

Structural insights into the catalytic mechanism of a sacrificial sulfur insertase of the N-type ATP pyrophosphatase family, LarE

Matthias Fellner^{a,1}, Benoît Desguin^{a,b,1}, Robert P. Hausinger^{a,c,2}, and Jian Hu^{a,d,2}

^aDepartment of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; ^bInstitute of Life Sciences, Université catholique de Louvain, B-1348 Louvain-La-Neuve, Belgium; ^cDepartment of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824; and ^dDepartment of Chemistry, Michigan State University, East Lansing, MI 48824

Edited by JoAnne Stubbe, Massachusetts Institute of Technology, Cambridge, MA, and approved July 19, 2017 (received for review March 25, 2017)

The *lar* operon in *Lactobacillus plantarum* encodes five Lar proteins (LarA/B/C/D/E) that collaboratively synthesize and incorporate a niacin-derived Ni-containing cofactor into LarA, an Ni-dependent lactate racemase. Previous studies have established that two molecules of LarE catalyze successive thiolation reactions by donating the sulfur atom of their exclusive cysteine residues to the substrate. However, the catalytic mechanism of this very unusual sulfur-sacrificing reaction remains elusive. In this work, we present the crystal structures of LarE in ligand-free and several ligand-bound forms, demonstrating that LarE is a member of the N-type ATP pyrophosphatase (PPase) family with a conserved N-terminal ATP PPase domain and a unique C-terminal domain harboring the putative catalytic site. Structural analysis, combined with structure-guided mutagenesis, leads us to propose a catalytic mechanism that establishes LarE as a paradigm for sulfur transfer through sacrificing its catalytic cysteine residue.

thiolation | ATP pyrophosphatase | crystal structure | Lar protein | catalysis

Lactic acid, composed of both L- and D-isomers, is a widespread organic compound produced during microbial fermentations via stereospecific lactate dehydrogenases. Certain bacteria possess the ability to interconvert the two enantiomers by using lactate racemase, which was only recently described in genetic (1), structural (2), synthetic modeling (3), and computational studies (4, 5).

LarA from *Lactobacillus plantarum* is responsible for lactate racemase activity. This Ni-dependent enzyme (1) contains a newly identified cofactor, pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN), that is covalently attached to an active-site lysine residue (2). Most interestingly, P2TMN binds an Ni atom using sulfur, carbon, and sulfur atoms of an SCS-pincer complex (2), making LarA the ninth discovered Ni-dependent enzyme (6).

Synthesis of Ni-bound P2TMN occurs through a pathway involving three proteins encoded in the *lar* operon (i.e., LarB, LarC, LarE) (1). LarB is a carboxylase/hydrolase that produces pyridinium-3,5-biscarboxylic acid mononucleotide (P2CMN) from nicotinic acid adenine dinucleotide (7). LarE then converts P2CMN into P2TMN through two successive sulfur transfer reactions. Finally LarC is thought to provide the Ni atom to generate the active form of the pincer cofactor of LarA (7) (Fig. 1).

The amino acid sequence of LarE suggests it has two domains. The N-terminal domain is homologous to N-type ATP pyrophosphatase (PPase) domains (8), containing a conserved SGGxDS motif that binds and hydrolyzes ATP to form AMP and pyrophosphate. Examples of enzymes with this domain are guanine monophosphate (GMP) synthetase, nicotinamide mononucleotide (NMN) synthetase, and nicotinamide adenine dinucleotide (NAD) synthetase. These enzymes activate substrate carboxyl or carbonyl groups by adenylation (AMPylation) (9). The C-terminal domain of LarE has no homology to any member of the N-type ATP PPase family, which use their C-terminal domains to recognize specific substrates and catalyze

their versatile reactions. We therefore hypothesize that LarE represents a new member of this enzyme family that acts as a sulfur insertase of P2CMN by sacrificing a sulfur atom from an invariant cysteine residue (Cys176) (7) that is located in the C-terminal domain. Structural studies of LarE could be used to clarify the detailed process of this unique reaction.

Here, we present crystal structures of LarE in different states (substrate-free, ATP-bound, AMP-bound, and substrate analog-bound), which enables us to propose a catalytic mechanism that is further supported by site-directed mutagenesis and enzymatic activity assays. Our study reveals that LarE is a member of the N-type ATP PPase family and demonstrates that its C-terminal catalytic site confers a unique catalytic mechanism.

Results

Quaternary Structure of LarE. The initial crystal structure of *L. plantarum* LarE (LarE_{Lp}) produced in *Lactococcus lactis* was solved by single-wavelength anomalous dispersion using a selenomethionine-substituted crystal at 3.3 Å. Although we later solved the crystal structures of LarE_{Lp} produced by *Escherichia coli* with several different space groups under different crystallization conditions, LarE always was shown to form a hexamer (Fig. 2). Size exclusion chromatography results also supported a

Significance

Thiolation reactions are essential steps in the synthesis of numerous biological metabolites. To make the novel sulfur-containing cofactor of LarA, an Ni-dependent lactic acid racemase, LarE catalyzes a critical sulfur transfer reaction to a nicotinic acid-derived substrate by converting the protein's cysteine residue to dehydroalanine. In this study, crystal structures of ligand-free and several ligand-bound forms of LarE provide a structural basis for a catalytic mechanism that is further supported by structure-guided mutagenesis and functional assays. This work establishes LarE as a sulfur insertase within the N-type ATP pyrophosphatase family and presents a paradigm for sulfur transfer through sacrificing a catalytic cysteine residue.

