-flattened map by using an O graphics program (16). β-Strands appeared clearly, and residues 37–70 were constructed with the initial assignment of Phe-56 for one molecule. The entire chain was then assigned for this molecule. The two molecules of the asymmetric unit are found to be related by a 140° rotation, and the second molecule was fitted in the electron density by

FIG. 2. Schematic ribbon representation of the overall folding of the H-protein, a component of the glycine decarboxylase complex. The lysine side chain and the lipoate moiety are marked with small spheres.

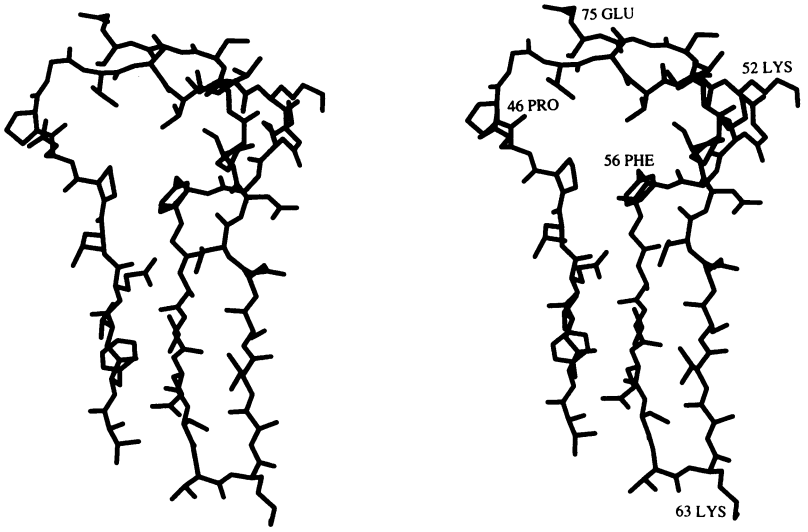

FIG. 3. Stereo representation of the three β-strands and the hairpin β-sheet containing the lysine-63 that binds the lipoamide group of the H-protein. Some of the most conserved amino acid residues are marked.

using this rotation. The model was refined with an X-PLOR computing package (15) by using the simulated annealing slow-cooling protocol, thus extending the resolution to 2.6 Å. During this stage the *R*-factor dropped from 49% to 25%. *B* factors were introduced for each residue backbone and side chain, and a final *R*-factor of 23% was obtained for 7494 reflections in the range of 15–2.6 Å ($F > 2\sigma$). Average overall *B* factors were 16.6 Å^2 and 15.8 Å^2 for the two protein molecules, respectively, and 33 Å^2 and 30 Å^2 for the two lipoate groups. The r.m.s. values on bonds are 0.025 Å in both cases, and the r.m.s. values on angles are 4.9° and 4.6°, respectively, for the two molecules. A Ramachandran plot shows all residues falling within the expected boundaries.

The two protein molecules per asymmetric unit have the same overall conformation shown on the schematic ribbon representation (17) in Fig. 2. The core of this structure consists mainly of two antiparallel β-sheets forming a "sandwich" structure, one with four strands including residues 13–19, 22–28, 74–80, and 101–105 and one with three strands including residues 37–43 and two adjacent antiparallel strands (55–61 and 65–70) joined by a loop (hairpin β-motif) in which the lipoylated lysine residue is situated. It is generally agreed that β strand sandwich structures have a rather rigid configuration because of the hydrophobic interior and the H-bonding network between β-strands (18). A stereo view of the hairpin motif containing the active site Lys-63 is shown in Fig. 3.

Table 2. Alignment of amino acid sequences of H-protein from pea-leaf mitochondria, *Arabidopsis thaliana* (ARATH), chicken, bovine, and human liver

