Structure of P-protein of the glycine cleavage system: implications for nonketotic hyperglycinemia

Tadashi Nakai^{1,2,*}, Noriko Nakagawa^{1,2},
Nobuko Maoka¹, Ryoji Masui^{1,2},
Seiki Kuramitsu^{1,2} and Nobuo Kamiya¹

¹RIKEN Harima Institute at SPring-8, Mikazuki, Sayo, Hyogo, Japan and

²Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka, Japan

The crystal structure of the P-protein of the glycine cleavage system from *Thermus thermophilus* HB8 has been determined. This is the first reported crystal structure of a P-protein, and it reveals that P-proteins do not involve the α_2 -type active dimer universally observed in the evolutionarily related pyridoxal 5'-phosphate (PLP)-dependent enzymes. Instead, novel $\alpha\beta$ -type dimers associate to form an $\alpha_2\beta_2$ tetramer, where the α - and β -subunits are structurally similar and appear to have arisen by gene duplication and subsequent divergence with a loss of one active site. The binding of PLP to the apoenzyme induces large open-closed conformational changes, with residues moving up to 13.5 Å. The structure of the complex formed by the holoenzyme bound to an inhibitor, (aminooxy)acetate, suggests residues that may be responsible for substrate recognition. The molecular surface around the lipoamide-binding channel shows conservation of positively charged residues, which are possibly involved in complex formation with the H-protein. These results provide insights into the molecular basis of nonketotic hyperglycinemia.

The EMBO Journal (2005) 24, 1523–1536. doi:10.1038/sj.emboj.7600632; Published online 24 March 2005

Subject Categories: structural biology; molecular biology of disease

Keywords: glycine cleavage system; glycine decarboxylase; nonketotic hyperglycinemia; P-protein; pyridoxal 5'-phosphate (PLP)-dependent enzyme

Introduction

The glycine cleavage system (GCS) is a multienzyme complex composed of four different components (P-, H-, T- and L-proteins). This system catalyzes the oxidative cleavage of glycine in a multistep reaction (Figure 1) (Motokawa *et al*, 1995). In almost all organisms, the GCS plays a crucial role in the degradation of glycine, and it has been studied extensively (Perham, 2000; Douce *et al*, 2001). In humans, a mutation in the GCS genes can lead to a dramatic accumula-

tion of glycine in the blood, resulting in severe neurological diseases, termed nonketotic hyperglycinemia (NKH) (Tada and Kure, 1993; Applegarth and Toone, 2001). In plants, large amounts of GCS proteins are present in leaf mitochondria, where they are involved in the photorespiratory pathway (Douce *et al*, 2001). *In vivo*, the GCS acts as a stable complex with an approximate ratio of 2P:27H:9T:1L; *in vitro*, it easily dissociates into its component proteins, and the H-protein acts as a mobile cosubstrate that commutes between the other three enzymes (Oliver *et al*, 1990). The three-dimensional structures of H-protein from pea (Pares *et al*, 1994; Cohen-Addad *et al*, 1995; Faure *et al*, 2000) and *Thermus thermophilus* HB8 (*Tth*) (Nakai *et al*, 2003a), T-protein from *Thermotoga maritima* (Lee *et al*, 2004), and L-protein from pea and six other species (Faure *et al*, 2000, and references therein) have been published, and only the structure of the P-protein has not yet been reported.

P-protein, also known as glycine decarboxylase (glycine: lipoylprotein oxidoreductase (decarboxylating and acceptor-aminomethylating), EC 1.4.4.2), is a pyridoxal 5'-phosphate (PLP)-dependent enzyme (PLP-enzyme). PLP-enzymes have been classified into fold-types I–V (Grishin *et al*, 1995; Jansonius, 1998; Mehta and Christen, 2000; Schneider *et al*, 2000), and P-protein belongs to fold-type I based on sequence similarities (Grishin *et al*, 1995). Fold-type I, of which aspartate aminotransferase (Ford *et al*, 1980) is the prototype, is the best-characterized structurally among the five types. All of the fold-type I enzyme structures determined so far have shown that the protein occurs as a homodimer (α_2 -type active dimer) with an internal two-fold axis or as multimers (α_4 , α_6 or α_{12}) (Supplementary Table I), where each α^I -subunit (α^I , ~50 kDa) contains one molecule of PLP. On the other hand, the subunit compositions of P-proteins from many species have been classified into two types: those from eukaryotes (e.g. human (Kume *et al*, 1991) and pea (Bourguignon *et al*, 1988)) and some of the P-proteins from prokaryotes (e.g. *Escherichia coli* (Okamura-Ikeda *et al*, 1993)) are in the homodimeric form (α_2^{NC}), while the rest of those from prokaryotes (e.g. *Clostridium acidurici* (Gariboldi and Drake, 1984), *Eubacterium acidaminophilum* (Freudenberg and Andreessen, 1989) and *Tth* (Nakai *et al*, 2003b)) are in the heterotetrameric form ($\alpha_2^N\beta_2^C$), where the α^N - and β^C -subunits (α^N and β^C , each ~50 kDa) correspond respectively to the N- and C-terminal halves of the ~100 kDa α^{NC} -subunit (α^{NC-N} and α^{NC-C} , each ~50 kDa) based on sequence similarities. Interestingly, while PLP is attached not to α^{NC-N} and α^N but to α^{NC-C} and β^C (Fujiwara *et al*, 1987), all of these proteins exhibit sequence similarities to α^I . Thus, the roles of α^{NC-C} and β^C appear to correspond to those of α^I , but those of α^{NC-N} and α^N have remained unknown.

The *Tth* P-protein used in this study has a total molecular mass of 200 kDa (α^N 47.1 kDa with 438 residues and β^C 52.7 kDa with 474 residues), where α^N and β^C have 31 and

*Corresponding author. Advanced Protein Crystallography Research Group, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan. Tel. +81 791 58 29 21; Fax: +81 791 58 28 26; E-mail: nakaix@spring8.or.jp

Received: 21 December 2004; accepted: 28 February 2005; published online: 24 March 2005

Figure 1 The multistep reaction catalyzed by the glycine cleavage system. First, P-protein catalyzes the decarboxylation of the glycine molecule concomitantly with the transfer of the residual amino-methyl group to a sulfur atom on the lipoyl group of the oxidized H-protein (H_{ox}), generating the aminomethylated H-protein (H_{am}). Next, the T-protein catalyzes the transfer of a methylene group from H_{am} to tetrahydrofolate (THF), resulting in the release of NH_3 and the generation of reduced H-protein (H_{red}). Finally, the dihydrolipoyl group of H_{red} is oxidized by L-protein and the lipoyl group of H_{ox} is regenerated, thereby completing the catalytic cycle. The figure is adapted from Douce *et al* (2001).

37% sequence identities with α^{NC-N} and α^{NC-C} of human P-protein, respectively. This suggests that *Tth* and human P-proteins have the same folding topology and similar three-dimensional structures. In humans, more than 80% of NKH patients have a specific defect in P-protein (Tada and Kure, 1993; Toone *et al*, 2000). Therefore, we expected the *Tth* P-protein structure to be a suitable model for understanding the molecular basis of NKH. Moreover, determination of the P-protein structure should greatly facilitate studies aimed at understanding the roles of active-site residues, the reaction mechanisms and the structural architecture of this multi-enzyme complex.