Author contributions: M.F., B.D., R.P.H., and J.H. designed research; M.F. and B.D. performed research; M.F., B.D., R.P.H., and J.H. analyzed data; and M.F., B.D., R.P.H., and J.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5UDQ for substrate-free, 5UNM for alternative substrate-free, 5UDR for NMN-bound, 5UDS for ATP-bound, 5UDT for AMP-bound, 5UDU for Mn-bound, 5UDV for Fe-bound, 5UDW for Ni-bound, and 5UDX for Zn-bound structures).

¹M.F. and B.D. contributed equally to this work.

²To whom correspondence may be addressed. Email: hausinger@msu.edu or hujian1@msu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1704967114/-DCSupplemental.

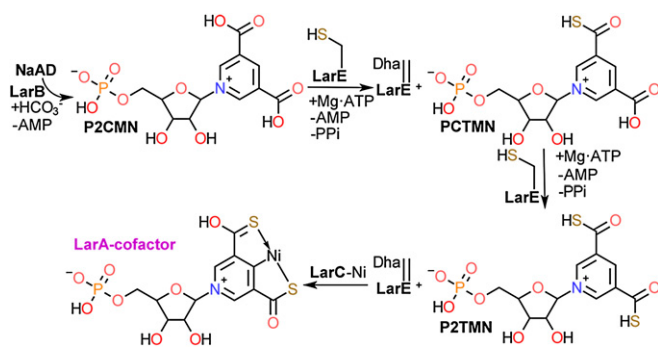


Fig. 1. Ni-pincer LarA-cofactor biosynthesis highlighting the sacrificial sulfur transfer reactions of LarE. Dha, dehydroalanine; PCTMN, pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide.

hexamer in solution (*SI Appendix*, Fig. S1). The hexamer is formed by two head-to-head trimers of three LarE subunits (trimer ABC $\approx$ trimer EFD; *SI Appendix*, Figs. S2 and S3 and Table S1) through the C-terminal “head domain.” A metal binding site was discovered at the center of the trimer bound to Asp231 from three chains (Fig. 2A). Crystal soaks revealed that this site can bind Mn, Fe, Ni, or Zn, with Ni and Zn also binding to surface residues (*SI Appendix*, Figs. S4 and S5). However, as Asp231 is not conserved (Fig. 3) and a D231R variant was as active as wild type (WT), it appears that this site is not catalytically relevant.

Domain Structure of LarE. LarE_{LP} possesses 276 residues plus 10 additional residues from the Strep-tag II and an Ala-Ser linker fused to the C-terminus. In the highest resolution structure [substrate-free protein at 2.09 Å, Protein Data Bank (PDB) ID code 5UDQ], residues 2–125 and 148–259 are modeled in all chains, except for a small gap (residues 172–175) in chain A. The protein contains 14 helices and seven strands (Fig. 4A) and, as predicted, consists of two domains: Residues 1–165 form the N-terminal ATP PPase domain containing the signature PP-loop motif (²⁶SGGxDG³¹), and residues 179–276 form the C-terminal head domain with an $\alpha\beta$ fold involved in oligomerization. The two domains are connected by a flexible linker (residues 166–178) containing the catalytic residue Cys176. A fragment (residues 126–147, including several highly conserved residues) is severely disordered, except for in chain B of an

alternative substrate-free structure (PDB ID code 5UNM), where it folds into an extended loop (*SI Appendix*, Fig. S6A). The C-terminus is also disordered beyond residue 258 or 259, except that chains A and E in some structures could be extended to residues 276 and 279, respectively. Overall, the head domain appears to be more ordered because the oligomeric interfaces prevent motion. In contrast, the ATP PPase domain shows significant shifts among the LarE subunits (*SI Appendix*, Fig. S7), which makes this domain and the flexible linker the dynamic regions of LarE.

We used several search engines (10, 11) to find existing structure matches to the LarE subunit in the PDB. No search provided an exact match, but several close matches to the LarE ATP PPase domain were identified in the ATP PPase family. In contrast, no match was found for the flexible linker or, especially, the head domain of LarE. The closest structural analog of LarE is *E. coli* GMP synthetase, which contains two larger subunits (9, 12): the glutamine aminotransferase domain and the ATP PPase domain, with only the latter appearing to match the LarE structure (*SI Appendix*, Fig. S8A). The ATP PPase domain of GMP synthetase contains two subdomains: The authentic ATP PPase domain (Fig. 4B) can be superimposed on the corresponding domain of LarE with an rmsd of 2.65 Å for the C α atoms, but its C-terminal dimerization domain cannot be satisfactorily superimposed on the head domain of LarE despite the similarity in secondary structure and overall fold (*SI Appendix*, Fig. S8B). Additional differences are associated with the linker connecting the two subdomains of GMP synthetase. This linker is characterized by several highly conserved proline residues, which are not present in LarE, whereas it does not contain a conserved Cys residue, which is absolutely required for catalysis by LarE. Collectively, the crystal structure suggests that LarE is a new member of the N-type ATP PPase family.

