



Cyclodipeptide Synthases of the NYH Subfamily Recognize tRNA Using an α -Helix Enriched with Positive Residues

Anastasia Croitoru, Morgan Babin, Hannu Myllykallio, Muriel Gondry, Alexey Aleksandrov

► To cite this version:

Anastasia Croitoru, Morgan Babin, Hannu Myllykallio, Muriel Gondry, Alexey Aleksandrov. Cyclodipeptide Synthases of the NYH Subfamily Recognize tRNA Using an α -Helix Enriched with Positive Residues. *Biochemistry*, 2021, 60 (1), pp.64-76. 10.1021/acs.biochem.0c00761 . hal-03084339

HAL Id: hal-03084339

<https://hal.science/hal-03084339>

Submitted on 23 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cyclodipeptide Synthases of the NYH Subfamily Recognize tRNA using an Alpha Helix Enriched with Positive Residues

Anastasia Croitoru¹, Morgan Babin², Hannu Myllykallio¹, Muriel Gondry², and Alexey Aleksandrov^{1*}

¹Laboratoire d'Optique et Biosciences (CNRS UMR7645, INSERM U1182), Ecole Polytechnique, Institut polytechnique de Paris, F-91128 Palaiseau, France

²Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France

[*Corresponding author: Alexey.Aleksandrov@polytechnique.edu](mailto:Alexey.Aleksandrov@polytechnique.edu)

Running title: tRNA recognition of the NYH-subfamily of cyclodipeptide synthases

Keywords: Cyclodipeptide Synthase, aa-tRNA utilizing enzymes, nonribosomal peptide biosynthesis, diketopiperazine, natural product biosynthesis, CDPS, AlbC, tRNA, molecular dynamics; binding free energy;

Abstract

Cyclodipeptide synthases (CDPSs) perform nonribosomal protein synthesis using two aminoacyl-tRNA substrates to produce cyclodipeptides. There is no available structural detail on the CDPS:tRNA interaction to date. Using AlbC, a CDPS that produces cyclo(L-Phe-L-Phe), the interaction between AlbC with its Phe-tRNA substrate was investigated. Simulations of models of the AlbC:tRNA complex, proposed by rigid body docking or homology modeling, demonstrated that interactions with residues of an AlbC alpha helix, $\alpha 4$, significantly contribute to the binding free energy of AlbC to tRNA. Individual residue contributions to the tRNA binding free energy of the discovered binding mode explain well available biochemical data, and the results of *in vivo* assay experiments performed in this work and guided by simulations. In molecular dynamics simulations the phenylalanyl group predominantly occupied the two positions observed in the experimental structure of AlbC in the dipeptide intermediate state, suggesting that tRNAs of the first and second substrates interact with AlbC in a similar manner. Overall, given the high sequence and structural similarity among the members of the CDPS NYH protein subfamily, the mechanism of the protein:tRNA interaction is expected to be pertinent to a wide range of tRNA interacting proteins.

Introduction

Cyclodipeptide synthases (CDPSs) are a family of enzymes that use two aminoacyl-tRNAs (aa-tRNAs) to synthesize cyclodipeptides (1, 2). Cyclodipeptides belong to the diketopiperazine family of secondary metabolites produced in many bacteria. This class of compounds is of broad interest due to its wide range of biological activities, including antibacterial, antifungal, antiviral, antiprion, antitumor, and immunosuppressive functions (3–10). AlbC was the first member of the CDPS family identified during the characterization of the biosynthetic pathway of the antibacterial cyclodipeptide albonoursin in *Streptomyces noursei* (11). It mainly catalyzes the formation of cyclo(L-Phe-L-Phe) (cFF) by incorporating two phenylalanines, and cyclo(L-Phe-L-Leu) (cFL) from phenylalanine and leucine, depending on the substrate availability (11, 12).

To date, seven CDPSs, given in Table S1 have been structurally characterized in tRNA free forms (13–16). These CDPSs contain a Rossmann-fold domain and exist as monomers, except *Nbra*-CDPS, which is supposedly a homodimer in solution (14). Interestingly, two residue sequences are mainly found in CDPSs at positions 40, 202 and 203 (AlbC numbering), NYH and XYP, which allowed to class CDPSs into two subfamilies based on residue types at these positions. AlbC belongs to the NYH subfamily. In the XYP subfamily, a proline substitutes the histidine beside the conserved tyrosine and a non-conserved residue is found at the asparagine position (17). NYH members incorporate 17 different amino acids into cyclodipeptides, except for His, Asp and Lys; XYP members also incorporate 17 amino acids, differing from NYH members by the exclusion of Arg (18, 19). Structural analysis of the two subfamilies demonstrated that the main difference is in the first half of their Rossmann fold, but the catalytic residues are identical in the two families and adopt similar positions. Moreover, mutations of these residues have similar effect on the function in both CDPS subfamilies strongly suggesting that CDPSs of the two subfamilies share the same catalytic mechanism. Thus, it was proposed that the XYP and NYH motifs appeared as alternative solutions to the same enzymatic problem but adopted for differences in the Rossmann fold in the two subfamilies (14).

The catalytic mechanism has been extensively studied experimentally for the structurally characterized CDPSs (13, 15, 16, 20, 21). In particular, AlbC was structurally characterized with a covalently attached dipeptide analogue corresponding to the reaction intermediate state before the final cyclisation step (20, 21). The catalytic cycle begins with the binding of the first aa-tRNA with the aminoacyl group accommodated in the deep and hydrophobic P1 pocket that contains the conserved catalytic residues (15). The subsequent transfer of the aminoacyl moiety to the conserved serine residue, Ser37, leads to the formation of an aminoacyl enzyme intermediate. For the second step, the tRNA^{Phe} part of the first substrate dissociates from AlbC and a second aa-tRNA binds to the enzyme with its aminoacyl group accommodated in the wider P2 cavity close to the P1 pocket. The phenylalanyl-AlbC reacts with the second aa-tRNA to form a dipeptidyl-AlbC intermediate (15, 20). In the last step, the cyclodipeptide product is obtained through intramolecular cyclization. Residues important for the

reaction in AlbC were identified through site-directed mutagenesis and biochemical studies (15, 20). These residues including Ser37, Tyr202, Tyr178, Glu182, Asn40, and His203. Tyr178 and Glu182 are involved in the stabilization of the aminoacyl moiety (named Phe1) of the first phenylalanyl-tRNA throughout the catalytic cycle as suggested by the crystal structure of the diphenylalanyl-enzyme intermediate mimic. The hydroxyl group of Tyr202 serves as a proton relay in the last step of cyclization reaction, as demonstrated recently by computer modelling (21). Three residues in AlbC, Asn40, Tyr178, and His203 help maintain the correct reactive conformation of the dipeptidyl group (20) during the last cyclization step.

The binding of AlbC to the first tRNA appears to be contributed from interactions of the substrate aminoacyl moiety in the pocket P1 and interactions of the tRNA moiety with the patch of basic residues on the helix $\alpha 4$ (15). Residues Arg80, Arg91, Lys94, Arg98, Arg99 and Arg102, all except Arg80 belonging to $\alpha 4$, were identified as important for the cyclodipeptide-synthesizing activity. Their mutation to alanine decreases considerably the production *in vivo* and *in vitro* (15, 20). The AlbC specificity for the first substrate is also contributed by interactions with the aminoacyl moiety, and not at the sequence of the tRNA. However, AlbC seems to handle differently its second substrate. In particular, Asn159, Arg160 and Asp163 of the $\alpha 6$ - $\alpha 7$ loop and Asp205 of the $\beta 6$ - $\alpha 8$ loop of AlbC together with the aminoacyl moiety and the G¹-C⁷² base pair of the acceptor arm appear to be important for interactions with the second aa-tRNA substrate (12). CDPSs show structural similarity to the catalytic domains of class Ic aminoacyl tRNA synthetases (aaRSs), suggesting that CDPSs probably evolved from these aaRSs (15, 22) or from a common ancestor. In class Ic aaRSs, the loops equivalent to the AlbC loop $\alpha 6$ - $\alpha 7$ are known to be implicated in tRNA binding (13), which made Moutiez et al suggest that the interaction of the second substrate with AlbC could be similar to that observed for tRNA binding to class-Ic aaRSs.

