Structure of the Thiazole Biosynthetic Enzyme THI1 from Arabidopsis thaliana*

Paulo H. C. Godoi†, Rodrigo S. Galhardo§, Douglas D. Luche*, Marie-Anne Van Sluys†, Carlos F. M. Menck*, and Glaucoius Oliva‡

From the †Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, CP 369, 13560-970, Brazil, the ‡Departamento de Físico-Química, Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP, 13560-970, Brazil, the ¶Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508-900, Brazil, and the §Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, 05508-900, Brazil

Thiamin pyrophosphate is an essential coenzyme in all organisms that depend on fermentation, respiration or photosynthesis to produce ATP. It is synthesized through two independent biosynthetic routes: one for the synthesis of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate (pyrimidine moiety) and another for the synthesis of 4-methyl-5-(β-hydroxyethyl) thiazole phosphate (thiazole moiety). Herein, we will describe the three-dimensional structure of THI1 protein from Arabidopsis thaliana determined by single wavelength anomalous diffraction to 1.6 Å resolution. The protein was produced using heterologous expression in bacteria, unexpectedly bound to 2-carboxylate-4-methyl-5-(ethyl adenosine 5’-diphosphate) thiazole, a potential intermediate of the thiazole biosynthesis in Eukaryotes. THI1 has a topology similar to dinucleotide binding domains and although details concerning its function are unknown, this work provides new clues about the thiazole biosynthesis in Eukaryotes.

Thiamin pyrophosphate (TPP3 or vitamin B1) is an essential co-factor to all organisms, taking part in biochemical reactions of carbon–carbon bond formation and breaking in positions adjacent to carbonyl groups in the metabolism of carbohydrates and the biosynthesis of branched-chain amino acids (1). TPP is produced in Bacteria, Archaea, as well as Eukaryotes belonging to the Fungi and Plantae groups. The entire animal kingdom does not synthesize thiamine but obtains it through dietary uptake. In humans, its deficiency may result in Beriberi and the Wernicke-Korsakoff syndrome, severe conditions that involve damage to the cardiovascular, muscular, gastrointestinal, and nervous systems (2).

The TPP biosynthesis occurs by the condensation of two intermediate precursors, generally referred to as the thiazole (4-methyl-5-[(β-hydroxyethyl)thiazole phosphate) and pyrimidine (2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties. In Bacteria, the chemical and molecular mechanisms involved in the biosynthesis of thiamin are better understood, requiring the combined action of many gene products. Six genes (thiF, thiS, thiG, thiH, or thiO, iscS, or NifS and dks) are required for the biosynthesis of 4-methyl-5-(β-hydroxyethyl)thiazole phosphate (3). Two genes (thiC and another yet unidentified gene) are required for the biosynthesis of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate (4), and thiamin phosphate synthase (thiE) is required for the condensation reaction of both moieties to form thiamin phosphate. The remaining genes are kinases, which are responsible for the phosphorylation of precursors (thiD, thiM, and pdxK) and to convert thiamin phosphate to its active form, thiamin pyrophosphate (thiL).

In bacteria, the thiazole moiety is synthesized sequentially (Fig. 1). The precursors are glyceraldehyde 3-phosphate (G3P), pyruvate, cysteine and glycine (Bacillus subtilis), or tyrosine (Escherichia coli). G3P and pyruvate are converted to deoxy-D-xylulose 5-phosphate (DXP) in a reaction catalyzed by the enzyme DXP synthase (dks). DXP contributes to the 4-methyl-5-hydroxyethyl group and the C-4 atom of thiazole. The sulfur atom originates from cysteine, but, in fact, the sulfur incorporation to the thiazole ring depends on the combined action of ThiF, ATP, IscS (a cysteine desulfurase), pyridoxal 5’-phosphate, cysteine, and ThiS resulting in the formation of an intermediate carrier protein, ThiS-thiocarboxylase (5). Thiazole synthase (ThiG), a key enzyme for this pathway, catalyzes the final thiazole ring formation though an imine intermediate with DXP, interacting with ThiS-thiocarboxylate to assist the sulfur transfer to the DXP carbon-3 (6) and most likely interacting with ThiH (E. coli) or ThiO (B. subtilis) to incorporate a nitrogen atom to the thiazole ring.