Here we report the crystal structures of three forms of *Tth* P-protein: the apoenzyme (P_{apo}) at 2.4 Å resolution, the holoenzyme (P_{holo}) at 2.1 Å resolution and P_{holo} in complex with a substrate analog inhibitor (aminooxy)acetate (AOA) ($P_{holo} \cdot AOA$) as an external aldimine intermediate model at 2.4 Å resolution. These structures reveal a novel $\alpha^N\beta^C$ -type active dimer and show that the enzyme undergoes large conformational changes with movements up to 13.5 Å upon binding of the cofactor. Neither of these characteristics have ever been observed in the other fold-type I enzymes. We discuss the molecular mechanisms of these enzymes in light of the new structural information, and in particular how the

Table I Data collection, refinement and model statistics

	P_{apo}	P_{holo}	$P_{holo} \cdot AOA$	$P_{holo} \cdot CH_3HgCl$
<i>Data collection</i>				
Beamline	BL45XU, SPring-8	BL45XU, SPring-8	BL44B2, SPring-8	BL44B2, SPring-8
Wavelength (Å)	0.90	1.00	1.00	1.00
Crystal form	I	II	II	II
Space group	$P3_121$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters $a/b/c$ (Å)	89.6/89.6/371.0	134.9/166.7/189.0	134.2/166.3/190.3	135.0/166.6/189.9
Resolution range ^a (Å)	50.0–2.40 (2.49–2.40)	50.0–2.10 (2.18–2.10)	50.0–2.40 (2.49–2.40)	50.0–3.38 (3.50–3.38)
No. of measured reflections	708 434	1 697 129	1 101 608	420 800
No. of unique reflections	69 153 (6807)	245 553 (23 611)	166 371 (16 003)	60 776 (5955)
Redundancy ^a	10.2 (9.8)	6.9 (5.7)	6.6 (5.7)	6.9 (6.4)
Completeness ^a (%)	99.9 (99.8)	99.0 (96.2)	99.4 (96.6)	99.8 (99.5)
$R_{merge}^{a,b}$ (%)	14.8 (47.3)	12.4 (67.3)	11.8 (57.8)	20.5 (47.9)
Average I/σ^a (I)	15.3 (5.2)	15.0 (2.6)	16.0 (3.1)	10.0 (3.9)
<i>Refinement</i>				
Resolution range ^a (Å)	50.0–2.40 (2.55–2.40)	50.0–2.10 (2.23–2.10)	50.0–2.40 (2.55–2.40)	
R_{work}^a	0.177 (0.214)	0.189 (0.227)	0.188 (0.234)	
R_{free}^a	0.237 (0.309)	0.229 (0.266)	0.230 (0.292)	
<i>Model statistics</i>				
<i>Deviations</i>				
Bond lengths (Å)	0.006	0.005	0.006	
Bond angles (deg)	1.3	1.3	1.3	
<i>Mean B-factors</i>				
Main-chain atoms ^c (Å ²)	27.89 (1816)	27.51 (3640)	34.62 (3640)	
Side-chain atoms (Å ²)	30.25	30.35	36.67	
PLP atoms ^c (Å ²)	—	31.43 (4)	50.25 (4)	
AOA atoms ^c (Å ²)	—	—	61.55 (4)	
Water atoms ^c (Å ²)	33.01 (1024)	34.58 (2145)	33.83 (1192)	
<i>Ramachandran plot^d</i>				
Favored	1345	2737	2725	
Additional allowed	145	242	253	
Generously allowed	7	22	23	
Disallowed	5	7	7	

^aValues in parentheses correspond to the reflections observed in the highest resolution shell.

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

^cValues in parentheses correspond to number of residues or molecules. Owing to poor electron density, the following N- or C-terminal residues of each heterodimer were not included in the final models: Ala438 α and Met439 β for all models, and Glu911 β –Gly912 β for the P_{apo} model.

^dThree residues of each heterodimer (Ser129 α , Arg312 α and Thr857 β) for all models fell in the disallowed regions, but the strained conformations of these residues were confirmed in the electron density maps.

mutations identified in NKH patients might lead to loss of enzymatic activity.

Results

Description of the structure

The P-protein has an $\alpha_2\beta_2$ -tetrameric structure with an overall size of approximately $114 \times 79 \times 82$ Å (Figure 2), where two pairs of α^N and β^C form intimate $\alpha^N\beta^C$ dimers that are related by a noncrystallographic two-fold axis. Thus, the tetramer can be described as a dimer of heterodimers. Furthermore, α^N and β^C of the $\alpha^N\beta^C$ dimer exhibit a similar folding topology (root-mean-square deviation (r.m.s.d.) of 2.01 Å for 395 C α atoms) and thus appear to be related by an approximate two-fold axis (Figure 3A), resulting in the $(\alpha^N\beta^C)_2$ tetramer, which mimics a homotetramer with pseudo-222 symmetry (Figure 2). The designation of the subunits as α^N , $\alpha^{N'}$, β^C and $\beta^{C'}$ (a prime after a Greek letter indicates the other subunit related by a two-fold axis) has been chosen such that the α^N - and β^C -subunits (or $\alpha^{N'}$ and $\beta^{C'}$) form intimate $\alpha^N\beta^C$ (or $\alpha^{N'}\beta^{C'}$) dimers, that is, the active dimeric form, as shown in Figure 2. Each subunit in the tetramer interacts with the other subunits through six interfaces, and the accessible surface areas buried in the α^N - $\alpha^{N'}$, α^N - β^C ($\alpha^{N'}$ - $\beta^{C'}$), α^N - $\beta^{C'}$ ($\alpha^{N'}\beta^C$) and β^C - $\beta^{C'}$ interfaces are 1512, 17 222 (17 125), 1433 (1418) and 2199 Å², respectively. Notably, the

α^N - β^C interface buries a large hydrophobic surface of 11 058 Å², which would likely render each subunit insoluble in aqueous solution if α^N and β^C were expressed separately (Nakai *et al*, 2003b). This may be the reason why coexpression of the two subunits is essential for obtaining functional recombinant P-protein (Nakai *et al*, 2003b).

The α^N -subunit is composed of three parts: an N-terminal arm (N-arm) (residues 1–65), a large α/β domain (residues 79–337) and a small $\alpha + \beta$ domain (residues 66–78 and 338–438) (Figures 3C and 4). The N-arm, which contains two α -helices, makes contact with the other three subunits. The N-arm is connected to the large domain via a loop belonging to the small domain. The large domain consists of a central seven-stranded β -sheet (a, g, f, e, d, b and c) plus nine α -helices, which surround the sheet on both the interior and surface of the protein, and a small two-stranded antiparallel β -sheet (h and i). The small domain consists of a four-stranded antiparallel β -sheet (a', b', c' and d') and three α -helices that make up part of the protein surface. The β^C -subunit is composed of four parts: an N-arm (residues 439–505), a large domain (residues 518–775), a small domain (residues 506–517 and 776–889) and a C-terminal arm (residues 890–912) (note that the residue numbering starts from 439, as described in the legend to Figure 4). The former three domains correspond to those of α^N and share common structural features described above except for a few short

Figure 2 Overall structures of P-protein and GluDC, the closest homolog with a known structure. Front view (A), top view (B) and side view (C) of $(\alpha^N\beta^C)_2$ -tetrameric P-protein, and side view of α_6 hexameric GluDC (D). PLP molecules are represented by yellow spheres.