Nevertheless, CDPSs and TyrRSs differ significantly. CDPSs do not have the C-terminal domain present in aaRSs and needed to recognize the anticodon, and also lack the ATP-binding motifs, since there is no need to activate amino acids. Two other protein families, FemX aminoacyl-transferases and aa-tRNA protein transferases bind aa-tRNAs to create peptide bonds, however they are structurally different from CDPSs since they possess a GCN5-related N-acetyltransferase (GNAT) fold (2, 23, 24). FemX from *Weissella viridescens* widely used as a model transferase for experimental purposes catalyzes the addition of alanine from Ala-tRNA^{Ala} to the UDP-MurNAc-pentapeptide (25, 26). It interacts with the acceptor arm of the tRNA moiety, as it is still able to interact with an artificial helix mimicking the acceptor arm of tRNA. Similar to CDPSs, the base pair G²-C⁷¹ was shown to be important for the tRNA recognition in addition to the aminoacyl moiety (23). In the second protein family, leucyl/phenylalanyl-tRNA protein transferase (L/F transferase) catalyzes peptide-bond formation by using Leu-tRNA^{Leu} (or Phe-tRNA^{Phe}) as a donor substrate and its terminal Arg (or Lys) as an acceptor substrate. Biochemical and structural studies indicate that L/F transferase interacts with two sequence

regions in the acceptor stem of tRNA, G³-C⁷⁰ base pair and a set of four nucleotides (C⁷², A⁴-U⁶⁹, C⁶⁸). Similar to FemX, L/F transferase can efficiently interact with a helical mimic of the tRNA acceptor stem. Moreover, similar to AlbC, L/F transferase has a strong preference for the CAG isoacceptor of Leu-tRNA^{Leu} (12, 24, 27). Finally, a recent crystallographic and biochemical study showed that the CDPS from *Candidatus Glomeribacter gigasporarum* belonging to the XYP subfamily interacts with the major groove of the acceptor stem of tRNA through the basic residues of β 2 and β 7 strands (14). However, these residues as demonstrated by biochemical experiments, are not implicated in interactions with tRNA in CPDPs of the NYH subfamily, demonstrating that CDPSs of the two subfamilies do not share the same mode of interaction with tRNA (12, 15).

Here we show for the first-time structural details of how NYH CDPSs interact with the tRNA moiety of their substrates. We studied the interaction between AlbC and its Phe-tRNA substrate using computational techniques, including molecular dynamics simulations (MD) and binding free energy calculations. We propose a model that explains previous biochemical and structural experiments. Based on this model, mutagenesis experiments were performed in this work that further corroborate the model. In particular, two residues untested previously were tested in *in vivo* assays demonstrating effects in agreement with the binding contributions of these residues in the model. In this model, the acceptor stem of the tRNA substrate interacts with the basic residues of an alpha helix, α 4, present in all CDPSs, while the phenylalanyl moiety predominantly occupies the two positions previously identified for the first and second substrate. The total charge of the alpha helix is well conserved in enzymes of the NYH subfamily, suggesting that the binding mode of tRNA is shared by the all NYH CDPSs. Moreover, residues that were proposed to be important for binding of the first and second tRNA substrates, were found to interact with tRNA in this model, strongly suggesting that the first and second tRNA substrates both interact with the α 4 helix in a similar binding mode.

Methods

Residue and charge conservation analysis

The sequences of CDPSs with the sequence identity lower than 90% were selected for analysis from reference (18). These sequences were further divided into two groups according to the signature sequences NYH and XYP identified previously (17). Structural alignments were used to guide the sequence alignment with Clustal Omega software (28) for each subfamily. For the NYH subfamily, AlbC (Protein Data Bank (PDB) code: 4Q24 (20)), Rv2275 (PDB: 2X9Q (16)) and YvmC (PDB: 3OQH (13)) were used. For the XYP subfamily, *Nbra*-CDPS (PDB: 5MLQ (14)), *Rgry*-CDPS (PDB: 5MLP (14)) and *Fdum*-CDPS (PDB: 5OCD (14)) were used. The structural alignment was performed with the PyMOL software (29). To understand better the residue conservation in the context of interactions with tRNA, the average total charge of different protein regions was calculated as following. From the NYH alignment, the region corresponding to the α 4 helix in AlbC (residues 81 to 106) was considered for

each sequence and the total charge of the region was calculated by subtracting the number of negatively charged residues, aspartates and glutamates from the number of positive residues, lysines and arginines. Histidines were assumed neutral. In the XYP alignment, the region corresponding to the $\alpha 4$ helix in *Nbra*-CDPS (residues 78 to 102) was used. A similar approach was used for helices $\alpha 5$ and $\alpha 6$.

AlbC:tRNA model building

The structure of tRNA^{Phe} was retrieved from the PDB, with reference code 4YCO (chain D) corresponding to dihydrouridine synthase in complex with tRNA^{Phe} from *E. coli* with resolution of 2.1 Å (30). The chosen tRNA^{Phe} structure is one of the most complete tRNA^{Phe} structures available in the PDB and contains 74 residues out of 76 with defined coordinates. This tRNA^{Phe} structure has no post-transcriptional modifications, however it is similar to the mature tRNA and can be efficiently used by AlbC as demonstrated previously (12). This tRNA contains two transversion, C3–G70 and G3–C70 in the acceptor stem, which were modeled as in the crystal structure (31). The structure 4YCO was obtained at very high concentrations of Mg ions (200 mM MgCl₂) (30), explaining the large content of magnesium ions found in the crystal structure (25 Mg ions total). Thus, out of eight magnesium ions present in the crystallographic structure chain D, four magnesium ions interacting with phosphate groups of tRNA were maintained in the model. The other four magnesium ions were not considered, since they interact with less than two phosphates of the tRNA backbone, and were not observed in the other crystal structure of tRNA^{Phe}, PDB: 1I9V (32). The aminoacyl group and missing residues 75 and 76 were initially built in the extended conformation and were energetically minimized with the rest of the tRNA structure fixed. The CHARMM36 force field was used for the protein (33, 34) and the TIP3P model for water (35–37). The phenylalanyl group was modelled using the force field model specifically developed as a part of this work. The details of the force field development as well as the force field model are given in the Supporting Information.

The crystal structure of AlbC, PDB entry 4Q24 (20) was used for the AlbC model. As previously described (21), the structure of the wild-type AlbC was built from the existing crystal structure by converting the Sγ of Cys37 into an oxygen and deleting the ZPK ligand (N-carbobenzoyloxy-L-Phe-methyl ketone). The same protein protonation state was assigned as in the previous study (21), except Glu182, which was in the deprotonated form. This glutamate acts as a catalytic base and becomes protonated only after the reaction with the first aa-tRNA (21).

In order to acquire models of the AlbC:tRNA interaction, rigid-body docking was applied with the ZDOCK server (38) proven to be a valuable tool for protein-RNA complexes docking (39). tRNA docking onto AlbC was done with ZDOCK 3.0.3 default parameters using the coordinates of AlbC and Phe-tRNA^{Phe}. Results from the top 2,000 ZDOCK predictions were filtered using the condition that aminoacylated adenosine of tRNA interacts with residues 35, 37, 178, 182, 202 of AlbC in order to position the aminoacyl moiety in the catalytic pocket. Predictions were only kept if all selected residues

were within 6 Å of the docked tRNA and resulted in 32 models. The first 10 top-score models were selected for further analysis. An additional model of the AlbC:tRNA interaction was created based on structural homology between CDPSs and synthetases as described in the Supporting Information.

Preliminary calculations on AlbC:tRNA interactions were performed using smaller models. In the small models, tRNA residues that are distant from the protein and, thus not expected to contribute to the protein binding were excluded from simulations. In particular, tRNA nucleotides with a distance to the protein longer than 26 Å were deleted, hence residues 23 to 45 of phe-tRNA^{Phe} were not considered.

Molecular Dynamics simulations

MD simulations were performed with the NAMD simulation package (40). Each protein-tRNA complex was immersed in a box of water, the sides of which were distant by at least 12 Å from any atom of the tRNA protein. Water molecules overlapping the protein and tRNA were removed. Periodic boundary conditions were applied, and the entire box was replicated periodically in all directions. All long-range electrostatic interactions were computed efficiently by the particle mesh Ewald method (41), while the short-range non-bonded interactions were calculated with a cut-off of 11 Å. An appropriate number of potassium counter-ions was included to render the system electrically neutral. MD simulations were performed at constant room temperature and pressure, after 31 ps of thermalization. The CHARMM36 force field (34, 42, 43) was used for the protein, tRNA and the modified version of the TIP3P water model (35).

For the small models, harmonic restraints with a force constant of 1 kcal·mol⁻¹·Å⁻² were applied on heavy atoms around 6 Å of the truncation region using the tRNA crystal structure as a reference. Initially molecular dynamics simulations were performed for 30 ns on each model and continued for at least 200 ns for three models (9, 2 and 7) characterized by the strongest interactions between the protein and tRNA.

