Crystal Structure of T-protein of the Glycine Cleavage System

COFACTOR BINDING, INSIGHTS INTO H-PROTEIN RECOGNITION, AND MOLECULAR BASIS FOR UNDERSTANDING NONKETOTIC HYPERGLYCINEMIA*

Received for publication, August 23, 2004, and in revised form, September 3, 2004
Published, JBC Papers in Press, September 7, 2004, DOI 10.1074/jbc.M409672200

Hyung Ho Lee‡§, Do Jin Kim‡§, Hyung Jun Ahn‡, Jun Yong Ha, and Se Won Suh¶

From the Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea

The glycine cleavage system catalyzes the oxidative decarboxylation of glycine in bacteria and in mitochondria of animals and plants. Its deficiency in human causes nonketotic hyperglycinemia, an inborn error of glycine metabolism. T-protein, one of the four components of the glycine cleavage system, is a tetrahydrofolate-dependent aminomethyltransferase. It catalyzes the transfer of the methylene carbon unit to tetrahydrofolate from the methyamine group covalently attached to the lipoamide arm of H-protein. To gain insight into the T-protein function at the molecular level, we have determined the first crystal structure of T-protein from Thermotoga maritima by the multiwavelength anomalous diffraction method of x-ray crystallography and refined four structures: the apoform; the tetrahydrofolate complex; the folic acid complex; and the lipoic acid complex. The overall fold of T-protein is similar to that of the C-terminal tetrahydrofolate-binding region (residues 421–830) of Arthrobacter globiformis dimethylglycine oxidase. Tetrahydrofolate (or folic acid) is bound near the center of the tripartite T-protein. Lipoic acid is bound adjacent to the tetrahydrofolate binding pocket, thus defining the interaction surface for H-protein binding. A homology model of the human T-protein provides the structural framework for understanding the molecular mechanisms underlying the development of nonketotic hyperglycinemia due to missense mutations of the human T-protein.

* This work was supported by a grant from the Korea Ministry of Science and Technology (NRL-2001, Grant M1-1004-00-0132). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: GCS, glycine cleavage system; NKH, nonketotic hyperglycinemia; MAD, multiwavelength anomalous diffraction; H4folate, tetrahydrofolate; SeMet, selenomethionine; Tm, Thermotoga maritima.

‡ Both authors contributed equally to this work.
§ Recipients of the BK21 fellowship.
¶ To whom correspondence should be addressed. Tel.: 82-2-880-6653; Fax: 82-2-889-1568; E-mail: sewonsuh@snu.ac.kr.

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.
length anomalous diffraction (MAD) method. Here we report the crystal structure of T-protein from *Thermotoga maritima* (Tm) in four forms: the apo; the folinic acid complex; the H$_4$folate complex; and the reduced lipoic acid complex. This study provides essential structural information on cofactor/ inhibitor binding and useful insights into H-protein recognition by T-protein. It also provides the structural framework for study provides essential structural information on cofactor/ligand complex; and the reduced lipoic acid complex. This study provides essential structural information on cofactor/inhibitor binding and useful insights into H-protein recognition by T-protein. 

### EXPERIMENTAL PROCEDURES

**Protein Expression and Purification**—The gei/T gene (TM0211) encoding Tm-T-protein was cloned into the expression vector pET-28a (+) (Novagen). The intact protein without any purification tag was overexpressed in *E. coli* BL34(DE3) cells using terrific broth culture medium. Protein expression was induced by 1 mM isopropyl-1-thio-β-D-galactopyranoside, and the cells were incubated for an additional 30 h at 15 °C following growth to mid-log phase at 37 °C. The cells were lysed by sonication in 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration. The protein was further purified by gel filtration on a HiLoad XK-16 Superdex 200 prep-grade column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2) and 200 mM NaCl. Following heat treatment at 80 °C for 10 min, the sample was centrifuged at 18,000 rpm for 60 min. The supernatant was applied to a HiLoad 26/10 Q-Sephrose column (Amersham Biosciences), which was previously equilibrated with 50 mM Tris-HCl (pH 7.2). Following elution with a gradient of NaCl in the same buffer, Tm-T-protein was eluted at 250–300 mM NaCl concentration.
protein grew up to approximate dimensions of $0.2 \times 0.2 \times 0.3$ mm within a few days. The crystals of the $\text{H}_4\text{folate}$ complex was obtained by soaking crystals of the SeMet-substituted protein in a $\text{H}_4\text{folate}$-saturated solution (20% (w/v) polyethylene glycol 3350, 200 mM sodium dihydrogen phosphate (pH 4.25), and 10 mM dithiothreitol) for 24 h before cryoprotection. To obtain the reduced lipoic acid complex, the crystals of the SeMet-substituted protein were soaked in 50% (v/v) dimethyl sulfoxide solution containing 10 mM dithiothreitol, which was previously saturated with lipoic acid, for 2 min before cryoprotection.