In Saccharomyces cerevisiae, as well as many other Eukaryotes, only one gene has been identified for the thiazole moiety biosynthesis (thi4 (7)) and one for the pyrimidine moiety biosynthesis (thi5 (8)). The five carbon precursor to thiazole...
Structure of the Thiazole Biosynthetic Enzyme THI1

FIGURE 1. Thiazole biosynthesis in bacteria. The central circle emphasizes that the incorporation of sulfur is intermediated by the protein ThiS-thiocarboxylase, while the incorporation of carbon-2 and nitrogen-3 atoms occurs in the active site of thiazole synthase through an imine intermediate with DXP and the residue Lys96 (THI1-Lys96).

originates from 2-pentulose (9, 10). Glycine (11) and cysteine (12) are also considered precursors.

THI4 is a conserved protein, with orthologues found in plants as well as in Archaea. In Arabidopsis thaliana, the ortholog to this protein was named THI1, since it was the first enzyme of the pathway to be identified in plants. Despite this, thi1 was initially identified by its ability to complement bacterial mutant strains defective in DNA damage repair mechanisms (13), while further experiments confirmed its ability to complement the thiamin auxotrophy in S. cerevisiae thi4 null mutant strain. Using the same model, it was noticed that THI1 complementation does not change the cell survival rate under DNA-damaging conditions, but it is able to restore its mitochondrial DNA stability (14). Therefore, thi1 and thi4 are credited for a dual role in the thiazole biosynthesis and mitochondrial DNA damage tolerance.

Herein, we describe the crystal structure of THI1 after heterologous expression in E. coli. The structure reveals that this protein assembles as an octamer and combines with a previously undescribed ligand, modeled to 2-carboxylate-4-methyl-5-β-(ethyl adenosine 5′-diphosphate) thiazole (AHZ), most likely an intermediate of the thiazole biosynthesis in Eukaryotes. Site-directed mutagenesis experiments and in vitro functional assays were also performed and support some features disclosed by the THI1 structure, including details of the dinucleotide binding domain and octamer formation for the thiazole biosynthesis.

EXPERIMENTAL PROCEDURES

Protein Preparation—The thi1 gene was subcloned in pET-28a (Novagen, Darmstadt, Germany) between BamHI and EcoRI sites. The resulting plasmid was used for the transformation of BL21(DE3), after selection with 30 μg/ml kanamycin. Cells were cultured in LB medium containing 30 μg/ml kanamycin at 37 °C until A600 nm of about 0.8. Promoter induction occurred by adding 0.4 mM isopropyl β-D-thiogalactopyranoside at 37 °C for 3 h. The cells were harvested at 4,000 × g for 30 min at 4 °C and stored at −20 °C.

Cells from 11 cultures were resuspended in 40 ml of buffer 50 mM NaH2PO4, 300 mM NaCl, 20 mM imidazole, pH 8.0, with 1 mM phenylmethylsulfonyl fluoride, 1 mM pepstatin-A, and 1 mM leupeptin (buffer A). To promote the cell lysis, 1 mg/ml lysozyme was added, and this mixture was kept on ice for 30 min. The cell lysate was sonicated and centrifuged at 10,000 × g for 30 min at 4 °C. The supernatant was applied into a 5-ml nickel-nitrilotriacetic acid chelate column (Qiagen, Hilden, Germany), previously equilibrated in buffer A at a flow rate of 1 ml/min. The contaminating proteins were removed in a linear gradient of 50 ml of buffer A containing 20–100 mM imidazole, while the elution was carried out in a linear gradient of 50 ml of buffer A containing 100–250 mM imidazole.