The model with the complete tRNA and protein molecules was built based on the tRNA position observed in model 7 characterized by the lowest binding free energy among the models. The crystal structure with reference code 4Q24 (20) was superposed on the snapshot of model 7 with the lowest protein:tRNA binding free energy using the protein backbone atoms and coordinates of tRNA residues were retained. The coordinates of the nucleotides missing in the truncated model were taken from the complete tRNA^{Phe} crystal structure with reference code 4YCO (44). However, the tRNA aminoacyl group in model 7 was misoriented relative to the experimental structure (PDB: 4Q24), as shown in Figure S1, suggesting that MD simulations of model 7 have not converged. In order to correct the position of the aminoacyl group in model 7, we used its position in model 9, which is close to the experimental position in the crystal structure (as demonstrated in Figure S1) as follows. The tRNA was superimposed on the lowest energy structure of model 9, and tRNA terminal residues 75 and 76 with

the aminoacyl group were retained for the final model. The model was then energetically minimized using 200 minimization steps in the CHARMM software with restraints applied on residues 75 and 76 of the nucleic acid (45). The model with the complete tRNA contained around 162,000 atoms and the small models contained around 127,000 atoms each. For the model with the complete tRNA molecule one μ second MD simulation was performed with the center of mass of the protein and tRNA heavy atoms weakly restrained to the origin of the system by a harmonic potential with a force constant of $0.1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ to prevent the drift of the AlbC:tRNA complex in MD simulations.

Binding free energy calculations

Binding free energy of Phe-tRNA^{Phe} to the AlbC protein was estimated as the difference of the total free energy of the complex and the free energies of the separated partners:

$$\Delta G_{prot:tRNA} = G_{prot:tRNA} - G_{prot} - G_{tRNA} \quad [\text{Eq 1}]$$

The free energy (G) has three contributions from polar interactions computed using the Poisson-Boltzmann model (G_{PB}), non-polar (G_{SA}) term, and vibrational entropy:

$$G = G_{PB} + G_{SA} - TS, \quad [\text{Eq 2}]$$

where T is room temperature. The Poisson-Boltzmann equation was solved numerically with the PBEQ module (46, 47) implemented in the CHARMM software version c41b1 (45). The dielectric constants for the solute and solvent volumes were set to 4 and 80, respectively. The solute-solvent dielectric boundary was defined as a molecular surface using a water probe with a radius of 1.4 \AA . The Poisson-Boltzmann equation was solved using a cubic grid and a finite difference algorithm. A two-step protocol was used with an initial calculation performed using a large box with a coarse grid providing the boundary conditions for a second calculation with a smaller box using a finer grid (48). The coarse and fine grid spacing were set to 0.8 and 0.4 \AA respectively, and the grid size was chosen to include the entire protein/tRNA complex in both calculations. A physiological ionic strength of 0.15 M monovalent ion concentration was used in addition to the 4 structural magnesium ions. The non-polar contribution was estimated by the term proportional to the solvent accessible surface area (SASA):

$$G_{SA} = \alpha \times \text{SASA}, \quad [\text{Eq 3}]$$

where the surface tension $\alpha = 6 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ (49).

Separate tRNA and protein structures were obtained by simply discarding the unwanted partner. Thus, structural relaxation upon the dissociation was not explicitly modelled, but included implicitly in the higher protein internal dielectric constant (50, 51). Calculations were performed for 300 snapshots taken each 100 ps from the 30 ns MD simulations for each small model. For models 9, 2 and 7, binding free energy calculations were performed for at least 2,000 snapshots taken each 100 ps from the MD

simulations. For the model with the complete tRNA molecule, free energy calculations were performed on 1,000 structures taken each 1 ns from the 1 μ s MD simulation.

The conformational entropy was estimated by normal-mode analysis (NMA) (52, 53) on ten snapshots taken each one-ns during the last 10 ns of MD simulations. To calculate normal modes, all water molecules were removed and the system containing the protein and tRNA was energetically minimized using the Adopted Basis Newton-Raphson minimizer implemented in CHARMM (45). The distant-dependent dielectric constant of four was used in this calculations. The tolerance applied to the average gradient of 0.0001 was used as convergence criterion. The error of the vibrational entropy calculations was estimated by dividing energies corresponding to the 10 snapshots into two batches and computing the difference.

To estimate long-range electrostatic effects of residue Asp95 in the catalytic center the Poisson-Boltzmann (PB) model was used (54). The electrostatic potential on atoms of the catalytic residues was computed using the PB model, and averaged over structures taken from 100 ns molecular dynamics simulations with the protein in the dipeptide-intermediate state. The same setup described above was used for Poisson-Boltzmann calculations. Calculations were repeated with zero charges on the Asp95 sidechain and the difference in the electrostatic potential was calculated to obtain the electrostatic potential due to the charge on Asp95 and the associated solvent response.

Individual residue contribution to the binding free energy

The individual contribution of protein residues to protein:tRNA binding was estimated by the component analysis, as described previously (55, 56). Only the electrostatic contribution was calculated, as it is expected to be dominant in interactions between the protein and tRNA. Charges on a sidechain up to C β of the residue were zeroed during binding free energy calculations. The energy contribution of the residue is calculated as the difference between the binding free energy of the wild-type complex and the energy of the complex with zero charges on the residue sidechain.

***In vivo* assays for wild-type AlbC and Lys46Ala, Asp95Ala variants**

The expression plasmid encoding AlbC was previously constructed (15). The plasmids encoding AlbC variants were generated by PCR mutagenesis according to the QuikChange method (Stratagene). The plasmids were used to transform *E. coli* BL21AI [pREP4] cells. The strains were grown at 37°C in the M9 minimum medium supplemented with trace elements and vitamins, 0.5% glycerol (0.5% glucose for starter cultures), 200 μ g/ml ampicillin and 25 μ g/ml kanamycin. CDPS expressions were induced by isopropyl- β -D thiogalactopyranoside (IPTG, 2 mM final concentration) and cultivations were continued for 18 h at 20°C. The cultures were centrifuged at 4,000 g for 45 min: supernatants and cell pellets were analyzed for cyclodipeptide-synthesizing activities and protein expression, respectively. Supernatants were acidified with trifluoroacetic acid at 2% v:v final and

submitted to LC-MS analyses as previously described (57). Cell pellets were frozen at - 80°C and broken as described in (1). The soluble protein fractions were separated from the insoluble fractions by centrifugation at 20,000 g for 45 min. Both fractions were analyzed by 12% SDS-PAGE with Coomassie blue staining.

Results

Conservation of the positive charge of CDPS α 4 helix belonging to the NYH family

The residue conservation was first analyzed in the context of interactions with tRNA. In particular, positive charges of protein residues may contribute strongly to tRNA binding. The electrostatic potential on the AlbC surface calculated using the experimental structure 4Q24 (20) is shown in Figure 1. The total charge of AlbC at pH of 8 is $-4 e$, however, the protein is strongly polarized at this pH, which is manifested in relatively large patches of positive and negative electrostatic potential on the protein surface in Figure 1. It was shown that proteins with such large polarization are implicated in binding to the ribosome or interact with nucleic acids (58). In particular, a significant region of positive potential is found on the α 4 helix. The α 4 helix is comprised of 26 residues, with 13 residues being ionized at pH of 8 including seven arginines, two lysines, three glutamates and one aspartate. The total charge of the α 4 helix at pH 8 is, thus $+5 e$. Potential on the protein surface, particularly in the catalytic pocket and around the α 4 helix, correlates well with the elevated level of charge conservation in the NYH subfamily of CDPSs (Figure 1 B).