**X-ray Data Collection and Structure Determination**—A crystal of the SeMet-substituted protein was frozen using a cryoprotectant solution containing 25% (v/v) glycerol in the crystallization mother liquor. X-ray diffraction data were collected at 100 K on a Bruker CCD area detector system at the Beamline-6B experimental station of Pohang Light Source. For each image, the crystal was rotated by 1° and the crystal-to-detector distance was set to 360 mm. The raw data were processed and scaled using the program suite HKL2000 (19). The SeMet-substituted crystal belongs to the space group $P_{2_1}2_12_1$, with unit cell parameters of $a = 52.61$ Å, $b = 54.16$ Å, and $c = 149.44$ Å. Table I summarizes the statistics of MAD data collection. All of the eleven expected selenium atoms of a monomer in each crystallographic asymmetric unit were located with the program SOLVE (20), and the selenium sites were used to calculate the phases with RESOLVE (21). Phasing statistics are summarized in Table I. X-ray diffraction data of the $\text{H}_4\text{folate}$ complex and folinic acid complex were collected as above. Data of the reduced lipoic acid-bound crystal were collected at 100 K on a Quantum 315 CCD detector (Area Detector Systems Corporation, Poway, CA) at the Beamline-5A experimental station of Photon Factory, whereas the data of the apoform were collected at the Beamline-18B experimental station on an ADSC Quantum 4R CCD detector. The raw data were processed and scaled using the program suite HKL2000 (19).

**Model Building and Refinement**—Excellent quality of the electron density map allowed automatic model building by the program RESOLVE (21), giving an initial model that accounted for ~70% of the backbone of the polypeptide chain with much of the sequence assigned. Subsequent manual model building was done using the program O (22). The model was refined with the program CNS (23), including the bulk solvent correction. 10% of the data were randomly set aside as the test data for the calculation of $R_{	ext{free}}$ (24). Several rounds of model building, simulated annealing, positional refinement, and individual $B$-factor refinement were performed. Subsequently, this model was used to refine structures of the apoform, the folinic acid-bound form, the $\text{H}_4\text{folate}$-bound form, and the lipoic acid-bound form. Refinement statistics are summarized in Table I. All of the models have excellent stereochemistry (Table I) as evaluated by the program PROCHECK (25).

**Homology Modeling of Human T-protein**—A structural model of the human GCS T-protein (Leu1–Phe394) was built by the homology modeling server (swissmodel.expasy.org/) using the crystal structure of Tm T-protein as template. The N-terminal (Met1–Val32) and C-terminal (Val395–Lys403) regions of the human T-protein were not modeled due to a lack of significant sequence homology.

