Published online 3 February 2011 *Nucleic Acids Research, 2011, Vol. 39, No. 10* ***4475–4489***
doi:10.1093/nar/gkr027

Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis

Ludovic Sauguet[1], Mireille Moutiez[1], Yan Li[1], Pascal Belin[1], Jérôme Seguin[1], Marie-Hélène Le Du[2], Robert Thai[1], Cédric Masson[1], Matthieu Fonvielle[1], Jean-Luc Pernodet[3], Jean-Baptiste Charbonnier[2,*] and Muriel Gondry[1,*]

[1]CEA, IBITECS, Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), F-91191 Gif-sur-Yvette, [2]CEA, IBITECS, Service de Bioénergétique, Biologie Structurale et Mécanismes (SB[2]SM), CNRS, URA2096, F-91191 Gif-sur-Yvette and [3]Institut de Génétique et Microbiologie, Univ. Paris-Sud 11, CNRS, UMR8621, F-91405 Orsay, France

Received November 5, 2010; Revised January 6, 2011; Accepted January 12, 2011

ABSTRACT

Cyclodipeptide synthases (CDPSs) belong to a newly defined family of enzymes that use aminoacyl-tRNAs (aa-tRNAs) as substrates to synthesize the two peptide bonds of various cyclodipeptides, which are the precursors of many natural products with noteworthy biological activities. Here, we describe the crystal structure of AlbC, a CDPS from *Streptomyces noursei*. The AlbC structure consists of a monomer containing a Rossmann-fold domain. Strikingly, it is highly similar to the catalytic domain of class-I aminoacyl-tRNA synthetases (aaRSs), especially class-Ic TyrRSs and TrpRSs. AlbC contains a deep pocket, highly conserved among CDPSs. Site-directed mutagenesis studies indicate that this pocket accommodates the aminoacyl moiety of the aa-tRNA substrate in a way similar to that used by TyrRSs to recognize their tyrosine substrates. These studies also suggest that the tRNA moiety of the aa-tRNA interacts with AlbC *via* at least one patch of basic residues, which is conserved among CDPSs but not present in class-Ic aaRSs. AlbC catalyses its two-substrate reaction *via* a ping-pong mechanism with a covalent intermediate in which L-Phe is shown to be transferred from Phe-tRNA^{Phe} to an active serine. These findings provide insight into the molecular bases of the interactions between CDPSs and their aa-tRNAs substrates, and the catalytic mechanism used by CDPSs to achieve the non-ribosomal synthesis of cyclodipeptides.

INTRODUCTION

Cyclodipeptide synthases (CDPSs) constitute a family of peptide-bond forming enzymes that use aminoacyl-tRNAs (aa-tRNAs) as substrates to form various cyclodipeptides (1). They are associated with cyclodipeptide-tailoring enzymes in biosynthetic pathways dedicated to the synthesis of diketopiperazines (DKPs) (2,3), a large class of natural products mainly produced by microorganisms. The CDPS family includes at least eight identified members found in various bacterial phyla. All members are similar in size (239–289 residues). They display a moderate sequence similarity, but conserved residues are found in 13 positions.

CDPSs use previously activated amino acids in the form of aa-tRNAs and catalyse the formation of the peptide bonds of the DKP scaffold (1) (Figure 1). Only two other enzymes families also use loaded tRNA as substrates to form peptide bonds (4–7): Fem aminoacyl transferases form cross-bridges in peptidoglycan biosynthesis (8), and

*To whom correspondence should be addressed. Tel: +33 169 08 76 47; Fax: +33 169 08 90 71; Email: muriel.gondry@cea.fr
Correpondence may also be addressed to Jean-Baptiste Charbonnier. Tel: +33 169 08 76 77; Fax: +33 169 08 47 12; Email: jb.charbonnier@cea.fr
Present address:
Ludovic Sauguet, Institut Pasteur, CNRS, URA2185, Unité de Dynamique Structurale des Macromolécules, 25, rue du Dr Roux, F-75015 Paris, France.

The authors wish it to be known that, in their opinion, the second and third authors should be regarded as joint First Authors.

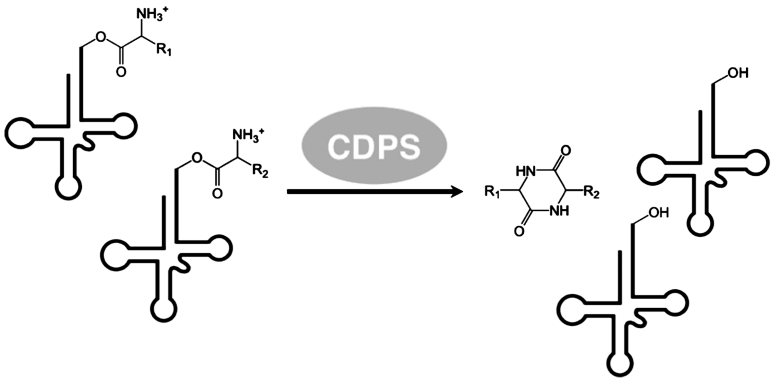

Figure 1. The reaction catalysed by CDPSs. CDPSs use activated amino acids, in the form of aa-tRNAs, as substrates to catalyse the formation of cyclodipeptides, also known as DKPs.

aa-tRNA protein transferases modify the N-terminal extremity of proteins for subsequent degradation (9). Fem aa-transferases and aa-tRNA protein transferases thus catalyse similar chemical reactions, in which the amino acid loaded on an aa-tRNA donor is transferred to the N-terminal amino group of a peptide or a protein acceptor. They both also belong to the same structural GCN5-related *N*-acetyltransferase (GNAT) protein superfamily (9–11). CDPSs differ from Fem aa-transferases and aa-tRNA protein transferases in that they use two aa-tRNAs to form two peptide bonds, whereas transferases use only one aa-tRNA to form only one peptide bond. CDPSs and transferases have also different structural organizations. Indeed, the crystal structure of a CDPS, namely Rv2275 of *Mycobacterium tuberculosis*, was recently determined (12), revealing that Rv2275 is a homodimer, each monomer being structurally related to class-Ic aminoacyl-tRNA synthetases (aaRSs) (13–16). Rv2275 uses two tyrosyl-tRNA^{Tyr} (Tyr-tRNA^{Tyr}) as substrates to produce cyclo(L-Tyr–L-Tyr) (cYY) (1).

Here, we solved at 1.9 Å resolution the crystal structure of *Streptomyces noursei* AlbC, a CDPS that uses mainly phenylalanyl-tRNA^{Phe} (Phe-tRNA^{Phe}) and leucyl-tRNA^{Leu} (Leu-tRNA^{Leu}) as substrates to synthesize cyclo(L-Phe–L-Leu) (cFL). The AlbC structure consists of a monomer. It contains a Rossmann-fold domain and shares similarity with the catalytic domain of class-Ic aaRSs, as previously reported for Rv2275 (12). The reaction catalysed by Rv2275 was suggested to proceed *via* a ping-pong kinetic mechanism with a covalent intermediate in which L-Tyr is transferred from Tyr-tRNA^{Tyr} to an active serine, but such a transfer was not conclusively demonstrated (12). Here, we showed that the serine residue is covalently bound to the aminoacyl transferred during the catalytic cycle, demonstrating that CDPSs use a mechanism involving a covalent aminoacyl-enzyme intermediate. Moreover, we also showed that the interaction between AlbC and its aa-tRNA substrates is driven by the binding of the substrate aminoacyl moiety in a deep pocket—similar to the tyrosine-binding pocket of TyrRSs—and the interaction of the tRNA moiety with at least one patch of basic residues. This provides clues about the molecular bases of the interactions between CDPSs and their aa-tRNAs substrates.

MATERIALS AND METHODS

Mutagenesis and purification of AlbC and variants

The expression plasmid encoding AlbC was constructed from the vector pQE60 (Qiagen). The resulting plasmid encodes the full-length AlbC (residues 1–239) with the C-terminal addition of two residues (RS) followed by a His_6 tag (1). The desired variants of AlbC were created *via* PCR mutagenesis according to the QuikChange method (Stratagene). Thirty-two variants were constructed (28 variants with a single mutation, and four variants with a double mutation). For 12 variants, the mutations are localized in the putative catalytic pocket (L33Y/L185D, G35A, S37A, S37C, Y178A, Y178F, E182A, E182Q, E182D, L200N, Y202A, Y202F), and for the remaining variants, the mutations are outside this pocket (H31A, W54A, F59A, G79A, R87A, K90A, R91A, K94A, R97A, R98A, R99A, R98A/R99A, R101A/R102A, S120A, Y128A, C149A, P184A, R214A/R215A, R220A, R231A). The C-terminal His_6-tagged proteins were produced and purified as described previously (1,17). Protein production was followed by SDS–PAGE analyses or western blotting according to standard protocols, and His-tagged proteins were visualized using alkaline phosphatase-conjugated anti-His antibody and BCIP/NBT solution (Sigma).

Crystallization and structure determination

Small crystals of AlbC ($80 \times 10 \times 10\,\mu m^3$) were grown by vapour diffusion in sitting drops at 17°C by mixing 2 µl of protein sample at 4 mg/ml with 2 µl reservoir solution. First crystalline precipitates were obtained by a reverse screening approach (18). By screening different pHs, crystals were obtained with reservoir containing 1.7 M Na/K phosphate at pH 6.0. A screening of different additives showed that the presence of reducing agents improves crystals size and the addition of a large excess of dithiothreitol (DTT) (100 mM) gave the best results. These crystals were cryocooled with different cryoprotectants, but they did not diffract better than 3.5 Å. Larger

crystals ($300 \times 40 \times 40\,\mu m^3$) were obtained by a batch method under silicon oil and microseeding from solutions of crushed AlbC crystals (19). The native crystals were cryocooled with 35% glycerol and diffracted up to 1.9 Å on Proxima1 (Soleil, France) and ID29 (ESRF, France) beamlines. The crystal structure was determined with selenomethionine-substituted AlbC by MAD (multi-wavelength anomalous dispersion) using phase information derived from anomalous scattering data collected at the Se K-edge. The crystals belong to the space group I4 with a single copy of the protein in the asymmetric unit. The positions of four selenium atoms were determined by SHELXD (20). Experimental phases were calculated in SHARP (21), and the tracing of an initial model in COOT that was subsequently completed by automatic building in ArpwARP (22) using the AlbC native data set.

The final model was refined in REFMAC5 (23) to R/R_{free} values of 0.186/0.236 at 1.9 Å resolution with one TLS parameter for the protein chain. In the final model, residues 1–27, 217, the eight C-terminal residues [$RS(His)_6$] introduced during cloning, and the side chain of residue R214 were not visible in the electron density map, they were not included in the crystallographic model. The MolProbity score (24) for the refined model is 1.91, in the 75th percentile of structures refined at comparable resolution. Almost all residues (97.6%) were in favoured regions of the Ramachandran plot, and 100.0% were in allowed regions.