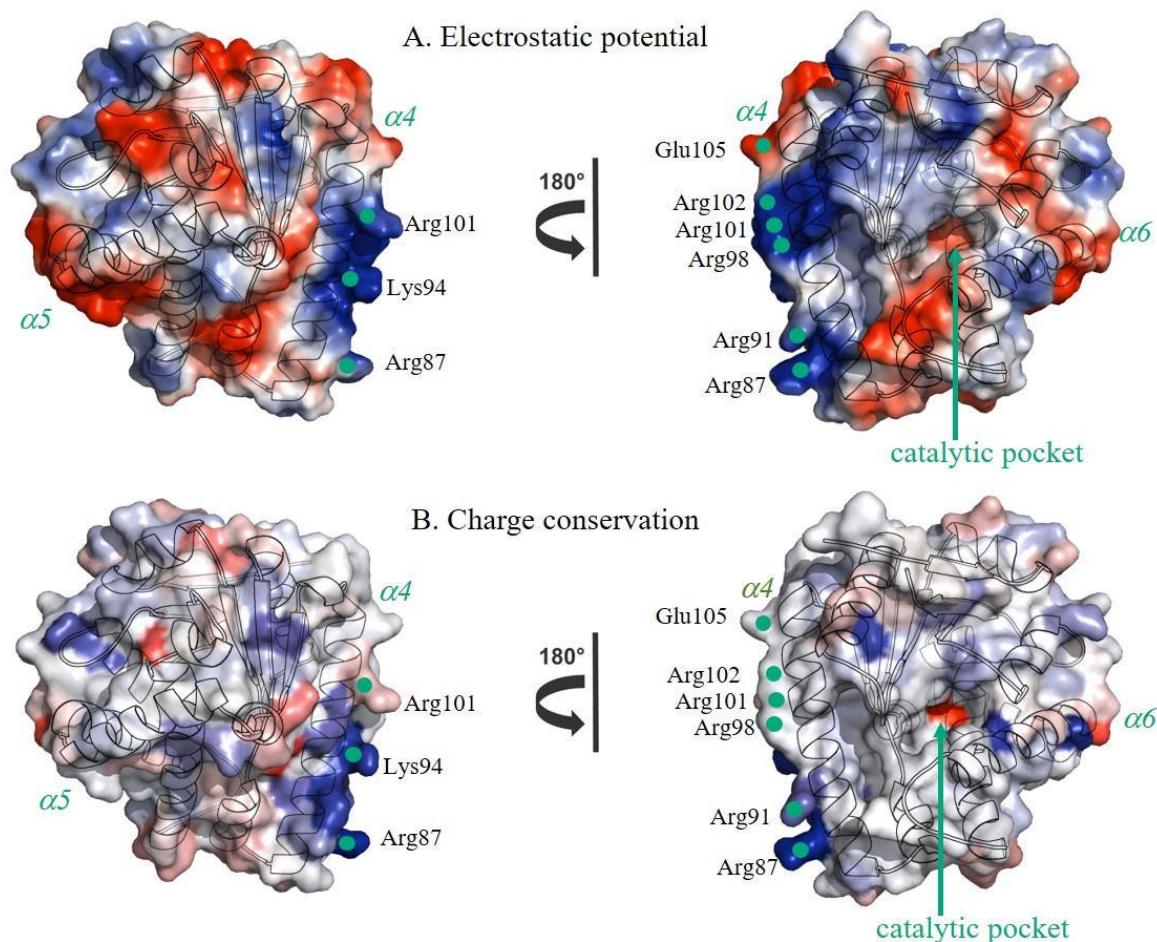


Figure 1. A) Electrostatic potential on the AlbC surface, and B) conservation of the residue charge in the NYH subfamily of CDPs. Green dots show the center of mass of indicated important residues. A) Positive and negative potentials are shown in blue and red respectively. B) Average residue charge obtained from the sequence alignment and considering lysines and arginines positively charged and aspartates and glutamates negatively charged.

The residue charge conservation shown in Figure 1 is clearly visible for the residues in the catalytic pocket, and in particular for Glu182 implicated in the enzymatic activity (15). Positively charged residues of the $\alpha4$ helix possess a relatively significant level of conservation of the ionization state but demonstrate no strict conservation. More precisely, the $\alpha4$ helix is enriched in arginines and lysines in CDPs in the entire NYH subfamily, however these residues have scattered positions in the alignments published previously (1, 14, 15). Thus, in the NYH subfamily (272 sequences), the total positive charge of the $\alpha4$ helix is conserved with the average charge of $+5.5 e$ and standard deviation (SD) of $1.5 e$ as shown in Figure 2. This suggests that the positive residues of the $\alpha4$ helix could be implicated in tRNA binding. In contrast to other solvent exposed helices, the $\alpha4$ helix is the longest helix and the only helix having significant positive charge. Other helices on the protein surface, $\alpha5$ and $\alpha6$ have the average total charge of $0.3 e$ (SD $1.7 e$) and $-0.2 e$ (SD $2.2 e$) respectively (Figure 2). Interestingly, in the XYP subfamily (231 sequences) the $\alpha4$ helix charge is small and slightly negative with an average of $-1.2 \pm 2.6 e$, while helices $\alpha5$ and $\alpha6$ are almost neutral, suggesting that the $\alpha4$ helix may accomplish different functions in tRNA binding in these two subfamilies.

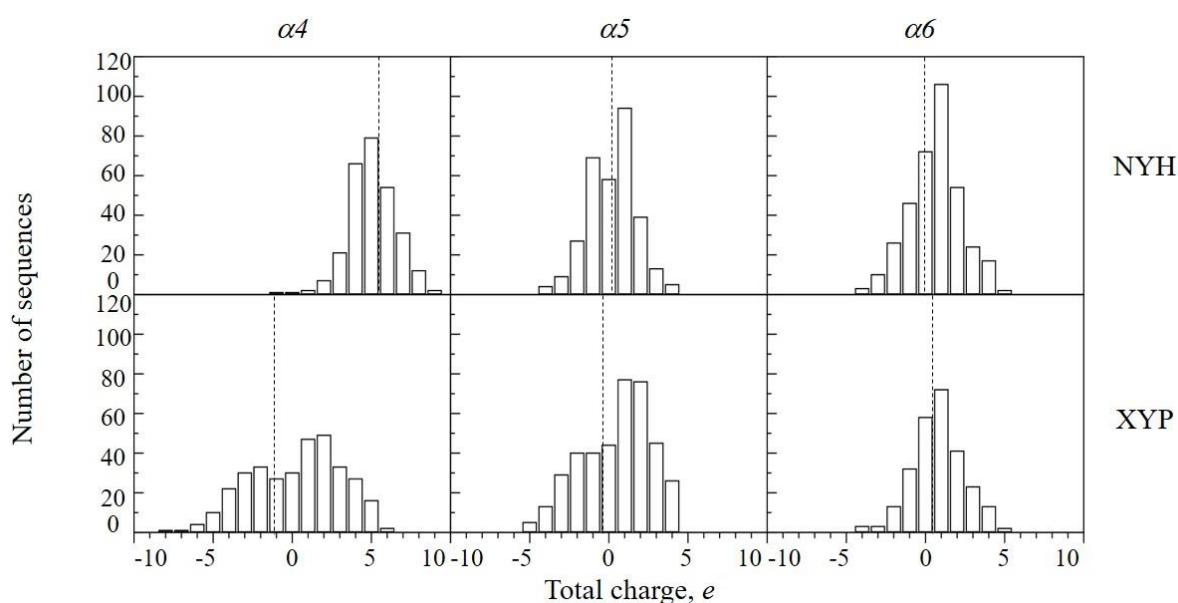


Figure 2. Distribution of the total charge of alpha helices $\alpha 4$, $\alpha 5$, and $\alpha 6$ in the NYH and XYP subfamilies of CDPs obtained after the sequence alignment. The dashed line represents the mean charge.

Rigid-body docking study of AlbC:tRNA interactions

In order to design initial models of the AlbC:tRNA complex, rigid-body docking of the first tRNA substrate onto AlbC was performed with the ZDOCK server (38). The models of the docked complex were filtered based on the proximity of the important residues in the catalytic site to tRNA as described in the Method section. However, the proximity requirement with the minimal distance set to 6 Å was not strict and no filtering restraints were applied to interactions with other protein residues including residues of the $\alpha 4$ helix. Thus, rigid-body docking provided 10 initial models for the AlbC:tRNA complex. One additional model for the first substrate binding was designed based on the structural similarity between AlbC and Tyrosyl-tRNA synthetase (15). Eleven models were then classified based on the Root Mean Square (RMS) deviation (Table S2) giving four groups of models shown in Figure 3. Among the considered eleven models, only one subgroup of models 5, 9 and the model based on TyrRS presented no interactions with the $\alpha 4$ (15). This group of models cannot explain the experimental evidence of the contribution of residues of the $\alpha 4$ helix to binding of first tRNA substrate. Nevertheless, these models were also investigated to elucidate the source of the difference between interactions of CDPs and TyrRS with tRNA substrates, and to probe a hypothesis that the $\alpha 4$ helix is involved in binding of the second tRNA substrate despite what was previously suggested (12). Additionally, the Poisson-Boltzmann binding free energy was computed for 2,000 structures generated with ZDOCK after structural relaxation as described in the Supporting Information. The results show that the binding free energy for strongly interacting AlbC and tRNA correlates well with the contribution of the $\alpha 4$ helix, as shown in Figure S2, and the five models with the lowest binding free energy are all similar and also similar to model 7, shown in Figure S2 and Figure 3.