Figure 3 Dimer and subunit structures of P-protein and GluDC. Stereoview of the $\alpha^N\beta^C$ dimer of the P-protein (A) and view of α^I_2 dimer of GluDC (B). Stereoviews of α^N (C) and β^C (D) of the P-protein. Superposition of α^N (green) and β^C (orange) of the P-protein, and α^I of GluDC (purple) (E). PLP molecules and their bound lysine residues are shown as ball-and-stick models. In panels C and D, topology diagrams are included, in which α -helices and β -strands are represented by circles and triangles, respectively. The diagrams were generated by Tops (Westhead *et al*, 1999) with manual modifications.

helices (Figures 3 and 4). Major structural features found only in β^C are as follows. The cofactor PLP is covalently attached to the side chain of Lys704 β . The C-terminal arm, which is found only in β^C and which contains an α -helix, makes contact with α^N .

Structural comparison of P-protein with other PLP-enzymes

A comparison of P-protein with structures in the Protein Data Bank using the program DALI (Holm and Sander, 1993) produced a list of 62 structures having Z-scores (indicating

A

B

Figure 4 Sequence alignment of P-proteins with secondary structure assignment of *Tth* P-protein: (A) α^N and (B) β^C . The amino-acid sequences were obtained from the Swiss-Prot database. The multiple alignments of these sequences with those of *Tth* P-protein (GCSA_THETH and GCSB_THETH) were performed using the program CLUSTAL W (Thompson *et al*, 1994). The figure was produced using ESPRIPT (Gouet *et al*, 1999). The sequence names correspond to the Swiss-Prot accession IDs, except for the *Tth* P-protein. The residue numbering of the human P-protein starts from 40 since the N-terminal presequence is omitted. The numbering of β^C of *Tth* P-protein starts from 439, as if α^N and β^C consisted of a single polypeptide, in order to facilitate comparisons between the $\alpha_2\beta_2$ -dimeric and $(\alpha^N\beta^C)_2$ -tetrameric P-proteins.

the degree of structural similarity) better than 14.1. (Supplementary Table I; the highest Z-score for excluded structures is 7.0). All of these structures belong to fold-type I, indicating that P-protein should be classified as a member of fold-type I. However, since all the fold-type I enzymes other than P-protein occur as an α_2^1 -type active dimer or its multiple, the $\alpha^N\beta^C$ -type active dimer observed in the P-protein is a remarkable exception.

The highest Z-scores for α^N and β^C of P-protein were calculated to be 28.3 and 31.5, respectively, and both were from comparisons with α^I of glutamate decarboxylase (GluDC) from *E. coli* (Capitani *et al*, 2003), indicating that GluDC is the closest homolog with a known three-dimensional structure. While GluDC forms an α_6^I hexamer (Figure 2D), unlike P-protein, the overall polypeptide fold of P-protein is indeed similar to that of GluDC (Figure 3E); the superimposition of α^N and β^C onto α^I of GluDC yields r.m.s.d.s of 3.8 and 3.0 Å, and sequence identities of 12 and 12% for 354 and 369 C α atoms, respectively. The most notable differences between these structures, other than subunit compositions, are found at the N-terminal half of each N-arm (Figure 3A and B). The α^N - and β^C -subunits of P-protein and α^I of GluDC have an N-arm that crosses over

and embraces the adjacent subunit. In GluDC, the N-terminal half (residues 1–30) of the N-arm mostly participates in interactions with the neighboring dimer, leading to trimerization of the α_2^I dimers; in P-protein, residues 1–39 α and 439–447 β participate in interactions with the adjacent subunit, contributing to dimerization of α^N and β^C . These regions also participate in interactions with the other dimer (α^N/β^C dimer) and thus participate in dimerization of $\alpha^N\beta^C$ dimers. Therefore, P-protein and GluDC are similar in that their N-termini are involved in assembly of active dimers, in spite of differences in the final assembly of subunits (Figure 3A and B).

Open-closed conformational changes upon binding of the cofactor PLP

Binding of PLP to P_{apo} induces an open-closed conformational change involving primarily two regions that are adjacent to the active site, with backbone shifts of up to 13.5 Å (Figure 5). The first region is a loop that includes an α -helix (helix 10) from the α^N large domain (residues 305–322 α), which we will refer to as the ‘mobile loop’. The second region is composed of a subdomain from the β^C large domain (residues 591–671 β), which we will refer to as the ‘mobile

Figure 5 Conformational changes upon binding of the cofactor PLP. (A) Stereoview of the superimposition of the P_{holo} and P_{apo} (semitransparent) structures by least-squares fitting of C α atoms in immobile regions (see text), where the displacements are represented by a color gradient from blue (<0.37 Å) to red (>2.50 Å). The orientation shows the protein rotated by 180° around a vertical axis with respect to Figure 3A. (B) Displacements for equivalent C α atoms between the two structures after the superimposition. The curves are colored by domain according to Figure 3A, except for the black curve indicating the displacements after the superimposition of only the mobile subdomains, in which the values are shifted by +6.0 Å unit in order to improve clarity.

subdomain'. The P-protein C α atoms, except for those of the mobile loop and mobile subdomain, plus a few other residues (described later), in the open and closed forms, were superimposed by least-squares fitting, giving an r.m.s.d. of 0.37 Å, with a maximum displacement of 1.15 Å for 797 C α atoms (Figure 5B). In contrast to the good agreement of the C α atoms of these residues, larger differences were observed in the loop and subdomain C α atoms, which gave r.m.s.d.s of 7.31 and 1.28 Å, with maximum displacements of 13.5 Å (Ile309 α) and 2.5 Å (Arg636 β), respectively.

The mobile loop in P_{apo} contains an α -helix (helix 10). Binding of PLP induces conformational changes in the loop involving shortening of helix 10 (from Arg306 α –Arg311 α to Gln308 α –Arg311 α) and formation of two ₁₀-helices (helices η 7 and η 8; Figure 4A). The mobile subdomain consists of three α -helices (helices 7, 8 and 9) and three β -strands (strands b, c and d). The motion of the subdomain can be approximated by a 4.0° rotation as a rigid body to close the active site, because the subdomain C α atoms of P_{apo} are superimposable onto those of P_{holo} with an r.m.s.d. of 0.30 Å and a maximum displacement of 1.13 Å for 81 C α atoms (Figure 5B). Besides the loop and the subdomain, four regions (Glu33 α –Pro38 α , Pro267 α –His268 α , Ala569 β –Gly570 β and Glu739 β –Glu740 β) exhibit prominent conformational changes (Figure 5B). The first three regions are located around the active site and interact with PLP directly or indirectly via other mobile regions (Figure 5A). Notably, the peptide bond preceding Pro267 α undergoes *trans*–*cis* isomerization. Ala569 β –Gly570 β move up to 4.1 Å, allowing their main-chain NH groups to form hydrogen bonds with the phosphate group of PLP. In contrast, the fourth region (Glu739 β –Glu740 β) lies at a molecular surface distant from the active site and is involved in crystal packing environments that differ between the P_{apo} and P_{holo} forms, resulting in different conformations in this region.