In the final electron density, remaining peaks that could not be attributed to water were modelled with three molecules of phosphate and a molecule of cyclo-dithiothreitol (cDTT), which were present in the crystallization solution as a precipitant agent and an additive, respectively (see above, cDTT comes from the oxidized fraction of the DTT used; Supplementary Figure S1). This observation suggests that the large quantity of DTT used during crystal optimisation served both to maintain AlbC under its reduced state, and to feed crystals with cDTT present in equilibrium with reduced DTT. Figures were prepared with PyMOL (PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). The programme Superpose (25) was used for structural alignment and Areaimol (26) for accessible surface area measurements. The coordinates and structure factors of *S. noursei* AlbC have been deposited in the Protein Data Bank (PDB) (accession code 3OQV).

Cyclodipeptide-synthesizing activity assays

Assays with supernatants of cultures of *Escherichia coli* cells expressing the genetic constructs encoding AlbC wild-type or an AlbC variant were analysed by LC-MS/MS as described previously (1). The *in vivo* activities are reported as percentage of the wild-type activity. Error bars represent the standard deviation for at least two independent experiments.

Detection and identification of the covalent aminoacyl-enzyme intermediate

To detect the covalent aminoacyl-enzyme intermediate, purified AlbC or relevant variants were incubated with [^{3}H]Phe-tRNA^{Phe}, the proteins were separated by SDS–PAGE before being transferred on a polyvinylidene fluoride (PVDF) membrane, which was analysed with a Beta-Imager^{TM} 2000 from Biospace (Paris, France). Protein concentrations were quantified by UV spectroscopy. The tritiated substrate was obtained as follows: a 100 µl reaction mixture containing 3.2 µM *E. coli* tRNA^{Phe} (Sigma), 0.32 mM Phe, 0.9 µM [^{3}H]Phe (Phenylalanine, L [2,3,4,5,6-^{3}H], Perkin-Elmer, 110 Ci/mmol), 2 mM ATP, 1 mM DTT, 20 mM $MgCl_2$, 50 mM KCl, 1 µM *E. coli* PheRS ($PheRS_{Ec}$) in 50 mM Hepes–KOH, pH 7.0 was incubated at 30°C for 30 min. Then, AlbC or AlbC variant was added at a final concentration of 1 µM. After 20 s incubation, the reaction was quenched by the addition of 5 µl Laemmli loading buffer [final concentration, 0.1% (w/v) bromophenol blue, 2% (w/v) SDS, 10% (v/v) glycerol, 50 mM Tris–HCl, pH 6.8 and 100 mM DTT] followed by heating the samples 2 min at 50°C. These samples were immediately resolved by SDS–PAGE electrophoresis followed by blotting. Blotting of proteins onto a PVDF membrane was achieved using a semi-dry transfer blot apparatus (Bio-Rad). After transfer (400 mA, 30 min), the membrane was dried before radioactivity analysis with the Beta-Imager^{TM} 2000. This apparatus allows an absolute counting of the tritium β particles, with a detection threshold of 0.007 cpm/mm^2 for tritium.

To identify the residue acylated and the nature of the acylation, the aminoacyl-enzyme intermediate was produced by a direct assay using purified Phe-tRNA^{Phe}. The aminoacylation reaction was performed as described above. Charged Phe-tRNA^{Phe} was recovered by phenol/chloroform extraction and ethanol precipitation [the loading of tRNA^{Phe} with Phe was estimated to be 70%, as previously described (1)]. Two micrograms of Phe-tRNA^{Phe} were mixed with 3 µg of AlbC (or variant) in 20 µl of the reaction buffer containing 50 mM Hepes–KOH pH 7.0 and 50 mM KCl. After 20 s of incubation, a 5 µl-aliquot was removed and submitted to trypsin digestion: the withdrawn aliquots were diluted up to 10 µl in 50 mM NH_4AcO_2, pH 8.0 and 1 µl of trypsin (400 ng/µl porcine trypsin, Promega) was added. Digestion was performed at 50°C up to 10 min. Sample digests were then acidified by 1.25% trifluoroacetic acid (TFA) in water, desalted and concentrated by C18 ziptip and spotted with 4-HCCA matrix solution (10 mg/ml in 49.8/49.8/0.4: CH_3CN/H_2O/TFA: v/v/v) on MALDI plate. MS and MS/MS spectra were registered using a 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Foster City, USA) in a positive reflectron mode. Each MS spectrum was the result of 1000 laser shots and calibration was applied by using trypsin autolysis fragments as internal standards. MS/MS analyses were performed under post-source decay (PSD) mode. Each MS/MS spectrum was the result of 10 000 shots. The sequence of the trypsic fragments was identified by using Data Explorer® processing software (Version 4.9, Applied Biosystems) to assign the potential chemical modifications of the serine or cysteine residues.

RESULTS

Crystal structure of *S. noursei* AlbC

We produced AlbC in *E. coli*, purified it to homogeneity (1) and crystallized the purified AlbC. We solved the crystal structure of AlbC by the multi-wavelength anomalous dispersion (MAD) method using selenomethionine-labelled protein. The native structure was refined to a final R-factor/R-free of 18.6/23.6% at a resolution of 1.9 Å (Table 1). The crystal structure of AlbC is a monomer that presents a compact α/β structure (Figure 2). It is composed of a central β-sheet with five parallel β strands (β3–β7) surrounded by ten α helices (α2–α9) (Figure 2A). The AlbC structure contains a Rossmann-fold domain (27) formed by strands β3–β5, helix α2 and the C-terminal part of helix α4. This domain is followed by a helical region composed of three main helices (α5–α7) that pack against the Rossmann fold. The structure ends with two β strands (β6 and β7) that extend the Rossmann β-sheet and two small connecting helices (α8 and α9). AlbC residues 1–27 were not visible in the electron density map. Crystal packing analysis showed the presence of large cavities near the residue 28 that are compatible with the presence of a flexible N-terminal extension (Supplementary Figure S2). Some regions of AlbC were visible in the electron density map, but they present higher B-factors than the rest of the protein (Supplementary Figure S3). These regions contain the residues 38–46 in loop β3–α2, the residues 105–111 in loop α4–β5, the residues 215–220 in loop α8–β7, and the C-terminal residues 236–239.

A striking feature of the structure of AlbC is the presence of a deep (14 Å long) and narrow (diameter ranging from 4 to 7 Å) surface-accessible pocket, which is delineated by helices α6, α7, α8 and β-strands β3, β4 and β6. The buried part of the pocket in AlbC is bordered mostly by side chains of nine hydrophobic residues: L33, V65, L119, L185, F186 and L200 at its deepest point and V67, M152 and I208 on the edges (Figure 2B). Four additional residues contribute to define the buried part of the pocket, G35, E182 and the main chains of Y202 and H203. Near the residue G35, the pocket presents a small bulge bordered by the hydrophobic residues I34, F43, L51, C201 and V230. Residues P207 and V156, the carboxylate moiety of E182 and the side chains of four polar amino acids—S37, N40, Y178, T206—define the entrance of the pocket (Figure 2B). Additional electron density was attributed to a molecule of phosphate and to a molecule of cDTT, which are located at the entrance and at the bottom, respectively, of the pocket (Figure 2B; 'Materials and Methods' section). The phosphate molecule, probably present as a $H_2PO_4^-$ ion at the pH used for crystallization (pH = 6), is within hydrogen bond distances of residues Y178 and E182, and to two water molecules. The cDTT is surrounded by hydrophobic residues in the most buried part of the pocket and its oxygens are within hydrogen bond distances of phosphate and water molecules.

Table 1. Data collection, phasing and refinement statistics

	Native	SeMet Crystal 1			SeMet Crystal 2
Data collection					
Space group	I4	I4			I4
Cell dimensions					
a, *b*, *c* (Å)	97.2, 97.2, 45.5	97.9, 97.9, 45.5			98.0, 98.0, 45.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0			90.0, 90.0, 90.0
		Peak	Inflection	Remote	Peak
Wavelength	0.9793	0.9792	0.9794	0.9757	0.9792
Resolution (Å)	48.6–1.90 (2.00–1.90)	69.0–2.70 (2.70–2.85)	69.0–3.00 (3.16–3.00)	69.0–3.50 (3.69–3.50)	69.0–3.00 (3.16–3.00)
R_{sym} or R_{merge}	8.70 (61.5)	8.10 (28.6)	9.80 (36.6)	9.80 (24.2)	9.10 (40.6)
$I/\sigma I$	13.0 (3.10)	13.8 (4.40)	11.1 (3.50)	11.4 (5.50)	12.9 (3.80)
Completeness (%)	99.7 (100.0)	99.3 (98.9)	99.1 (99.8)	98.5 (99.7)	99.9 (100.0)
Redundancy	7.4 (6.9)	4.0 (3.8)	3.7 (3.9)	3.4 (3.6)	5.6 (5.7)
Refinement					
Resolution (Å)	1.90				
No. reflections	15963				
R_{work}/R_{free}	18.6/23.6				
No. atoms					
Protein	1621				
Ligand/ion	23				
Water	107				
B-factors (Å^2)					
Protein	29.3				
Ligand/ion	55.8				
Water	36.2				
Rmsd					
Bond lengths (Å)	0.006				
Bond angles (°)	0.793				

Values in parentheses are for highest resolution shell.