**RESULTS AND DISCUSSION**

**Overall Structure**—We have determined the crystal structure of Tm T-protein by the MAD method and refined four structures: (i) the apoform at 1.84 Å; (ii) the complex with (S)-folinic acid (5-formyl-5,6,7,8-tetrahydrofolic acid) at 2.0 Å; (iii) the complex with $\text{H}_4\text{folate}$ at 2.4 Å; and (iv) the complex with reduced lipoic acid at 1.95 Å. The latter three models account for residues 1–362 of one T-protein monomer in an
asymmetric unit, whereas Arg\textsuperscript{362} is additionally missing from the apostructure. The missing residues have no electron density. Tm T-protein is monomeric and is oblate-shaped with approximate dimensions of 55×50×30 Å (Fig. 1). It is tripartite. Its three domains are positioned in a cloverleaf-like arrangement. Domain 1 (residues 1–51 and 140–240) consists of a predominantly antiparallel, six-stranded \( \beta \)-sheet that contains a single Greek-key motif packed on one side by three \( \alpha \)-helices (\( \alpha_1, \alpha_5, \) and \( \alpha_6 \)) and on the other side by two \( \alpha \)-helices (\( \alpha_2 \) and \( \alpha_7 \)). Domain 2 (residues 52–139 and 241–280), which includes a long excursion from domain 1, has a five-stranded antiparallel \( \beta \)-sheet with flanking \( \alpha \)-helices. The two antiparallel \( \beta \)-sheets from domains 1 and 2 are loosely packed against each other. The C-terminal domain 3 (residues 281–362) forms a distorted six-stranded jelly roll that packs perpendicular with the \( \beta \)-sheets of domains 1 and 2. The C-terminal tail (residues 354–362) of domain 3 covers part of domain 1 (Fig. 1, front side) on the opposite side of the N terminus.

DALI structural similarity searches (26) with the Tm T-protein apostructure identified two close relatives: the C-terminal H4folate-binding region (residues 421–830) of Arthrobacter globiformis dimethylglycine oxidase (PDB code 1PJ5; a root mean square (r.m.s.) deviation of 1.53 Å for 361 equivalent C\textsubscript{\text{\( \alpha \)}} positions, a Z-score of 47.3, and a sequence identity of 24%) (27) and the \textit{E. coli} Ygfz protein of unknown biological function (PDB code 1NRK) (an r.m.s. deviation of 3.2 Å for 299 equivalent Ca positions, a Z-score of 26.5, and a sequence identity of 14%).

**Binding Mode of H4folate and Folinic Acid**—H4folate-bound and folinic acid-bound structures of Tm T-protein are virtually identical with an r.m.s. deviation of 0.20 Å for 361 Ca atoms (Met\textsuperscript{1}-Arg\textsuperscript{361}). H4folate and folinic acid are bound near the center of Tm T-protein in essentially identical manners (Figs. 1 and 2). Both H4folate and folinic acid adopt a kinked conformation (Fig. 2). The H4folate-bound structure is also nearly identical to the apostructure with an r.m.s. deviation of 0.23 Å for 361 C\textsubscript{\text{\( \alpha \)}} atoms (Met\textsuperscript{1}-Arg\textsuperscript{361}), and the central hole has similar solvent accessible pocket volumes in both the apostructure and the H4folate complex structure. This finding suggests that Tm T-protein has a rigid H4folate binding pocket. The mouth opening of the central hole is more open on the C-terminal side or the glutamate tail side (Fig. 1A, front side) than the N-terminal side or the H-protein interaction side (Fig. 1A, back side). Interactions of T-protein with H-protein and the glutamate tail of H4folate are further discussed below.