Because the C-terminal subdomain of GMP synthetase provides the key residues for binding the substrate xanthine monophosphate (XMP), particularly for interaction with the phosphate and the ribose moieties, we hypothesized that the C-terminal domain of LarE could exert a similar function. Of potential relevance, we noticed that a strong tetrahedral electron density appeared in the head domain of all LarE chains, associated with a positively charged surface composed of Ser180, Arg212, Arg214, Arg259, and, in some cases, Cys176 (Fig. 4C). Because this density appeared in all structures, even when no phosphate or sulfate was added during crystallization, we believe this region represents a

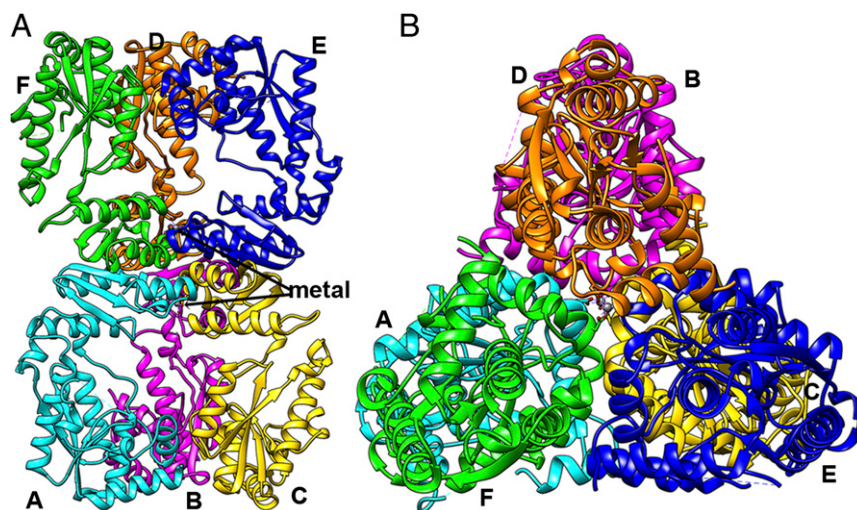


Fig. 2. Quaternary structure and metal binding of LarE. (A) Side view of the quaternary hexameric structure of LarE. Individual chains are colored as A = cyan, B = magenta, C = yellow, D = orange, E = blue, and F = green. A metal (gray sphere) binding site (along with a water molecule represented as a red sphere) is depicted for each trimer. (B) Top view of the hexamer.

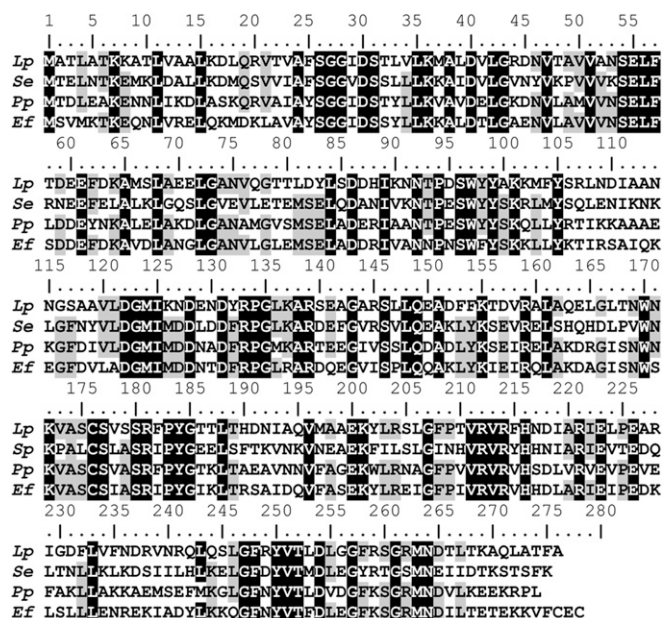


Fig. 3. Sequence alignment of LarE homologs. Numbering is based on LarE_{Lp} sequence. Residues are depicted white on black for invariant residues and black on gray for conserved ones. Ef, *Enterococcus faecalis*; Lp, *L. plantarum*; Pp, *Pediococcus pentosaceus*; Se, *Staphylococcus epidermidis*.

high-affinity binding site for either phosphate or sulfate that is carried over from the purification. Comparison of LarE with the XMP-bound GMP synthetase structure (PDB ID code 2VXO) (12) reveals that the phosphate group of XMP is located close to (6.6 Å in the match shown) this tetrahedral density (SI Appendix, Fig. S8C); therefore, we further postulate that this putative anion binding site corresponds to the binding site for the phosphate group of P2CMN, the substrate of LarE. Consistent with this proposal, the residues at this anion binding site, and even those in the secondary coordination sphere, are highly or universally conserved (Fig. 3), strongly indicating their key roles in LarE function. The occupation of this site by either phosphate or sulfate appeared to be crucial for crystallization, suggestive of a role in structural stabilization of LarE. This is probably due to the involved residues being part of a hydrogen bonding network that spans across the head domain.

ATP/AMP Binding to the ATP PPase Domain of LarE. The ATP-bound structure of LarE was obtained by soaking the substrate-free crystals with ATP. The structure (Fig. 5A) reveals that ATP binds to the PP-loop in a conserved manner as found in other members of the N-type PPase protein family: The γ -phosphate of ATP is coordinated by the invariant residues of the SGGxDS motif (Ser26, Gly27, Gly28, Asp30, and Ser31), the ribose moiety forms hydrogen bonds with the backbone atoms of Gly123 and Ala24, and the adenine moiety is stabilized by hydrogen bonding with the backbone nitrogen and oxygen of Ala52. In addition, Mg²⁺ is associated with the three phosphate groups of ATP, stabilizing a highly bent conformation. As a result, the bound ATP molecule is oriented with the α -phosphate positioned toward the C-terminal domain of LarE, where the substrate P2CMN is supposed to bind. Consistent with the importance of the SGGxDS motif for LarE function, a D30A LarE variant showed that activity was abolished (7).

