Published in final edited form as:

J Mol Biol. 2004 October 15; 343(2): 395–406. doi:10.1016/j.jmb.2004.08.037.

The Structure and Ligand Binding Properties of the *B. subtilis* YkoF Gene Product, a Member of a Novel Family of Thiamin/HMP-binding Proteins

Yancho Devedjiev¹, Yogesh Surendranath¹, Urszula Derewenda¹, Alexandra Gabrys¹, David R. Cooper¹, Rong-guang Zhang², Lour Lezondra², Andrzej Joachimiak², and Zygmunt S. Derewenda^{1,*}

R. Huber

¹Department of Molecular Physiology and Biological Physics, University of Virginia Charlottesville, VA 22908-0736 USA

²Biosciences Division and Structural Biology Center Argonne National Laboratory 9700 South Cass Avenue Building 202, Argonne, IL 60439 USA

Abstract

The crystal structure of the *Bacillus subtilis* YkoF gene product, a protein involved in the hydroxymethyl pyrimidine (HMP) salvage pathway, was solved by the multiwavelength anomalous dispersion (MAD) method and refined with data extending to 1.65 Å resolution. The atomic model of the protein shows a homodimeric association of two polypeptide chains, each containing an internal repeat of a ferredoxin-like $\beta\alpha\beta\alpha\beta$ fold, as seen in the ACT and RAM-domains. Each repeat shows a remarkable similarity to two members of the COG0011 domain family, the MTH1187 and YBL001c proteins, the crystal structures of which were recently solved by the Northeast Structural Genomics Consortium. Two YkoF monomers form a tightly associated dimer, in which the amino acid residues forming the interface are conserved among family members. A putative small-ligand binding site was located within each repeat in a position analogous to the serine-binding site of the ACT-domain of the *Escherichia coli* phosphoglycerate dehydrogenase. Genetic data suggested that this could be a thiamin or HMP-binding site. Calorimetric data confirmed that YkoF binds two thiamin molecules with varying affinities and a thiamine–YkoF complex was obtained by co-crystallization. The atomic model of the complex was refined using data to 2.3 Å resolution and revealed a unique H-bonding pattern that constitutes the molecular basis of specificity for the HMP moiety of thiamin.

Keywords

protein structure; macromolecular crystallography; surface engineering; thiamin/HMP binding; ACT/RAM domain family

Introduction

Vitamin B1 (thiamin pyrophosphate) is a cofactor of a number of important enzymes in carbohydrate metabolism. It can be synthesized by micro-organisms, fungi and plants, but not

vertebrates.¹ Comparative genetics of the thiamin biosynthesis pathway in prokaryotes has been investigated recently and a number of new genes were identified.² Aside from genes coding for the enzymes directly involved in biosynthesis, clusters of genes involved in the active transport of vitamin B1 were also identified. In a number of Gram-positive bacteria, including *Bacillus subtilis*, a unique thiamin-related ABC transporter consists of four genes: *YkoF*, *YkoE*, *YkoD*, and *YkoC*. The *YkoE* and *YkoC* gene products are two transmembrane components while the *YkoD* protein is an ATPase component.² No substrate-binding component for this system has been identified, and the specific function of the *YkoF* gene remains unknown.

The *YkoF* protein was selected as one of the targets of the Midwest Center for Structural Genomics[†]. However, it initially failed to crystallize in the high-throughput pipeline. We have therefore applied the new method of surface conformational entropy reduction,³ and easily obtained high quality crystals from a double mutant, K33A/K34A. The crystal structure was solved by the multiwavelength anomalous dispersion (MAD) method using the selenomethionine (SeMet) substituted protein, and refined at 1.65 Å resolution. Unexpectedly, the structure revealed an internal repeat of an ACT domain-like fold, which is known to bind small ligands. Based on genetic data, we hypothesized that *YkoF* could be a hydroxymethyl pyrimidine (HMP), or a thiamin-binding protein. Co-crystals with thiamin were obtained and the structure was solved and refined at 2.3 Å resolution. The structure confirmed our hypothesis, showing thiamin bound to each of the internal repeats, albeit with slightly different stereochemistries. Calorimetric studies showed that the two sites have markedly different affinities. Here we describe the structure of the *YkoF* protein and the mechanism by which it selectively binds the HMP moiety of thiamin. We also show how the surface mutagenesis strategy formed an intermolecular Ca²⁺ binding site, essential for the formation of the crystal lattice, thereby generating high-quality crystals.

Results and Discussion

Crystallization

We reassessed the propensity of the wild-type *YkoF* to crystallize using the commercially available Hampton I & II (Hampton Research), Wizard I & II (Emerald Biostructures), and Sigma (Sigma-Aldrich) sparse matrix screens, with a range of protein concentrations. All these attempts failed to yield crystals. Based on the surface entropy reduction concept,³ two double mutants were designed to generate surface patches potentially suitable for formation of crystal contacts. Both mutants, K112A/E114A and K33A/K34A, were used in all five screens, as described above for the wild-type protein. The K112A/E114A mutant remained recalcitrant to crystallization, but the K33A/K34A mutant readily crystallized in several conditions. Optimization revealed a pronounced preference for divalent ions in the crystallization conditions, with Ca²⁺ yielding the best crystals. When this work was completed, it was discovered that the second tier of crystallization screens at MCSG yielded crystals of wild-type, SeMet-labeled *YkoF*, which were used for independent structure determination.