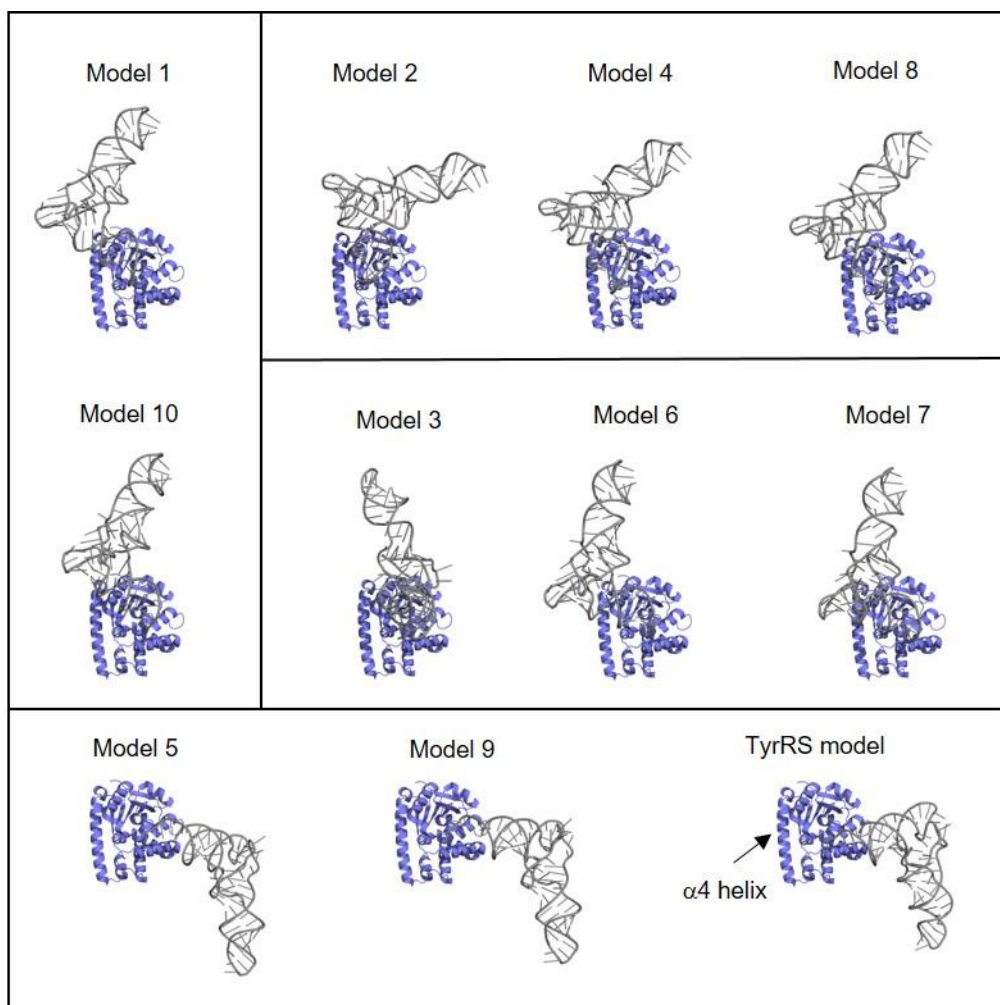


Figure 3. Initial models of the AlbC:Phe-tRNA^{Phe} complex derived by docking and homology modeling. Ten models were obtained by rigid-body docking, and one additional model was based on the experimental structure of the TyrRS:tRNA complex. The models were superimposed using the protein atoms and classified according to the structural similarity. The model number corresponds to their docking scoring rank.

MD simulations were performed for each of eleven models in the explicit solvent. In preliminary calculations on the docked structures, Phe-tRNA^{Phe} residues distant from AlbC by at least 26 Å were not present, as these distant residues do not contribute to AlbC binding. Indeed, in these calculations the contribution of distant groups is approximated as the contribution of the solvent occupying their space. We estimated this contribution of the truncated region using the Poisson-Boltzmann model as described in the Method section and found that it contributes less than 0.1 kcal·mol⁻¹ in all cases in agreement with the previous studies (59). To maintain the tRNA structure near the truncated region, harmonic restraints were applied on heavy atoms around the truncated region, and initially 30 ns of MD simulations were performed to evaluate binding free energy between AlbC and tRNA. In the MD simulations tRNA remained positioned close to the protein and the aminoacyl group remained in the AlbC catalytic pocket in all models. Similar to their starting poses, TyrRS based model and models 5 and 9 did not interact with the $\alpha 4$ helix during the simulation, while in the remaining

models tRNA preferred to interact with the $\alpha 4$ helix. To evaluate the binding free energy, Poisson-Boltzmann/Surface Area (PB/SA) free energy calculations were performed using the structures drawn from the MD simulations. The vibrational entropy was estimated using Normal Mode Analysis. The results are summarized in Table S3. Model 7 presented the lowest binding free energy, i.e. with the strongest interaction between the protein and tRNA as shown in Figure S3, including the entropy estimate lowered further the total binding free energy for model 7 relative to other models. Model 2 demonstrated a lower binding free energy towards the end of the simulation suggesting that the MD simulation has not converged for this model with 30 ns. Model 9 was included in subsequent analysis since it represents the TyrRS binding pose. Thus, based on the preliminary calculations, MD simulations were continued for representative models 2, 7 and 9 for 300 ns for models 2 and 7 and 200 ns for model 9 respectively. The binding free energy and RMS deviation shown in Figure S4 did not demonstrate large fluctuations after 20 ns of simulations for models 7 and 9, suggesting that these simulations converged. The average PB/SA binding free energies observed for models 2, and 7 in MD simulations were -41.0 (SD: $7 \text{ kcal}\cdot\text{mol}^{-1}$) $\text{kcal}\cdot\text{mol}^{-1}$ and -47.0 (SD 7 : $\text{kcal}\cdot\text{mol}^{-1}$) $\text{kcal}\cdot\text{mol}^{-1}$ respectively. The average binding free energy for model 9 is -20.5 (SD: $3 \text{ kcal}\cdot\text{mol}^{-1}$) $\text{kcal}\cdot\text{mol}^{-1}$. Including the vibrational entropy estimate lowered the binding free energy for model 9 relative to model 7 and 2, however, model 7 still has a total binding free energy $14.2 \text{ kcal}\cdot\text{mol}^{-1}$ lower in comparison with model 9. Thus, the models based on TyrRS are characterized by significantly weaker interactions with AlbC due to the lack of interaction with the $\alpha 4$ helix, since the $\alpha 4$ helix is one of structural elements providing the strongest contribution to the first tRNA binding. In contrast to NYH CDPSs, the total charge of the $\alpha 4$ helix in TyrRS from *Thermus thermophilus* is zero (60), suggesting that the binding mode of the NYH CDPS and TyrRS can be different.

Individual residue contributions to the total binding free energy for each model, given in Table S4 were further correlated with the experimental data. In the four best models, Glu182 contributes favorably to aa-tRNA binding with an average of $-5.7 \text{ kcal}\cdot\text{mol}^{-1}$ through ionic interactions with the charged amino group of the phenylalanyl moiety (Figure S5). Glu182 is a conserved residue in all CDPSs and has an important role in the catalytic reaction by participating in the reaction and maintaining the correct orientation of the reactant groups (15, 21). In both models 2 and 7, residues Lys94, Arg98, Arg99 and Arg102 contribute strongly to the binding free energy, while Lys46 contributes mostly in model 2 and Arg91 in model 7. Biochemical experiments indicate that Arg91, Arg98, Arg99 and Arg102 are important for cyclodipeptide formation (12, 15). The mutation of Lys46 into an alanine does not have any significant effect on the cyclodipeptide production and protein expression (Table S5 and Figure S6), suggesting that this residue is not implicated in AlbC:tRNA interaction, however a positive residue, lysine or arginine is frequently found at this position in the CDPS NYH subfamily. In Model 9 based on the TyrRS:tRNA structure, suggested previously to explain the second substrate binding, Arg231 has the strongest contribution to the total binding free energy of

-8.3 kcal·mol⁻¹, but residues Asp163 and Asp205 do not contribute practically to interactions with the tRNA substrate. Importantly, biochemical experiments demonstrate that the contribution of Arg231 to tRNA binding is negligible and Asp163 and Asp205, both implicated in tRNA interactions (2, 15). Overall, this demonstrates that model 9 is unlikely for the complex with the first or second tRNA substrate.

The individual residue contributions for models 2 and 7 are very similar. In model 2, residues 94-102 of the α 4 helix interact with the phosphate backbone of nucleotides 65-67, while in model 7 interactions occur mainly between residues 91-103 and nucleotides 62-65, as shown in Figure S5. Thus, models 2 and 7 representative of two groups of structures predicted by rigid-body docking converged to similar binding modes relying on the α 4:tRNA interaction.

Interactions with the α 4 helix maintain in long Molecular Dynamics simulations

Based on the preliminary calculations described above, a model with the complete tRNA molecule was built using the tRNA position observed in model 7, which is characterized by the strongest interaction between the protein and tRNA. In the model with the complete tRNA, the orientation of the aminoacyl group in model 7 was improved using the information available from the experimental structure (PDB: 4Q24). In particular, the orientation in the model 9, where the structure of the aminoacyl group is close to its position in the experimental structure of AlbC complexed with the dipeptide analogue, as shown in Figure S1, was used as described in the Method section. The model was subjected to a long MD simulation of one μ s with no restraints applied to the protein and tRNA atoms. The RMS deviation is shown in Figure S7. The average backbone RMS deviation for the protein referenced to the crystal structures of the protein (PDB reference code 4Q24) was of 1.8 Å with the standard deviation of 0.2 Å. In agreement with previous studies, tRNA demonstrates higher structural fluctuations than the protein with the backbone average RMS deviation of 4.0 (SD 1.8) Å referenced to the crystal structure of tRNA^{Phe} (PDB reference code 4YCO). However, the position of AlbC relative to tRNA is well maintained during the entire simulation with the tRNA acceptor arm preserving the contact with α 4 helix (Figure 4A). Moreover, the covariance of atomic displacements of the two partners indicates that the AlbC catalytic site and α 4 helix fluctuations are correlated with those of the tRNA during the MD simulations indicating that AlbC interacts strongly with tRNA through its catalytic site and α 4 helix (Figure 4B).