	S	N	V	L	D	G	L	K	Y	A	P	S	H	E	W	V	K	H	E	G	S	V	A	T	I	G	I	T	D	Pea
	S	T	V	L	E	G	L	K	Y	A	N	S	H	E	W	V	K	H	E	G	S	V	A	T	I	G	I	T	A	ARATH
	S	A	R	-	-	-	-	K	F	T	D	K	H	E	W	I	S	V	E	N	G	I	G	T	V	G	I	S	N	Chicken
	S	V	R	-	-	-	-	K	F	T	E	K	H	E	W	V	T	T	E	N	G	V	G	T	V	G	I	S	N	Bovine
	S	V	R	-	-	-	-	K	F	T	E	K	H	E	W	V	T	T	E	N	G	I	G	T	V	G	I	S	N	Human
30	H	A	Q	D	H	L	G	E	V	V	F	V	E	L	P	E	P	G	V	S	V	T	K	G	K	G	F	G	A	V	Pea
30	H	A	Q	D	H	L	G	E	V	V	F	V	E	L	P	E	D	N	T	S	V	S	K	E	K	S	F	G	A	V	ARATH
26	F	A	Q	E	A	L	G	D	V	V	Y	C	S	L	P	E	I	G	T	K	L	N	K	D	D	E	F	G	A	L	Chicken
26	F	A	Q	E	A	L	G	D	V	V	Y	C	S	L	P	E	V	G	T	K	L	N	K	Q	E	E	F	G	A	L	Bovine
26	F	A	Q	E	A	L	G	D	V	V	Y	C	S	L	P	E	V	G	T	K	L	N	K	Q	D	E	F	G	A	L	Human
				★																											
60	E	S	V	K	A	T	S	D	V	N	S	P	I	S	G	E	V	I	E	V	N	T	G	L	T	G	K	P	G	L	Pea
60	E	S	V	K	A	T	S	E	I	L	S	P	I	S	G	E	I	I	E	V	N	K	K	L	T	E	S	P	G	L	ARATH
56	E	S	V	K	A	A	S	E	L	Y	S	P	L	T	G	E	V	T	D	I	N	A	A	L	A	D	N	P	G	L	Chicken
56	E	S	V	K	A	A	S	E	L	Y	S	P	L	S	G	E	V	T	E	I	N	K	A	L	A	E	N	P	G	L	Bovine
56	E	S	V	K	A	A	S	E	L	Y	S	P	L	S	G	E	V	T	E	I	N	E	A	L	A	E	N	P	G	L	Human
90	I	N	S	S	P	Y	E	D	G	W	M	I	K	I	K	P	T	S	P	D	E	L	E	S	L	L	G	A	K	E	Pea
90	I	N	S	S	P	Y	E	D	G	W	M	I	K	V	K	P	S	S	P	A	E	L	E	S	L	M	G	P	K	E	ARATH
86	V	N	K	S	C	Y	Q	D	G	W	L	I	K	M	T	V	E	K	P	A	E	L	D	E	L	M	S	E	D	A	Chicken
86	V	N	K	S	C	Y	E	D	G	W	L	I	K	M	T	F	S	N	P	S	E	L	D	E	L	M	S	E	E	A	Bovine
86	V	N	K	S	C	Y	E	D	G	W	L	I	K	M	T	L	S	N	P	S	E	L	D	E	L	M	S	E	E	A	Human
120	Y	T	K	F	C	E	-	-	E	E	D	A	A	H																	Pea
120	Y	T	K	F	C	E	-	-	E	E	D	A	A	H																	ARATH
116	Y	E	K	Y	I	K	S	I	E	D																					Chicken
116	Y	E	K	Y	I	K	S	I	E	E																					Bovine
116	Y	E	K	Y	I	K	S	I	E	E																					Human

The lipoyl-lysine (K63) is marked with a star. The sequences were derived from EMBL/GenBank.

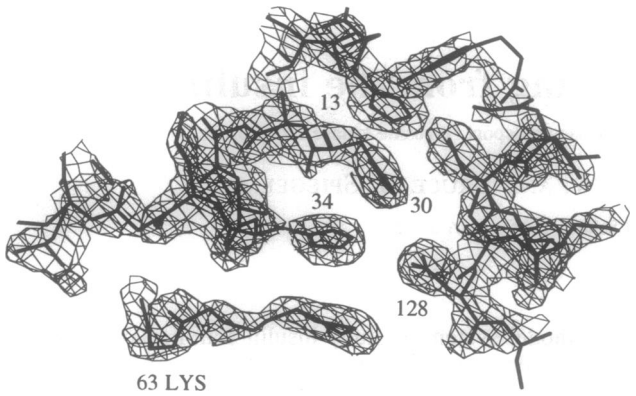

FIG. 4. $2F_o - F_c$ electron density map contoured at 1 σ level of the region surrounding the lipoate group. The lipoate moiety is at the surface of the molecule extending towards His-34 and Asp-128.

Interestingly, the constitutive residues of the antiparallel β-sheets, and especially those of the hairpin β-motif bearing the lipoate group, tend to be invariant in all of the H-proteins (Table 2), suggesting that they all possess considerable structural similarity. The schematic diagram in Fig. 2 also indicates the presence of two short α-helical segments: one between residues 29 and 35 and one at the C-terminal end starting at residue 119. A $2F_o-F_c$ map of the lipoate region is shown in Fig. 4. Fig. 2 indicates that the Lys-63 chain-side and lipoate moiety are situated at the surface of the molecule in the solvent region. In other words, the lipoate group is not buried and has freedom to interact readily with the other proteins (P-, T-, and L-proteins) in the glycine decarboxylase complex. However, in the crystal structure, the lipoate arm is locked down because of intermolecular interactions (which are not the same for the two independent molecules of the asymmetric unit), giving a clear view of its configuration. Fig. 4 indicates that the lipoate group is in the vicinity of residues His-34 and Asp-128.