Quality of the atomic models

The final crystallographic model of the K33A/K34A mutant was refined at 1.65 Å resolution to an *R* factor of 17.2% and *R*_{free} of 22.4%. It consists of two independent monomers of *YkoF* denoted A and B and containing a total of 366 residues, 390 water molecules, two acetate ions and one Ca²⁺. In the A molecule, eight N-terminal amino acid residues, three C-terminal residues, as well as three more residues located in the long linker connecting two internal repeats within the molecule, are not seen in the electron density map. The corresponding

[†]<http://www.mcsg.anl.gov>

numbers for the B molecule are eight, four and eight. The quality of the model was assessed by PROCHECK⁴ and shows that 88.3% of the residues are in the most favored Ramachandran regions with no residues in the disallowed region. Crystallographic details for the mutant structure, as well as that of the complex with thiamin, are shown in Table 1.

YkoF monomer contains a tandem of ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ motifs

Each of the two independent YkoF monomers in the asymmetric unit folds into an eight-stranded, antiparallel β -sheet, with the strands arranged in the order 23148576. The four connecting α -helices are stacked against one face of the β -sheet, leaving the other exposed. The monomer has an internal approximate 2-fold symmetry axis, reflecting an internal tandem repeat of a $\beta\alpha\beta\beta\alpha\beta$ tertiary motif. The repeat is not readily identifiable at the amino acid sequence level. The two ferredoxin-like motifs form a side-to-side contiguous β -sheet *via* an antiparallel interaction between β -strands 4 and 8. Thus, the topology of the β -sheet may be described as 2314/4'1'3'2' to reflect the pseudosymmetry of the tandem repeat (Figure 1(A)).

The first motif, consisting of residues 9 to 85, can be superposed onto the second motif consisting of residues 116 to 190 with a root-mean-square difference between the 74 C $^{\alpha}$ pairs of 1.7 Å (Figure 1(B)). The most significant differences are associated with the conformation of the C termini of the structurally homologous helices A and A'. Helix A is shorter, consisting of only three turns, while helix A' in the second motif has four turns followed by a short linker to the subsequent β -strand. The two motifs are connected by a 30 residues long linker, which begins at the C terminus of β -strand 4, and folds over helix B to connect to the first β -strand of the second repeat. This long loop has no secondary structure elements and shows significant flexibility, exemplified by lack of interpretable electron density for some of the residues in its center, and significantly higher isotropic displacement parameters (*B* factors) than those observed for the rest of the structure.

Each of the two $\beta\alpha\beta\beta\alpha\beta$ repeats contains its own well-defined, hydrophobic core. In the N-terminal motif the residues that make up the core include Phe15, Leu32, Val40, Leu51, Ile65 and Met78, while in the second repeat the core includes Phe119, Cys117, Leu156, Leu167 and Val170. The two cores are connected by a constellation of aromatic residues which lie at the interface of the two repeats, i.e. Phe13, Phe59, Tyr66, Phe82, Phe164, and Phe171 (Figure 2).

Dimer architecture

The two molecules in the asymmetric unit form a head-to-tail homodimer with an extensive face-to-face interface between the β -sheets. Each monomer buries 2504 Å² out of a total of 10,305 Å² of solvent accessible surface. Thus, the assembled homodimer has an overall spherical appearance, with each monomer contributing one hemisphere (Figure 3). A tight interface suggests a high association constant for the dimer, and gel filtration experiments (data not shown) confirmed that the protein exists as a dimer in solution.

Van der Waals interactions between hydrophobic side-chains, direct H-bonds and water-mediated H-bonds contribute to the stability of the homodimer. Most notable are the stacking interactions between symmetry-related residues Tyr18^A and Tyr18^B as well as between Tyr122^A and Tyr122^B. Close interactions are also seen between Met20^A and Phe24^B; Phe24^A and Met20^B; Val77^A and Tyr18^B; Tyr18^A and Val77^B. A water-filled channel, 40 Å long, runs through the center of the dimer.

The HMP/thiamin binding sites

A close inspection of the YkoF crystal structure revealed in each molecule the presence of two symmetrically located putative small molecule binding sites. They are found in analogous locations in the two ferredoxin-like motifs: each is at the end of the central β -sheet in between

the hairpin formed by β -strands 2 and 3 and helix A (Figure 1(A)). We found this very suggestive, particularly because the location of the putative binding site within each repeat coincides with the serine-binding site found in the structurally similar ACT-domain of 3PGDH.⁵ In the 1.65 Å resolution structure of the double mutant, the site in the N-terminal repeat is occupied by a clearly resolved acetate ion and five water molecules, while the one in the C-terminal repeat contains residual electron density which cannot be unequivocally interpreted.