The amino acid sequence added to the N terminus, including the His6-tag, was found to be inhibitory to the crystallization process (data not shown). Therefore, to prepare samples for crystallization, samples purified in nickel-nitrilotriacetic acid were dialyzed against 100 mM Tris-HCl, pH 7.2 (buffer B), further concentrated to 10 mg/ml using Centricon 10 (Millipore, Billerica, MA) and papain digested in a reaction containing 1 ml of THI1 concentrated to 10 mg/ml, 10 μl of 500 mM EDTA, pH 8.0, 10 μl of 300 μM 2-mercaptoethanol, and 5 μl of 10 mg/ml papain (Roche Applied Science, Basel, Switzerland) for 30 min at 37 °C. The limited proteolysis reaction resulted in the elimination of the first 31 amino acids, as confirmed by Edman sequencing (data not shown), including the histidine tag and other amino acids added by the plasmid construct. The THI1-digested protein was further purified by size exclusion chromatography using a Superdex 200 XK16 column (GE Healthcare, Chalfont St. Giles, UK) previously equilibrated in buffer B at a flow rate of 0.5 ml/min. Samples were dialyzed against 20 mM TES, 1 mM MgCl2, pH 7.4, and concentrated to 10 mg/ml using Centricon 10 prior to the crystallization experiments.

Crystallization and Data Collection—THI1 crystals were obtained by hanging-drop vapor diffusion against a well solution of 100 mM MES, pH 6.0, 40% (v/v) 2-methyl-2,4-pentanediol, and 15% (v/v) heptane-1,2,3-triol. Crystals were seen after 2 weeks at 291 K as small plates of 0.1 × 0.1 × 0.01 mm.

A two-wavelength data collection was measured at beamline ID29 of the European Synchrotron Radiation Facility in Grenoble, France, using a Rontec Silicon Drift Diode fluorescence detector for the absorption edge scan and an ADSC.
Quantum 210 CCD area detector. One crystal diffracting to 2.35 Å was used to collect data for phasing by single-wavelength anomalous dispersion (SAD) at the zinc peak wavelength (λ = 1.28269 Å), revealed after a broad fluorescence scan. A second crystal was measured to further extend the resolution to 1.6 Å. Both crystals were perfectly isomorphous, collected in a cold nitrogen stream at 100 K.

Crystals belong to space group F222 with unit cell dimensions of a = 102.356 Å, b = 133.147 Å, and c = 142.301 Å. Two molecules are found in the asymmetric unit, as seen by the self-rotation function calculated by GLRF (15) revealing a two-order non-crystallographic symmetry axis in φ = 0°, ψ = 45°, and κ = 180° (data not shown). The Matthews coefficient (16) was calculated to 1.9 Å³/Da with a solvent content of 34.5%.

The data set used in SAD phasing was indexed and integrated by DENZO (17) and further reduced by SOLVE (18). The native SAD phasing using the zinc peak wavelength. Two zinc ions were added by CNS at peaks above 4 σ. Ligand molecule AHZ was clearly found early during the refinement as implemented in CNS (22). The results of refinement are summarized in Table 1. In our final model, 92.8% of residues are found in the most favored region and the remaining 7.2% are in the additional allowed region of the Ramachandran plot. The figures were prepared with PyMol, Topdraw (25), and Ligplot (26).

S. cerevisiae Strains and Media—The S. cerevisiae strains used in this work were W303 (mata, ade2-1, trp1-1, leu2-3-112, can1-100, ura3-1, his3-11-115) and KBY5 (as W303, THI4: URA3 (7)). Selection of transformants with pG1 constructs was performed in yeast nitrogen base (BD Biosciences) lacking tryptophan. For the thiamin auxotrophy experiment, cells were grown in YPD (1% yeast extract, 2% peptone, and 2% dextrose) until saturation, centrifuged, and resuspended in 3 volumes of sterile water. One aliquot was spread in minimal medium plates with the required supplements, including or not thiamin, and incubated for 3 days at 30 °C.