Side chains of Asp\textsuperscript{96}, Tyr\textsuperscript{100}, Tyr\textsuperscript{169}, Tyr\textsuperscript{188}, Glu\textsuperscript{195}, Arg\textsuperscript{227}, and Arg\textsuperscript{362} as well as the carbonyl oxygen of Val\textsuperscript{110} interact directly with H4folate (Fig. 2A). Five of these residues (Asp\textsuperscript{96}, Tyr\textsuperscript{100}, Tyr\textsuperscript{188}, Glu\textsuperscript{195}, and Arg\textsuperscript{227}) are well conserved among bacterial T-proteins (Fig. 3). Tyr\textsuperscript{83}, Tyr\textsuperscript{168}, Tyr\textsuperscript{236}, Leu\textsuperscript{237}, and
Tyr\textsuperscript{239} interact with H\textsubscript{4}folate indirectly through water molecules or through protein main chain atoms. Two of them (Tyr\textsuperscript{83} and Tyr\textsuperscript{239}) are well conserved among bacterial T-proteins (Fig. 3). The pterin group of H\textsubscript{4}folate is largely buried in the central cavity. The side chain O\textsubscript{2}/H\textsubscript{9254} atom of Asp\textsuperscript{96} makes a hydrogen bond with the N10 nitrogen atom of the pterin group (3.05 Å). The pterin N8 atom interacts with the carbonyl oxygen of Val\textsuperscript{110}. Both the hydroxyl oxygen atom of Tyr\textsuperscript{100} and the O\textsubscript{1} atom of Glu\textsuperscript{195} contact the NA\textsubscript{2} atom of H\textsubscript{4}folate. The side chain of Arg\textsuperscript{227} is close to the O4 atom of H\textsubscript{4}folate. The N\textsubscript{2}/H\textsubscript{9257} atom makes a direct hydrogen bond (3.31 Å), whereas the N\textsubscript{1} atom is hydrogen-bonded through a water molecule (4.12 Å). The oxygen atom of Tyr\textsuperscript{169} is hydrogen-bonded to the carbonyl oxygen of the glutamyl group of H\textsubscript{4}folate (2.98 Å), whereas the side chain of Arg\textsuperscript{362} makes a salt bridge with the oxygen O\textsubscript{1} atom of the glutamyl group of H\textsubscript{4}folate (3.11 Å). This mode of H\textsubscript{4}folate binding to Tm T-protein resembles that of folinic acid binding to the C-terminal region (residues 430–827) of Arthrobacter globiformis dimethylglycine oxidase (27). Some other structural features around H\textsubscript{4}folate are noteworthy (Fig. 2). fig. Arrows above the sequences denote α-helices and cylinders β-strands. Blue circles above the sequence indicate the residues that interact with H\textsubscript{4}folate with the exception of Asn\textsuperscript{112}, which interacts with folinic acid only. Orange squares below the sequences represent the residues that are close to the bound lipolic acid. Red triangles below the sequences are the missense mutation sites of the human T-protein associated with NKH. Three signature sequence motifs are enclosed by colored boxes: TGYTGXXGE motif (residues 186–195) in magenta; PXGLGARDXXRhEXAXXXLG motif (residues 221 and 240) in blue; and GXXhT/S/T/IGXXSPTL motif (residues 306–317) in green, respectively. The N-terminal region (residues 14–35) of domain 1, which plays a crucial role in H-protein interaction, is also enclosed by a dotted orange box. This figure was drawn with ClustalX (38) and GeneDoc (39).
FIG. 4. Binding of the glutamate tail of H4folate and lipoic acid and a model of the complex between T-protein and H-protein. A, positively charged residues around the glutamate tail of H4folate. The molecular surface is colored according to the electrostatic potential. The positive electrostatic potential is colored in blue, and the negative potential is in red. This figure is drawn with GRASP (40). B, surface diagram showing the conserved residues around the lipoic acid binding pocket. This view is obtained by an ~150° rotation of A. The carboxylate group of the bound lipoic acid sticks out of the pocket. The residues that are strictly conserved in bacterial T-proteins are colored in green, and semi-conserved residues are in yellow. An orange circle denotes the N-terminal portion (residues 1–51) of domain 1. C, stereoview of the T-protein active site. H4folate has been incorporated into the structure of the lipoic acid complex. Two sulfur atoms of lipoic acid are labeled. Side chains of the residues lining the active site are shown. Black dotted lines depict hydrogen bonds. Blue balls represent water molecules. D, a model of the complex between T-protein and H-protein. T-protein is represented by the electrostatic potential at the molecular surface. The backbone of T. thermophilus H-protein (Protein Data Bank code 1ONL) is drawn in blue tubes with a green arrow indicating the probable direction of Lys80 to which the lipoyl moiety is attached. A half-transparent red dot near Lys289 (corresponding to E. coli Lys268) of Tm T-protein indicates the location of Glu92 (corresponding to E. coli Asp91) of T. thermophilus H-protein. This figure is drawn with GRASP (40).

bonded to Asp96 (2.89 and 3.17 Å to the Oδ1 and Oδ2 atoms of Asp96, respectively).