The three-dimensional structure of the lipoylated domain of the H-protein from the pea leaf glycine cleavage system has been recently predicted (19) on the basis of (*i*) the structures of the lipoyl domains of the *Bacillus stearothermophilus* (20) and *Escherichia coli* (19) 2-oxo acid dehydrogenase multienzyme complexes (determined by means of multidimensional nuclear magnetic resonance spectroscopy), and (*ii*) information related to interactions that stabilize the folded state. The predicted model shows a β-strand structure between residues 38 and 69. The most conserved part is the active-site region around the lipoic acid cofactor, showing a β-hairpin loop containing a single β-turn. However as pointed out judiciously by Brocklehurst and Perham (19), further speculation must await the determination of the structure of the H-protein. The results presented in this paper indicate that the x-ray structure of the H-protein is strikingly similar to the predicted three-dimensional structure of the lipoyl domain of lipoate-containing proteins. These are in contrast with another previous prediction, which indicated that the lipoate moiety resides in an α-helical region (11). Interestingly, the predicted three-dimensional structure of the biotinyl domain of pyruvate carboxylase (the biotin cofactor in biotin-dependent carboxylases is attached by an amide linkage to the N^6-amino group of a specific lysine residue) (19, 21, 22) also discloses striking similarities with the β-sheet configuration that we have found in the region of residues 38–79 in the H-protein crystal structure. From the structural analogies and predictions mentioned above, we could make the assumption that the hairpin β-motif configuration observed in the H-protein crystal structure around lipoic acid is also present in all proteins that contain lipoic acid or biotin. In the H-protein, such a structure might play a crucial role in the movement of the lipoic acid prosthetic group from the site where it acquires a methylamine group from glycine to the site where it donates this group to $H_4PteGlu_n$ and finally to a site where it donates the electrons to FAD. In other words, such a structure might contribute to the movement of a lipoyl–lysine swinging arm—i.e., to activate site coupling as discussed by Perham (23). Further crystallographic studies, including high-resolution analysis of the H-protein and its methylamine-charged form (24), crystallization of the other proteins of the complex, and studies in solution using small-angle scattering techniques are needed to improve our knowledge of the structural features of the glycine decarboxylase complex at both global and atomic level.

This paper is dedicated to Prof. Jean-Pierre Ebel. We thank O. Dideberg for helpful assistance in interpreting the electron-density map; S. Cusack and C. Berthet for many valuable discussions; J. Gagnon for the sequence alignment using DNASTAR; and M. S. Lehmann for giving access to Institut Laue Langevin facilities. X-ray data collection was performed at the Institut de Biologie Structurale and at European Molecular Biology Laboratory (Grenoble outstation). This work was supported by a grant from Centre National de la Recherche Scientifique (Interface Chimie-Biologie, contract 900057).

1. Husic, D. W., Husik, H. D. & Tolbert, N. E. (1987) *CRC Crit. Rev. Plant Sci.* **5**, 45–100.
2. Douce, R. & Neuburger, M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 371–414.
3. Bourguignon, J., Neuburger, M. & Douce, R. (1988) *Biochem. J.* **255**, 169–178.
4. Klein, S. M. & Sagers, R. D. (1966) *J. Biol. Chem.* **241**, 197–205.
5. Kochi, H. & Kikuchi, G. (1969) *Arch. Biochem. Biophys.* **132**, 359–369.
6. Kikuchi, G. (1973) *Mol. Cell Biochem.* **1**, 169–187.
7. Tada, K. & Hayasaka, K. (1987) *Eur. J. Pediatr.* **146**, 221–227.
8. Oliver, D. J., Neuburger, M., Bourguignon, J. & Douce, R. (1990) *Plant Physiol.* **94**, 833–839.
9. Mérand, V., Forest, E., Gagnon, J., Monnet, C., Thibault, P., Neuburger, M. & Douce, R. (1993) *Biol. Mass Spectrometry* **22**, 447–453.
10. Macherel, D., Lebrun, M., Gagnon, J., Neuburger, M. & Douce, R. (1992) *Biochem. J.* **286**, 627–630.
11. Fujiwara, K., Okamura-Ikeda, K. & Motokawa, Y. (1986) *J. Biol. Chem.* **261**, 8836–8841.
12. Douce, R., Bourguignon, J., Brouquisse, R. & Neuburger, M. (1987) *Methods Enzymol.* **148**, 403–415.
13. Sieker, L., Cohen-Addad, C., Neuburger, M. & Douce, R. (1991) *J. Mol. Biol.* **220**, 223–224.
14. Messerschmidt, A. & Pflugath, J. W. (1987) *J. Appl. Crystallogr.* **20**, 306–315.
15. Brünger, A. T. (1992) X-PLOR (Yale Univ., New Haven, CT), Version 3.1.
16. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110–119.
17. Carson, M. (1987) *J. Mol. Graph.* **5**, 103–106.
18. Branden, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York), p. 1.
19. Brocklehurst, S. M. & Perham, R. N. (1993) *Protein Sci.* **2**, 626–639.
20. Dardel, F., Davis, A. L., Laue, E. D. & Perham, R. N. (1993) *J. Mol. Biol.* **229**, 1037–1048.
21. Lim, F., Morris, C. P., Occhiodoro, F. & Wallace, J. C. (1988) *J. Biol. Chem.* **263**, 11493–11497.
22. Hale, G., Wallis, N. G. & Perham, R. N. (1992) *Proc. R. Soc. London B* **248**, 247–253.
23. Perham, R. N. (1991) *Biochemistry* **30**, 8501–8512.
24. Neuburger, M., Jourdain, A. & Douce, R. (1991) *Biochem. J.* **278**, 765–769.