The determination of respiratory mutant induction was performed with cells grown in Difco minimal medium containing thiamin. Analysis of MMS-induced respiratory mutants was performed in cells grown in minimal medium to 10⁷ cells/ml, centrifuged, and resuspended in sterile water and either treated with 0.2% MMS or mock-treated for 1 h at 30 °C under shaking. After dilution in sterile water, cells were plated in YPD and incubated at 30 °C for 2–3 days. After colony formation, a 3-ml inoculum was poured onto the plates, which were incubated at room temperature for 1 h. Respiratory mutants (full white or sectored colonies (27)) were scored, and the percentage relative to the total number of colonies was calculated. The induction of respiratory mutants was also measured after cell growth at 37 °C. In these experiments, the yeast cells were cultivated in minimal medium until the stationary phase and then inoculated in minimal medium with 1/100 dilutions. Cells were then grown to A600nm = 1.0 both at 30 and 37 °C, diluted in sterile water, and plated on YPD. After colony formation, respiratory mutants were detected as described above.

Site-directed Mutagenesis of thi1—We performed PCR-mediated site-directed mutagenesis of the thi1 cDNA using primers with the desired mutations. The introduced mutations led to single amino acid substitutions in the encoded protein. Two complementary mutagenic primers were used with external primers to amplify fragments of thi1 containing the desired mutations. Both fragments were used in a further reaction with only the external primers to generate full-length cDNAs containing the introduced mutation.

All PCR reactions were carried out with the high fidelity PfX enzyme (Invitrogen), using the cloned thi1 cDNA as a template.
Structure of the Thiazole Biosynthetic Enzyme THI1

After production of full-length mutant cDNAs, they were subcloned in the pGEM-T easy vector (Promega, Madison, WI) and sequenced to ensure that the correct mutation was introduced and that no additional mutations were generated during amplification. The selected cDNAs were transferred to the pG1 yeast expression vector (28) into BamHI and SalI sites. The expression vector (Ala140Val in this work).

FIGURE 2. Monomeric structure of THI1. A, ribbon drawing of THI1 in complex with 2-carboxylate-4-methyl-5-[β-ethyl (adenosine 5’-diphosphate)] thiazole (blue sticks), Zn²⁺ (gray sphere), and two water molecules (red spheres). B, topology diagram of THI1 depicting the two halves variation of mononucleotide-binding motifs, colored in magenta and blue. The first α-helix is in yellow, B-subdomain is in green, and the β-meander is colored red.

RESULTS

Overall Monomeric Structure of THI1—The monomeric THI1 protein presents a long α-helix (residues 18–38) followed by a globular αβ domain with a three-layer (ααβ) sandwich architecture and a topology similar to FAD/NAD(P) binding domain (residues 39 to 284, Fig. 2). The globular domain is made up of a central six-stranded β-sheet with β5-β2-β1-β9-β11-β10 topology, β10 being the only antiparallel strand. The central β-sheet is flanked on one side by a twisted antiparallel three-stranded β-sheet formed by β6-β7-β8 and helix α5 and by a four α-helix bundle containing α4-α2-α9-α7 on the other side. Helix α6 is connecting β9 to α7, and α8 is connecting β11 to α9. This fold resembles a FAD binding domain with the classical two-halves β₁-α₁-β₂-α₂-β₃ topology, containing a variation of the Rossmann fold and a substitution to the αC-helix crossover usually found in the classical dinucleotide binding motif and the glutathione reductase structural family (30). In the first half, THI1 possesses an additional variation due to the insertion of the sheet β₃-β₄ and α₃ (B-subdomain) connecting β2 to α4. The connection between the halves, formed by the twisted sheet β₆-β₇-β₈, is homologous to the β-meander found in the glutathione reductase family but also includes the helix α5 and the loop between α5 and β₈ (L-loop) that, in turn, make contact with a symmetry-related ligand binding site. The second half topology is less similar to the Rossmann fold. It presents an inversion of strands (β₅-β₇-β₆) with a cross-connection between α₁-β₁ and β₂-α₂, resulting in a β₁-α₁-β₂-β₃-α₂ topology. The helices α₁, α₂, in turn, are split to become α₆-α₇ and α₈-α₉, respectively.

A three-dimensional structure search over DALI (31), CE (32), and CATH (33) data bases suggest that THI1 is homologous to many dinucleotide-binding proteins with a higher similarity to FAD-binding proteins such as flavocytochrome-c₃, fumarate reductase, thioredoxin reductase, and glycine oxidase (see supplemental Table 1). The highest scoring flavocytochrome-c₃ from Shewanella frigidimarina has 40 identical residues in 182 matching residues with a root mean square deviation of 1.48 Å, as calculated by LSQMAN (34). In addition, a significant similarity is found to the glycine oxidase from B. subtilis, an enzyme responsible for the incorporation of the thiazole ring nitrogen-3 atom.