Cofactor binding to proteins often results in conformational changes. Another PLP-enzyme, 8-amino-7-oxononanoate synthase, also shows similar conformational changes (Alexeev *et al*, 1998), including movement of two regions corresponding to the mobile loop and subdomain of the P-protein. However, the conformational changes in 8-amino-7-oxononanoate synthase involve smaller backbone shifts of up to 5.3 Å (except for the C-terminus) as compared with 13.5 Å in the P-protein. This is likely due to the absence of refolding of the helices, as observed in the P-protein. The mobile loop of the P-protein also corresponds to a disordered loop found in the crystal structure of DOPA decarboxylase (Burkhard *et al*, 2001). The disordered loop includes highly conserved residues that are essential for catalytic activity (Ishii *et al*, 1996). Bertoldi *et al* (2002) have reported that one of the loop residues is responsible for protonation of the decarboxylated reaction intermediates to form biological amines. It is also notable that in P-protein, highly conserved residues are located in the mobile loop (Figure 4). These data imply that the mobile loop may also play a catalytic role in the latter catalytic steps carried out by the P-protein (Supplementary Figure 1C–F).

The active site

The ($\alpha^N\beta^C$)₂ P-protein tetramer has two active-site pockets at the α^N – β^C and $\alpha^{N'}$ – $\beta^{C'}$ interfaces (Figure 2). Each pocket is situated also at the domain interface of β^C (Figure 3D). The

residues constituting the pocket can be considered as two regions. The first region is the bottom of the active-site pocket, which consists of residues at one end of the seven-stranded β -sheet from the β^C large domain (Gly570 β , Ala571 β , Glu574 β , His604 β , Ser606 β , Thr648 β , Thr652 β , Asp677 β , Ala679 β , Asn680 β , His699 β , Asn701 β , His703 β and Lys704 β) (Figure 6A). The second region is the top of the pocket, which consists of residues from the β^C small domain (Ser511 β and Cys512 β) and from the α^N large domain (Tyr95 α , Thr96 α , Tyr98 α , Gln308 α , Tyr309 α , Thr320 α and Thr321 α) (Figure 6B). Of the 23 residues comprising these two regions, 20 are strictly conserved (Figure 4), suggesting that structure and function of the active site are essentially conserved among the P-proteins.

The cofactor PLP is bound to the active-site pocket by extensive noncovalent interactions with all the bottom residues of the pocket and with some of the top residues (Ser511 β , Tyr95 α , Thr320 α and Thr321 α). PLP also forms an internal aldimine bond (Schiff base linkage) with the catalytic residue Lys704 β (Figure 6B and Supplementary Figure 1A). The pyridine ring of PLP is sandwiched by the methyl group of Ala679 β and the imidazole ring of His604 β . The pyridine ring of the cofactor makes an angle of 14.9° with the imidazole ring of His604 β . The phosphate group of PLP is involved in a number of hydrogen bonds and one ion pair, and acts as an anchor to fix the cofactor to the active site. The negative charge of the phosphate group is likely to be balanced by the positive charge of His703 β and the dipole of helix 6 of β^C , which is positioned with its N-terminus close to the phosphate group. In addition, Asp677 β forms an ion pair with the protonated nitrogen atom of the pyridine ring of PLP, as in other PLP-enzymes, and Ser511 β forms hydrogen bonds with the O3 atom of PLP.

Binding of substrate

The substrate-binding site of the P-protein consists mostly of the top residues of the active-site pocket (Tyr95 α , Thr96 α , Tyr98 α , Gln308 α , Tyr309 α , Thr320 α , Ser511 β , Cys512 β and His604 β) (Figure 6); in the P_{holo}·AOA complex, the AOA moiety makes van der Waals contacts with these residues. AOA is a substrate analog in which the amino (NH₂) group of glycine is replaced by an aminooxy (ONH₂) group.

In the substrate-free form, seven water molecules (W1–W7) in the substrate-binding site are involved in hydrogen-bond networks, where W1–W4 together with the OH group of Tyr95 α form a linear hydrogen-bond network (Figure 6B). This network is connected to the phosphate group of PLP through W1 and to the side-chain C=O of Gln308 α and the main-chain NH of Gly605 β through W4.

Upon binding of AOA, two water molecules (W1 and W2) are liberated from the substrate-binding site and the cofactor makes a new covalent bond with AOA in place of Lys704 β (Figure 6C and Supplementary Figure 1A and B). This may be an aldimine bond, based on the good agreement between the substrate model and electron density (Figure 6C). In this case, this bond corresponds to the external aldimine shown in Supplementary Figure 1B. The released ϵ -amino group of Lys704 β makes an ion pair with the phosphate group of PLP. While a new hydrogen-bond network is formed between AOA and the surrounding residues, the active-site residues, except for Lys704 β , do not change their positions, and the interactions among them are retained upon AOA binding; the

Figure 6 Structure of the active-site pocket of P-protein. Stereoviews of the cofactor-binding site for P_{holo} (A); the substrate-binding site for P_{holo} (B) and for $P_{\text{holo}} \cdot \text{AOA}$ (C); and superposition of the corresponding sites between P_{holo} (cyan) and $P_{\text{holo}} \cdot \text{AOA}$ (magenta) (D). Panels B and C include the $F_o - F_c$ simulated-annealing omit map contoured at 3.5σ around Lys704 β , PLP and AOA. After removing them and their neighboring residues, a simulated-annealing map was calculated with a starting temperature of 1000 K. In panels A–C, carbon atoms of each subunit are colored according to Figure 2, whereas the α and β designations in residue labels are omitted. Water molecules are shown as red spheres. Dashed lines represent hydrogen bonds.

corresponding active-site residues of P_{holo} and $P_{\text{holo}} \cdot \text{AOA}$ were superimposed with an r.m.s.d. of 0.15 Å and with a maximum displacement of 0.55 Å (Figure 6D).

The carboxylate of AOA occupies the space formed by the liberated water molecules. One of the carboxylate oxygen atoms of AOA forms hydrogen bonds with the OH group of Tyr95 α and with the OH group of Thr96 α through W6, and the other forms hydrogen bonds with the OH group of Tyr98 α , with the NH₂ group of Gln308 α and with W3. W3 is also involved in the hydrogen-bond network described above. The N1 atom of AOA forms a hydrogen bond with the imidazole N ϵ 2 atom of His604 β . The pyridine ring of the cofactor and the imidazole ring of His604 β rotate by 18.3 and 6.3° toward the solvent side, respectively, compared with their positions in the free enzyme (Figure 6D).

AOA is fixed at the substrate-binding site by extensive interactions with the active-site residues. The AOA and glycine structures are nearly identical except for the one extra oxygen atom proximate to the amino group, which suggests that the binding mode of AOA is similar to that of glycine but is not strictly identical, due to displacement (up to ~1.5 Å) of the distal carboxylate. Nevertheless, the pocket residues surrounding AOA are likely to be involved in substrate recognition. This is supported by the facts that all residues involved in the above interactions with AOA are highly conserved (Figure 4) and that the orientation of the carboxylate is appropriate for decarboxylation. The latter fact is based on the model for control of the reaction specificity of PLP-enzymes (Dunathan, 1966); the scissile bond of the external aldimine intermediate must be oriented perpendicular to the π system of the cofactor. In the $P_{\text{holo}} \cdot \text{AOA}$ structure, the carboxylate of the inhibitor is in an orientation approximately orthogonal to the plane of the cofactor ring (Figure 6C and Supplementary Figure 1B).