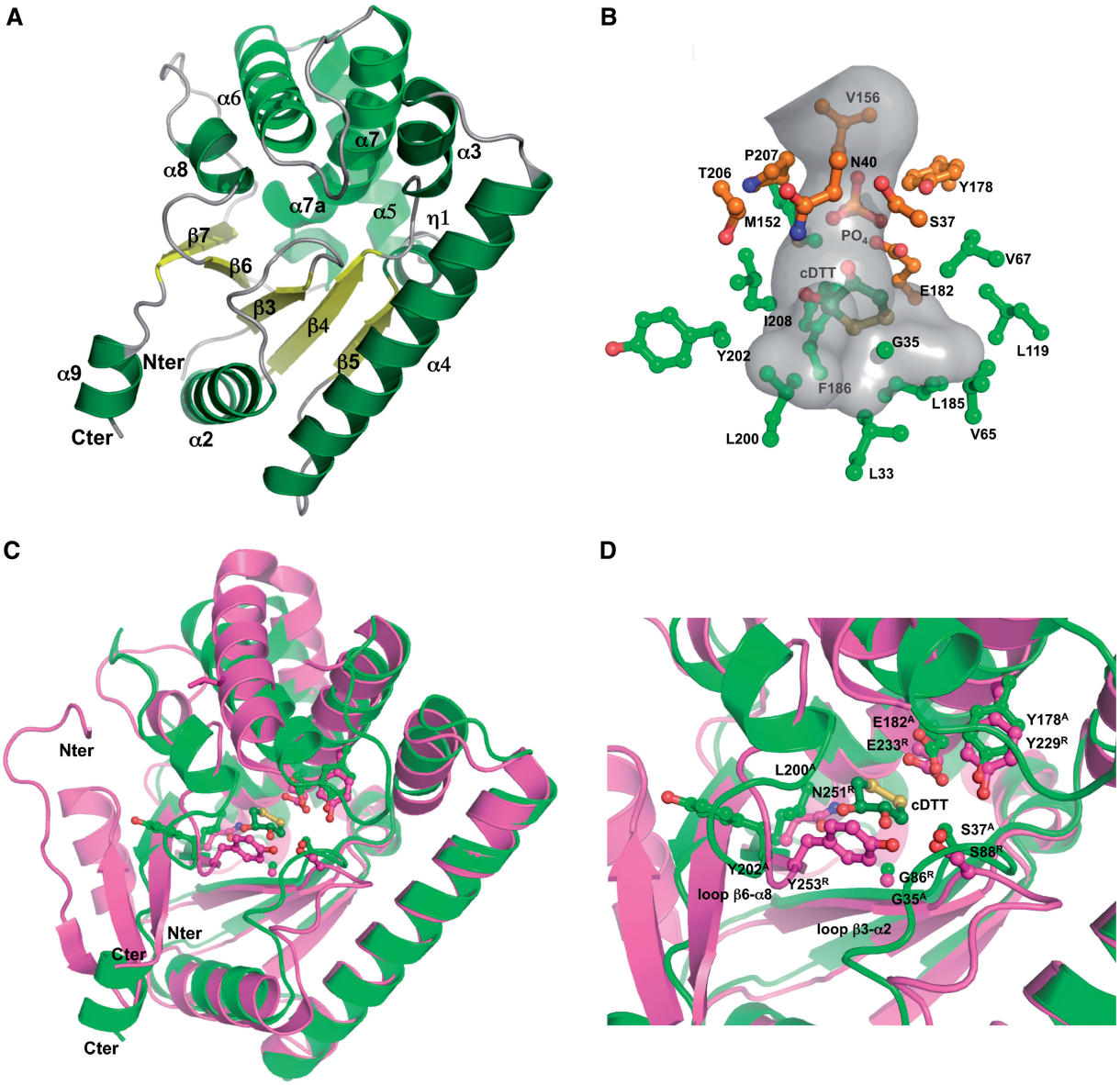

Figure 2. Crystal structure of *S. noursei* AlbC and comparison with that of *M. tuberculosis* Rv2275. (**A**) Overall structure of AlbC (region 28–236) in cartoon mode, with α helices, β-strands, and loops coloured in green, yellow and grey, respectively. AlbC secondary-structure elements are numbered according to Rv2275 numbering (12). The Rossmann-fold domain is constituted by helix α2, the C-terminal part of helix 4 and the β-sheet (β3–β5) that is extended by two additional strands β6 and β7. A helical region (in light green), composed of helices α5, α6 and α7, packs against the Rossmann fold. (**B**) Structure of the AlbC surface-accessible pocket. Residues that delineate the AlbC pocket are shown in ball and stick representation. The most buried part of the pocket is mainly constituted of hydrophobic residues (coloured in green). The entrance of the pocket is mainly delineated by polar residues (coloured in orange). The boundaries of the pocket, calculated using the program CAVER (58), are defined by a semi-transparent grey surface. The molecules of phosphate and cDTT modelled in the electron density are shown in orange and dark green, respectively. (**C**) Superimposition of the structures of AlbC and Rv2275 in cartoon representation. AlbC and Rv2275 are coloured in green and magenta, respectively. Both proteins superpose well. Most significant structural differences concern N- and C-terminus regions, and also relative positions of helices α5, α6 and α8. (**D**) Superimposition of the surface-accessible pockets of AlbC and Rv2275, with an orientation similar to that shown in Figure 2C. Residues represented in ball and stick are essential for CDPS activity. AlbC and Rv2275 residues are indicated by A and R letters, respectively. The superimposition highlights the structural deviations between loops β3–α2 and β6–α2, and residues $Y202^{A}$ and $Y253^{R}$ located at the entrance of the active sites.

The three-dimensional (3D) superimposition of AlbC and Rv2275 (PDB id: 1j1u) [root mean square deviation (rmsd) value of 2.27 Å over 192 Cα] revealed that the two enzymes superimpose well despite sequence divergence (AlbC shares 25.9 % identity and 40.6 % similarity with Rv2275; Figure 2C). Almost all secondary-structure elements superimpose well, in particular the Rossmann-fold domain and the two additional strands β6 and β7. However, there are some major differences between the two enzymes. First, the region 50–77 of Rv2275 is structured (strands β1, β2 and helix α1) while the corresponding region of AlbC (1–27) is not observed in the

electron density maps. At opposite, two regions of AlbC, the loop α6–α7 and the C-terminal helix α9 are structured in AlbC, and the corresponding regions are not observed for Rv2275. Second, the axes of helices α5, α6 and α8 are tilted by some degrees, so that residues located at these helices extremities superimpose poorly. Finally, important deviations are observed between loops, in particular loops β3–α2, α4–β5, β6–α8 and α8–β7.

Both enzymes possess a similar surface-accessible pocket that contains six residues conserved in the CDPS family. Four of these residues, G35 (AlbC numbering) belonging to the strand β3, and Y178, E182 and P184 belonging to the helix α7, are remarkably well superimposed. The two other conserved residues S37 and Y202, respectively, belong to loops β3–α2 and β6–α8 that have different conformations as mentioned above. In particular, the side chain of Y202 adopts a conformation directed towards solvent in AlbC whereas it is directed towards the active site entrance and residue S37 in Rv2275 (Figure 2D).

Structural similarity of AlbC with the catalytic domain of the class-I aminoacyl-tRNA synthetases

We used the Dali server (28) to compare the structure of AlbC to protein structures in the PDB. Due to the Rossmann-fold domain, AlbC was found to share structural similarity with more than 300 proteins with a high *Z*-score ($Z > 4$) (Supplementary Table S1). Most of these proteins are class-I aaRSs, enzymes that catalyse the activation of amino acids and their transfer to cognate tRNAs to form aa-tRNAs. The 120 solutions with the highest *Z*-scores, from 10.7 to 7.3, were all obtained with the closely related class-Ic TyrRSs and TrpRSs, as previously reported for Rv2275 (12). The greatest similarity is with the archaeal *Archaeglobus fulgidus* (PDB id: 2cyb) and *Methanococcus jannaschii* TyrRSs ($TyrRS_{Mj}$) (PDB id: 1zh6, 1j1u), and with the eukaryotic *Entamoeba histolytica* TrpRS ($TrpRS_{Eh}$) (PDB id: 3hzr) (Supplementary Table S1). The 3D superimposition of AlbC and $TyrRS_{Mj}$ (rmsd value of 3.5 Å over 160 Cα) (Figure 3) or $TrpRS_{Eh}$ (rmsd value of 3.0 Å over 163 Cα) (Supplementary Figure S4) revealed that AlbC superimposes well with the N-terminal catalytic domains of the two aaRSs, despite low-sequence similarity (AlbC shares 15.5 and 15.8% similarity with the catalytic domains of $TyrRS_{Mj}$ and $TrpRS_{Eh}$, respectively). Note that these enzymes possess C-terminal tRNA-binding domains whereas AlbC does not. The catalytic domains of each of the two aaRSs are composed of a Rossmann-fold domain and a connective-polypeptide 1 (CP1) domain that are both found in AlbC. The overall position of the two domains is conserved between AlbC and the two aaRSs and is mediated by substantial conservation of residues at the interface between the two domains. Most secondary-structure elements are well conserved, and the central β-sheets are well superimposed, but the helices that pack against these sheets present slight differences in length and positioning (Figure 3; Supplementary Figure S4). Moreover, the pocket in AlbC and the aminoacyl binding pockets in the two aaRSs are positioned similarly relative to the

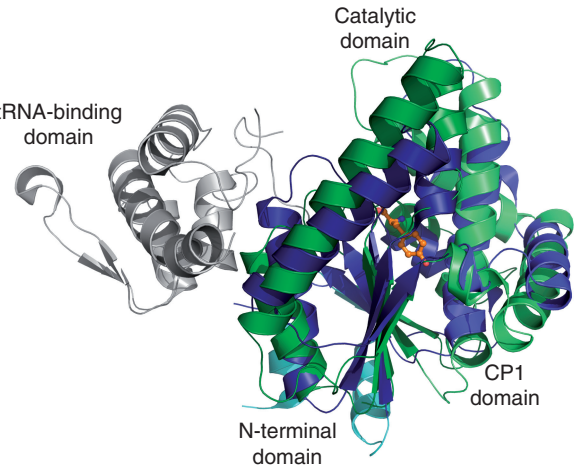

Figure 3. Superimposition of the structures of AlbC and $TyrRS_{Mj}$ in complex with its L-tyrosine substrate (PDB id: 1j1u). The two enzymes are shown in cartoon mode, and L-tyrosine is in ball and stick coloured in orange. The Rossmann-fold and the CP1 domains of $TyrRS_{Mj}$ are coloured in dark and light blue, respectively, and the corresponding domains of AlbC are coloured in dark and light green, respectively. The Rossmann-fold and CP1 domains of the two proteins have similar rmsd values (3.25 Å over 96 Cα and 3.26 Å over 53 Cα, respectively). $TyrRS_{Mj}$ possesses two additional regions not present in AlbC: a C-terminal domain involved in tRNA-binding and anticodon recognition coloured in grey, and a short N-terminal region coloured in light blue.

β-sheet forming part of the Rossmann-fold domain (29–33).

However, there are some major differences between AlbC and the two aaRSs. The hydrophobic regions of the CP1 domains of TyrRS or TrpRS involved in the homodimerization of these enzymes (34,35) are not found in AlbC. In addition, when we mimicked this association for two AlbC monomers by superimposing them on the crystal structure of $TyrRS_{Mj}$, we did not find any complementary dimer interface, and we observed major steric hindrances between helices α5, α6 and α7 of each AlbC monomer (Supplementary Figure S5A). However, the crystal structure of Rv2275 was obtained as a homodimer. But, the Rv2275 residues that make up the interface belong to other secondary-structural elements than those involved in the homodimerization of the two aaRSs (12). Noted that these residues are not conserved in CDPSs (Figure 4 and Supplementary Figure S5B). All these observations are consistent with AlbC being found as a monomer both in solution (1) and in crystal forms. Another major difference between CDPSs and class-I aaRSs is that CDPSs do not have the two consensus motifs that are conserved in class-I aaRSs and are involved in ATP binding (36–38) (Figure 4).