LarE cocrystallized with ATP/Mg²⁺ formed AMP-bound LarE crystals (Fig. 5B). This result demonstrates the presence of ATP PPase activity and also suggests that AMP stabilizes the LarE structure more than ATP. AMP is bound to LarE similar to how

ATP is bound, while the pyrophosphate is released and no longer observed in the structure. As a result, the AMP molecule lacks the direct interaction with the PP-loop residues seen for ATP. The AMP-bound structure of LarE is nearly identical to that of the ligand-free protein, except that the previously disordered fragment (residues 126–147) harboring several highly conserved residues now folds into a helix, with Arg133 coordinating the α -phosphate of AMP (Fig. 5C). Interestingly, this segment appears to be unique to LarE, and is very different in GMP synthetases or other ATP PPases. In the *E. coli* GMP synthetase structure (PDB ID code 1GPM) (9), the equivalent segment is also disordered, while the *Homo sapiens* GMP synthetase (PDB ID code 2VXO) (12) shows a very different fold (SI Appendix, Fig. S6). We therefore postulate that this segment plays a specific role in LarE catalysis. For instance, Arg133 may stabilize the intermediate of the AMPylation reaction by neutralizing the negative charge developed in the transition state.

Substrate Binding to the C-Terminal Head Domain of LarE. We are particularly interested in how and where P2CMN binds to the protein. Because P2CMN is not commercially available, we conducted exhaustive trials to obtain a substrate homolog-bound structure and managed to solve the costructure of LarE with NMN by soaking with substrate-free crystals. NMN resembles P2CMN, but one carboxyl group in P2CMN is replaced by an amide group and the second carboxyl group is absent. The NMN-bound structure (Fig. 6A) revealed that Ser180, Arg212, Arg214, and Arg259 bind the phosphate moiety of NMN, confirming our previous hypothesis that the highly conserved anion binding site contributes to substrate binding. As shown in the *F_o-F_c* omit map, the strong electron density of the sugar ring enables us to model the structure without ambiguity, showing that the hydroxyl groups on the sugar ring form hydrogen bonds with an invariant Tyr184. The nicotinamide moiety appears to be more flexible with higher B-factors and less density fit compared with the rest of the molecule, although it may be at least partially stabilized by stacking with the side chain of Trp97. Potential reasons for this flexibility include

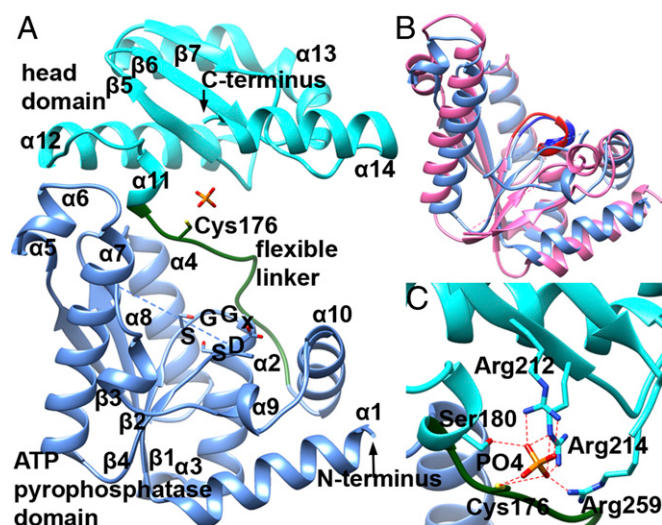


Fig. 4. Domain structure and phosphate binding in LarE. (A) Domain structure of LarE. The head domain is shown in cyan, a flexible linker with Cys176 is shown in dark green, and the ATP PPase domain is shown in light blue with the conserved ²⁶SGGxDS³¹ PP-loop motif illustrated. Helices are labeled α 1– α 14, and sheets are labeled β 1– β 7. The disordered fragment (residues 126–147) is depicted as a dashed line. (B) ATP PPase domain of LarE in light blue is aligned with the ATP PPase domain of *E. coli* GMP synthetase (PDB ID code 1GPM) (9) in pink. The PP-loop residues are colored darker. (C) Close-up view of a phosphate molecule bound to the LarE head domain. H-bonds are illustrated as red dashed lines.

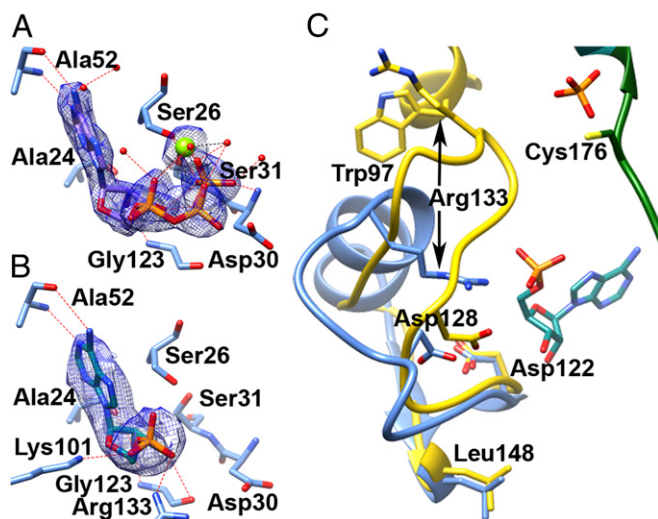


Fig. 5. Comparison of ATP-bound, AMP-bound, and substrate-free LarE structures. ATP-bound structure of LarE (*A*; PDB ID code 5UD5) compared with AMP-bound structure of LarE (*B*; PDB ID code 5UDT). The $2F_o - F_c$ maps are shown as blue meshes at 1σ . Interacting residues are shown in stick mode. Carbon atoms of ATP and AMP are depicted in purple and teal, respectively. An Mg atom is shown as a green sphere, with its chelation represented as black dashed lines. Hydrogen bonds are indicated as red dashed lines. (*C*) Comparison of residues 122–148 loop in the absence (alternative substrate-free LarE PDB ID code 5UNM ribbon in yellow) and presence of AMP.