Given the genetic data implicating the *YkoF* gene in HMP transport for thiamin biosynthesis in *B. subtilis* and some related Gram-positive bacteria, we hypothesized that the protein binds HMP and/or thiamin. Co-crystallization experiments with thiamin yielded good quality crystals, which allowed for characterization of the structure of the complex. Inspection of the 2.3 Å resolution electron density map revealed the presence of thiamin molecules in both putative binding sites in each of the two molecules in the asymmetric unit (Figure 4).

The two identical binding pockets within the N-terminal motifs of each of the monomers are lined with the hydrophobic side-chains of Leu17, Tyr18, Phe24, and Leu32. In addition, the benzene ring of Phe15 forms the bottom of the pocket against which the C(2') methyl substituent of the pyrimidine ring rests. The specificity of the interaction with the HMP moiety of thiamin is mediated by hydrogen bonds to all the pyrimidine nitrogen atoms. The main chain of Leu17 interacts with the pyrimidine ring so that the amide donates a hydrogen bond to N(3'), while the carbonyl oxygen accepts one from the N(4')-amino group. The second hydrogen of the amino group is donated to a water molecule that is also in close contact to the C(2) atom of the thiazole moiety: 3.1 Å. The C(2)-bound proton is quite acidic as it is adjacent to two electron-withdrawing atoms, i.e. N(3) and S(1), and it is therefore capable of donating a hydrogen bond to water forming a weak C-H...O bond, stabilizing thiamin's conformation. A similar interaction involving a Cl⁻ has been observed in the crystal structure of 3-benzyl-5-(2-hydroxyethyl)-4-methyl-1,3-thiazolium chloride.⁶ Finally, the side-chain hydroxyl of Thr49 donates a hydrogen bond to the N(1') atom of the pyrimidine ring. The conformation of Thr49 is stabilized by an additional H-bond with Thr44, so that the hydroxyl is poised favorably to bind the ligand.

The binding pockets within the C-terminal motifs are very similar to those in the N-terminal motifs. Their walls are formed by Ala120, Leu121, Tyr129, Met130, Ile133, Val137, and Tyr152 with Phe119 at the bottom. Although the pocket is formed primarily by one protein molecule, two residues from the B molecule of the homodimer, Cys86, and His180, also contribute to the thiamin-binding site and are in van der Waals contact with the thiazole ring. Specificity of binding is enforced by hydrogen bonds analogous to those observed in the N-terminal motif. Thus, the main chain of Leu121, which is analogous to Leu17, anchors N(3') and the N(4')-amino group, while the hydroxyl of Ser154 is within an H-bonding distance of N(1'). The conspicuous difference between the two sites is the absence of a water molecule which bridges the N(4')-amino group to C(2) of the thiazole.

To assess the affinities of the two sites for thiamin, we carried out microcalorimetric titration experiments, which confirmed the presence of two sites with different dissociation constants (Figure 5). The high-affinity site has a K_D in the low micromolar range, approximately 10 μ M, whereas the second site has a much lower affinity, with a K_D of about 250 μ M. In both cases the affinity arises primarily from enthalpic changes, although in the case of the high-affinity site there is an unfavorable change in entropy, in contrast to a favorable entropic term in the low affinity site. Assuming that this difference can be rationalized by bound water molecules, the high affinity site would correspond to the N-terminal domain, while the low affinity site is likely to be the one residing in the C terminus.

The conformation of thiamin is defined by the two dihedral angles around the C(7') methylene bridge, i.e. the ϕ_T and ϕ_P angles defined as [C(5')–C(7')–N(3)–C(2)] and [N(3)–C(7')–C(5')–C(4')].⁷ In the YkoF–thiamin complex, the ϕ_T and ϕ_P angles of both thiamin molecules in the high-affinity sites are $\sim 110^\circ$ and $\sim 75^\circ$, respectively. Those in the lower-affinity sites are marginally different, i.e. $\sim 98^\circ$ and $\sim 50^\circ$. These values are close to the so-called V conformation ($\phi_T=90^\circ$ and $\phi_P=90^\circ$), in which the N(4')-amino group is adjacent to the reactive C(2) on the thiazolium ring, and which was originally proposed to be the active form of the enzyme-bound coenzyme.⁸ The V conformation, which is unusual for free thiamin and its C(2)-substituted derivatives with the exception of thiamin thiazolone⁹ is, however, observed in complexes with enzymes that utilize thiamine diphosphate for catalysis.¹⁰ The V conformation is typically supported by a bulky hydrophobic side-chain which is in van der Waals contact with both aromatic rings of thiamin.¹¹ In YkoF, this role appears to be assumed by Leu28 and Leu133. The functional significance of this is unclear, unless YkoF has some catalytic function that is neither suggested by the fold, nor readily identifiable from the structure.