The pocket of AlbC, a binding site for the aminoacyl moiety of the substrate

Using 3D superimposition, we compared the pocket in AlbC to the conserved amino acid-binding pockets in $TyrRS_{Mj}$ and $TrpRS_{Eh}$ (29–33). For example, the AlbC pocket was compared to that of the $TyrRS_{Mj}$ complexed with its tyrosine substrate (Figure 5), which forms a

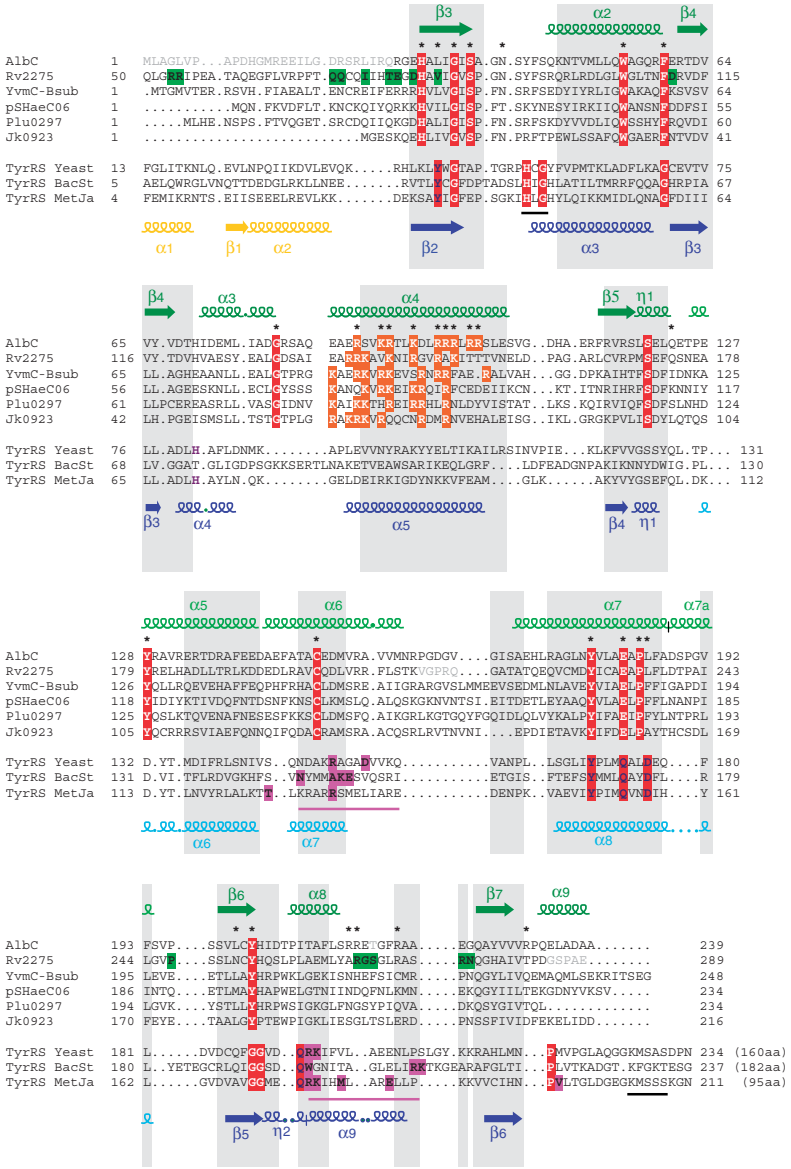

Figure 4. Structure-based sequence alignment of six CDPSs and three TyrRSs from the three primary kingdoms. The secondary-structural elements of AlbC and the $TyrRS_{Mj}$ are indicated above and below the alignment, respectively. The Rossmann-fold and the CP1 domains of AlbC are shown in dark and light green, respectively. The N-terminal region, the Rossmann-fold domain, and the CP1 domain of the $TyrRS_{Mj}$ are shown in yellow, dark and light blue, respectively. The core regions of AlbC and the $TyrRS_{Mj}$ displaying greatest structural similarity (rmsd value of 2.36 Å over 116 Cα) are shown with a gray background. The residues conserved among CDPSs, and those conserved among TyrRSs are indicated by a red background. Black stars above the AlbC sequence indicate the residues that were modified by site-directed mutagenesis in this study. The basic residues of CDPSs, which define a patch of positively charged residues belonging to the helix α4, are indicated by an orange background. The residues that are disordered in the crystal structure of AlbC and Rv2275 are shown in grey. The residues of TyrRSs previously shown to be involved in tyrosine recognition are indicated in dark blue; the position of the two clusters involved in acceptor stem recognition (41) are indicated by pink lines below the TyrRSs sequences, and the residues involved in this recognition are indicated by a pink background. Black lines below the TyrRSs sequences indicate the class I signature motifs, HIGH and KMSKS, which are involved in ATP binding. Rv2275 residues that are involved in the interface of the dimer are shown with a green background. Protein sequence alignments were done according to structural alignments for AlbC and TyrRSs and were corrected manually for CDPS based on previously described alignments (1). Two CDPSs are not mentioned, namely YvmC from *Bacillus thuringiensis* and YvmC from *B. licheniformis*, because they are very close to YvmC from *B. subtilis*. Abbreviations: Bsub, *B. subtilis*; BacSt, *Geobacillus stearothermophilus*; MetJa, *M. jannaschii*.

hydrogen bond network with the conserved residues Y32, Y151, Q155, D158 and Q173, and interacts with the conserved residue G34 (29–31) (PDB id: 1j1u) (Figure 5A). The two pockets are positioned similarly relative to the Rossmann-fold. Each pocket contains some residues that are highly conserved in the corresponding family and that are remarkably well superimposed: the AlbC residues G35, Y178 and E182 superimpose with the $TyrRS_{Mj}$ residues G34, Y151 and Q155, the rmsd values on all atoms being 1.30, 1.29 and 1.95 Å, respectively (Figures 4 and 5B). These structural similarities suggest that the pocket of AlbC could accommodate the aminoacyl

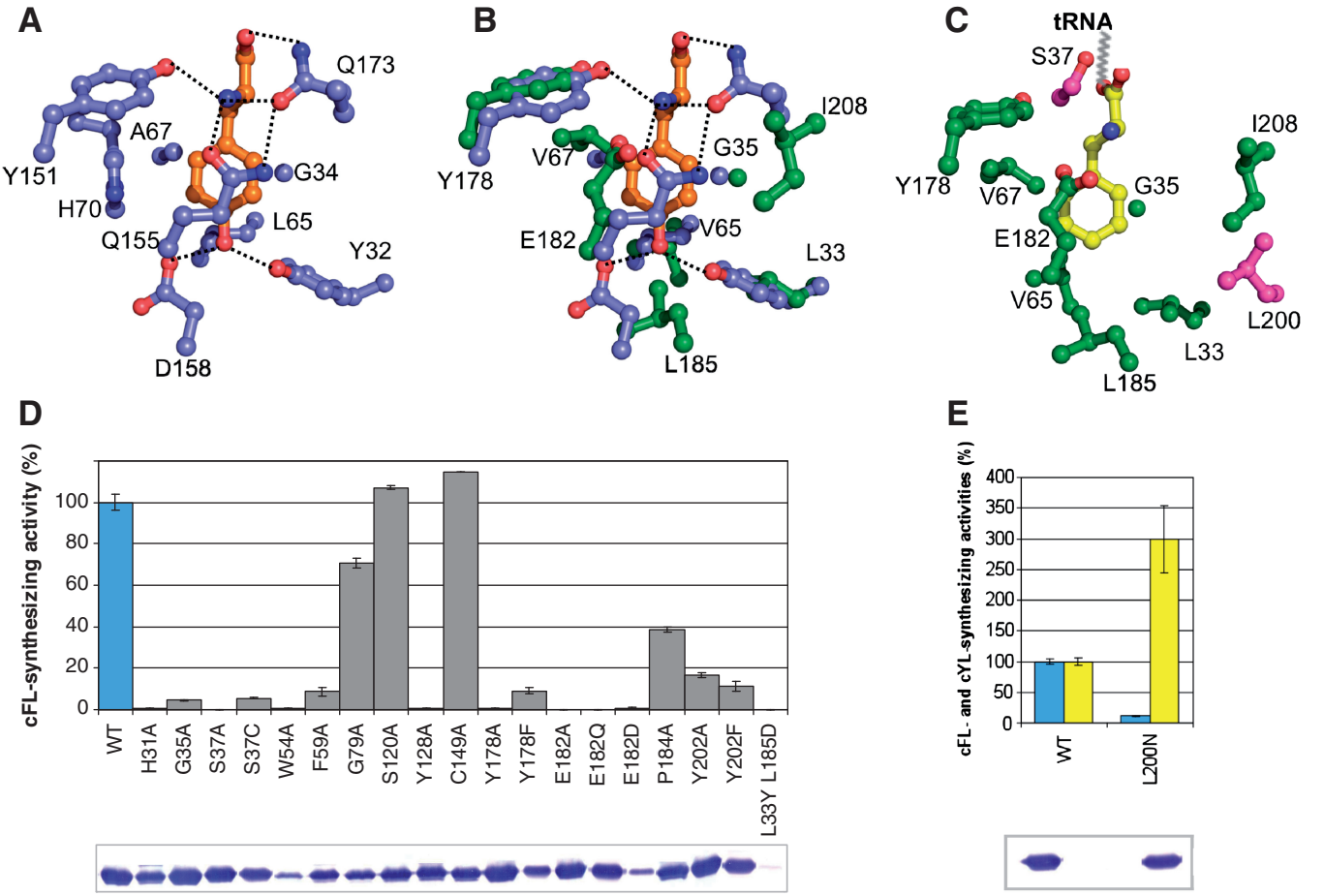

Figure 5. Comparison of the AlbC and $TyrRS_{Mj}$ pockets. (**A**) The tyrosine-binding pocket of $TyrRS_{Mj}$ in complex with its L-tyrosine substrate. TyrRS residues and L-tyrosine are represented in ball and stick, and coloured in blue and orange, respectively, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. (**B**) Superimposition of the TyrRS residues (A) with the corresponding AlbC residues. AlbC residues are represented in ball and stick, and coloured in green, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. Residues lining the two pockets superimpose with an rmsd value of 1.49 Å over 32 main chain atoms. (**C**) The AlbC pocket in the presence of the phenylalanyl moiety of a $Phe\text{-}tRNA^{Phe}$ substrate. The phenylalanyl moiety position is mimicked on that of the L-tyrosine in the pocket of $TyrRS_{Mj}$, as shown in (A) and (B), except that the hydroxyl group has been removed. The conserved residue S37 and the residue L200, which have no corresponding residues in TyrRS, are shown in pink, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. (**D** and **E**) Site-directed mutagenesis study of AlbC. Cyclodipeptide-synthesizing activities are shown with error bars. The corresponding western blots indicating amounts of the proteins are also shown. (D) cFL-synthesizing activities of the wild-type AlbC (in blue) and of each of the variants in which a residue conserved among CDPSs is substituted (in grey). (E) cFL- and cYL-synthesizing activities (in blue and yellow, respectively) of the wild-type AlbC and the variant L200N. The wild-type AlbC synthesizes $33.5 \pm 2.5\,mg\,l^{-1}$ of cFL and $4.9 \pm 0.5\,mg\,l^{-1}$ of cYL (Supplementary Figure S6).

moiety of an aa-tRNA substrate in the same way that TyrRSs bind their tyrosine substrates. To test this possibility, we independently substituted each residue in AlbC suspected to interact with the aminoacyl moiety of the substrate by alanine and, where relevant, other amino acids. We produced the resulting variants in *E. coli* and compared their *in vivo* synthesis of cFL, the major cyclodipeptide produced by AlbC, to that of the wild-type enzyme (1). The substitution of Y178 with alanine resulted in an almost inactive variant (Y178A), and its substitution with phenylalanine gave a poorly active variant (Y178F), despite being produced in amounts similar to that of the wild-type enzyme (Figure 5D). The variants E182A and E182Q were inactive, and the variant E182D had a weak but detectable activity although its production was strongly affected. The variant G35A was poorly active despite being abundantly produced (Figure 5D). Thus, the hydroxyl group of Y178 and the carboxylate group of E182 are essential for AlbC activity, suggesting that these groups may mediate hydrogen bonds with the main chain amino group of the substrate, in a way similar to what is described for $TyrRS_{Mj}$. The conserved residue G35 is also essential for the accommodation of the CDPS substrate. The conserved residues Y32 and D158 in $TyrRS_{Mj}$ form hydrogen bonds with the hydroxyl group of the tyrosine substrate (Figure 5A), and superimpose with the hydrophobic residues L33 and L185, respectively, in AlbC (Figure 5B). This is consistent with hydrophobic phenylalanyl or leucyl moieties in the AlbC substrates. To try to accommodate a $tyrosyl\text{-}tRNA^{Tyr}$ substrate in the AlbC pocket, we attempted to construct an L33Y/L185D AlbC variant, but the genetic construct was not expressed (Figure 5D).