poor stabilization of the amide group compared with the two carboxyl groups in authentic P2CMN substrate or a requirement for flexibility to perform the reaction with ATP and Cys176. An alignment of LarE sequences from multiple species indicates that all of the residues in the identified substrate binding site are highly conserved or invariant, strongly supporting their key roles in LarE function (Fig. 3). Notably, the NMN-coordinating residues are only conserved in LarE sequences, and not in GMP synthetases. A detailed structural comparison shows the latter enzyme uses a different set of conserved residues for binding XMP (*SI Appendix, Fig. S8C*). To further validate this substrate binding site experimentally, we conducted site-directed mutagenesis to substitute selected LarE residues and measured the activity of the W97A, S180A, and R212A variants (Fig. 6B), with C176A LarE affecting a known essential residue (7) and D231A LarE affecting a known nonessential residue. As expected, the variant proteins affecting the substrate binding site exhibited no detectable activity, confirming the importance of this newly identified structural region of LarE.

Implications of the Structural Findings on the Catalytic Mechanism of LarE. The NMN-bound structure enables us to model a P2CMN molecule at the likely substrate binding site. By overlaying this model structure with the AMP-bound structure (Fig. 7A), we are able to propose a catalytic mechanism of LarE (Fig. 8) that includes three successive steps: (i) AMPylation of the substrate, (ii) formation of a substrate-LarE adduct, and (iii) sulfur transfer from protein to substrate. In the first step, a carboxyl group of P2CMN carries out a nucleophilic attack on the α -phosphate of ATP, bound at the PP-loop motif, leading to release of the pyrophosphate product. The current structure model strongly supports this reaction because the modeled carboxylic acid group of P2CMN is only 4.6 Å from the α -phosphate of ATP. The two highly conserved and positively charged residues (Lys101 and Arg133) may facilitate this reaction through direct association with the α -phosphate, reducing the negative charge of the intermediate in the transition state. Notably, the flexible fragment (residues 126–147) unique in LarE undergoes a coil-to-helix

transition upon AMP binding, which exposes the otherwise blocked substrate binding site in the head domain in the substrate-free structure (Fig. 5C and *SI Appendix*, Fig. S6A and B). The AMP-dependent conformational change implies that this highly dynamic structural element plays an additional role in coordinating ATP and substrate binding. In the second step, the AMPylated substrate is suggested to be attacked by Cys176, located only 5.1 Å away, forming a substrate/LarE adduct and releasing AMP. Cys176 is known to be the residue that donates its sulfur atom to P2CMN, and a substrate/LarE adduct was detected previously using mass spectrometry (7). In the third step, LarE sacrifices the sulfur atom of its Cys176 residue for substrate thiolation, leaving the enzyme with a dehydroalanine residue at this position as demonstrated by mass spectrometry (7). Our structure model suggests that Arg181 or Glu61, which are on the opposite side of the Cys side chain from the substrate and located 4.7 or 5.7 Å from this residue, respectively, may act as the base that abstracts the C_α proton of Cys176. Notably, both residues are highly conserved in LarE sequences from a variety of species. We conducted site-directed mutagenesis on the codons for Arg181, Glu61, and Glu200, with the latter forming two hydrogen bonds with Arg181 (Fig. 7A). Both the R181K and E200Q variants showed no activity, whereas the E61Q variant showed only a slight decrease in activity (Fig. 7B). Therefore, the sulfur transfer most likely occurs through a β-elimination reaction catalyzed by Arg181, assisted by Glu200. Although Arg residues generally are not considered as good candidates for general bases due to their high pK_a (~12), catalytic Arg residues have been suggested in enzymes from distinct families (13). A common feature of these enzymes is that the Arg is hydrogen-bonded with a carboxylate residue. Due to pK_a mismatch, it is unlikely that the Arg pK_a can be efficiently reduced by a carboxylate group, but it has been proposed that proton exchange promoted by the negatively charged carboxylate group may facilitate the Arg-involved reactions (13). Thus, while E200Q should maintain a hydrogen bond with Arg181, the completely abolished activity of this variant reinforces the opinion that the negative charge of the carboxylate is critical for activation of the catalytic Arg. Admittedly, formation of the thioester intermediate may lead to a large conformational change in the active site so that the C_α of Cys176 may be approached and deprotonated by a yet-to-be-identified residue for elimination reaction. To complete the conversion of P2CMN to P2TMN, the initial product must be relocated to a second LarE molecule and the three reaction steps must be repeated to allow for a second sulfur transfer. Transfer of the intermediate containing a single sulfur atom may be facilitated by the high-order oligomerization of LarE.