Similarity to other proteins

A search for homologous and/or structurally related proteins using PSI-BLAST,¹² DALI¹³ and the MetaServer of the Polish Bioinformatics Site,^{14–16} revealed only two other protein sequences with a significant level of amino acid sequence similarity to YkoF, and a more distant relationship to the COG0011 and a related DUF77 domain families, with a total of 66 known sequences. Crystal structures of two members of the COG0011 family, i.e. the yeast protein YBL001c and the *Methanobacterium thermoautotrophicum* protein MTH1187, have been recently determined.¹⁷

The two close homologues, from *Oceanobacillus iheyensis* and *Mesorhizobium loti* contain both $\beta\alpha\beta\beta\alpha\beta$ repeats (Figure 6). A majority of the residues involved in YkoF in ligand binding, in the hydrophobic core and in the dimer interface are all highly conserved among the three proteins. This pattern suggests that all three proteins share the same function. In contrast, all members of the COG0011 family, including the two proteins with known crystal structures,¹⁷ appear to contain a single repeat. However, we note that the putative binding cavity is largely preserved in both MTH1187 and YBL001c, and Thr47, which corresponds to Thr49 in YkoF and is involved in HMP recognition, is completely conserved.

The ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ fold is also observed among the ACT domains,¹⁸ which are related to the regulatory domain of the *Escherichia coli* 3-phosphoglycerate dehydrogenase (3PGDH),⁵ as well as in RAM domains involved in the regulation of amino acid metabolism in prokaryotes.¹⁹ All these structurally related proteins mediate allosteric regulation and ligand binding, and typically undergo oligomerization, albeit the architecture of the resulting oligomers varies significantly (Figure 7). YkoF is unique in that it contains two $\beta\alpha\beta\beta\alpha\beta$ motifs arranged side-by-side so that the terminal β -strands of both form an antiparallel sheet. Furthermore, YkoF generates a dimer with a face-to-face association of the exposed faces of the tandem's eight-stranded β -sheets, resulting in a pseudotetrameric arrangement of the constituent motifs. In contrast, the single $\beta\alpha\beta\beta\alpha\beta$ motif of the ACT domain of 3PGDH forms a homodimer by a side-by-side association involving strand β_2 rather than β_4 . The RAM domains show yet another homodimerization pattern forming a β -barrel by a face-to-face association of two $\beta\alpha\beta\beta\alpha\beta$ monomers. The only protein that we found to contain an internal repeat of the ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ fold is the C-terminal, regulatory domain of the pyridoxal phosphate-dependent allosteric threonine deaminase,²⁰ but even here the side-to-side arrangement of the $\beta\alpha\beta\beta\alpha\beta$ motifs is mediated by the antiparallel association of the β_2 and β_2' strands, thus reversing their order as compared to YkoF. A dimer is then formed so that the β -sheets face each other, but they do not interact directly as they are separated by the residues from the inter- $\beta\alpha\beta\beta\alpha\beta$ linkers.

The crystal contacts

Each molecule in the homodimer is involved in two different inter-dimer contacts that synergistically form the crystal lattice. Along the Z axis, the dimers come into contact so that molecule A and molecule B form a symmetric interface mediated by the surface patches which contain the two mutated sites, Ala33 and Ala34, as well as Thr35 (Figure 8(A)). The solvent accessible main-chain carbonyl groups from Ala33 and Thr35 from symmetry-related molecules create a Ca^{2+} site complemented by three water molecules to generate a distorted pentagonal bipyramid (Figure 8(B)). This contact rationalizes our observation that the crystallization step was dependent on divalent cations, notably Ca^{2+} .

The second inter-dimer crystal contact involves residues Val147^A, Lys141^A, and Glu148^A from one dimer, which form van der Waals contacts with Lys141^B, Asp162^B and Glu143^B from the symmetry-related dimer. Three H-bonds (between Val147^A O–N^{ε2} Glu143^B, Asp157^A O^δ–N^{ε2} Glu176 and Lys141^A N^ζ–O^{γ1} Asp162) contribute to the stability of the contact.

Following the completion of this study we have learned of the independent structure determination of the wild-type YkoF based on crystals that were prepared in a second tier of screening at the Midwest Center. This structure, refined at 2.2 Å resolution, has been deposited with the Protein Data Bank (PDB) and released (1S7H). The model consists of four independent monomers and 471 water molecules, yielding an *R* factor of 22.1% and *R*_{free} of 28.1%. Its comparison with the K33A/K34A double mutant structure shows that there is no difference between the two structures, with a representative root-mean-square difference between the C^α atoms of molecule A in one and molecule A in the other is only 0.43 Å (the corresponding difference between molecules A and B in the 1.6 Å structure is 0.15 Å). Close inspection of the mutated patch reveals that its main-chain structure is virtually identical, except for very minor conformational shifts due to crystal packing and calcium binding. In the wild-type structure Lys33 protrudes from the surface, clearly impeding intermolecular contact at this site.

Conclusion

The crystal structure of the YkoF protein revealed an internally duplicated ACT-like fold. Genetic data along with structural considerations made it possible to suggest a potential function for the protein as that of an HMP/thiamin transporter or storage molecule. Indeed, this was confirmed experimentally, and the structure of the complex revealed a specific binding site with unique hydrogen bonding of the HMP moiety. Thus, YkoF represents a novel thiamin-binding fold, which is not related to that found in thiamin diphosphate-dependent enzymes.²¹ It is likely that YkoF is involved in a unique transport system associated with the vitamin B1 biosynthesis pathway in certain Gram-positive bacteria. The structural study paves the way for biochemical and genetic work aimed at elucidating the physiological function of this protein.