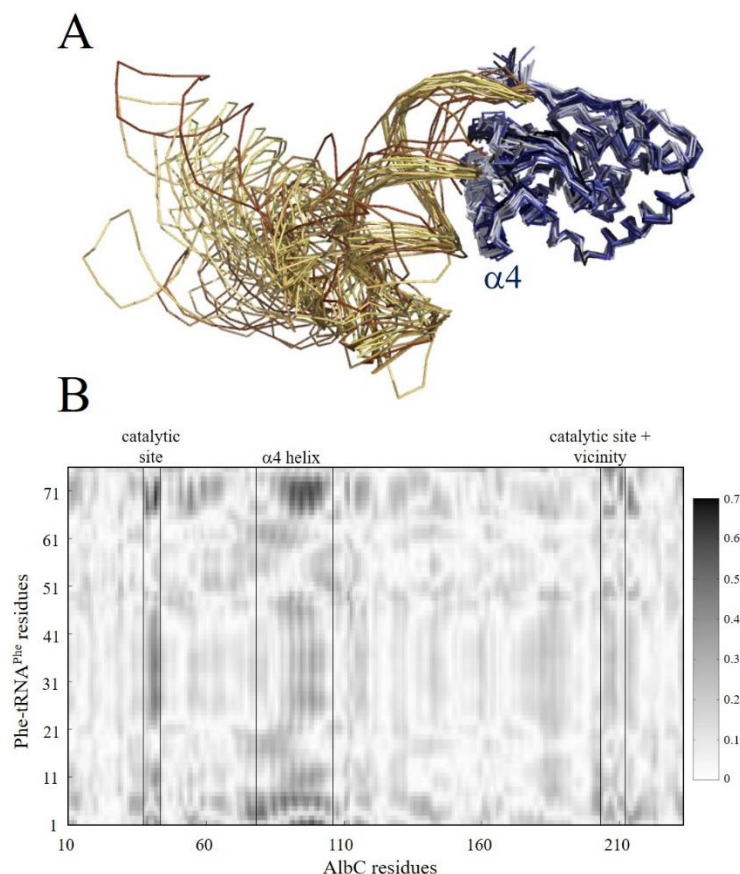


Figure 4. A) Interactions between AlbC and tRNA observed in MD simulations of the AlbC:tRNA complex. Superimposed 20 snapshots are shown taken each 50 ns from one μ s MD simulations. B) Absolute covariance for the atomic displacements of the protein backbone C α 's and tRNA backbone phosphate groups. Darker areas correspond to large values of the absolute covariance.

The binding free energy of AlbC to tRNA observed in MD simulations is shown in Figure S8. The binding free energy fluctuates near the average value of $-40.3 \text{ kcal} \cdot \text{mol}^{-1}$, but fluctuations are small with the standard deviation of $5.0 \text{ kcal} \cdot \text{mol}^{-1}$. Overall, small variations in the binding free energy and RMS deviation given above suggest that MD simulations of the AlbC:tRNA complex have mostly converged. The individual residue contributions to the total binding free energy of AlbC to tRNA are given in Table 1. Residues contributing to the total binding free energy less than $-7 \text{ kcal} \cdot \text{mol}^{-1}$ are Arg91, Lys94, Arg98 and Arg99 all belong to the $\alpha 4$ helix. Positively charged residues of the $\alpha 4$ helix, Arg80, Arg87, Arg91, Lys94, Arg98, Arg99 and Arg102 create multiple salt bridges with the phosphate backbone as shown in Figure 5A. Interestingly, in MD simulations both Arg80 and Arg99 stay very close to the tRNA substrate (5.4 \AA and 4.0 \AA respectively Figure 5A), and thus can contribute through long-range electrostatic interactions. Arg80 interacts through long-range electrostatic interactions with Asp163 and Arg99 is relatively close to C72 nucleotide both involved in the second substrate binding (12) suggesting that second substrate binding pose could be similar to that of the first substrate.

Asp95 belonging to the same helix makes a positive contribution to the total binding free energy of $4.4 \text{ kcal} \cdot \text{mol}^{-1}$ decreasing tRNA binding. To further study the role of Asp95 we performed 200 ns of

MD simulation of the Asp95Ala variant with the complete tRNA. The binding free energy of the AlbC Asp95Ala variant to tRNA observed in MD simulations has small fluctuations shown in Figure S9 with an average value of $-49.2 \text{ kcal}\cdot\text{mol}^{-1}$ (SD: $6.3 \text{ kcal}\cdot\text{mol}^{-1}$). Compared to the wild type MD simulations, the average binding free energy of the mutant is $9 \text{ kcal}\cdot\text{mol}^{-1}$ lower suggesting that Asp95 has a destabilizing role in the AlbC:tRNA interaction, due to repulsive electrostatic interactions between the Asp95 carboxylate group and tRNA phosphates. This mutation was further tested using *in vivo* experiments. The Asp95Ala variant was shown to be approximately two-three times less efficient than the wild-type protein (Table S5 and Figure S6), which demonstrate that Asp95 is implicated in interactions with tRNA. However, the effect of the Asp95Ala mutation found experimentally is reversed in comparison to the simulations results. We further tested if Asp95 can affect the enzymatic activity through long-range electrostatic effects. In particular the electrostatic potential due to the charge of Asp95 was estimated as described in the Method section using MD simulations of the AlbC in the dipeptide state. Asp95 is far away from the catalytic center: the shortest distance in the experimental structure 4Q24 between the dipeptide and Asp95 ($C\gamma$ of Asp95 and C of Phe1) is 16.4 \AA , and the distance between $C\delta$ of the catalytic Glu182 and $C\gamma$ of Asp95 is 20.3 \AA . In agreement with the long distance from the catalytic center and the fact that this residue is solvent exposed, the results show that the electrostatic effects due to Asp95 in the catalytic center are very small. For example, the interaction free energy between Asp95 and the bound dipeptide is just $-0.04 (0.01) \text{ kcal}\cdot\text{mol}^{-1}$, while electrostatic potential due to Asp95 on the closest atom $O\gamma$ of Ser37, is just $0.09 (0.01) \text{ kcal}\cdot\text{mol}^{-1}\cdot e^{-1}$. Overall, the long range electrostatic effects of Asp95 can be neglected. Finally, we tested if allosteric effects can be implicated with the Asp95Ala mutation. Figure S10 compares average structures observed in MD simulation with Asp95 and Asp95Ala with AlbC in the dipeptide state and in complex with tRNA. The positions of the catalytic residues, surface residues participating in interactions with tRNA, as well as the tRNA phosphates are all practically identical demonstrating the absence of allosteric effects. Based on these insights, and the fact that computed effect of Asp95Ala is particularly strong, we propose that the Asp95Ala mutation increases the strength of interactions with Phe-tRNA^{Phe}, but also with other tRNA forms including tRNA without the aminoacyl group. This can have a total negative effect on the cyclodipeptide production in the cellular context, where AlbC should specifically bind to the correct tRNAs (Phe-tRNA^{Phe}) and dissociate from tRNA when the aminoacyl group is transferred to the protein to become available for the next catalytic cycles.

Table 1. Contributions of individual residues to the binding free energy in $\text{kcal}\cdot\text{mol}^{-1}$. Average values over 100 frames are given. The standard deviation is given in parentheses. The energies are given only for residues with the absolute contribution to the total binding free energy larger than $1.0 \text{ kcal}\cdot\text{mol}^{-1}$ (except Arg231).

Residue	Calculated	Experimental
Lys46	-1.9 (3.0)	no effect*
Arg80	-2.6 (1.2)	strong
Arg87	-3.4 (1.6)	no effect

Arg91	-11.3 (3.3)	strong
Lys94	-9.2 (2.6)	strong
Asp95	+4.4 (0.9)/+8.9 (6.3) [§]	strong*
Arg98	-7.4 (2.2)	strong
Arg99	-9.0 (2.8)	strong
Arg102	-3.1 (2.1)	strong
Glu182	-6.4 (1.2)	strong
Arg231	-0.4 (0.2)	no effect

**In vivo* assay results from the current study (Table S5), while the experimental data for the remaining residues was compiled from the previous studies (12, 15, 20). [§]Results from the component analysis of MD simulations of the wild type protein and binding free energy difference computed using MD simulations with the Asp95Ala variant are given, respectively.