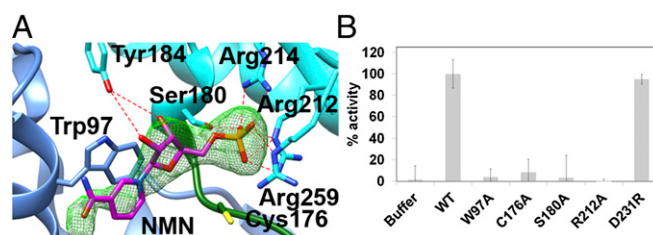


Fig. 6. Substrate analog NMN binding to the C-terminal head domain of LarE. (A) NMN-bound structure of LarE (PDB ID code 5UDR). The $F_o - F_c$ map is shown as green mesh at 2.5 σ for the model before adding the ligand. Carbon atoms of NMN are depicted in magenta. Hydrogen bonds are illustrated as red dashed lines. (B) Activity measurement of LarE variants compared with WT and with the addition of just buffer without LarE protein. The relative activity was the mean of three to nine repeats from up to three independent purifications, and the error bar represents the SD.

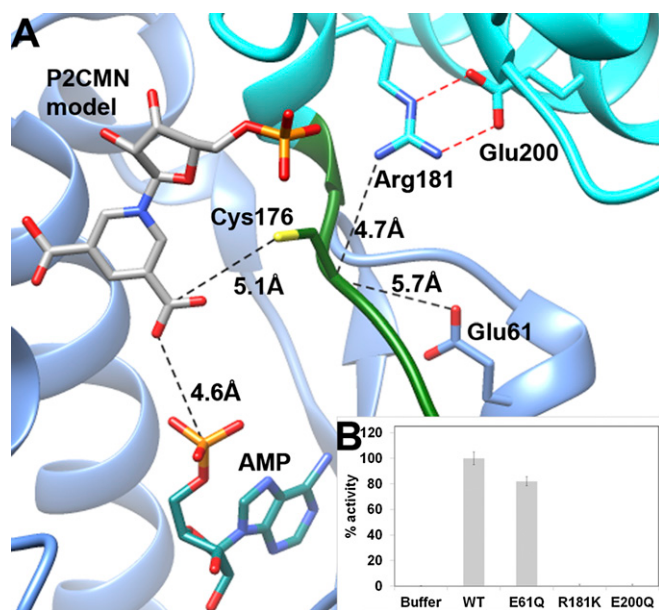


Fig. 7. Structural model of P2CMN-bound LarE. (A) Carbon atoms of P2CMN and AMP are illustrated in gray and teal, respectively; the black dashed lines represent distances between the P2CMN, AMP, and sulfur atom of Cys176, as well as the distances from the C $_{\alpha}$ of Cys176 to Glu61 and Arg181. Red dashed lines indicate hydrogen bonds between Arg181 and Glu200. (B) Activity measurement of LarE variants compared with WT and with the addition of just buffer without LarE protein. The relative activity was the mean of four repeats, and the error bar represents the SD.

Discussion

LarE, one of four proteins needed for LarA's lactate racemase activity, catalyzes a key step in the biosynthesis of the P2TMN cofactor. In this work, we solved the crystal structures of LarE from *L. plantarum*, providing structural insights into the catalytic mechanism of this enzyme (Fig. 8). ATP binds to the conserved N-terminal ATP PPase domain primarily through the signature PP-loop motif, whereas the unique C-terminal domain defines an important functional region for substrate binding and thiolation.

The N-terminal domain of LarE forms an $\alpha/\beta/\alpha$ fold containing the invariant PP-loop motif for ATP binding and hydrolysis, indicating that it belongs to the adenine nucleotide α -hydrolase-like superfamily. This superfamily can be divided into six families according to their structural features (8): N-type ATP PPase, PP-loop ATPase, phosphoadenylyl sulfate reductase, α subunit of electron transfer flavoprotein, universal stress protein-like protein, and a thiamine biosynthesis protein. Because the N-terminal domain of LarE shows remarkable similarity to the corresponding region of the ATP PPase domain of GMP synthetase, a representative of the N-type ATP PPase family, we therefore categorize LarE as a member of the N-type ATP PPase family. Detailed structural comparison of LarE with GMP synthetase reveals that the sulfur insertase has an additional highly conserved segment (residues 126–147) in the N-terminal ATP PPase domain and also coordinates the substrate at the C-terminal head domain in a markedly distinct way from GMP synthetase or any other known ATP PPase. These structural features strongly support the proposal that LarE is a new member of the N-type ATP PPase family.

The most remarkable feature making LarE stand out from the other enzymes with ATP PPase activity is the nucleophilic attack on the AMPylated substrate by the universally conserved Cys176. In the enzyme family of N-type ATP PPases, the AMPylated substrate is nucleophilically attacked by either an ammonia generated from glutamine (e.g., NAD synthetase, GMP synthetase, asparagine synthetase) or the amino group of a reactant

(e.g., β -lactam synthetase, argininosuccinate synthetase). However, LarE is very unusual in using a Cys residue of the enzyme itself as the nucleophile for attacking the AMPylated substrate. In the current structural model, the sulfur atom of Cys176 is only 5.1 Å away from the carboxylic acid group of P2CMN (Fig. 7), providing a strong structural basis for the formation of the Cys176-mediated LarE-P2CMN adduct, which has been detected in previous mass spectrometry experiments (7). Still unclear is the question of how the thiol group of Cys176 is activated to act as a nucleophile. One possibility is that the phosphate group of the substrate, which is close enough to form a hydrogen bond with the thiol, may play a role as a base, thus increasing the nucleophilicity by at least partially deprotonating the thiol group.