This study provides yet another illustration of the general applicability of crystallization by surface entropy reduction.³ The YkoF structure is striking in that regard, because the key crystal contact is formed by an intermolecular Ca^{2+} -binding site, generated by the mutated epitopes. The crystals of the mutant grow easily and quickly, in contrast to the wild-type form, which required considerable effort and crystallized after time-consuming, extensive custom screens. Furthermore, as has been shown in other cases,²² the low-entropy contacts generate better quality crystals than those obtained for the wild-type protein. In the case of YkoF the mutant crystals diffract to 1.6 Å resolution, in contrast to the 2.2 Å resolution obtained for the wild-type form. A comparison of the two independently solved and refined structures further validates the approach, as it shows that the Lys→Ala mutations do not affect the structure of the protein.

Experimental Procedures

Protein expression and purification

The open reading frame of the *B. subtilis* YkoF protein was amplified from genomic DNA with a recombinant KOD HiFi DNA polymerase (Novagen) from *Thermococcus kodakaraensis* using conditions and reagents provided by the vendor (Novagen). The gene was cloned into a pMCSG7 vector²³ using a modified ligation independent cloning protocol.²⁴ This process generated an expression clone producing a fusion protein with an N-terminal His₆ tag and a recognition site for the tobacco etch virus (TEV) protease. Mutations were introduced into this vector using the QuikChange™ site-directed mutagenesis kit (Stratagene) and confirmed by direct sequencing.

For crystallization at the Midwest Center for Structural Genomics (MCSG), the wild-type protein was over-produced in the BL21-Gold (DE3) cells (Stratagene) harboring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The cells were grown in LB medium at 37 °C to an A_{600 nm} of ~0.6 and protein expression was induced with 0.4 mM isopropylthiogalactoside (IPTG). After induction, the cells were incubated overnight with shaking at 15 °C and harvested. A SeMet derivative of the expressed protein was prepared as described earlier.²⁵ The protein was purified by resuspension of IPTG-induced bacterial cells in binding buffer (500 mM NaCl, 5% (v/v) glycerol, 20 mM Hepes (pH 8.0), 10 mM imidazole, 10 mM β-mercaptoethanol). The cells were lysed by adding lysozyme to 1 mg/ml in the presence of a protease inhibitor mixture (Sigma P8849) (0.25 ml/5 g cells) and sonication for two to three minutes. After centrifugation for 30 minutes at 17,000 rpm (Beckman) and passage through a 0.2 μm filter, the lysate was applied manually to Ni-NTA Superflow resin (Qiagen) and unbound proteins removed by washing with ten volumes of binding buffer. The protein was eluted from the column with 250 mM imidazole, and the fusion tag cleaved with recombinant His-tagged TEV protease. Target protein was purified from the His tag, undigested protein and TEV protease by application of the solution to a second Ni-NTA column. The buffer in the purified protein was exchanged with 10 mM Tris-HCl (pH 7.6), 50 mM NaCl on a PD-10 column (Pharmacia) and concentrated using a BioMax concentrator (Millipore). Before crystallization, any particulate matter was removed from the sample by centrifugation for 20 minutes at 14,000 rpm at 4 °C.

For subsequent studies at the University of Virginia the wild-type protein as well the two mutants were expressed in *E. coli* BL21 strain. Cell cultures were grown in regular LB broth until an A of ~0.7 and induced with 1 mM IPTG for 12 hours. Protein expression was conducted at 28 °C. Cells were harvested and lysed in buffer containing 50 mM Tris-HCl (pH 8.0) and 300 mM NaCl. The protein was initially purified using nickel affinity chromatography (Ni-NTA agarose column, Qiagen) and subjected to rTEV proteolysis at 10 °C for 36 hours to cleave the His tag. Samples were run again through a Ni-affinity column to isolate pure, untagged protein, which was dialyzed for 12 hours against buffer consisting of 20 mM Tris-HCl (pH 8.0). Pure native protein was concentrated to 14–15 mg/ml. β-Mercaptoethanol was added to a final concentration of 2.5 mM and protein samples were stored at 4 °C. To generate the SeMet labeled protein samples, *B. subtilis* YkoF K33,34A mutant protein was expressed in *E. coli* B834 methionine auxotrophic cells. Cell cultures were grown in the presence of selenomethionine until an A of ~1.0 and induced with 1 mM (IPTG). Protein was harvested after 48 hours and purified using the procedure described above.