The glycine oxidase from B. subtilis (ThiO (35)) catalyzes the FAD-dependent oxidation of glycine to imine. THI1 and ThiO possess similar FAD binding domains. Based on the THI1 and
ThiO structural alignment, 20 identical residues are found out of a total 111 matching residues with a root mean square deviation of 1.34 Å, involving residues in the central six-stranded β-sheet, the β-meander and four of the six-helix bundle in ThiO (helices 7, 1 and 8 and 11; see the supplemental data). The mechanism of binding adenine, ribose, and the α-phosphate oxygen atoms are similar, but the β-phosphate oxygen atoms establish additional interactions with the THI1 helix-8, absent in ThiO. The most probable active site region, where the group 2-carboxylate-4-methyl thiazole is found, is distinct from ThiO. It is also clear that THI1 cannot bind FAD due to a limited space to accommodate its isoalloxazine extremity, while recent evidence has shown that the yeast orthologue Thi4 protein possesses a NAD-binding site (36), a result that we also have recently confirmed for THI1.4

Ligand Binding Site—The structure of THI1 reveals the presence of a ligand not previously described, observed in the region where a bound dinucleotide was expected. This ligand has a clear electron density that was modeled to 2-carboxylate-4-methyl-5-(β-hydroxyethyl)thiazole covalently bound to adenosine diphosphate by its 5-(β-hydroxyethyl) group (AHZ, Fig. 3).

The AHZ molecule adopts an elongated conformation with the adenine moiety in a reverse chain direction motif (37). As expected for the members of general dinucleotide binding proteins, THI1 contains the characteristic GxGxxG signature (Gly46-Ala47-Gly48-Ser49-Ala50-Gly51) in the A1-loop between the first β-strand (β1) and the next α-helix (α2), as well as a glutamic acid at the end of the second β-strand (β2, Glu70), bound to the ribose diol group through a bidentate hydrogen bond. Additionally, Gln71 establishes a second hydrogen bond to the diol group (O2*). On the adenine end, the N-1 and N-6 atoms form hydrogen bonds with the backbone NH and the CO group of Ala143, respectively. Hydrophobic interactions are also found between the adenine delocalized π-electrons and the residues Phe193, atoms Cβ and Cγ of Gln71 (A face), Val45, and Cys187 (B face). Thus, THI1 follows the pattern-2 interaction (38) with the adenine AHZ.

4 D. D. Luche, P. H. C. Godoi, S. M. Chabregas, M. C. Silva-Filho, G. Oliva, and C. F. M. Menck, manuscript in preparation.
The diphosphate binding site is formed through the interactions across the loops β1-a2 (A1-loop), β2-β3 (A2-loop), β9-a6 (zinc-loop), and helices α2 and α8. The α-phosphate oxygen atoms are mostly attached by indirect interactions through three water molecules and the amine NH atom of Gly, while the β-phosphate oxygen atoms orient themselves in-line with the dipolar momentum of helices α2 and α8, interacting with two amine nitrogen atom of Ala (α2) and Met (α8) and two other water molecules.

The 2-carboxylate-4-methyl-5-β-hydroxyethylthiazole moiety of AHZ is oriented to accommodate its sulfur atom closer to a hydrophobic pocket containing three methionine residues, Met (4.3 Å), Met (3.8 Å), and Met (5.15 Å). The carboxylate group at C-2 interacts via hydrogen bonds to Arg (2.51 NE1 (3.15 Å), NH (2.71 Å), and to a zinc ion (2.07 Å). This zinc ion is strongly attached to the loop between α8-α9 (zinc-loop), closer to the surface, through bonds with the N2 thiazole atom (2.10 Å), His (Ne2 (2.09 Å), Asp (O81 (2.05 Å, monodentate) and two water molecules completing the coordination sphere in a distorted octahedral geometry.