The LarE-P2CMN adduct decomposes by breaking the C-S bond of the Cys, a striking step that inactivates LarE by replacing the catalytic Cys with a dehydroalanine residue, limiting the enzyme to a single cycle of reaction. Under physiological conditions, the sulfur transfer reaction is most likely catalyzed by a base-catalyzed β -elimination mechanism involving deprotonation of the C $_{\alpha}$ of Cys176. Arg181 is a potential residue playing such a role. Proton exchange of Arg181 with a carboxylate residue, which would be required in order for the residue to act as a general base (13), could be realized by Glu200. The location of Arg181 on the opposite side of the thiol group from the substrate would be close to the C $_{\alpha}$ of Cys176 if a 1- to 2-Å displacement of the flexible linker harboring Cys176 were to occur. Because of the key roles of the "176CSxxSR¹⁸¹" motif, we suggest that proteins with an N-terminal ATP PPase domain and a C-terminal domain containing this motif are examples of LarE, as described here (SI Appendix, Figs. S9 and S10). In a previous bioinformatic study (1) analyzing 1,087 available bacterial and archaeal genomes, 270 (25%) contained a LarE homolog for which nine (from Lactobacillales or Bacillales) are closely related to *L. plantarum* LarE (SI Appendix, Fig. S11A). The remaining 261 homologs found in Actinobacteria, α -Proteobacteria, Clostridia, Cyanobacteria, δ -Proteobacteria, Euryarchaeota, and Spirochaetes harbor a slightly different, but highly conserved, "Cxx(S/T)R" motif, and a "CxxC" motif replacing the conserved Trp97 at the active site of *L. plantarum* LarE (SI Appendix, Fig. S11A). As many of the corresponding *larE* genes were found in an operonic structure with *larB* and *larC* (1), they most probably also encode LarE proteins performing the same overall reaction. One example is the LarE_{Ti} homolog from *Thermoanaerobacterium thermosaccharolyticum*, where the LarA *T. thermosaccharolyticum* (LarA_{Ti}) homolog has already been demonstrated to be a lactate racemase (1). LarE_{LP} and LarE_{Ti} share 38% sequence identity, and a model of LarE_{Ti} based on the *L. plantarum* structure shows an active site with three conserved cysteines instead of one (SI Appendix, Fig. S11B). It will be interesting to test whether these

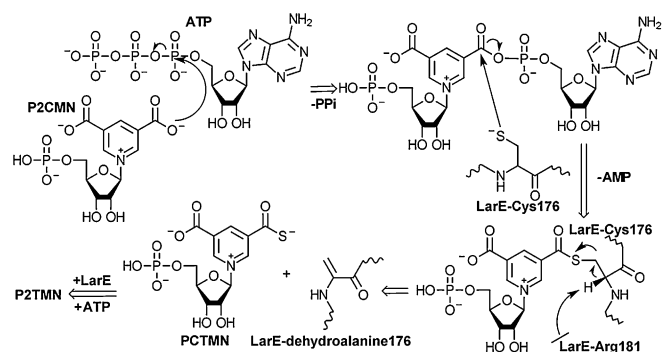


Fig. 8. Proposed LarE mechanistic reaction scheme. PCTMN, pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide.

homologs also sacrifice their catalytic cysteine or if they use an alternative sulfur donor.

Taken together, the series of LarE structures presented in this work provides a structural basis supporting a novel catalytic mechanism of LarE, a sulfur insertase assigned as a new member in the N-type ATP PPase family. LarE is therefore the second identified enzyme sacrificing its catalytic Cys after thiamine thiazole synthase of *Saccharomyces cerevisiae*, THI4p (14). Conversion of the catalytic Cys into dehydroalanine makes LarE a single-turnover enzyme, which is fundamentally different from the conversion of Cys into formylglycine by the well-studied sulfatases, where the posttranslational modification of the catalytic cysteine is absolutely required for catalysis (15). Although we have demonstrated LarE does not require other enzymes for its reaction in the in vitro assay (7), we cannot rule out that LarE might be regenerated by other enzymes in vivo. This unsolved issue and others, such as the detailed processes of sulfur transfer and substrate relocation to a second subunit for catalyzing the second sulfur insertion, warrant further investigation of this very unusual enzyme.

Materials and Methods

Genes, Plasmids, and Cloning. Plasmid pGIR076, containing *L. plantarum* *larE* along with a sequence encoding a C-terminal *Strep* II tag (ASWSHPQFEK; IBA), was constructed by digestion of plasmid pGIR072 (1) with *Nco*I and *Hind*III ligation into pBADHisA. Site-directed mutagenesis of the gene encoding *Strep* II-tagged WT LarE was performed using a QuikChange mutagenesis kit (Agilent). All of the constructs have been verified by DNA sequencing. Plasmid and primer details are provided in *SI Appendix, Table S2*.

***L. lactis* LarE Overexpression, Purification, and Crystallization.** *L. lactis* cells were transformed with pGIR072, pGIR082, or pGIR700; grown in M17 broth supplemented with glucose; and (for those containing pGIR072) used for LarE purification as previously described (1, 7). Details about these and the *E. coli* strains are provided in *SI Appendix, Table S2*. The selenomethionine-containing LarE was obtained by a method described elsewhere (16); details are provided in *SI Appendix*. The crystallization condition is shown in *SI Appendix, Table S3*.