Crystallization

Initial screen at the MCSG using commercial formulations and the wild-type protein yielded no results. Similarly, a search for suitable crystallization conditions for the wild-type protein was carried out using the protein obtained at the University of Virginia, Hampton I & II

(Hampton Research), Wizard I & II (Emerald Biostructures), and Sigma (Sigma-Aldrich) screens at different protein concentrations, but did not yield any crystals. For the K(33,34)A mutant protein, initial crystals were obtained using Wizard II (Emerald Biostructures) and Hampton II (Hampton Research) screens. Crystals for data collection were grown by microseeding at 20 °C using the sitting-drop vapor-diffusion method. A concentrated protein solution (1 μ l of 10–15 mg/ml) was mixed with an equivalent amount of mother liquid containing 20% (w/v) polyethylene glycol 8000, 100 mM 2-[*N*-morpholino]ethanesulfonic acid (pH 6.5), and 200 mM calcium acetate with suspended seeds. Crystals typically formed in one to two days and belonged to space group $P2_12_12_1$ with cell dimensions $a=60.92$ Å, $b=83.28$ Å, $c=85.90$ Å. Assuming a molecular mass of 22 kDa and two molecules in the asymmetric unit, the Matthews' coefficient²⁶ is 2.47 Å³/Da, i.e. within normal range for globular proteins. SeMet crystals, diffracting to 1.6 Å at the APS, were obtained using the sitting-drop vapor-diffusion method by mixing 2 μ l of concentrated protein with an equivalent amount of mother liquid (20% polyethylene glycol 8000, 100 mM 2-[*N*-morpholino]ethanesulfonic acid (pH 6.5), 200 mM calcium acetate) containing micro crystal seeds of the unlabeled protein diluted 10,000 times. Large crystals with maximal dimensions 0.5 mm $\times$ 0.2 mm $\times$ 0.15 mm appeared in two days and were of the same space group as the unlabeled protein. The complex of YkoF with thiamin was prepared by co-crystallization of the K33A,K34A mutant protein with 2.5 mM thiamin by microseeding.

Independent of this work, a second screen was conducted at MCSG. Oval-shaped microcrystals of the wild-type protein were obtained using 0.16 M MgCl₂, 0.08 M Tris-HCl (pH 8.5), and 24% PEG 4000. Further refinement of crystallization conditions was carried out with SeMet derivative and improved original crystals. The best crystals were grown using hanging-drop vapor-diffusion at ambient temperatures using equal volume of reservoir and protein solution at 30 mg/ml in 50 mM Hepes (pH 8.0), 500 mM NaCl, 2 mM DTT equilibrated against reservoir containing 0.2 M MgCl₂, 0.1 M Tris-HCl (pH 8.5) and 25% PEG 3350. The crystals grew to their maximum sizes of approximately 0.15 mm $\times$ 0.15 mm $\times$ 0.1 mm. The cryo-conditions were obtained by addition of 10% ethylene glycol or glycerol to the reservoir. The best quality crystals diffracted to 2.2 Å resolution. Crystals were mounted on cryo-loops (Hampton Research) and flash-frozen in liquid nitrogen. Crystals belonged to the space group $P2_12_12_1$ with the cell parameters $a=169.74$ Å, $b=55.08$ Å, $c=85.49$ Å. All other data relevant to this result were deposited with the PDB together with the coordinates (entry 1S7H).

Data collection

Crystals of the K(33,34)A mutant were frozen by immersion in liquid nitrogen using cryo-protecting solution consisting of the mother liquor with addition of 20% (v/v) ethylene glycol. Data were collected at APS, SER-CAT beamline 22ID at $\lambda=0.97942$, $\lambda=0.97953$, and $\lambda=0.96419$ corresponding to peak, inflection point and remote wavelengths at Se edge, respectively. Data were processed and reduced with HKL2000.²⁷ Data for the complex of YkoF with thiamin were collected with the use of Enraf-Nonius FR591 X-ray generator equipped with purple confocal mirrors (MSC), and the R-Axis IV detector (MSC).

Structure solution

SHELXS and SHELXD²⁸ were used to identify a substructure of 12 selenium atoms out of 14 present in the asymmetric unit. The data were phased with MLPHARE²⁹ and the phases were improved by density modification with DM,³⁰ which yielded an overall figure of merit of 0.84. An initial model was built using Arp-Warp³¹ and consisted of 331 residues. The structure was refined with REFMAC³² to R factor 17.1 (R -free 22.4) at 1.65 Å resolution.

Isothermal titration calorimetry

Isothermal calorimetric measurements were performed at 21 °C on a MicroCal-ITC (MicroCal, Inc.; Northampton, MA). In this experiment, 10 µl aliquots of 4.5 mM thiamin in 50 mM phosphate buffer (pH 6.5) were injected from a 300 µl syringe into an isothermal sample chamber containing 1.43 ml of 0.3 mM YkoF protein solution in the same buffer. The experiment was accompanied by the corresponding control experiment in which 10 µl aliquots of 4.5 mM thiamin, were injected into the buffer alone. The duration of each injection was 5.0 seconds, and the delay between injections was 240 seconds. The initial delay prior to the first injection was 60 seconds. The data were analyzed with the use of Origin software (MicroCal, Inc.; Northampton, MA). The heat associated with each thiamin-buffer injection was subtracted from the corresponding heat associated with each thiamin-protein injection to yield the heat of thiamin binding for the corresponding injection. The protein solution for the ITC experiment was extensively dialyzed against 6.0 l phosphate buffer (pH 6.5), to completely remove the original Tris-HCl buffer in which the protein was stored after purification. All buffers contained 2.0 mM β-mercaptoethanol.

Protein Data Bank accession codes

The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank[†] (PDB ID code 1S99 for the mutant structure, 1SBR for the complex).