The lipoamide-binding channel and the surface properties neighboring its entrance

After formation of the external aldimine between PLP and glycine, P-protein catalyzes a second reaction: the decarboxylation and reductive aminomethylation of lipoamide (Supplementary Figure 1). The lipoamide is presented to P-protein in the form of a 16 Å long lipoyl-lysine arm of H-protein (Figure 7). Molecular recognition and interaction between H- and P-proteins have been suggested by biochemical experiments, in which H-apoprotein lacking lipoamide was found to behave as a competitive inhibitor relative to H-protein in the second reaction (Neuburger *et al*, 2000). The active site of P-protein is connected to the molecular surface by an ~18 Å deep channel with a broad entrance facing the solvent (Figure 7). The channel is lined by residues from both α^{N} and β^{C} (Figure 7D), most of which are hydrophobic residues that could interact with the aliphatic chain of the lipoyl-lysine. In the model structure (Figure 7D), the disulfide of the lipoamide lies close to the side chains of Tyr95 α , Tyr98 α , Gln308 α and His604 β . While these residues should be involved in the substrate recognition as described above, one of the residues may also act as a proton donor in the reduction of the disulfide bond (Supplementary Figure 1D). Note that Gln308 α is located at the mobile loop, and its counterpart in DOPA decarboxylase is important for catalysis, as described above.

The molecular surface neighboring the entrance of the channel possesses two remarkable properties. As shown in Figure 7B and C, the molecular surface of the P-protein reveals that conserved residues cluster around the entrance of channel and that several of them are positively charged (Lys146 α , Arg280 α , Arg306 α , Arg311 α , Arg312 α , Lys314 α , Lys316 α , Arg894 β and Lys902 β). In contrast, negatively charged residues are situated around the lipoyl-lysine arm of the H-protein, and most of them are highly conserved (Nakai *et al*, 2003a). The data imply that these residues play a critical role in the interaction between the P- and H-proteins.

Discussion

The structure of P-protein presented here reveals that the enzyme is present as an $(\alpha^{\text{N}}\beta^{\text{C}})_2$ tetramer containing a unique arrangement of $\alpha^{\text{N}}\beta^{\text{C}}$ -type active dimers. The molecular mass of the tetramer is consistent with dynamic light scattering results in aqueous solution (Nakai *et al*, 2003b), indicating that the tetramer is the native oligomeric state. P-protein has been classified as a member of fold-type I PLP-enzymes, based on sequence similarities (Grishin *et al*, 1995), and this is now confirmed experimentally. All other fold-type I enzymes with known structures are intimate α_2^{I} dimers, each with an internal two-fold axis, or loose multiples thereof, where an α_2^{I} -type dimer has two active sites at the $\alpha^{\text{I}}\text{--}\alpha^{\text{I}'}$ interface. In contrast, two intimate $\alpha^{\text{N}}\beta^{\text{C}}$ dimers in the P-protein are related by a two-fold axis to form an $(\alpha^{\text{N}}\beta^{\text{C}})_2$ tetramer, in which each $\alpha^{\text{N}}\beta^{\text{C}}$ -type dimer has only one active site at the $\alpha^{\text{N}}\text{--}\beta^{\text{C}}$ interface. The α^{N} - and β^{C} -subunits show 24% sequence identity and have similar structures, and so the $\alpha^{\text{N}}\beta^{\text{C}}$ dimer appears to have an approximate two-fold axis (Figure 3A) and to mimic an α_2^{I} dimer such as those seen in GluDC (Figure 3B). These data suggest that the $\alpha^{\text{N}}\beta^{\text{C}}$ -type active dimer of the P-protein arose by gene duplication of a homodimeric ancestor, after which the ancestral P-protein has structurally diverged such that the protein has been specifically adapted for use as a multienzyme complex component, even though this involved the loss of one active site. More specifically, in the other fold-type I enzymes including the presumed homodimeric ancestor, the pocket size of the active site has remained fixed so that only small substrates have been recognized (Supplementary Table I). In the P-protein, the pocket size had to expand to recognize macromolecular substrates such as H-protein, and this may have required major rebuilding on a molecular scale, involving the loss of one active site, and thereby leading to an asymmetric active dimeric form for the enzyme.

In contrast to the *Tth* P-protein, the human P-protein occurs in a homodimeric form (α_2^{NC}). Our results suggest that the human P-protein consists of 'dimer-like active monomers', each of which corresponds to the $\alpha^{\text{N}}\beta^{\text{C}}$ -type active dimer seen in the *Tth* P-protein. Figure 8 shows a model structure of the human P-protein. The structures of $\alpha^{\text{NC}}\text{--N}$ and $\alpha^{\text{NC}}\text{--C}$ of the human P-protein model are essentially the same as those of α^{N} and β^{C} of the *Tth* P-protein, respectively. The two halves of the model are connected by a 17-residue stretch of residues from Ser492 to Ile508. The corresponding 31-residue linker region in the *Tth* P-protein stretches from residues 437 α to 467 β (Figure 4). These regions in the human and *Tth* P-proteins do not exhibit significant sequence identity, and the different lengths of the linker regions indicate

Figure 7 Lipoamide-binding channel and the molecular surface properties neighboring its entrance. (A) Ribbon diagrams with molecular surfaces of the P-protein (upper) and the H-protein (lower), where PLP and the lipoyl-lysine arm are represented by yellow and cyan spheres, respectively. (B) Electrostatic potential surfaces, onto which the negative (red) and positive (blue) charges are mapped. (C) Space-filling models, colored as follows: red, fully conserved residues; yellow, conservation of 'strong' group residues; green, conservation of 'weaker' group residues, where the groups are according to the CLUSTAL W (Thompson *et al*, 1994) manual. (D) Lipoamide-binding channel viewed from its entrance: molecular surface (left) and the residues constituting the channel (right, stereoview). The model structure for the lipoyl-lysine arm was manually docked into the channel: first, the aminomethylated lipoyl-lysine was connected to PLP via an aldimine bond, as shown in Supplementary Figure 1E, in place of AOA and positioned within the channel, avoiding significant steric clashes; then, the aminomethyl group was removed.

that the structures are locally divergent here. From the viewpoint of molecular evolution, we suggest that a dimer-like active monomer (α^{NC}) has evolved by gene fusion of the two subunits of the $\alpha^{\text{N}}\beta^{\text{C}}$ -type active dimer. The two subunits had arisen by gene duplication and evolutionary divergence, as described in the previous paragraph. This hypothesis is supported by the higher sequence identity between the $(\alpha^{\text{N}}\beta^{\text{C}})_2$ -tetrameric and α^{NC} -dimeric P-proteins; α^{NC} -N and α^{NC} -C share ~ 31 and $\sim 37\%$ sequence identities with α^{N} and β^{C} , respectively, compared to the sequence identities of ~ 24 and $\sim 13\%$, which resulted from alignments between α^{N} and β^{C} , and between α^{NC} -N and α^{NC} -C, respectively. Many mutations in the human P-protein gene have been identified in NKH patients: a three-base deletion resulting in deletion of