In the folinic acid complex, the same residues that contact H4folate interact with folinic acid. Additionally, the oxygen atom of the N5-formyl group of folinic acid interacts with the side chain of Asn112 (Fig. 2B). H4folate lacks a formyl group and thus does not interact with Asn112 (Fig. 2A). The N5-formyl group is also in proximity of the side chain of Tyr188 with its aromatic ring nearly perpendicular to the p-aminobenzoic acid group (Fig. 2B). In the case of the folinic acid complex, we observe an additional binding of the second folinic acid at a narrow cleft on the surface of domain 2 near the interface between domains 1 and 2 (Fig. 1A). The second folinic acid takes an extended conformation, and the residues that interact with the second folinic acid (the side chains of Glu196, Glu160, Lys173, Ile275, and Glu180, main chain atoms of Met197 and Leu198) are not highly conserved. Furthermore, the average B-factor of the second folinic acid (41.8 Å²) is much higher than the first (18.1 Å²), suggesting a weaker binding of the second folinic acid. This secondary binding of folinic acid is probably a crystallization artifact.

It was reported that a pool of polyglutamate forms of folate is dominated by tetraglutamate (25%) and pentaglutamate (55%) in the pea leaf mitochondria (28). The binding affinity of H4folate polyglutamates for pea leaf T-protein was found to increase with increasing number of glutamates up to six residues (29). This observation may be explained by the presence of a surface patch with highly positive electrostatic potential due to clustering of nine positively charged residues (Lys80, Arg185, Lys280, Lys328, Lys352, Lys354, Arg357, Arg361, and Arg362) in the vicinity of the glutamate tail of H4folate bound to Tm T-protein (Fig. 4A). Five of them (Lys80, Arg185, Lys280, Lys328, and Arg362) are well conserved in bacterial T-proteins. The first glutamate moiety is bound to Tm T-protein in essentially identical manners in both complexes of H4folate and folinic acid (Fig. 2).

Lipoic Acid Binding Reveals Insights into H-protein Recognition—To obtain information on the binding site of the lipoyl arm of H-protein, we have determined the structure of Tm T-protein complexed with the reduced form of lipoic acid. The lipoic acid-bound structure of Tm T-protein is essentially identical to other forms. The r.m.s. deviations for 361 Cα atoms (Met1–Arg261) are 0.13, 0.19, and 0.25 Å for comparing the lipoic acid-bound structure against the apo, H4folate-bound, and folinic acid-bound structures, respectively. This finding suggests that the lipoic acid-binding site of Tm T-protein is relatively rigid, similar to the H4folate binding pocket, and the structure of Tm T-protein changes little upon binding H4folate or lipoic acid. Assuming that the ternary complex with H4folate and lipoic acid is structurally similar
to the binary complexes with H₄folate, folic acid, or lipoic acid, we incorporated H₄folate of the H₄folate complex into the model of the lipoic acid complex to facilitate discussion (Figs. 1 and 4C).

Lipoic acid is bound in an 15-Å deep pocket adjacent to the H₄folate binding pocket with its carboxylate group pointing toward the bulk solvent (Figs. 1 and 4C). The S6 sulfur atom makes a hydrogen bond with the side chain of Asp228 at a distance of 2.90 Å. Arg227 (N9 atom) is close to the S8 sulfur atom (3.23 Å). Phe20, Tyr188, Leu224, and Leu238 as well as the aliphatic part of the Arg227 side chain surround the aliphatic part of lipoic acid (Fig. 4C). Leu224, Arg227, Asp228, and Leu238 are all strictly conserved in bacterial T-proteins, and they belong to the highly conserved sequence motif PXXGLGAR-DXXSPTL between positions 221 and 240 in domain 1 of Tm T-protein (boxed in blue in Fig. 3B), where two possible residues are grouped within parentheses. This signature motif in domain 3 is not directly involved in catalysis or the binding of H₄folate and lipoic acid. It appears to be important for H-protein recognition (further discussed below).