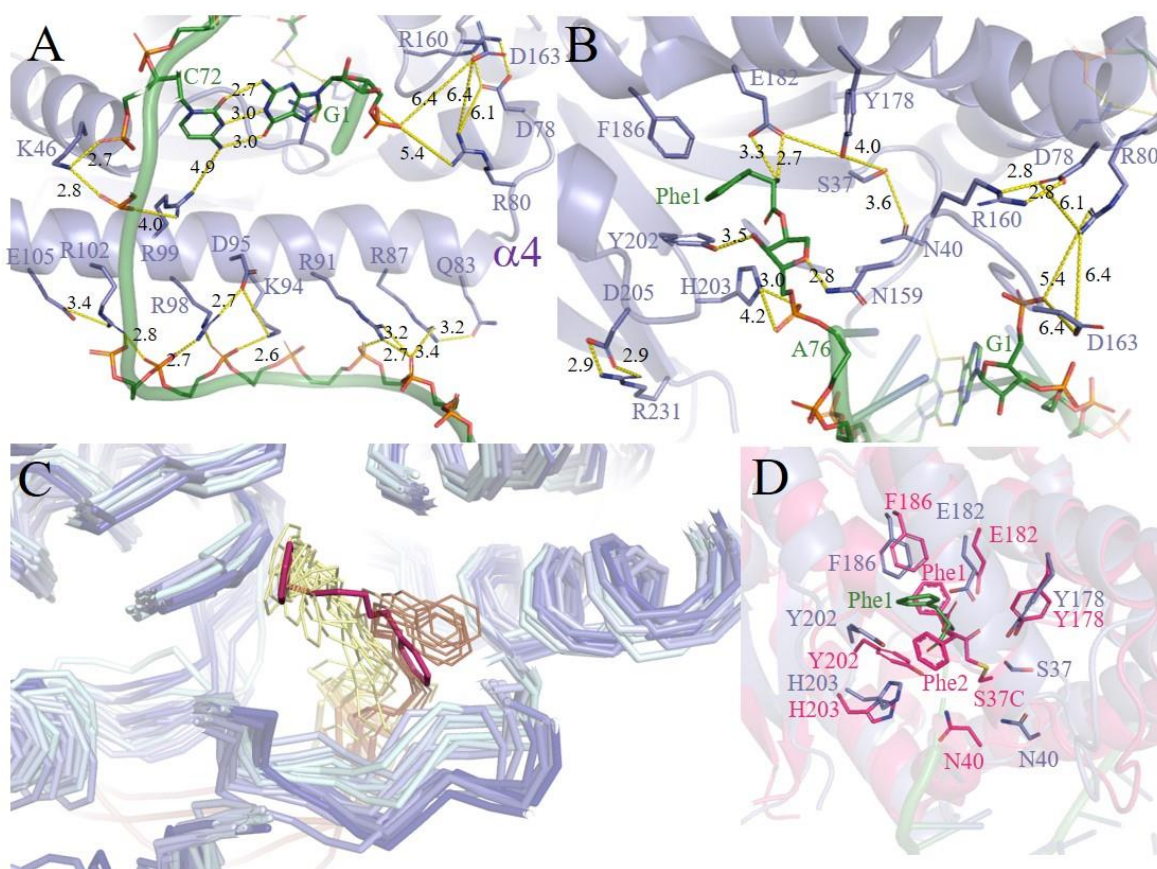


Figure 5. Interactions observed in one μ s MD simulation of the AlbC:Phe-tRNA^{Phe} complex. The protein and Phe-tRNA^{Phe} are colored in blue and green, respectively. The important inter-atom distances are shown in Å. The conformation with the binding free energy close to the average value in Figure S8 is shown. A) $\alpha 4$ helix and tRNA major groove interaction. B) Close-up view of the catalytic center. C) Superimposed 20 snapshots taken each 50 ns from the MD simulation demonstrating two predominant conformations of Phe1 colored in yellow and orange respectively. The structure of the dipeptide intermediate in the crystal structure (PDB reference code 4Q24) is colored in pink. D) Superposition of the conformation from the MD simulation with the experimental structure 4Q24 (20).

Glu182 has one of the largest contributions to tRNA binding via strong electrostatic interactions with the protonated amine group of Phe1 and helps orientate and stabilize the aminoacyl in the catalytic

pocket, shown in Figure 5B. Important catalytic residues Tyr202, His203, Tyr178, and Asn40 help stabilize the aminoacyl moiety through a network of hydrogen bonds. Overall, the structure of the catalytic center observed in the MD simulations is in good agreement with the crystal structure with the dipeptide substrate analogue (PDB code 4Q24). Importantly, residues Tyr178, Glu182, and His203 essential for the catalytic function are found in a similar position as in the crystallographic structure, shown in Figure 5D. However, two important residues, Asn40 and Ser37 are found in different orientations compared to the crystallographic structure (PDB code 4Q24). In the crystal structure, Asn40 is solvent exposed and close to Ser37 with the C β -C β distance of 3.9 Å. Arguably, this is an artefact of rigid-body docking, where the protein can be trapped in a higher energy conformation, which cannot be relaxed in the presence of tRNA. Simulations of the dissociation of the protein and tRNA is complex and beyond the scope of the current study. However, the effect of the misorientation of Ser37 and Asn40 is expected to be small, since these residues do not contribute significantly to tRNA binding, thus cannot change the overall binding position of tRNA relative to the protein.

It was previously suggested (12) that residues Asn159, Arg160, Asp163 and Asp205 contribute to the second substrate binding. Importantly, the model proposed in this work can rationalize all these contributions as shown in Figure 5B. In particular, Arg160 and Asp205 create salt bridges with Asp78 and Arg231 respectively stabilizing loops α 6- α 7 and β 6- α 8; Asn159 makes a hydrogen bond with the ribose of A76 and Asp163 has a repulsive electrostatic effect on G1 phosphate group. Overall, this suggest that the binding poses for the first and second substrate may be similar. Notably, the phenylalanine group of Phe-tRNA^{Phe} occupied in MD simulations predominantly two positions, shown in Figure 5C. The two dominant orientations are compatible with the first and second phenylalanine residues of the dipeptide analogue in the crystal structure (Figure 5). This demonstrates that the first substrate can bind by its aminoacyl group into both binding pockets, P1 and P2. These pockets are occupied by the first and second phenylalanine groups respectively in the protein crystal structure with the dipeptide analogue.

Overall, the new model presents an extended interaction between Phe-tRNA^{Phe} and AlbC through α 4 helix consistent with positive charge conservation of the helix in NYH subfamily and consistent with experimental data.

Effect of tRNA interactions on the AlbC conformation

Finally, in order to determine whether AlbC can change its conformation during its function cycle, 100 ns MD simulations were performed for AlbC in the apo and intermediate states to compare with the MD simulation of the AlbC:tRNA complex. In particular, for the apo form two independent simulations were performed using the crystal structures with PDB codes 4Q24 (20) and 3OQV (15), respectively. For AlbC in the intermediate state with Phe1 attached to Ser37, the starting structure was obtained from the structure with the dipeptide analogue by deleting Phe2 of the dipeptide analogue. The

RMS deviation between backbone atoms of the available crystal, 4Q24 and the average structure in the MD simulations is given in Table 2.

Table 2. The RMS deviation computed between the average structure from MD simulations and the crystal structure (PDB reference code 4Q24). The RMS deviation was computed after the structures were superimposed on the protein C α atoms, excluding atoms of the N- and C-termini (residues 1-14 and 231-239 respectively) due to their high structural fluctuations.

State	RMS deviation, Å
Apo form modelled	1.05
Apo form 4Q24	1.05
AlbC-Phe1 IS	1.05
AlbC:tRNA	1.33

The MD simulations of the apo form starting from two different crystal structures converge to a very similar protein conformation with 1.05 Å RMS deviation from the crystal structure with dipeptide analogue. Indeed, the two crystal structures of AlbC in the apo state and in the dipeptide analogue state used for the initial structures for MD simulations are very similar, with the RMS deviation of 0.9 Å compute for 169 C α (20). The RMS deviation for AlbC in the intermediate state, with Phe1 is practically the same as for the apo protein demonstrating that the bound Phe1 does not perturb the protein conformation. However, the RMS deviation computed for AlbC in the tRNA:AlbC complex is 1.33 Å, higher than 1.05 Å, demonstrating that for tRNA interactions the protein adopts a slightly different conformation. In particular, flexible loops between α 2 and β 3 (CL1) and between α 6 and α 7 are mainly contributing to the RMS deviation. However, the position of the backbone of the β strands and α helices, including α 4 is very similar. The α 2- β 3 loop, so called CL1 (20) contributes to the structure of catalytic site, and in all available crystal structures in the tRNA-free form the conformation of this loop is very similar, suggesting that its flexibility may be involved in tRNA binding.