All these data are consistent with the pocket of AlbC being the binding site of the aminoacyl moiety of an aa-tRNA, but do not prove it. We therefore replaced a residue in the AlbC pocket suspected to interact with the aminoacyl moiety of the substrate to see if this may modify the enzyme's specificity. The sequence alignment of CDPSs (Figure 4), and the specificity determined for each characterized CDPS (1), led us to focus on Rv2275 from *M. tuberculosis* that mainly synthesizes cyclo(L-Tyr–L-Tyr) (cYY) (1,3). The residue L200 located near the bottom of the AlbC pocket (Figure 5C) corresponds to an asparagine in Rv2275 (Figure 4). This asparagine residue may form a hydrogen bond with the hydroxyl group of the tyrosyl moiety of a tyrosyl-tRNA^{Tyr} substrate. We constructed the variant L200N, and tested its *in vivo* specificity: it synthesized mainly cYL instead of cFL and thus showed some of the specificity of Rv2275 (Figure 5E, Supplementary Figure S6). This indicates that the substitution of one residue located in the AlbC pocket is sufficient to change the nature of one of the aminoacyl residue incorporated into the cyclodipeptides it produces. This also demonstrates that the AlbC pocket is the binding site of the phenylalanyl moiety of the Phe-tRNA^{Phe} substrate. But, this result does not provide data on the binding of the leucyl moiety of the other substrate, which is required to synthesize a molecule of cFL.

Formation of a covalent intermediate during the catalytic cycle of AlbC

The superimposition of the pockets of AlbC and TyrRS$_{Mj}$ revealed significant similarities (see above), but also intriguing differences. The strictly conserved residue S37 of AlbC has no corresponding conserved residue in class Ic-aaRSs (Figures 4 and 5B). The position of the phenylalanyl moiety of a Phe-tRNA^{Phe} in the AlbC pocket, as predicted by analogy with the tyrosine substrate in the TyrRS$_{Mj}$ pocket (Figure 5C), indicates that the Oγ atom of S37 would be located in the proximity of the carbonyl carbon of the esterified Phe-tRNA^{Phe} substrate. This organization together with the catalysed reaction—the formation of a peptide bond from aminoacyl-tRNA—suggests that S37 may be a catalytic residue serving as a nucleophile. This is consistent with the fact that substitution of S37 with alanine resulted in an inactive variant (S37A), but its substitution with the nucleophilic cysteine (S37C) did not completely abolish the enzymatic activity (Figure 5D). So, S37 may attack the carbonyl carbon of the esterified Phe-tRNA^{Phe} substrate, leading to the covalent phenylalanylation of S37. The formation of a covalent intermediate in which L-Tyr is transferred from Tyr-tRNA^{Tyr} to an active serine (S88) was also suggested for Rv2275, using radioactive labelling (12). Moreover, the S88 residue of Rv2275 was suggested to be activated by the conserved Y253 residue, which corresponds to Y202 in AlbC (12). To verify the existence of a covalent intermediate, we further analysed AlbC and the five variants S37A, S37C, Y178F, E182Q and Y202F. The enzymes were incubated in the presence of [^{3}H]Phe-tRNA^{Phe}, and they were separated by SDS–PAGE before being transferred on a PVDF membrane, which was analysed with a radioimager. As observed for Rv2275, wild-type AlbC retained a high amount of the radiolabel, whereas the variant S37A did not retain any radiolabel. The variant S37C presented behaviour similar to that of the wild-type enzyme, confirming the formation of a covalent intermediate and the probable status of S37 as the point of covalent attachment (Figure 6). The variants Y178F and E182Q retained significant amounts of the radiolabel although much less than the wild-type enzyme (Figure 6). This result, combined with the substantial and complete loss of enzyme activity for the variants Y178F and E182Q, respectively (Figure 5D), argues that Y178 and E182 may play a crucial role in the good positioning of the aminoacyl moiety of the substrate in the active site—essential for an efficient formation of the covalent intermediate—concerning the variant Y202F, it retained ~11% of the enzymatic activity (Figure 5D), and at least as much radiolabel as the wild-type enzyme (Figure 6). As the removal of the hydroxyl group in the variant Y202F does not prevent the formation of the covalent intermediate, this demonstrated that Y202 is not responsible for the S37 activation in AlbC.

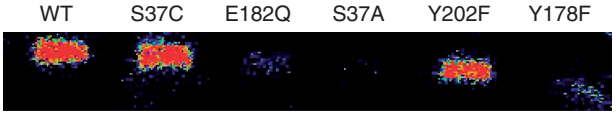

Figure 6. Covalent labelling of AlbC and variants by tritiated Phe transferred from [^{3}H]Phe-tRNA^{Phe}. Enzymes were incubated with [^{3}H]Phe-tRNA^{Phe}, as described in 'Materials and Methods' section, separated on SDS–PAGE, then transferred onto a PVDF membrane that was analysed with a radioimager.

In order to unambiguously identify the residue acylated and the nature of the acylation, we performed peptide mass fingerprint (PMF) analyses on trypsin-digested AlbC, previously incubated or not with Phe-tRNA^{Phe}. AlbC without substrate incubation gave two identified fragments containing the residue S37 (fragments [29–46] and [27–46]) (Supplementary Figure S7A). The corresponding m/z of these fragments (calculated m/z 1878.94 and 2163.09, respectively) were selected and isolated as precursor ion for PSD MS/MS sequence characterization (Supplementary Figure S7B and C). The same experiment performed on AlbC incubated with Phe-tRNA^{Phe} gave two additional fragments (Figure 7A). The m/z of these two fragments were 2026.01 (observed m/z 2026.00) and 2310.16 (observed m/z 2310.17), and corresponded to the values expected for the addition of a phenylalanyl moiety (addition of 147.07) on fragments [29–46] and [27–46], respectively. These fragments, sequenced unambiguously by both b- and y-ions series, contained a phenylalanyl moiety on S37 (Figure 7B and Supplementary Figure S8). This demonstrated the formation of a covalent intermediate, in which L-Phe is transferred from Phe-tRNA^{Phe} to S37 during the catalytic cycle. Moreover, the same experiment was performed on relevant AlbC variants. No modified fragments were observed for the variant S37A (Supplementary Figure S9). For the variants E182Q and

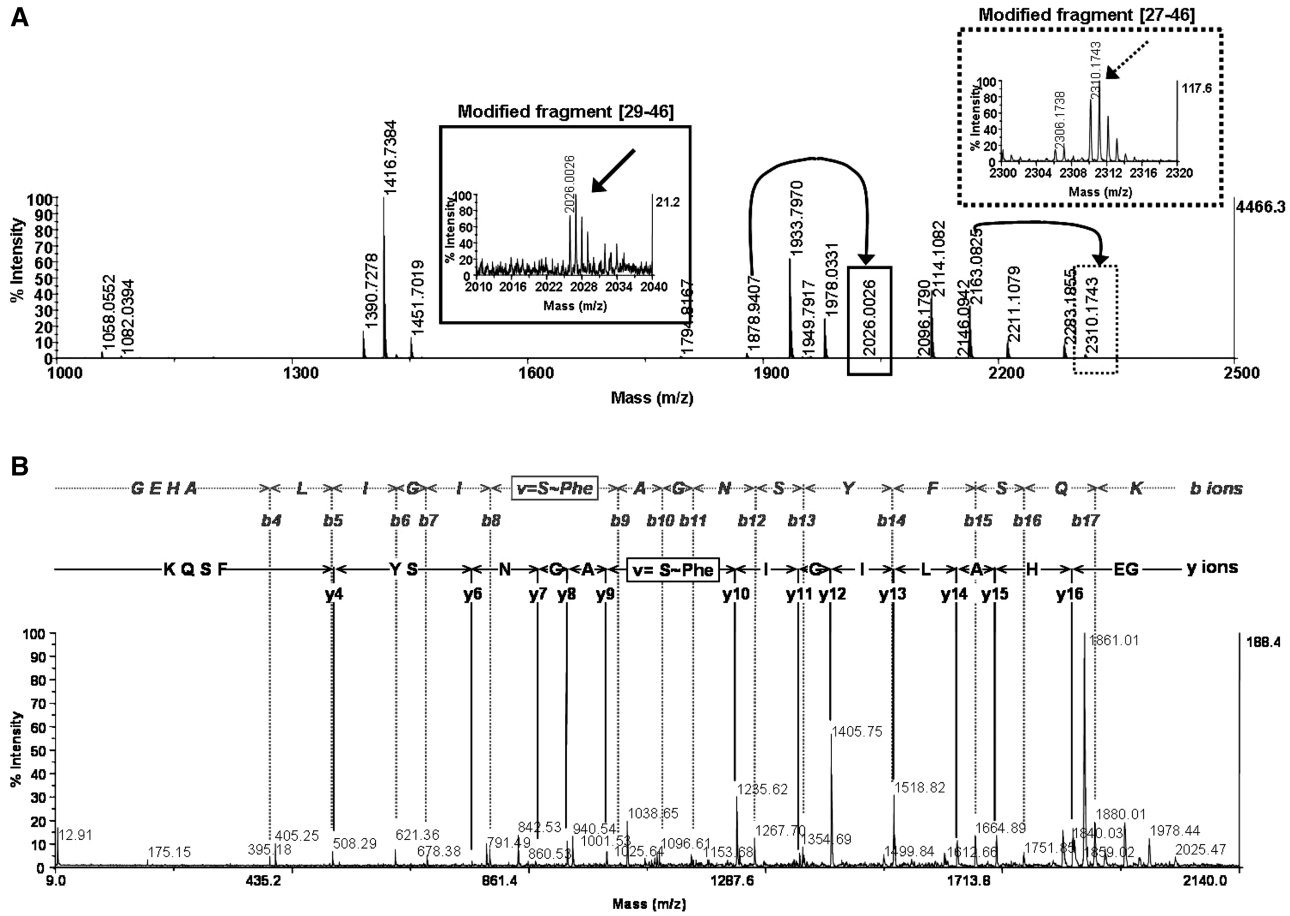

Figure 7. Identification of the covalent aminoacyl-enzyme intermediate. (**A**) Peptide mass fingerprint (PMF) analysis of trypsin-digested AlbC pre-incubated with Phe-tRNA^{Phe}. Arrows indicate the formation of two additional fragments compared to PMF analysis of trypsin-digested AlbC not incubated with Phe-tRNA^{Phe} (Supplementary Figure S7): one at m/z 2026.00 (continuous line frame) corresponding to phenylalanylated fragment [29–46] at m/z 1878.94, and the other at m/z 2310.17 (dashed line frame) corresponding to phenylalanylated fragment [27–46] at m/z 2163.08. (**B**) PSD-MS/MS spectrum of m/z 2026.00. Both b- and y-ions series identify the amino acids sequence of the modified AlbC fragment [29–46], and the residue 37 as the phenylalanylated residue. PSD-MS/MS spectrum of m/z 2310.17 is shown in Supplementary Figure S8.