Two structurally conserved water molecules are bound between lipoic acid and H₄folate (Wat47 and Wat114 in Fig. 4C). They are present in all four structures except in the folic acid complex where Wat114 is absent, because the N5-formyl group occupies the site of Wat114. Wat114 is hydrogen-bonded to the S8 atom of lipoic acid at a distance of 2.83 Å and to the O81 atom of Asn112 at 2.88 Å. Wat47 is hydrogen-bonded to the backbone nitrogen atom of Tyr239 at a distance of 2.74 Å and to the O81 atom of Asp96 at 2.84 Å. If a methylamine group were covalently attached to the S8 atom of lipoic acid, Asn112 (O81 oxygen) and Tyr239 (O91 oxygen) would be within hydrogen-bonding distances from the methylamine group. Asn112 and Tyr239 are strictly conserved in T-proteins (Fig. 3).
The binding site of lipoic acid clearly suggests the possible interface for interaction with H-protein. The carboxylate group of lipoic acid bound in the pocket points toward the solvent on the N-terminal side with the other end pointing toward the pterin group of \( \text{H}_4 \text{folate} \) (Figs. 1B and 4B). There is a deep cleft between domains 1 and 3 on this side of Tm T-protein (Fig. 4D). Along this cleft around lipoic acid, many conserved residues are clustered, including Phe20, Leu224, Arg227, Asp228, Leu238, and Tyr239 (Fig. 4B). This cleft appears to be the site of H-protein-binding deletions. Deletions of the N-terminal 4, 7, 11, and 16 residues from the \( E. \ coli \) T-protein led to reduction in the activity to 42, 9, 4, and 0%, respectively, relative to the wild-type enzyme (9, 10). Our Tm T-protein structure revealed that the N-terminal region of domain 1 (boxed in Fig. 4B) contributes to one side of this cleft and that removal of the N-terminal residues 1–13 would seriously affect the folding of the sequence region 14–35 (boxed in orange in Fig. 3; highlighted in thick orange lines in Fig. 1B), causing the distortion of the H-protein binding surface. Therefore, the loss of \( E. \ coli \) T-protein activity caused by a deletion of the N-terminal 16 residues is due to disruption of the H-protein-binding site. It is also apparent that the signature sequence motif GXX(T/S)/S/T/GXXSPTL in domain 3 (boxed in green in Fig. 3; highlighted in thick green lines in Fig. 1B) provides another side of the H-protein-binding cleft (Fig. 1B).

Cross-linking experiments indicated that Lys288 of \( E. \ coli \) T-protein is close to Asp43 of \( E. \ coli \) H-protein when they form a 1:1 complex (9). \( E. \ coli \) Lys288 corresponds to Tm Lys289 and is conserved as lysine or arginine in bacterial T-proteins (Fig. 3). The location of Lys289 in Tm T-protein is shown in Fig. 4, B and D. The lipoam moiety is covalently attached to Lys288 (indicated by a green arrow in Fig. 4D) of \( T. \ thermophilus \) H-protein (16), which has a 57% sequence identity with Tm H-protein. All of these pieces of information allowed us to build a crude but reasonable model of the complex between T-protein and H-protein (Fig. 4D). In this model, H-protein is positioned along the cleft between domains 1 and 3 of T-protein on the N-terminal side, Glu42 (corresponding to \( E. \ coli \) Asp43) of \( T. \ thermophilus \) H-protein (marked by a half-transparent red dot in Fig. 4D) is close to Lys289 of Tm T-protein (corresponding to \( E. \ coli \) Lys288), and Lys63 of \( T. \ thermophilus \) H-protein points toward the lipoic acid binding pocket (as indicated by a green arrow in Fig. 4D). It was suggested that H-protein undergoes a small conformational change upon binding to T-protein so that the lipoam arm carrying the methylamine group is exposed (6). Upon interaction with H-protein, T-protein may also undergo a small structural alteration such as limited domain rearrangement. To characterize the possible structural changes in both T-protein and H-protein that accompany the complex formation, further structural studies are required.