***E. coli* Arctic-Express LarE Overexpression, Purification, and Crystallization.** Plasmid pGIR076 and pBAD-based plasmids carrying the ampicillin gene were transformed into *E. coli* Arctic-Express (DE3) cells (Agilent). LarE transformation and expression followed the Arctic-Express and pBAD protocols, and *Strep-tactin* purification was performed according to the IBA manual. Details and slight modifications are provided in *SI Appendix*. Protein concentrations were measured using the Bradford (Bio-Rad) method. To assess the oligomeric state in solution, LarE was analyzed by chromatography on a Superdex-200 increase 10/300 GL column at 4 °C (GE Healthcare, Inc.) that had been equilibrated with 100 mM Tris-HCl (pH 7.5) and 300 mM NaCl, and its elution profile was compared with the standards. Size exclusion purification was not required for the activity assays or crystal formation. The crystallization and soaking conditions are listed in detail in *SI Appendix, Table S3*.

Diffraction Data Collection, Structure Determination, and Analysis. X-ray diffraction data were collected at the Advanced Photon Source LS-CAT beamlines (21-ID-D, 21-ID-F, and 21-ID-G) and at beamlines 5.0.1 and 5.0.2 at the Berkeley Center for Structural Biology Advanced Light Source facility. Datasets were processed with xdsapp2.0 (17) or iMosflm (18), and merging and scaling were done using aimless (19). The phase of a selenomethionine-substituted crystal was solved using single-wavelength anomalous dispersion in Phenix (20) at 3.3 Å, and molecular replacement was used for all subsequent datasets. Model building and refinement were conducted in Coot (21) and Phenix (20). Statistics for the datasets are listed in *SI Appendix, Tables S4–S7*. Structure figures were created with UCSF Chimera (22). The hexamer was examined by PDBsum (23); details are provided in *SI Appendix, Fig. S2 and Table S1*. Sequence alignments and phylogenetic tree representation used Clustal (24) and Dendroscope 3 (25), and homology modeling used SWISS-MODEL (26) (Fig. 3 and *SI Appendix, Figs. S9–S11*).

LarE Activity Assay. LarE activity was assessed by measuring the lactate racemase activity of LarA_{Tr}, a LarA homolog from *T. thermosaccharolyticum*. The method was modified based on a protocol described earlier (1, 7) (details are provided in *SI Appendix*).

ACKNOWLEDGMENTS. We thank Joel Rankin for assistance with establishing LarE purification from Arctic-Express cells and developing the LarE assay, Dexin Sui for assistance with mutagenesis, and Tuo Zhang for help with crystallographic work. This work was supported by National Science Foundation Grant CHE-1516126 (to R.P.H. and J.H.).

- Desguin B, et al. (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* 5:3615.
- Desguin B, et al. (2015) METALLOPROTEINS. A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* 349:66–69.
- Xu T, Wodrich MD, Scopelliti R, Corminboeuf C, Hu X (2017) Nickel pincer model of the active site of lactate racemase involves ligand participation in hydride transfer. *Proc Natl Acad Sci USA* 114:1242–1245.
- Zhang X, Chung LW (2017) Alternative mechanistic strategy for enzyme catalysis in a Ni-dependent lactate racemase (LarA): Intermediate destabilization by the cofactor. *Chemistry* 23:3623–3630.
- Wang B, Shaik S (2017) The nickel-pincer complex in lactate racemase is an electron relay and sink that acts through proton-coupled electron transfer. *Angew Chem Int Ed Engl*, 10.1002/anie.201612065.
- Boer JL, Mulrooney SB, Hausinger RP (2014) Nickel-dependent metalloenzymes. *Arch Biochem Biophys* 544:142–152.
- Desguin B, Soumillion P, Hols P, Hausinger RP (2016) Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc Natl Acad Sci USA* 113:5598–5603.
- Andreeva A, et al. (2008) Data growth and its impact on the SCOP database: New developments. *Nucleic Acids Res* 36:D419–D425.
- Tesmer JJ, Klem TJ, Deras ML, Davison VJ, Smith JL (1996) The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. *Nat Struct Biol* 3:74–86.
- Wiederstein M, Gruber M, Frank K, Melo F, Sippl MJ (2014) Structure-based characterization of multiprotein complexes. *Structure* 22:1063–1070.
- Madej T, et al. (2014) MMDb and VAST+: Tracking structural similarities between macromolecular complexes. *Nucleic Acids Res* 42:D297–D303.
- Welin M, et al. (2013) Substrate specificity and oligomerization of human GMP synthetase. *J Mol Biol* 425:4323–4333.
- Guillén Schlippe YV, Hedstrom L (2005) A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* 433:266–278.
- Chatterjee A, et al. (2011) *Saccharomyces cerevisiae* THI4p is a suicide thiamine thiazole synthase. *Nature* 478:542–546.
- Bojarová P, Williams SJ (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: A sulfation fascination. *Curr Opin Chem Biol* 12:573–581.
- Zhang G, Mills DA, Block DE (2009) Development of chemically defined media supporting high-cell-density growth of lactococci, enterococci, and streptococci. *Appl Environ Microbiol* 75:1080–1087.
- Sparta KM, Krug M, Heinemann U, Mueller U, Weiss MS (2016) Xdsapp2.0. *J Appl Crystallogr* 49:1085–1092.
- Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67:271–281.
- Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69:1204–1214.
- Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
- Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
- de Beer TA, Berka K, Thornton JM, Laskowski RA (2014) PDBsum additions. *Nucleic Acids Res* 42:D292–D296.
- Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Huson DH, Scornavacca C (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61:1061–1067.
- Biasini M, et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:W252–W258.