Y178F, phenylalanylated fragments could be significantly detected, but in much less amounts than those obtained with the wild-type enzyme (Supplementary Figures S10 and S11), in agreement with radiolabelling experiments (Figure 6). These fragments were formed with amounts insufficient for PSD-MS/MS analyses (Supplementary Figures S10 and S11). For the variant Y202F, the phenylalanylated fragments were observed in proportions similar to those obtained with the wild-type enzyme, and were unambiguously sequenced (Supplementary Figure S12). The results confirm that Y202 is not responsible for S37 activation. For the variant S37C, the phenylalanylation of the fragments [27–49] and [29–46] (respectively, calculated m/z of 2040.97 and 2326.13 instead of 1894.90 and 2179.06) was detected (Supplementary Figure S13A). MS/MS sequencing of both m/z demonstrated the formation of a thioester link between the phenylalanyl moiety and the thiol group of C37 (Supplementary Figure S13B and C). All these data allow the proposal of a mechanism for the formation of a covalent intermediate during the catalytic cycle of the CDPS AlbC (see Discussion section).

Distribution and role of the residues conserved among CDPSs but located outside the active site

Seven of the 13 residues conserved among CDPSs are outside the active site: H31, W54, F59, G79, S120, Y128 and C149 (Figure 4 and Supplementary Figure S14). Each of G79, S120 and C149 was substituted with alanine without any significant effect on the activity or production of the corresponding variants amounts (Figure 5D). Variants H31A, W54A and F59A expressed very substantially lower activities than the wild-type enzyme, and were produced in significantly (H31A and F59A) or substantially (W54A) smaller amounts (Figure 5D). The residues H31, W54 and F59 are located in strand β3, helix α2 and loop α2–β4, respectively. The residue H31 is fully buried, and its side chain makes two hydrogen bonds with the residues D63 and S198 that are located in neighbouring strands. The residues W54 and F59 are highly buried and their side chains interact with hydrophobic residues from strand β6 (Supplementary Figure S14A and B). The substitution of each of these three residues with alanine should lead to a significant structural destabilization,

which is consistent with our results showing that the production of the corresponding variants is strongly affected. The substitution of Y128 by alanine did not impair the production of the corresponding variant, but abolished cFL synthesis (Figure 5D). The residue Y128, which is located at the beginning of helix α5, is fully buried and its side chain makes two hydrogen bonds with Q123 from helix η1 and R132 from helix α5 (Supplementary Figure S14A and C). The substitution of Y128 by alanine should alter the relative positions of these helices and neighbouring interactions, resulting in a loss of AlbC activity.

Recognition of the tRNA moiety of the aa-tRNA substrates by AlbC

The electrostatic potential surface of AlbC presents a highly biased distribution of charged residues, which could interact with the tRNA substrate by forming salt bridges with the phosphates of the tRNA backbone or hydrogen bonds with the nucleotide bases. In particular, there is a large patch of positively charged residues covering a surface of 1040 $Å^2$ (Figure 8A). This patch is mostly composed of residues belonging to the helix α4, which contains nine basic residues. These residues, R87, K90, R91, K94, R97, R98, R99, R101 and R102, protrude toward the solvent, and a similar pattern of basic residues is found in all CDPSs (Figures 4 and 8B). We substituted with alanine each of the basic residues in helix α4. All the resulting variants were produced with yields similar to that of the wild-type enzyme but their cFL-synthesizing activities were diverse (Figure 8C). The variants R98A, R99A and R98A/R99A displayed only 19, 8 and <2% of the wild-type activity, respectively. Five other variants (K90A, K94A, R97A, R101A and R102A) had activities that were 53–72% of the wild-type enzyme. The activities of the two remaining variants, R87A and R91A, were similar to that of the wild type. These results showed that the basic residues located in the C-terminal half of the helix α4, especially R98 and R99, are important for cFL production, probably because they interact with the tRNA moiety of the substrate. A second patch of basic residues, consisting of R214, R215 and R220 from loop α8–β7 and R231 from loop β7–α9, is present on AlbC (Figure 8A). We substituted these residues and showed that the corresponding variants displayed properties similar to the wild-type enzyme (Figure 8C), indicating that this patch is not involved in tRNA binding. This is consistent with the fact that this patch is not conserved among CDPSs (Figure 4). Our results suggest that the interaction between AlbC and its aa-tRNA is driven by the binding of the substrate aminoacyl moiety in its deep pocket and the interaction of the tRNA moiety with at least one patch of basic residues on helix α4.

DISCUSSION

We report the 3D structure of AlbC, a member of the recently identified CDPS family. AlbC presents a compact α/β structure with a Rossmann-fold domain, as it was also recently reported for the CDPS Rv2275 (12). We show that, although having many structural similarities, AlbC and Rv2275 have significant differences that raise intriguing questions regarding the catalytic mechanism used by the CDPSs to synthesize cyclodipeptides.

AlbC, and Rv2275, share a high structural similarity with the catalytic domain of class-Ic aaRSs. aaRSs catalyse specific esterifications of particular amino acids to their cognate tRNAs in a two-step reaction. First, amino acids are activated with ATP to form enzyme-bound aminoacyl-adenylate intermediates. Second, the aminoacyl moieties are transferred to tRNAs to form aa-tRNA products. The structural similarities concern only the catalytic domains of class-I aaRSs, which are responsible for the binding of amino acid and ATP substrates. The amino acid-binding pockets of class-I aaRSs superimpose well with the AlbC pocket and we show that this pocket interacts with the aminoacyl moiety of the aminoacyl-tRNA substrate. However, AlbC does not possess the HIGH motif involved in ATP binding that is conserved among all catalytic domains of class-I aa-RSs (36–38). Also, AlbC does not have the loop connecting the catalytic domain to the C-terminal tRNA-binding domain, which contains the second ATP-binding motif —KMSKS—in class-I aa-RSs. This is consistent with the fact that CDPSs do not activate amino acids using ATP, but use already activated amino acids in the forms of aa-tRNAs (1). Another significant difference is that AlbC does not possess the C-terminal domain of aaRSs, which is crucial for their interaction with tRNAs (39).

The AlbC specificity seems to be directed mainly at the aminoacyl moiety, and not at the sequence of the tRNA moiety of the substrate. Indeed, the substitution of one residue of the aminoacyl-binding pocket of AlbC by the corresponding residue of Rv2275 gave a variant that, in *E. coli*, mainly bound tyrosyl-$tRNA^{Tyr}$ rather than phenylalanyl-$tRNA^{Phe}$. Comparing the acceptor stems of the two tRNA moieties revealed that they share the sequences of the single-strand extremity ($A^{76}C^{75}C^{74}A^{73}$) and the first base pair (C^{72}–G^{1}), suggesting that some specificity determinants could lay in these conserved sequences of the tRNA moiety of the substrate. We previously showed that AlbC is able to efficiently use various *E. coli* aa-tRNAs to synthesize various cyclodipeptides (1). We compared the sequences of all *E. coli* aa-tRNAs that can serve as substrates for AlbC (predominantly phenylalanyl- and leucyl-tRNAs, and to a lesser extent methionyl-, tyrosyl-, alanyl- and valyl-tRNAs). For all these tRNA families, we found molecules that also share the same sequence: $A^{76}C^{75}C^{74}A^{73}C^{72}$-$G^{1}$. This might suggest that sequences of the single-strand extremity and the first base pair, but not those of the double-stranded acceptor stem, are specifically recognized by AlbC. Noted that aa-tRNA protein transferases recognize the aa-tRNA in a tRNA sequence-independent manner, the double-stranded acceptor stem being fully dispensable (40).

An interesting question concerns the tRNA-binding mode of AlbC. Could it be similar to that described for class-Ic aaRSs? Recognition of tRNAs by TyrRSs and TrpRSs mainly involves their C-terminal tRNA-binding domains, essential for tRNA binding (39). However, several residues belonging to helices α7 and α9 of their

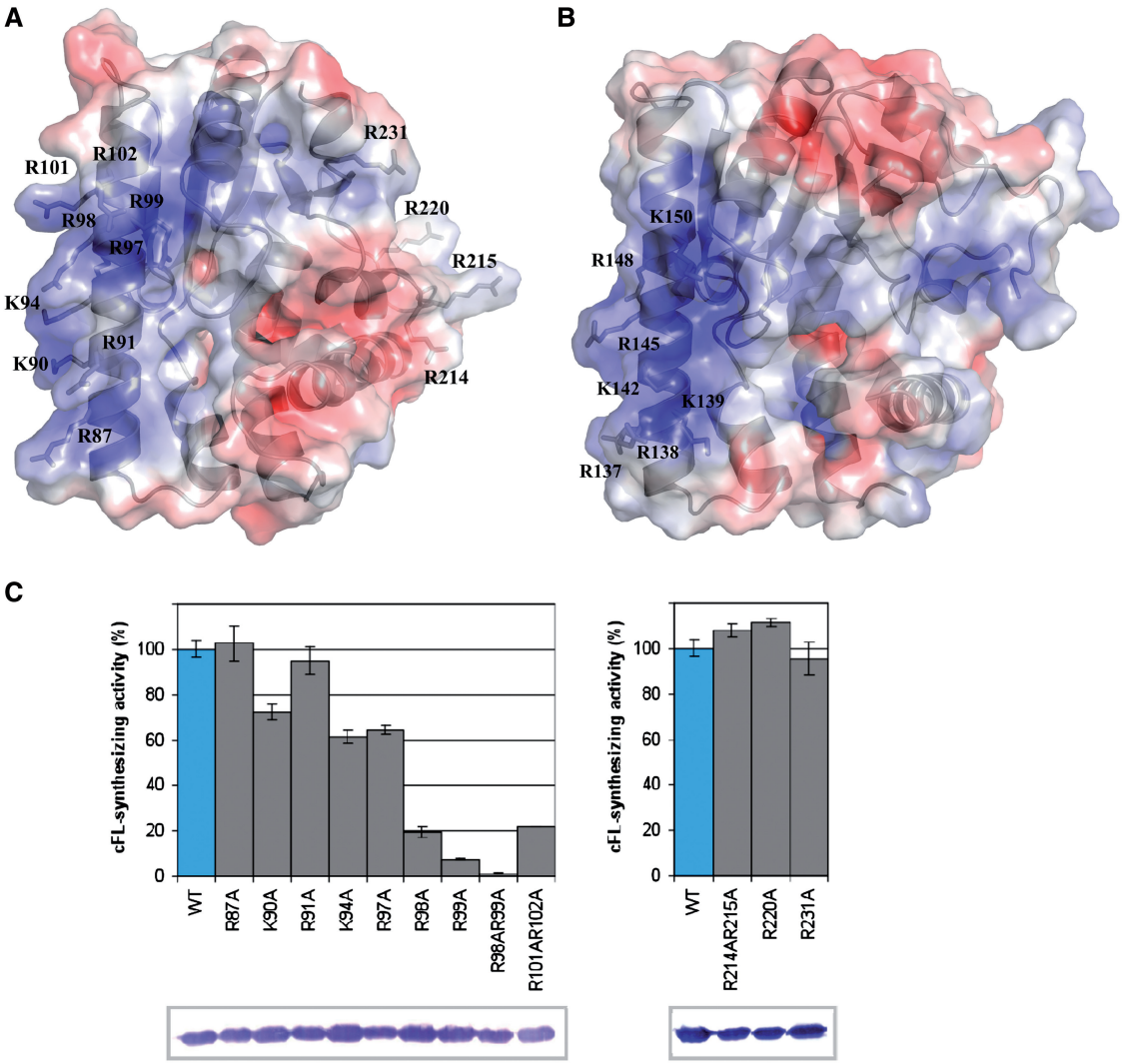

Figure 8. Basic patch residues contributing to CDPS activity. (**A**) Electrostatic surface potential of AlbC, mapped on its solvent-accessible surface at contouring levels of ±5 kTe–1. Positive charge is in blue, negative charge in red. The potential was calculated using APBS in PyMol (59) (http://www.pymol.org/) from a derived model of AlbC where the missing side chain of residue R215 and the missing residue T217 were added. AlbC structure is shown in grey and the side chains of all mutated basic residues are identified as sticks. (**B**) Electrostatic surface potential of Rv2275 generated like that of AlbC. (**C**) Site-directed mutagenesis study of AlbC. cFL-synthesizing activity of the wild-type AlbC (in blue) and of each of the variants in which a basic residue is substituted (in grey) are shown with error bars. The corresponding western blots indicating amounts of the proteins are also shown. Left, the residues belonging to the helix α4; right, other basic residues.