Structural Understanding of Nonketotic Hyperglycinemia—A considerable level of sequence similarity exists between the human T-protein and Tm T-protein (Fig. 5A); 33% identity between the human T-protein (Leu33–Phe394) and Tm T-protein (Met1–Phe394). Thus, the homology-modeled structure of the human T-protein is highly similar to that of Tm T-protein in its core. Only the surface regions (Ser68–His71, Gly171–Ala178, Gly217–Val220, His242–Ile247, Leu292–Ala299, and Gln311–Arg315 of human) that are not directly associated with the catalytic machinery show significant structural deviations. A number of mutations in the human T-protein gene have been identified among NKH patients, including missense mutations that lead to the following amino acid substitutions: G269D, G47R, and R320H (30); D276H (31); H42R (32); E211K, C95V (4), and N145I (33); and R296H, V212A, and Y225C (34). The spatial locations of these mutations are shown in Fig. 6A. All of these residues with the exception of Arg296 do not belong to the above-listed variable surface regions. The correspondence of the human T-protein mutation sites to Tm T-protein is...
as follows: His42 (human)/His10 (Tm) and Gly47 (human)/Ala15 (Tm) are semi-conserved among both bacterial and human T-proteins (Figs. 3 and 5A). His10 (Tm), Ala15 (Tm), and Asp242 (Tm) are slightly separated from either the H$_4$folate binding pocket or the lipoic acid-binding site (Fig. 1B). Mutations at these three sites would have no direct effect on the binding of H$_4$folate or lipoic acid, but they appear to alter the H-protein binding interface. Interestingly, the residues interacting with His10 (Tm) and Asp242 (Tm) (Ala15, Pro26, Tyr29, and Asp17; Arg231, Thr309, and Ser310) are well conserved in the T-protein family (Fig. 3). Details of the interactions around His10 (Tm) and Asp242 (Tm) are shown in Fig. 5, B and C, respectively. In Tm T-protein, Asp17 and Tyr29 stabilize the imidazole ring of His10 by hydrogen bonding and the side chain of Pro26 lies close to the imidazole ring of His10. The side chain of Ala15 points toward the hydrogen-bonding network around His10 (Fig. 5B). Thus, the H4R2 and G4R7 mutations in the human T-protein could possibly disrupt the network involving these residues, thus altering the proper conformation of the N-terminal portion of domain 1, which provides the interface for interaction with H-protein as discussed above. In Tm T-protein, Arg231, Thr309, and Ser310 are clustered around Asp242, forming a hydrogen-bonding network and a salt bridge (Fig. 5C). Arg231 is part of the signature motif PXGLGARDXXRHEAXXLYG in domain 1, whereas Thr309 and Ser310 belong to the GXh/TS/TG/S/TS/GXX- SPTL motif in domain 3 (Fig. 3). Mutation of Asp242 (human)/Asp242 (Tm) into histidine could possibly perturb the hydrogen-bonding network around this residue and might alter the surface features that are crucial for H-protein binding.

Val76 (Tm), corresponding to Val212 (human), makes a hydrophobic core with neighboring residues (Leu155, Val159, Val163, and Leu166 of Tm), of which Leu155, Val159, and Val163 are conserved in human, whereas Leu166 is replaced with Ile233 in human (Fig. 5A). Thus, it is expected that the mutation of Val212 (human)/Val216 (Tm) into alanine would destabilize this hydrophobic core. Leu281 (Tm), corresponding to Arg220 (human), is conserved among bacterial T-proteins only. It is not part of the H$_4$folate- or lipoic acid-binding site but is close to the conserved Asp242 (Tm). The environment around Asp242 (Tm) is shown in Fig. 5C. In the human T-protein model, Arg220 makes an additional salt bridge with Asp276 (Fig. 6B). This interaction may be necessary for the proper function or stability of the human T-protein, thus explaining why the R320H mutation causes the impaired T-protein activity.

Acknowledgments—We thank Dr. H. S. Lee and staff at Beamline BL-6B of Pohang Light Source for assistance during data collection. We also thank Prof. N. Sakabe and staff for assistance during data collection at Photon Factory, Beamlines BL-5A and BL-18B.