Acknowledgments

We thank all members of the SBC and SER-CAT at APS (Argonne National Laboratory) for their help conducting experiments. We thank Dr Frank Collart (Argonne National Laboratory) for providing an over-expressing clone of the YkoF protein. This work was supported by grants from the National Institutes of Health, National Institute of General Medical Sciences to A.J. (GM62414), and Z.S.D. (GM62615). Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline and at the Structural Biology Center 19ID at the Advanced Photon Source (APS), Argonne National Laboratory. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Abbreviations used

HMP	hydroxymethyl pyrimidine
MAD	multiwavelength anomalous dispersion
PDB	Protein Data Bank
IPTG	isopropylthiogalactoside

References

- Schowen, R. Thiamin-dependent enzymes.. In: Sinnott, L., editor. Comprehensive Catalysis. Academic Press; San Diego: 1998. p. 217-266.
- Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. J. Biol. Chem 2002;277:48949–48959. [PubMed: 12376536]
- Derewenda ZS. Rational protein crystallization by mutational surface engineering. Structure 2004;12:1–20. [PubMed: 14725756]
- Laskowski RA, McArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallog 1993;26:282–291.
- Schuller DJ, Grant GA, Banaszak LJ. The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. Nature Struct. Biol 1995;2:69–76. [PubMed: 7719856]

[†]www.rcsb.org

6. Shin W, Chae CH. Structure of 3-benzyl-5-(2-hydroxyethyl)-4-methyl-1,3-thiazolium chloride. *Acta Crystallog. sect. C* 1993;49:68–70.
7. Pletcher J, Sax M, Blank G, Wood M. Stereochemistry of intermediates in thiamine catalysis. 2. Crystal structure of dl-2-(alpha-hydroxybenzyl)thiamine chloride hydrochloride trihydrate. *J. Am. Chem. Soc* 1977;99:1396–1403. [PubMed: 838999]
8. Schellenberger A. The amino group and steric factors in thiamin catalysis. *Ann. N.Y. Acad. Sci* 1982;378:51–62. [PubMed: 6952788]
9. Shin W, Kim YC. Crystal structure of thiamin thiazolone: a possible transition-state analogue with an intramolecular N–H...O hydrogen bond in the V form. *J. Am. Chem. Soc* 1986;108:7078–7082.
10. Jordan F. Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Nature Prod. Rep* 2003;20:184–201. [PubMed: 12735696]
11. Guo F, Zhang D, Kahyaoglu A, Farid RS, Jordan F. Is a hydrophobic amino acid required to maintain the reactive V conformation of thiamin at the active center of thiamin diphosphate-requiring enzymes? Experimental and computational studies of isoleucine 415 of yeast pyruvate decarboxylase. *Biochemistry* 1998;37:13379–13391. [PubMed: 9748345]
12. Altschul SF, Koonin EV. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem. Sci* 1998;23:444–447. [PubMed: 9852764]
13. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol* 1993;233:123–138. [PubMed: 8377180]
14. von Grotthuss M, Pas J, Wyrwicz L, Ginalski K, Rychlewski L. Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins: Struct. Funct. Genet* 2003;53:418–423. [PubMed: 14579330]
15. Ginalski K, Rychlewski L. Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins: Struct. Funct. Genet* 2003;53:410–417. [PubMed: 14579329]
16. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 2003;19:1015–1018. [PubMed: 12761065]
17. Tao X, Khayat R, Christendat D, Savchenko A, Xu X, Goldsmith-Fischman S. Crystal structures of MTH1187 and its yeast ortholog YBL001c. *Proteins: Struct. Funct. Genet* 2003;52:478–480. [PubMed: 12866058]
18. Chipman DM, Shaanan B. The ACT domain family. *Curr. Opin. Struct. Biol* 2001;11:694–700. [PubMed: 11751050]
19. Ettema TJ, Brinkman AB, Tani TH, Rafferty JB, Van Der Oost J. A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *J. Biol. Chem* 2002;277:37464–37468. [PubMed: 12138170]
20. Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE, Chinchilla D, Eisenstein E. Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. *Structure* 1998;6:465–475. [PubMed: 9562556]
21. Muller YA, Lindqvist Y, Furey W, Schulz GE, Jordan F, Schneider G. A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure* 1993;1:95–103. [PubMed: 8069629]
22. Munshi S, Hall DL, Kornienko M, Darke PL, Kuo LC. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 Å resolution. *Acta Crystallog. sect. D* 2003;59:1725–1730.
23. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelley MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif* 2002;25:1–7. [PubMed: 12071692]
24. Dieckman L, Gu M, Stols L, Donnelley MI, Collart FR. High throughput methods for gene cloning and expression. *Protein Expr. Purif* 2002;25:8–15. [PubMed: 12071693]
25. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A. Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallog. sect. D* 1999;55:1168–1173.
26. Matthews B. Solvent content of protein crystals. *J. Mol. Biol* 1968;33:491–497. [PubMed: 5700707]
27. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.

28. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallog. D* 2002;58:1772–1779.
29. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D* 1994;50:760–763.
30. Cowtan K, Main P. Miscellaneous algorithms for density modification. *Acta Crystallog. sect D* 1998;54:487–493.
31. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol* 1999;6:458–463. [PubMed: 10331874]
32. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D* 1997;53:240–255.