N-terminal catalytic domains interact with the acceptor stems of tRNAs (Figure 4) (29,32,41–46). AlbC does not possess a C-terminal tRNA-binding domain like those in aaRSs nor any residue identical to those of the N-terminal catalytic domain involved in the tRNA interaction (Figure 4). Otherwise, we showed that a patch of basic residues, which is conserved among CDPSs but not present in class-I aa-RSs, is important for AlbC activity. This patch may interact with the phosphate backbone of the tRNA body, as proposed for aa-tRNA-protein transferases and Fem aa-transferases (47). Comparing the position of the basic patch in AlbC with that of the acceptor stem-binding region in $TyrRS_{Mj}$ revealed that they are different (Supplementary Figure S15). All these data might suggest that the tRNA-binding mode of AlbC is different from that of class-I aa-RSs, but this issue remains to be addressed.

We wondered how AlbC uses two tRNA substrates to form cyclodipeptides. One possibility is that AlbC acts as a homodimer, each monomer interacting with one aa-tRNA. Although AlbC was obtained as a monomer both in solution (1) and in crystal forms, we cannot exclude that aa-tRNA binding induces dimerization. Noted that class-Ic aaRSs act as functional dimers, but they are also obtained as dimers in crystals (44). Class-Ic aaRSs dimers are mainly formed by the association of a hydrophobic surface involving residues of the CP1 domain of each catalytic domain (34,35). The positions of these residues are not conserved in AlbC. In addition, when we mimicked this association for two AlbC monomers by superimposing them on different crystal structures of TyrRSs or TrpRSs, we observed major steric hindrances between helices α5, α6 and α7 of each AlbC monomer. Therefore, it is unlikely that AlbC dimerizes in a manner

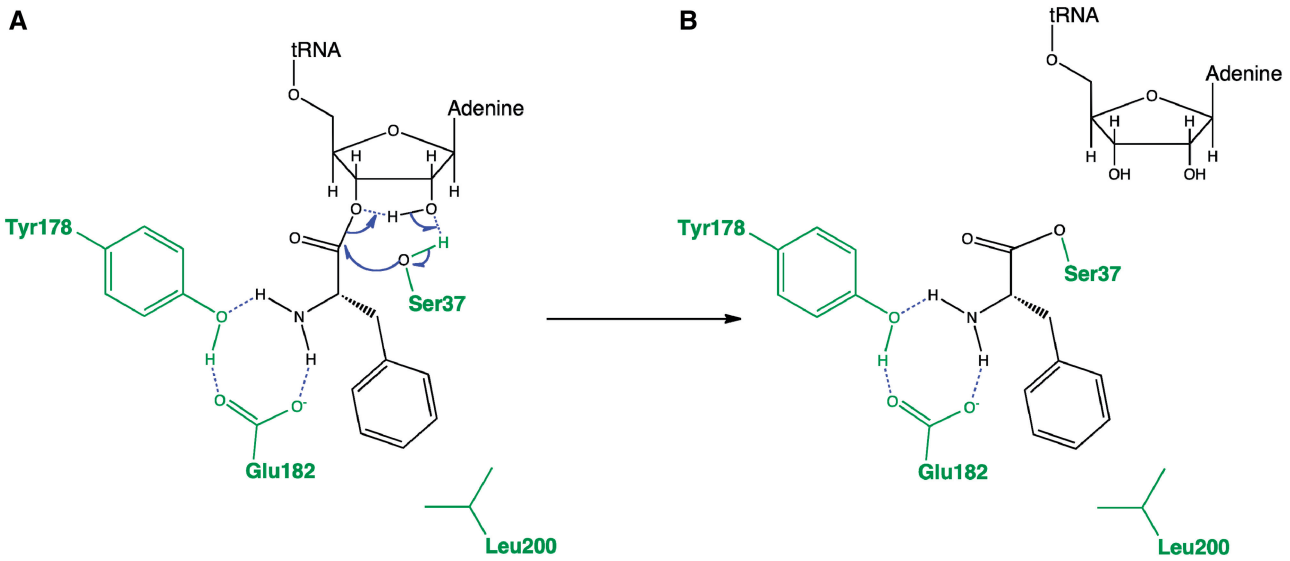

Figure 9. Proposed mechanism of covalent phenylalanyl-enzyme formation for AlbC. (**A**) Proposed anchoring of the Phe-tRNA^{Phe} substrate at the AlbC active site. The residues S37, Y178 and E182, which are essential for the covalent intermediate formation (Figures 6 and 7), are represented in green. The residue L200, shown to interact with the phenylalanyl moiety of the substrate (Figure 5E), is also represented in green. As no residue in the AlbC active site is likely to deprotonate the hydroxyl group of S37, the S37 activation is proposed to be achieved by a concerted proton shuttling mechanism involving the two adjacent vicinal hydroxyls of the nucleotide A^{76} of the tRNA moiety. (**B**) The resulting covalent phenylalanyl-enzyme.

similar to class-Ic aaRSs, and furthermore, any such dimerization would result in a physical separation of the two AlbC catalytic pockets, inconsistent with the formation of two peptide bonds. This is indeed what was observed for Rv2275, which is found as a homodimer: the residues that make up the interface belong to other secondary structural elements than those involved in the dimerization of class-Ic aaRSs, and the two Rv2275 catalytic pockets are not at the dimer interface, but are physically separated (12). Moreover, it is unlikely that AlbC dimerizes in a manner similar to Rv2275 since the residues that are involved in the Rv2275 interface are not conserved in AlbC and other CDPSs. As it seems more probable that AlbC is active as a monomer, we considered whether AlbC catalyses its two-substrate reaction using a ternary complex or a ping-pong mechanism. The size of the AlbC pocket does not appear appropriate to accommodate the aminoacyl moieties of two aa-tRNA substrates concomitantly. Thus, one aa-tRNA may bind to AlbC and its aminoacyl moiety would be stored as a covalent intermediate, i.e. as an aminoacyl-enzyme intermediate. The formation of a covalent intermediate resulting from the transfer of L-Tyr from Tyr-tRNA^{Tyr} to an active serine was suggested for Rv2275, but such a transfer was not conclusively demonstrated (12). We demonstrated that this transfer occurs in AlbC because we showed that the active serine residue is covalently bound to the phenylalanyl transferred from Phe-tRNA^{Phe} during the catalytic cycle. An intriguing point that remains to be determined concerns the activation of the catalytic serine, i.e. the deprotonation of its hydroxyl group serving as the nucleophile. For Rv2275, it was suggested that the catalytic serine (S88) residue is activated by a conserved tyrosine residue (Y253). This is consistent with the hydrogen bonding of Y253 to the side chain hydroxyl group of S88, and the much lower (by a factor of 200) catalytic activity of the Y253F variant (12). For AlbC, the corresponding tyrosine residue (Y202) adopts a different conformation in the AlbC structure, and the Y202F variant displays only a 10-fold decrease in catalytic activity. We also unambiguously showed that the formation of the covalent aminoacyl intermediate is not affected by the substitution of Y202 with phenylalanine, demonstrating that Y202 is not responsible for the activation of S37 in AlbC. As no other residue in the vicinity of S37 can account for its activation, an attractive hypothesis is that the serine activation results from a concerted proton shuttling mechanism involving the two adjacent vicinal hydroxyls of the nucleotide A^{76} of the tRNA moiety (Figure 9), as it has been proposed for the ribosomal peptididyl transferase (48) and the FemX aa-transferase of *Weissella viridescens* (49). At this stage, we cannot exclude some rearrangements in AlbC active site during the different steps of the reaction. The analysis of B-factors around AlbC active site suggests that the loop β3–α2, which precedes the position S37, may present some flexibility. The crystal structure of AlbC complexed with different intermediates of the reaction should clarify this point. But, whatever how the serine activation is performed, we demonstrated that CDPSs use a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate. The aminoacyl-enzyme may then react with the aminoacyl moiety of the second aa-tRNA to form a dipeptidyl-enzyme or dipeptidyl-tRNA intermediate, which may then undergo intramolecular cyclization leading to the formation of the cyclodipeptide product.

In conclusion, we report the crystal structure of AlbC, a CDPS from *S. noursei*. It confirms that CDPSs are structurally similar to the catalytic domain of class-I aaRSs. This structural similarity has provided clues about the

molecular bases of the interactions between CDPSs and their aa-tRNAs substrates. Moreover, we demonstrate that CDPSs catalyse their two-substrate reactions via a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate. CDPSs provide a novel example of the exaptation of aaRSs in processes not related to ribosomal protein synthesis. Some atypical aaRSs and aaRS-like proteins have already been shown to be involved in amino acid biosynthesis, RNA modification (50,51), and in non-ribosomal peptide synthesis (52,53). These proteins include enzymes having the catalytic domain of an aaRS and lacking its tRNA-binding domain. Thus, the archaeal Asn synthetase, a paralog of AsnRS, does not form Asn-tRNA^{Asn}, but synthesizes Asn by amidation of the β-carboxyl group of aspartate with ammonia (54). YadB (renamed glutamyl-queuosine tRNA^{Asp} synthetase), a paralog of GluRS, catalyses glutamylation of queuosine in the wobble position of tRNA^{Asp} (55,56). Amino acid:[carrier protein] ligases, which are SerRS homologues, catalyse the ATP-dependent activation of particular amino acids and their transfer to the phosphopantetheine prosthetic group of carrier proteins (57). Finally, CDPSs differ from these other aaRS-like proteins in that they have not only diverged from conventional aaRSs, but they have also acquired new active site residues, converting them into self-contained cyclodipeptide-forming enzymes, which can use aa-tRNAs as substrates.

ACCESSION NUMBER

PDB, 3OQV.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank P. Cuniasse, J. Baillon, and S. Lautru for helpful discussions. We thank E. Stura, A. Ducruix, and M. Ries-Kaut for helpful advice on crystallogenesis, and R. Guerois for help in bioinformatic. We thank S. Bregant and B. Czarny for advice on the use of the Beta-imager. We thank M. Delarue for the critical reading of the article. We are grateful to J. Marquez for access to the P-CUBE HTX platform (EMBL, Grenoble) and to F. Fenaille for access to mass spectrometry facilities (IBITECS/SPI).