Quaternary Structure—The THI1 protein is found as an octamer in solution as observed by size exclusion chromatography (data not shown) and by dynamic light scattering, with a hydrodynamic radius of 6.24 nm and a calculated molecular mass of 244 kDa. The crystal reveals an asymmetric unit containing two monomers related by a 2-fold non-crystallographic symmetry axis, where a 222-point symmetry generates the octamer.

The octamer is tightly packed as a two-layer ring torus structure with approximate dimensions of 98 × 70 × 23 Å, enclosing a 23-Å pore diameter (Fig. 4). Each ring torus contains four molecules formed by the globular domains, while the N-terminal loop (residues 7–18) and the α1 helix are found distal to the geometric center of the torus, interconnecting both rings. Each monomer contacts another four molecules, two neighbor molecules on its own ring and another two molecules on the next ring torus layer. Such interface involves a total buried area of 6.439 Å² per monomer.

Interactions across the same ring are almost exclusively established by residues distributed on the β-mean-der but also include residues from the N-terminal loop and the α1-helix from the neighbor layer, while interactions across rings involve a larger contact area that represents 53.5% over the total buried area.

Site-directed Mutagenesis of thi1—The alignment of THI1 against 6 Plantae, 12 Fungi, 1 Eubacteria, and 16 Archaea orthologues reveals a series of highly conserved residues that may play an essential role in the enzymatic activity of THI1 (Fig. 5 and supplemental data). These residues are mostly clustered around the ligand binding site (Fig. 6). While some interact directly with AHZ, others are seen in the interface formed between β6, the L-loop and the closest symmetry-related molecule, next to the thiazole ring (see supplemental data).

Seven highly conserved amino acids among THI1 orthologues were targeted by site-directed mutagenesis. To assay the biological activities THI1, heterologous complementation tests were performed with the S. cerevisiae strain carrying a disruption in the thi1 orthologue, thik. The Ala → Val mutation, isolated from an A. thaliana thiamin auxotrophic mutant (29) was also tested. Mutants Gly → Val, Ala → Val, His → Phe, and Glu → Gly resulted in disruption of the thiamin biosynthetic route, as evidenced by the lack of complementation of yeast thiamin auxotrophy but Asp → Gly, Asp → Gly, Trp → Leu, and Lys → Met mutations had no effect upon this activity.

It was previously shown that the yeast thi4 strain exhibits an increased frequency of respiratory mutants when exposed to UV radiation or MMS treatment (14). Therefore, the same
mutant cDNAs constructed were tested for their ability to restore wild-type levels of respiratory mutants after MMS treatment in the \( \text{thi}4 \) yeast strain. Results show that only the His\(^{167} \rightarrow \text{Phe} \) mutation affected the mitochondrial stability mediated by \( \text{THI1} \), in contrast to the other mutant cDNAs, which can complement the yeast strain as efficiently as the wild-type allele. Taken together, these results indicate that the His\(^{167} \rightarrow \text{Phe} \) mutation suppresses both activities of the protein, while Ala\(^{140} \rightarrow \text{Val}, \text{Gly}\(^{77} \rightarrow \text{Val} \), and Glu\(^{222} \rightarrow \text{Gly} \) mutations selectively affected the thiamine biosynthetic activity of \( \text{THI1} \).

**DISCUSSION**

**THI1 Structure in Complex with a Putative Thiazole Biosynthesis Intermediate**—The molecular events leading to the biosynthesis of thiazole and pyrimidine moieties of thiamine remain to be understood in Eukaryotes. As mentioned earlier, \( \text{THI1} \) homologues are the only known enzymes to take part in the thiazole
biosynthesis in a diversity of unicellular organisms belonging to Archaea and some Fungi, as well as highly organized multicellular organisms, such as Plantae. The absence of bacterial homologues suggests that the thiazole biosyntheses are dissimilar and evolved independently in Archaea and Bacteria. Thus, it is possible that Archaea, Fungi, and Plantae share a common ancestor. The only known exception is the Eubacteria T. maritima, where a lateral gene transfer from Archaea was suggested (39).