Phe756 (F756del) (Kure *et al*, 1991), 10 missense mutations (A283P, A313P, P329T, R410 K, P700A, G762R (Toone *et al*, 2002), R515S (Toone *et al*, 2000), S564I (Kure *et al*, 1992), G761R (Kure *et al*, *Am J Hum Genet* **65**, A425 (1999), meeting abstract) and A802V (Korman *et al*, 2004)) and two mutations (Takayanagi *et al*, 2000; Kure *et al*, 2002) resulting in highly altered polypeptides. Most of these 13 mutations have been found in single patients, but three are recurrent mutations (Applegarth and Toone, 2001): S564I and G761R in 78% (42/54) of alleles of Finnish patients and R515S in 5% of alleles of non-Finnish patients. Meanwhile, our results support the hypothesis that the human and *Tth* P-proteins share a similar three-dimensional structure. They also allow us to identify functionally important residues involved in recogni-

Figure 8 Stereoview of the NKH mutations mapped on the model structure of the human P-protein. The 11 point mutations associated with the disease are shown in magenta. The orientation is the same as that in Figure 3A. The linker region connecting $\alpha^{\text{NC-N}}$ and $\alpha^{\text{NC-C}}$ is represented by a cyan curve with spheres placed at intervals of 3.8 Å (the mean length between two successive C α atoms), while the corresponding region of *Tth* P-protein is shown in semitransparent gray. The other regions are colored by domain according to Figure 3A. In constructing the model, we started with the *Tth* P-protein backbone, and then added a linker region along the molecular surface, so as to select the shortest distance between the α^{N} C-terminus (Leu437 α in *Tth*) and one of the residues in the N-terminal region of β^{C} or $\beta^{\text{C'}}$, assuming that the overall arrangement of the human α_2^{NC} dimer subunits is similar to that of the *Tth* ($\alpha^{\text{N}}\beta^{\text{C}}$)₂ tetramer. The resulting linker region starts from Leu437 α and leads to Leu467 β , and is composed of 13 residues. Our model linker is shorter than the actual linker in the human P-protein, which consists of 17 residues connecting Ser492 with Ile508. The human P-protein linker thus may contain helical and other regions that are similar to the corresponding region in the *Tth* P-protein.

tion of the cofactor, substrate and possibly the H-protein, and show that most of these residues are well conserved among these P-proteins. Hence, the catalytic mechanisms should be essentially the same for both the *Tth* and human P-proteins.

We will now discuss the possible molecular mechanisms of the human P-protein, with our arguments based on the *Tth* P-protein structure, and we will also explore how pathogenic mutations (Figure 8) lead to loss of enzymatic activity. The *Tth* P-protein residue Phe706 β (Phe756 in the human P-protein; note that the corresponding residue number in the human P-protein is shown in parentheses) is fully conserved (Figure 4), and is situated close to the active site, where it is only two residues from the catalytic residue Lys704 β ; thus, F756del would disrupt the active site. Ala221 α (Ala283) and Ala251 α (Ala313) lie in the middle of helix 8 and strand f, respectively, and A283P or A313P would therefore break these secondary structures. Pro267 α (Pro329) is fully conserved and is involved in the conformational changes upon binding of PLP, suggesting that P329T affects binding of the cofactor. Pro650 β (Pro700) is a fully conserved residue in the conserved sequence motif MhTxPxT, where 'x' indicates any amino acid and 'h' is a hydrophobic residue. This proline residue adopts the *cis*-conformation and lies adjacent to the active site, and therefore P700A is likely to affect the active-site structure. Arg474 β (Arg515) is a fully conserved residue located at the $\alpha^{\text{N}}\text{--}\beta^{\text{C}}$ interface. The side chain of Arg474 β forms hydrogen bonds with the main-chain C=O groups of Glu119 α and Leu120 α , both of which are situated at the C-terminus of helix 5. The dipole of this helix is well balanced with the positive charge of Arg474 β . Hence, R515S would destabilize the corresponding interface. Pro518 β (Ser564) is located in tightly packed structure proximate to the con-

served sequence motif PLGSCTMKhN constituting the active site, suggesting that S564I affects the active-site structure. Gly711 β (Gly761) and Gly712 β (Gly762) are fully conserved residues in the sequence motif PHGGGGPG. Both residues lie in a tightly packed structure at the $\alpha^{\text{N}}\text{--}\beta^{\text{C}}$ interface neighboring the active site. In G761R or G762R, the introduction of bulky side chains likely hampers the subunit association at the corresponding interface and disrupts the active site. Therefore, there are compelling structural arguments that can explain the observed loss of enzymatic P-protein activity associated with each of these mutations. There is, however, no structural explanation for the effect of the other two mutations: A802V is situated at a region where the sequences of the two proteins are locally divergent, and the residue altered by the mutation R410 K aligns with Lys348 α in the *Tth* P-protein. Nevertheless, in most of the cases, the *Tth* P-protein structure provides a molecular basis for understanding how the NKH mutations lead to loss of enzymatic activity.

In conclusion, the *Tth* P-protein structure greatly aids in understanding the molecular pathology of NKH in humans. Plausible arguments have been presented for the molecular basis of point mutations associated with this disease, including three recurrent mutations (Applegarth and Toone, 2001). The structural information will be useful in interpreting other NKH mutations that may be identified in the future. Furthermore, our work on P-protein provides a structural characterization of the final GCS component with an unknown structure. It gives new insights into the molecular evolution of P-proteins, provides a structural basis for understanding the architecture and function of this fascinating complex, and should stimulate further structure/function studies.

Materials and methods

Crystallization

The expression of *Tth* P-protein by *E. coli*, and purification and crystallization of the expressed protein have been reported elsewhere (Nakai *et al*, 2003b). Briefly, the purification was accomplished by heat treatment and four-step column chromatography using 20 mM Tris-HCl buffer (pH 8.0). The purified protein was crystallized by vapor diffusion at 298 K using 30% (w/v) polyethylene glycol (PEG) 3350 and 300 mM KSCN as the precipitant (form I in Table I). However, since the crystallization rate of form I proved to be insufficient for structure determination by using heavy-atom derivatives, we searched for new crystallization conditions to find a new crystal form more suitable for structure determination (form II in Table I).

The purification conditions for form II were essentially the same as those for form I except that 20 mM Tris-HCl buffer (pH 8.0) was replaced with 10 mM sodium phosphate buffer (pH 7.0). The purified protein was crystallized by vapor diffusion at 277 K using 22% (w/v) PEG 3350, 300 mM Li₂SO₄ and 100 mM MES-NaOH (pH 6.5) as the precipitant. A droplet containing 2 μ l protein solution (10 mg ml⁻¹ protein, 20 mM HEPES-NaOH (pH 8.0), 200 μ M PLP and 1 mM DTT) was mixed with an equal volume of reservoir solution and equilibrated against 400 μ l of reservoir solution to give crystals of form II. Finally, the structures of forms I and II were revealed to be P_{apo} and P_{holo}, respectively.