REFERENCES

1. Douce, R., Bourguignon, J., Neuburger, M., and Rebeille, F. (2001) Trends Plant Sci. 6, 167–177

2. Tada, K., and Kure, S. (1993) J. Inherit. Metab. Dis. 16, 691–703

3. Sakata, Y., Oyada, Y., Satoh, K., Kojima, K., Hisanaga, K., Shinka, T., Suzuki, Y., Aoki, Y., Satoh, K., Kondo, H., Mutsatara, Y., and Kure, S. (2001) Brain Res. Mol. Brain Res. 94, 113–120

4. Toone, J. R., Applegarth, D. A., Cowlick-Mackie, M. B., and James, E. R. (2000) Mol. Genet. Metab. 70, 116–121

5. Toone, J. R., Applegarth, D. A., Cowlick-Mackie, M. B., and James, E. R. (2000) Hum. Mutat. 17, 76–76

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

7. Motokawa, Y., Fujiiwara, K., and Okamura-Ikeda, K. (1995) in Biotinoa in Health and Disease (Packer, L., and Cadenas, E., eds) pp. 389–407, Marcel Dekker, Inc., New York

8. Okamura-Ikeda, K., Ohmori, Y., Fujiiwara, K., and Motokawa, Y. (1993) Eur. J. Biochem. 216, 539–548

9. Okamura-Ikeda, K., Fujiiwara, K., and Motokawa, Y. (1999) Eur. J. Biochem. 264, 446–452

10. Okamura-Ikeda, K., Kameoka, N., Fujiiwara, K., and Motokawa, Y. (2003) J. Biol. Chem. 278, 10067–10072

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

12. Parex, S., Cohen-Addad, C., Sieker, L., Neuburger, M., and Douce, R. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 4850–4853

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

14. Faure, C., Cohen-Addad, C., Faure, S., Sieker, L., Neuburger, M., and Douce, R. (1995) Nature Struct. Biol. 2, 307–326

15. Faure, M., Bourguignon, J., Neuburger, M., Macherel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C., and Douce, R. (2000) Eur. J. Biochem. 267, 2990–2998

16. Nakai, T., Nakagawa, N., Maoka, N., Masui, R., Kuramitsu, S., and Kamiya, N. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 554–557

17. Nakai, T., Inoue, J., Neuburger, M., and Kamiya, N. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 1619–1618

18. Lokanath, N. K., Kuroishi, C., Ookaizaki, N., and Kunishima, N. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 1450–1452

19. Olewinski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326

20. Tervilinger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861

21. Tervilinger, T. C. (2000) Acta Crystallogr. Sect. D Biol. Crystallogr. 56, 38–44

22. Jonath, T. A., Zou, J. Y., and Coulston, N. S. (2003) EMBO J. 22, 4038–4048

23. Besson, V., Rebeille, F., Neuburger, M., Douce, R., and Cossins, A. E. (1993) Biochem. J. 292, 425–430

24. Rebeille, F., Neuburger, M., and Douce, R. (1994) Biochem. J. 302, 223–228

25. Nanao, K., Okamura-Ikeda, K., Motokawa, Y., Danske, D. M., Baumgartner, E. R., Takada, G., and Hayasaka, K. (1998) Hum. Genet. 95, 655–658

26. Kure, S., Shinka, T., Sakata, Y., Osamu, N., Takayanagi, M., Tada, K., Matsutara, Y., and Narisawa, K. (1998) Hum. Genet. 102, 430–434
33. Toone, J. R., Applegarth, D. A., Coulter-Mackie, M. B., and James, E. R. (2001) Mol. Genet. Metab. 72, 322–325
34. Toone, J. R., Applegarth, D. A., Levy, H. L., Coulter-Mackie, M. B., and Lee, G. (2003) Mol. Genet. Metab. 79, 272–280
35. Hutchinson, E. G., and Thornton, J. M. (1996) Protein Sci. 5, 212–220
36. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
37. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
38. Thompson, J. D., Gibson, T. J., Pilewski, F., Jeanmougin, F., and Higgins, D. G. (1997) Nucleic Acids Res. 25, 4876–4882
39. Nicholas, K. B., Nicholas, H. B., Jr., and Deerfield D. W., II (1997) Emnet. News 4, 14
40. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296