The structure of THI1 was solved by single-wavelength anomalous dispersion using zinc ions fortuitously bound to THI1. Its overall folding is similar to dinucleotide binding domains with a higher similarity to FAD binding enzymes. However, recent evidence has shown that THI1 binds to NAD(H) but not ATP (36), while FAD binding is unknown. Although we have seen that THI1 shares some structural similarity with ThiO, a FAD-dependent enzyme required for the thiazole biosynthesis in B. subtilis, the significant alignment is restricted to the conserved dinucleotide binding architecture, specifically, elements responsible for the recognition and binding to the adenosine group. It is evident that the remaining elements, responsible for the recognition and binding of nicotinamide ribose in NAD(H) or riboflavin in FAD, are unique. We also observed that FAD binding is unlikely in an octameric THI1 due to a limited space to accommodate its isoalloxazine group. NAD binding instead of ATP also suggests that a molecular mechanism for the recognition of nicotinamide ribose is present in THI1.

During the initial refinement process of THI1, a clear electron density was noticed and modeled to AHZ. It is not clear how the thiazole moiety was formed in our experiments but it is certain that the most fundamental precursors, such as glyceraldehyde-3-phosphate, pyruvate, glycine, tyrosine, and cysteine were available in the growth medium, as well as all the enzymes required for the thiazole biosynthesis in E. coli. Despite this, the hypothesis that AHZ could only have been captured by the THI1 octamer is unlikely since this compound has never been described in Bacteria. Thus, AHZ is most likely a result of the THI1 biological activity.

Taken together, these data suggest that the adenosine group in AHZ must have been produced from a NAD molecule acting as a substrate to donate AMP. Similar reactions, where NAD is used as a donor group, can be found in bacterial NAD-dependent DNA ligases, enzymes that utilize NAD to form a phosphamide-linked AMP, with the amino group of an active lysine side chain (40). Additionally, we observed that AHZ is found deeply buried inside the octamer, implying that NAD binding must occur in a lesser complex structure before the octamer assembly, while a dynamic oligomeric state may be relevant for its biological function. If this is the case, it is also possible that a binding partner be required in assisting its disassembly. Such hypothesis would explain the remarkable stability of octameric THI1 in complex with AHZ, crystallized after its expression in E. coli, limited digestion with papain, dialysis and purification (see “Experimental Procedures”). In support of this, it has been shown that cyclophilin NcCyP41 is required for the proper function of CyPPB37, a THI1 orthologue from Neurospora crassa (41). Cyclophilins are enzymes that interconvert cis/trans isomers, accelerating slow folding processes and conformational changes in proteins; however, an orthologue of NcCyP41 is to be found in A. thaliana.

Another interesting feature of AHZ is the presence of a carboxylate group linked at the carbon-2 atom of the thiazole ring, in agreement with a previous work showing that a carboxylated thiazole moiety is an intermediate for the thiazole biosynthesis (11). Despite this, we do not have experimental evidence that the thiazole ring was entirely synthesized by THI1, although it was suggested that THI1 homologues are functionally equivalent to ThiH (or ThiO) and ThiG, the last two enzymes responsible for the incorporation of carbon-2, nitrogen-3, and for the thiazole ring closure (42 and Fig. 1). The region where the 2-carboxylate thiazole end of AHZ is located most likely encompasses the THI1 active site, together with the zinc ion and the L-loop of a symmetry-related molecule that belongs to the same ring torus.

THI1 Mutants Impair the Thiazole Biosynthesis—To gain further data on the THI1 biological activity, we have designed and shown that the mutations Gly77 → Val, His167 → Phe, and Glu222 → Gly have an effect on thiazole biosynthesis, as well as the previously described mutation Ala140 → Val (29).

Gly77 is an invariant residue in all known thi1 orthologues. It is located on the A2-loop and although this residue does not participate in any specific interaction with the ligand but allows AHZ to position itself in close proximity to THI1, assuming a conformation (Ψ = 51.9 and θ = −132.7 degrees) that is energetically unfavorable to any other amino acid. A change to valine will force a local structural change that will likely prevent AHZ precursors to bind THI1. In addition, as seen in an
octameric state, the increase of volume may also disallow the octamer formation through a steric hindrance. In fact, any mutation that changes any one of the glycines in this loop (GGG signature) to any other amino acid will result in a change to the A2-loop conformation followed by disruption of precursors of AHZ interaction.