Crystal soaking and data collection

A crystal of P_{holo} was soaked for 3 h in artificial mother liquor containing an inhibitor, AOA (26% (w/v) PEG 3350, 300 mM Li₂SO₄, 5 mM AOA, 100 mM MES-NaOH (pH 6.5)), generating a crystal of P_{holo}·AOA. A mercury derivative was prepared by soaking a crystal of P_{holo} in the same liquor containing 1 mM methylmercury chloride (CH₃HgCl) for 3 h. All soaking experiments were performed at 277 K.

Prior to data collection, each crystal of P_{holo}, P_{holo}·AOA and the mercury derivative was soaked for a few seconds in a cryoprotectant solution containing 18% (w/v) PEG 400, 22% (w/v) PEG 3350, 300 mM Li₂SO₄ and 100 mM MES-NaOH (pH 6.5) and then transferred to a nylon loop and flash-cooled in a nitrogen gas stream at 90 K. The data collection preparation for P_{apo} was the same as above, except that its cryoprotectant solution contained 35% (w/v) PEG 3350 and 300 mM KSCN. X-ray diffraction data were collected at 90 K using synchrotron radiation from BL45XU (Yamamoto *et al*, 1998) or BL44B2 (Adachi *et al*, 2001) at SPring-8 (Table I). Data processing was completed using the program HKL2000 (Otwinowski and Minor, 1997) (Table I).

Structure determination

The structure of P_{holo} was solved by single isomorphous replacement with anomalous scattering (SIRAS) with a CH₃HgCl derivative. A total of 11 mercury sites of the derivative were identified and refined at 3.38 Å resolution by the program SOLVE (Terwilliger, 2002), which gave an overall figure of merit of 0.32. The SOLVE phases were subsequently input to the program RESOLVE (Terwilliger, 2002) for density modification and automatic model building. RESOLVE gave an overall figure of merit of 0.48 and built 79% of the peptide backbone in the asymmetric unit including 62% of the side chains. The electron density map after density modification was of good quality, and the remaining model was built manually into the map using the program Xfit within the software package XtalView (McRee, 1992).

References

- Adachi S, Oguchi T, Tanida H, Park SY, Shimizu H, Miyatake H, Kamiya N, Shiro Y, Inoue Y, Ueki T, Iizuka T (2001) The RIKEN structural biology beamline II (BL44B2) at the SPring-8. *Nucl Instrum Methods A* **467**: 711–714
- Alexeev D, Alexeeva M, Baxter RL, Campopiano DJ, Webster SP, Sawyer L (1998) The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *J Mol Biol* **284**: 401–419

The structure of P_{apo} was solved by molecular replacement using the program AMoRe (Navaza, 1994), using the refined coordinates of P_{holo} as the search model. The model was constructed from the ($\alpha^N\beta^C$)₂ tetramer, in which all the PLP and water molecules were removed. A rotational search using data from 15 to 4 Å resolution gave a distinct solution, which was followed by a translational solution showing an ($\alpha^N\beta^C$)₂ tetramer in the asymmetric unit with an *R*_{factor} of 34.1%.

Structure refinement

The structure of P_{holo} was refined by simulated annealing and energy minimization with the program CNS (Brunger *et al*, 1998). After several rounds of refinement and manual rebuilding, the difference Fourier map clearly exhibited residual electron density corresponding to the bound PLP. The electron density of PLP was connected to the ϵ -amino group of Lys704 β (Figure 6B), indicating that PLP is covalently bound to Lys704 β via an aldimine bond, like those observed for other PLP-enzymes. Water-molecule picking using the program CNS and further model building and refinement cycles produced the final model (Table I). The structure of P_{holo}·AOA was refined using the same procedure as that used for P_{holo}. The refined coordinates of P_{holo}, except for the PLP and water molecules, were used as the initial model. When the *R*_{work} was below ~25%, the difference Fourier map clearly exhibited residual electron density corresponding to the AOA and PLP molecules. The electron density of AOA was connected to the C4' atom of PLP and the density corresponding to the linkage between PLP and Lys704 β found in P_{holo} was absent (Figure 6C). This result indicates that the internal aldimine between PLP and Lys704 β is cleaved and a new external aldimine is formed between PLP and AOA. The structure of P_{apo} was refined using the same procedure as P_{holo}, except that the coordinates obtained by molecular replacement were used as the initial model. No electron density for PLP was observed, showing that the crystals represented the apo-form of the enzyme.

Analysis of the stereochemistry showed that every model was of good quality, with more than 99.0% of the residues falling in the allowed regions (Table I). Eight proline residues of each heterodimer (205 α , 267 α , 373 α , 650 β , 715 β , 733 β , 845 β and 887 β), except for Pro267 α in P_{apo}, were found in the *cis*-conformation.

Graphical representations of protein models were generated by MOLSCRIPT (Kraulis, 1991) (Figures 2, 3 and 5–8), Xfit (McRee, 1992) (Figure 6) and GRASP (Nicholls *et al*, 1991) (Figure 7), and rendered with Raster3d (Merritt and Bacon, 1997) (Figures 2, 3 and 5–8).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We thank Dr H Naitow, Dr Y Kawano, Mr T Matsu and Mr H Nakajima (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, by a JSPS Research Fellowship for Young Scientists (no. 15-03803) 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. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 1WYT, 1WYU and 1WYV for P_{apo}, P_{holo} and P_{holo}·AOA, respectively.

- Applegarth DA, Toone JR (2001) Nonketotic hyperglycinemia (glycine encephalopathy): laboratory diagnosis. *Mol Genet Metab* **74**: 139–146
- Bertoldi M, Gonsalvi M, Contestabile R, Voltattorni CB (2002) Mutation of tyrosine 332 to phenylalanine converts dopa decarboxylase into a decarboxylation-dependent oxidative deaminase. *J Biol Chem* **277**: 36357–36362
- Bourguignon J, Neuburger M, Douce R (1988) Resolution and characterization of the glycine-cleavage reaction in pea leaf