FUNDING

Commissariat à l'Energie Atomique; Centre National de la Recherche Scientifique, Université Paris-Sud 11; Agence Nationale de la Recherche (ANR-10-1501; CyDiPepS); Commissariat à l'Energie Atomique, doctoral fellowships (to L.S. and Y.L.). Funding for open access charge: Commissariat à l'Energie Atomique, IBITECS/SIMOPRO.

Conflict of interest statement. None declared.

REFERENCES

1. Gondry,M., Sauguet,L., Belin,P., Thai,R., Amouroux,R., Tellier,C., Tuphile,K., Jacquet,M., Braud,S., Courçon,M. *et al.* (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.*, **5**, 414–420.
2. Gondry,M., Lautru,S., Fusai,G., Meunier,G., Ménez,A. and Genet,R. (2001) Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. *Eur. J. Biochem.*, **268**, 1712–1721.
3. Belin,P., Le Du,M.H., Fielding,A., Lequin,O., Jacquet,M., Charbonnier,J.B., Lecoq,A., Thai,R., Courçon,M., Masson,C. *et al.* (2009) Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **106**, 7426–7431.
4. von Döhren,H. (2009) Charged tRNAs charge into secondary metabolism. *Nat. Chem. Biol.*, **5**, 374–375.
5. RajBhandary,U.L. and Söll,D. (2008) Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics. *Proc. Natl Acad. Sci. USA*, **105**, 5285–5286.
6. Banerjee,R., Chen,S., Dare,K., Gilreath,M., Praetorius-Ibba,M., Raina,M., Reynolds,N.M., Rogers,T., Roy,H., Yadavalli,S.S. *et al.* (2010) tRNAs: cellular barcodes for amino acids. *FEBS Lett.*, **584**, 387–395.
7. Francklyn,C.S. and Minajigi,A. (2010) tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett.*, **584**, 366–375.
8. Mainardi,J.L., Villet,R., Bugg,T.D., Mayer,C. and Arthur,M. (2008) Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol. Rev.*, **32**, 386–408.
9. Watanabe,K., Toh,Y., Suto,K., Shimizu,Y., Oka,N., Wada,T. and Tomita,K. (2007) Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. *Nature*, **449**, 867–871.
10. Benson,T.E., Prince,D.B., Mutchler,V.T., Curry,K.A., Ho,A.M., Sarver,R.W., Hagadorn,J.C., Choi,G.H. and Garlick,R.L. (2002) X-ray crystal structure of *Staphylococcus aureus* FemA. *Structure*, **10**, 1107–1115.
11. Biarrotte-Sorin,S., Maillard,A.P., Delettré,J., Sougakoff,W., Arthur,M. and Mayer,C. (2004) Crystal structures of *Weissella viridescens* FemX and its complex with UDP-MurNAc-pentapeptide: insights into FemABX family substrates recognition. *Structure*, **12**, 257–267.
12. Vetting,M.W., Hegde,S.S. and Blanchard,J.S. (2010) The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. *Nat. Chem. Biol.*, **6**, 797–799.
13. Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
14. Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature*, **347**, 249–255.
15. Arnez,J.G. and Moras,D. (1997) Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.*, **22**, 211–216.
16. Ibba,M. and Söll,D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.*, **69**, 617–650.
17. Braud,S., Moutiez,M., Belin,P., Abello,N., Drevet,P., Zinn-Justin,S., Courçon,M., Masson,C., Dassa,J., Charbonnier,J.B. *et al.* (2005) Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J. Proteome Res.*, **4**, 2137–2147.
18. Stura,E.A., Satterthwait,A.C., Calvo,J.C., Kaslow,D.C. and Wilson,I.A. (1994) Reverse screening. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 448–455.
19. Ducruix,A. and Giegé,R. (1999) *Crystallization of Nucleic Acids and Proteins: A Practical Approach*. Oxford University Press, Oxford.
20. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
21. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.

22. Langer,G., Cohen,S.X., Lamzin,V.S. and Perrakis,A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.
23. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
24. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. 3rd, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
25. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
26. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
27. Rao,S.T. and Rossmann,M.G. (1973) Comparison of super-secondary structures in proteins. *J. Mol. Biol.*, **76**, 241–256.
28. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
29. Kobayashi,T., Nureki,O., Ishitani,R., Yaremchuk,A., Tukalo,M., Cusack,S., Sakamoto,K. and Yokoyama,S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.*, **10**, 425–432.
30. Zhang,Y., Wang,L., Schultz,P.G. and Wilson,I.A. (2005) Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci.*, **14**, 1340–1349.
31. Kuratani,M., Sakai,H., Takahashi,M., Yanagisawa,T., Kobayashi,T., Murayama,K., Chen,L., Liu,Z.J., Wang,B.C., Kuroishi,C. *et al.* (2006) Crystal structures of tyrosyl-tRNA synthetases from Archaea. *J. Mol. Biol.*, **355**, 395–408.
32. Shen,N., Guo,L., Yang,B., Jin,Y. and Ding,J. (2006) Structure of human tryptophanyl-tRNA synthetase in complex with $tRNA^{Trp}$ reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res.*, **34**, 3246–3258.
33. Zhou,M., Dong,X., Shen,N., Zhong,C. and Ding,J. (2010) Crystal structures of *Saccharomyces cerevisiae* tryptophanyl-tRNA synthetase: new insights into the mechanism of tryptophan activation and implications for anti-fungal drug design. *Nucleic Acids Res.*, **38**, 3399–3413.
34. Brick,P. and Blow,D.M. (1987) Crystal structure of a deletion mutant of a tyrosyl-tRNA synthetase complexed with tyrosine. *J. Mol. Biol.*, **194**, 287–297.
35. Doublié,S., Bricogne,G., Gilmore,C. and Carter,C.W. Jr (1995) Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*, **3**, 17–31.
36. Webster,T., Tsai,H., Kula,M., Mackie,G.A. and Schimmel,P. (1984) Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science*, **226**, 1315–1317.
37. Ludmerer,S.W. and Schimmel,P. (1987) Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **262**, 10801–10806.
38. Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
39. Waye,M.M., Winter,G., Wilkinson,A.J. and Fersht,A.R. (1983) Deletion mutagenesis using an 'M13 splint': the N-terminal structural domain of tyrosyl-tRNA synthetase (*B. stearothermophilus*) catalyses the formation of tyrosyl adenylate. *EMBO J.*, **2**, 1827–1829.
40. Abramochkin,G. and Shrader,T.E. (1996) Aminoacyl-tRNA recognition by the leucyl/phenylalanyl-tRNA-protein transferase. *J. Biol. Chem.*, **271**, 22901–22907.
41. Nair,S., Ribas de Pouplana,L., Houman,F., Avruch,A., Shen,X. and Schimmel,P. (1997) Species-specific tRNA recognition in relation to tRNA synthetase contact residues. *J. Mol. Biol.*, **269**, 1–9.
42. Bedouelle,H. and Winter,G. (1986) A model of synthetase/transfer RNA interaction as deduced by protein engineering. *Nature*, **320**, 371–373.
43. Labouze,E. and Bedouelle,H. (1989) Structural and kinetic bases for the recognition of $tRNA^{Tyr}$ by tyrosyl-tRNA synthetase. *J. Mol. Biol.*, **205**, 729–735.
44. Yaremchuk,A., Kriklivyi,I., Tukalo,M. and Cusack,S. (2002) Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.*, **21**, 3829–3840.
45. Tsunoda,M., Kusakabe,Y., Tanaka,N., Ohno,S., Nakamura,M., Senda,T., Moriguchi,T., Asai,N., Sekine,M., Yokogawa,T. *et al.* (2007) Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms. *Nucleic Acids Res.*, **35**, 4289–4300.
46. Yang,X.L., Otero,F.J., Ewalt,K.L., Liu,J., Swairjo,M.A., Köhrer,C., RajBhandary,U.L., Skene,R.J., McRee,D.E. and Schimmel,P. (2006) Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J.*, **25**, 2919–2929.
47. Suto,K., Shimizu,Y., Watanabe,K., Ueda,T., Fukai,S., Nureki,O. and Tomita,K. (2006) Crystal structures of leucyl/phenylalanyl-tRNA-protein transferase and its complex with an aminoacyl-tRNA analog. *Embo J.*, **25**, 5942–5950.
48. Beringer,M. and Rodnina,M.V. (2007) The ribosomal peptidyl transferase. *Mol. Cell*, **26**, 311–321.
49. Fonvielle,M., Chemama,M., Lecerf,M., Villet,R., Busca,P., Bouhss,A., Ethève-Quelquejeu,M. and Arthur,M. (2010) Decoding the logic of the tRNA regiospecificity of nonribosomal FemX(Wv) aminoacyl transferase. *Angew. Chem. Int. Ed. Engl.*, **49**, 5115–5119.
50. Schimmel,P. and Ribas De Pouplana,L. (2000) Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci.*, **25**, 207–209.
51. Francklyn,C., Perona,J.J., Puetz,J. and Hou,Y.M. (2002) Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. *RNA*, **8**, 1363–1372.
52. Sareen,D., Steffek,M., Newton,G.L. and Fahey,R.C. (2002) ATP-dependent L-cysteine:1D-myo-inosityl 2-amino-2-deoxy-alpha-D-glucopyranoside ligase, mycothiol biosynthesis enzyme MshC, is related to class I cysteinyl-tRNA synthetases. *Biochemistry*, **41**, 6885–6890.
53. Aravind,L., de Souza,R.F. and Iyer,L.M. (2010) Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. *Biol. Direct*, **5**, 48.
54. Roy,H., Becker,H.D., Reinbolt,J. and Kern,D. (2003) When contemporary aminoacyl-tRNA synthetases invent their cognate amino acid metabolism. *Proc. Natl Acad. Sci. USA*, **100**, 9837–9842.
55. Dubois,D.Y., Blaise,M., Becker,H.D., Campanacci,V., Keith,G., Giegé,R., Cambillau,C., Lapointe,J. and Kern,D. (2004) An aminoacyl-tRNA synthetase-like protein encoded by the *Escherichia coli yadB* gene glutamylates specifically $tRNA^{Asp}$. *Proc. Natl Acad. Sci. USA*, **101**, 7530–7535.
56. Blaise,M., Olieric,V., Sauter,C., Lorber,B., Roy,B., Karmakar,S., Banerjee,R., Becker,H.D. and Kern,D. (2008) Crystal structure of glutamyl-queuosine $tRNA^{Asp}$ synthetase complexed with L-glutamate: structural elements mediating tRNA-independent activation of glutamate and glutamylation of $tRNA^{Asp}$ anticodon. *J. Mol. Biol.*, **381**, 1224–1237.
57. Mocibob,M., Ivic,N., Bilokapic,S., Maier,T., Luic,M., Ban,N. and Weygand-Durasevic,I. (2010) Homologs of aminoacyl-tRNA synthetases acylate carrier proteins and provide a link between ribosomal and nonribosomal peptide synthesis. *Proc. Natl Acad. Sci. USA*, **107**, 14585–14590.
58. Petrek,M., Otyepka,M., Banás,P., Kosinová,P., Koca,J. and Damborský,J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**, 316.
59. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.