Ala\textsuperscript{140} is located on the surface, in a short loop formed between \(\beta_5\) and \(\beta_6\). Although it is not located in proximity to the most probable active site, this residue resides in a highly packed interface between four molecules. It is not clear if the volume addition caused by the mutation to valine will interfere with the octamer formation but this change will certainly require a local rearrangement that may displace the L-loop and most likely the conformation of Glu\textsuperscript{70}, responsible for the AHZ ribose diol group interaction to THI1. Therefore, the precursors binding and/or AHZ formation may be impaired.

Glu\textsuperscript{222} interacts simultaneously with Ala\textsuperscript{258} (NH) and Arg\textsuperscript{92} (N\(\gamma\)), anchoring the \(\alpha_7\) to \(\alpha_9\) and, on the B-subdomain, \(\beta_3\). This region holds many interactions between symmetrical molecules and therefore, the mutation Glu\textsuperscript{222} \(\rightarrow\) Gly might cause a local dynamic disorder which could prevent or reduce the affinity for the octamer assembly.

Finally, His\textsuperscript{167} \(\rightarrow\) Phe is a mutation that disrupts both the thiazole biosynthesis and mitochondrial DNA stability mediated by \(hii1\). In this case, however, we cannot be certain that THI1 will be present or properly folded. All other mutants conserve at least one biological activity that confirms its gene expression and likelihood that the protein is properly folded.

A Function to Ascertain—Another important question relates to the hypothesis of multiple cellular functions for THI1, as suggested before (14, 43). The fact that some of the mutations prevent the thiazole biosynthetic activity of THI1 but not its mediated mitochondrial DNA stability support this suggestion and indicate that both functions can be dissociated.

In summary, the crystal structure of THI1 revealed an unexpected binding partner that we suggest to be a putative inter-mEDIATE of the thiazole biosynthesis in Eukaryotes. Even though 2-carboxylate thiazole was expected to be an intermediate to the thiazole biosynthesis, our data show that the molecular mechanism leading to its formation is dissimilar from that described in Bacteria. Our data also suggests that NAD is involved in the formation of AHZ, acting as a substrate for the reaction rather than solely as a cofactor and that this activity should require a dynamic mechanism of oligomerization, most likely assisted by a cyclophilin. Site-directed mutations targeted on the dinucleotide binding site, the putative active site, or on the octamer interfaces resulted in loss of function, as observed by functional complementation in yeast.

Acknowledgments—We thank Dr. William Shephard for his support during the data collection (European Synchrotron Radiation Facility ID29 Beamline, Grenoble, France). Also, we thank Katelyn and Harold Thureson for grammatical revisions.

Addendum—The identification and characterization of AHZ has been recently published (44) while this article was under revision.
Structure of the Thiazole Biosynthetic Enzyme THI1

36. Faou, P., and Tropschug, M. (2004) J. Mol. Biol. 344, 1147–1157
37. Denessiouk, K. A., Rantanen, V.-V., and Johnson, M. S. (2001) Proteins 44, 282–291
38. Carugo, O., and Argos, P. (1997) Proteins 28, 10–28
39. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999) Nature 399, 323–329
40. Wilkinson, A., Day, J., and Bowater, R. (2001) Mol. Microbiol. 40, 1241–1248
41. Faou, P., and Tropschug, M. (2003) J. Mol. Biol. 333, 831–844
42. Morett, E., Korbel, J. O., Rajan, E., Saab-Rincon, G., Olvera, L., Olvera, M., Schmidt, S., Snel, B., and Bork, P. (2003) Nat. Biotechnol. 21, 790–795
43. Medina-Silva, R., Barros, M. P., Galhardo, R. S., Netto, L. E., Colepicolo, P., and Menck, C. F. (2006) Res. Microbiol. 157, 275–281
44. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) J. Am. Chem. Soc. 128, 7158–7159