- mitochondria. Properties of the forward reaction catalysed by glycine decarboxylase and serine hydroxymethyltransferase. *Biochem J* **255**: 169–178
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Burkhard P, Dominici P, Borri-Voltattorni C, Jansonius JN, Malashkevich VN (2001) Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. *Nat Struct Biol* **8**: 963–967
- Capitani G, De Biase D, Aurizi C, Gut H, Bossa F, Grutter MG (2003) Crystal structure and functional analysis of *Escherichia coli* glutamate decarboxylase. *EMBO J* **22**: 4027–4037
- Cohen-Addad C, Pares S, Sieker L, Neuburger M, Douce R (1995) The lipoamide arm in the glycine decarboxylase complex is not freely swinging. *Nat Struct Biol* **2**: 63–68
- Douce R, Bourguignon J, Neuburger M, Rebeille F (2001) The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci* **6**: 167–176
- Dunathan HC (1966) Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc Natl Acad Sci USA* **55**: 712–716
- Faure M, Bourguignon J, Neuburger M, MacHerel D, Sieker L, Ober R, Kahn R, Cohen-Addad C, Douce R (2000) Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system 2. Crystal structures of H- and L-proteins. *Eur J Biochem* **267**: 2890–2898
- Ford GC, Eichele G, Jansonius JN (1980) Three-dimensional structure of a pyridoxal-phosphate-dependent enzyme, mitochondrial aspartate aminotransferase. *Proc Natl Acad Sci USA* **77**: 2559–2563
- Freudenberg W, Andreesen JR (1989) Purification and partial characterization of the glycine decarboxylase multienzyme complex from *Eubacterium acidaminophilum*. *J Bacteriol* **171**: 2209–2215
- Fujiwara K, Okamura-Ikeda K, Motokawa Y (1987) Amino acid sequence of the phosphopyridoxyl peptide from P-protein of the chicken liver glycine cleavage system. *Biochem Biophys Res Commun* **149**: 621–627
- Gariboldi RT, Drake HL (1984) Glycine synthase of the purinolytic bacterium, *Clostridium acidurici*. Purification of the glycine-CO₂ exchange system. *J Biol Chem* **259**: 6085–6089
- Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305–308
- Grishin NV, Phillips MA, Goldsmith EJ (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci* **4**: 1291–1304
- Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**: 123–138
- Ishii S, Mizuguchi H, Nishino J, Hayashi H, Kagamiyama H (1996) Functionally important residues of aromatic L-amino acid decarboxylase probed by sequence alignment and site-directed mutagenesis. *J Biochem (Tokyo)* **120**: 369–376
- Jansonius JN (1998) Structure, evolution and action of vitamin B₆-dependent enzymes. *Curr Opin Struct Biol* **8**: 759–769
- Korman SH, Boneh A, Ichinohe A, Kojima K, Sato K, Ergaz Z, Gomori JM, Gutman A, Kure S (2004) Persistent NKH with transient or absent symptoms and a homozygous GLDC mutation. *Ann Neurol* **56**: 139–143
- Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**: 946–950
- Kume A, Koyata H, Sakakibara T, Ishiguro Y, Kure S, Hiraga K (1991) The glycine cleavage system. Molecular cloning of the chicken and human glycine decarboxylase cDNAs and some characteristics involved in the deduced protein structures. *J Biol Chem* **266**: 3323–3329
- Kure S, Kojima K, Ichinohe A, Maeda T, Kalmanchev R, Fekete G, Berg SZ, Filiano J, Aoki Y, Suzuki Y, Izumi T, Matsubara Y (2002) Heterozygous GLDC and GCSH gene mutations in transient neonatal hyperglycinemia. *Ann Neurol* **52**: 643–646
- Kure S, Narisawa K, Tada K (1991) Structural and expression analyses of normal and mutant mRNA encoding glycine decarboxylase: three-base deletion in mRNA causes nonketotic hyperglycinemia. *Biochem Biophys Res Commun* **174**: 1176–1182
- Kure S, Takayanagi M, Narisawa K, Tada K, Leisti J (1992) Identification of a common mutation in Finnish patients with nonketotic hyperglycinemia. *J Clin Invest* **90**: 160–164
- Lee HH, Kim DJ, Ahn HJ, Ha JY, Suh SW (2004) Crystal structure of T-protein of the glycine cleavage system: cofactor binding, insights into H-protein recognition, and molecular basis for understanding nonketotic hyperglycinemia. *J Biol Chem* **279**: 50514–50523
- McRee DE (1992) A visual protein crystallographic software system for X11/XView. *J Mol Graphics* **10**: 44–46
- Mehta PK, Christen P (2000) The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv Enzymol Relat Areas Mol Biol* **74**: 129–184
- Merritt EA, Bacon DJ (1997) Raster3D photorealistic molecular graphics. *Methods Enzymol* **277**: 505–524
- Motokawa Y, Fujiwara K, Okamura-Ikeda K (1995) Role of lipoic acid in the glycine cleavage system and lipoylation of H-protein. In *Bithiols in Health and Disease*, Packer L, Cadenas E (eds) pp 389–407. New York: Marcel Dekker
- Nakai T, Ishijima J, Masui R, Kuramitsu S, Kamiya N (2003a) Structure of *Thermus thermophilus* HB8 H-protein of the glycine-cleavage system, resolved by a six-dimensional molecular-replacement method. *Acta Crystallogr D* **59**: 1610–1618
- Nakai T, Nakagawa N, Masui R, Kuramitsu S, Kamiya N (2003b) Coexpression, purification, crystallization and preliminary X-ray characterization of glycine decarboxylase (P-protein) of the glycine-cleavage system from *Thermus thermophilus* HB8. *Acta Crystallogr D* **59**: 554–557
- Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50**: 157–163
- Neuburger M, Polidori AM, Pietre E, Faure M, Jourdain A, Bourguignon J, Pucci B, Douce R (2000) Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system. 1. Biochemical studies. *Eur J Biochem* **267**: 2882–2889
- Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296
- Okamura-Ikeda K, Ohmura Y, Fujiwara K, Motokawa Y (1993) Cloning and nucleotide sequence of the *gcv* operon encoding the *Escherichia coli* glycine-cleavage system. *Eur J Biochem* **216**: 539–548
- Oliver DJ, Neuburger M, Bourguignon J, Douce R (1990) Interaction between the component enzymes of the glycine decarboxylase multienzyme complex. *Plant Physiol* **94**: 833–839
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326
- Pares S, Cohen-Addad C, Sieker L, Neuburger M, Douce R (1994) 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. *Proc Natl Acad Sci USA* **91**: 4850–4853
- Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**: 961–1004
- Schneider G, Kack H, Lindqvist Y (2000) The manifold of vitamin B₆ dependent enzymes. *Struct Fold Des* **8**: R1–R6
- Tada K, Kure S (1993) Non-ketotic hyperglycinaemia: molecular lesion, diagnosis and pathophysiology. *J Inher Metab Dis* **16**: 691–703
- Takayanagi M, Kure S, Sakata Y, Kurihara Y, Ohya Y, Kajita M, Tada K, Matsubara Y, Narisawa K (2000) Human glycine decarboxylase gene (GLDC) and its highly conserved processed pseudogene (ψ GLDC): their structure and expression, and the identification of a large deletion in a family with nonketotic hyperglycinemia. *Hum Genet* **106**: 298–305
- Terwilliger TC (2002) Automated structure solution, density modification and model building. *Acta Crystallogr D* **58**: 1937–1940
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680

- Toone JR, Applegarth DA, Coulter-Mackie MB, James ER (2000) Biochemical and molecular investigations of patients with nonketotic hyperglycinemia. *Mol Genet Metab* **70**: 116–121
- Toone JR, Applegarth DA, Kure S, Coulter-Mackie MB, Sazegar P, Kojima K, Ichinohe A (2002) Novel mutations in the P-protein (glycine decarboxylase) gene in patients with glycine encephalopathy (non-ketotic hyperglycinemia). *Mol Genet Metab* **76**: 243–249
- Westhead DR, Slidel TW, Flores TP, Thornton JM (1999) Protein structural topology: automated analysis and diagrammatic representation. *Protein Sci* **8**: 897–904
- Yamamoto M, Kumasaka T, Fujisawa T, Ueki T (1998) Trichromatic concept at SPring-8 RIKEN beamline I. *J Synchrotron Radiat* **5**: 222–225