Figure 1.

The tertiary structure of the YkoF molecule. (A) A diagram showing the overall fold, with the N-terminal $\beta\alpha\beta\beta\alpha\beta$ module colored in red and the C-terminal module shown in blue. The missing residues within the linker between the two modules are represented by a dotted fragment. The non-crystallographic dyad between the two motifs is indicated by its symbol, and the two thiamin molecules are shown as stick models in the respective binding sites (see the text for further details). (B) A superposition of the two $\beta\alpha\beta\beta\alpha\beta$ modules; note the different lengths of helices A and a different disposition of the antiparallel strands $\beta 2$ and $\beta 3$ resulting in conformational differences of the two ligand-binding sites.

Figure 2. The hydrophobic core structure of YkoF. The residues within the N-terminal module are shown in red, within C-terminal module in magenta, while those within the inter-module interface are yellow.

Figure 3.
A diagrammatic representation of the quaternary structure of YkoF highlighting the key residues involved in the intermolecular interface.

Figure 4.

The thiamin-binding sites of YkoF. (A) The high affinity site in the N-terminal motif, and (B) the low affinity site in the C-terminal motif in molecule A. The omit electron density map is contoured at 2.8σ level. Hydrogen bonds are shown as broken lines, the dotted line indicates a putative weak C–H...O interaction stabilizing the thiazole group in the high-affinity site. The corresponding electron density in molecule B is very similar to that shown here.

Figure 5.

Calorimetric titration of YkoF with thiamin. Raw data (upper plot), and a plot of the integrated heat *versus* thiamin/protein ratio (lower plot). The fitted line is calculated assuming two binding sites per YkoF monomer, with the following thermodynamic parameters: site 1, $n=1.1$, $K_d=253$ μM , $\Delta H=-4.1$ kcal/mol, $T\Delta S=0.8$ kcal/mol; and site 2, $n=0.8$, $K_d=9.1$ μM , $\Delta H=-7.6$ kcal/mol, $T\Delta S=-0.84$ kcal/mol.

Figure 6.

Sequence alignment of the three YkoF proteins from bacterial species. Residues in the hydrophobic core are labeled with an asterisk, those involved in dimer formation are labeled with crosses and the residues important for thiamin recognition are indicated by green boxes.

Figure 7. Representative modes of oligomerization in the ACT/RAM superfamily. (A) The arrangement of modules in YkoF; (B) the ACT domains of ̢-3-phosphoglycerate dehydrogenase (1PSD.pdb) and (C) the RAM domains of the Lrp-like transcriptional regulator (1I1G.pdb).

Figure 8.

The packing of the YkoF molecules in the crystal lattice. (A) The arrangement of non-crystallographic dimers, interacting *via* the engineered Ca-mediated crystal contact along the *b* axis; (B) the coordination of Ca^{2+} at the crystal contact containing the two mutated residues.

Table 1

Crystallographic data for the YkoF K33A,K34A mutant with and without bound thiamin

YkoF K33A/K34A mutant	Complex with thiamin			
Wavelength (Å)	0.97942	0.97953	0.96416	1.5418
Resolution (Å)	34.0–2.0 (2.14–2.0) ^a	34.0–2.0 (2.14–2.0)	34.0–1.65 (1.71–1.65)	30.0–2.3 (2.38–2.30)
Unique reflections	56,617 ^b (4827)	56,713 ^b (4811)	81,862 ^b (2581)	16,589 ^c (941)
Completeness (%)	99.4 (100)	99.5 (100)	82.2 (25.6)	85.8 (49.7)
Redundancy	4.3 (4.3)	4.2 (4.2)	2.3 (1.4)	4.7 (3.2)
<i>I</i> /σ	22.6 (14.2)	22.4 (12.6)	16.8 (1.5)	19.5 (2.8)
<i>R</i> _{merge} (%) ^d	4.4 (12.8)	4.5 (14.0)	5.1 (31.6)	7.5 (30.6)
Phasing power	0.62		0.68	
Figure of merit		0.474		
DM ^e		0.84		
<i>Refinement</i>				
Reflections (working/test)			47,795/1265	15,723/836
<i>R</i> _{work} (last shell) ^f			17.1% (28%)	21.1% (32%)
<i>R</i> _{free} (last shell) ^f			22.3% (29%)	27.4% (39%)
r.m.s. deviations from ideal stereochemistry				
Bond lengths (Å)			0.016	0.019
Bond angles (°)			1.5	1.8
Average <i>B</i> factor				
All atoms			27.8	52.7

^aThe values in parenthesis are for the last data shell.^bThe values are for the unmerged Friedel pairs.^cThe values are for the merged Friedel pairs.^d $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I$, where *I* is the intensity of the *i*th observation and $\langle I \rangle$ is the mean intensity of the reflection.^eDensity modification and histogram matching.²⁹^f $R_{\text{work}} = \sum \|F_{\text{obs}} - |F_{\text{calc}}|\| / \sum |F_{\text{obs}}|$ where all reflections belong to a test set of randomly selected data. *R*_{free} is an analogous factor calculated for a set of randomly selected data.