Discussion

In this work, using a synergy of *in vivo* experiments and simulations, the AlbC:Phe-tRNA^{Phe} complex was investigated. Eleven initial models were obtained by rigid docking and through homology modelling based on the sequence and structure similarity with TyrRS, without imposing strict restraints on possible protein:tRNA interactions. The models were classified into four groups of possible tRNA binding positions. In all studied models, the aminoacyl moiety interaction with AlbC was remained in preliminary MD simulations indicating that the AlbC:tRNA binding relies on the aminoacyl moiety interaction with the active site residues as suggested previously (12). Furthermore, the models from the two groups with the lowest protein:tRNA binding free energies converged to very similar structures of the AlbC:tRNA complex in MD simulations, which suggests that this binding pose is the solution to the binding problem. In all models with the lowest energy, AlbC interacts with the tRNA via the α 4 helix in agreement with the biochemical experiments. Notably, the analysis of the charge conservation showed that the α 4 helix has the total charge well conserved in the entire NYH subfamily. The average

charge of the $\alpha 4$ helix is $+5.5 e$ (SD: $1.5 e$), suggesting that the binding mode of tRNA is shared by the NYH CDPs.

The model with the lowest binding free energy was then used to study the protein:tRNA complex in long 1 μs of MD simulations. The model was stable as indicated by the RMS deviation and the protein:tRNA binding free energy. Analysis of individual residue contributions identified residues, Lys46, Arg80, Arg87, Arg91, Lys94, Arg98, Arg99 and Arg102 as strongly contributing to the complex formation and stability. Asp95 was also identified by simulations to be implicated in tRNA interactions, and further validated by *in vivo* experiments in this work. Arg80, Arg87, Arg91, Lys94, Arg98, Arg99 and Arg102 have been previously studied by mutations demonstrating that all, except Arg87, are necessary for AlbC function (12, 15). All these residues except Arg80 belong to the $\alpha 4$ helix. In MD simulations, the $\alpha 4$ helix makes multiple ionic interactions by positively charged residues with the phosphate groups of both strands delimiting major groove of the tRNA acceptor stem. Similar interactions were observed in other families of enzymes using aa-tRNA to form peptide bonds. Both FemX aminoacyl-transferases and aa-tRNA protein transferases recognize the cognate aa-tRNA via its acceptor stem (23, 24). Moreover, a proposed model for L/F transferase suggests that tRNA recognition occurs from a positive cluster located on a small solvent exposed α helix (24, 61).

However, FemX aminoacyl-transferases and aa-tRNA protein transferases interaction with tRNA highly depends on tRNA sequence of the acceptor stem (23, 24). Interestingly, it was suggested that AlbC:tRNA interaction was different for the first or the second tRNA (12). For the binding of the first substrate, the interaction with the basic patch of $\alpha 4$ helix is essential. The binding of the second substrate is highly dependent on both the aminoacyl moiety and the tRNA sequence itself. It would involve the loop $\alpha 6 - \alpha 7$ delimiting the P2 pocket but not $\alpha 4$ helix (12). AlbC distinguishes between the pairs G^1-C^{72} and C^1-G^{72} , similar to TyrRSs and FemX aminoacyl-transferases (62–64). This indicates that the modelled AlbC:Phe-tRNA^{Phe} interaction corresponds to the pose of the first substrate as in our model, and there is no specific interaction with G^1-C^{72} base pair, in agreement with the experimental evidence. However, in the 1 μs MD simulation the phenylalanine of the bound tRNA inside the catalytic site predominantly occupied one of two positions compatible with Phe1 and Phe2 observed in the crystal structure (PDB reference 4Q24 (20)) in the intermediate state. The conformational transitions between the Phe1 and Phe2 positions was accomplished without significant conformational rearrangements of the protein atoms. In MD simulations, the phenylalanine group of tRNA spend around twice more time in the Phe1 position than in the Phe2 position in agreement with the previous suggestion that the Phe1 is more specific for the phenylalanine than Phe2, since the second substrate type is less strict in AlbC. This suggests that with the same tRNA binding mode, the phenylalanine group, in principle can be poised for the second step of the catalytic reaction, and both tRNA's for the first and second steps of the enzymatic reaction bind in a similar mode to AlbC. This conjecture will be tested in future studies.

Interestingly, the previous experiments demonstrate a significant decrease in the enzymatic activity or a complete loss for *Nbra*-CDPS and AlbC respectively, when the XYP and NYH motifs in these CDPSs were converted to the NYH and XYP motifs respectively (14). This clearly demonstrates that the difference in the XYP and NYH motifs is just a reflection of more profound differences between the two subfamilies. We could propose that this distinction may be contributed by the difference in interactions of CDPSs with tRNA. Indeed, the overall positive charge of the $\alpha 4$ helix, proposed in this work to be important for the NYH CDPS interactions with tRNA is not observed in the XYP family suggesting that the $\alpha 4$ helix is not significantly involved in the tRNA recognition in agreement with recent crystallographic and biochemical studies (14).

Overall, the simulation results are corroborated by biochemical experiments. In particular, the AlbC variants with basic residues belonging to $\alpha 4$ helix to be important for tRNA binding identified through mutation and conservation in sequence alignment are also the key residues for AlbC:Phe-tRNA^{Phe} in our simulations. Overall, the mechanism of the tRNA recognition by CDPS AlbC discovered in this work is expected to be pertinent for other members of the CDPS NYH subfamily, since the CDPS:tRNA interaction involves the CDPS secondary structure with a conserved positive charge in the subfamily. This is also in line with the fact that CDPS:tRNA binding involves the tRNA acceptor stem, which is observed for other non-canonical enzymes such as FemX aminoacyl-transferases and aa-tRNA protein transferases (23, 27).

Supporting Information. Details of homology modeling, Poisson-Boltzmann (PB) binding free energies for ZDOCK docking structures, force field parametrization, results for force field parametrization; supplementary tables S1 to S9 and supplementary figures S1 to S9; force field parameters.

ACCESSION CODE: UniProtKB Q8GED7

Acknowledgements: This work was supported by the French National Research Agency grant ANR-18-CE44-0002 to AA. CINES (Grant 2018-A0040710436) is acknowledged for the generous allocation of computer time. We thank Alexandre Couëtoux for performing mutagenesis experiments and Nicolas Canu for his help with LC-MS experiments.

References:

1. Gondry, M., Sauguet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., Tuphile, K., Jacquet, M., Braud, S., Courçon, M., Masson, C., Dubois, S., Lautru, S., Lecoq, A., Hashimoto, S., Genet, R., and Pernodet, J.-L. (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.* **5**, 414–420
2. Moutiez, M., Belin, P., and Gondry, M. (2017) Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem. Rev.* **117**, 5578–5618
3. Bellezza, I., Peirce, M. J., and Minelli, A. (2014) Cyclic dipeptides: from bugs to brain. *Trends Mol. Med.* **20**, 551–558
4. Bolognesi, M. L., Ai Tran, H. N., Staderini, M., Monaco, A., López-Cobeñas, A., Bongarzone, S., Biarnés, X., López-Alvarado, P., Cabezas, N., Caramelli, M., Carloni, P., Menéndez, J. C., and Legname, G. (2010) Discovery of a Class of Diketopiperazines as Antiprion Compounds. *ChemMedChem.* **5**, 1324–1334
5. Furukawa, T., Akutagawa, T., Funatani, H., Uchida, T., Hotta, Y., Niwa, M., and Takaya, Y. (2012) Cyclic dipeptides exhibit potency for scavenging radicals. *Bioorg. Med. Chem.* **20**, 2002–2009
6. Kumar, N., Gorantla, J. N., Mohandas, C., Nambisan, B., and Lankalapalli, R. S. (2013) Isolation and antifungal properties of cyclo(D-Tyr-L-Leu) diketopiperazine isolated from *Bacillus* sp. associated with rhabditid entomopathogenic nematode. *Nat. Prod. Res.* **27**, 2168–2172
7. Kwak, M.-K., Liu, R., Kwon, J.-O., Kim, M.-K., Kim, A. H., and Kang, S.-O. (2013) Cyclic dipeptides from lactic acid bacteria inhibit proliferation of the influenza A virus. *J. Microbiol.* **51**, 836–843
8. Nicholson, B., Lloyd, G. K., Miller, B. R., Palladino, M. A., Kiso, Y., Hayashi, Y., and Neuteboom, S. T. C. (2006) NPI-2358 is a tubulin-depolymerizing agent: in-vitro evidence for activity as a tumor vascular-disrupting agent. *Anticancer. Drugs.* **17**, 25–31
9. Mishra, A., Choi, J., Choi, S.-J., and Baek, K.-H. (2017) Cyclodipeptides: An Overview of Their Biosynthesis and Biological Activity. *Molecules.* **22**, 1796
10. Rhee, K.-H. (2004) Cyclic dipeptides exhibit synergistic, broad spectrum antimicrobial effects and have anti-mutagenic properties. *Int. J. Antimicrob. Agents.* **24**, 423–427
11. Lautru, S., Gondry, M., Genet, R., and Pernodet, J.-L. (2002) The Alboursin Gene Cluster of *S. noursei*. *Chem. Biol.* **9**, 1355–1364
12. Moutiez, M., Seguin, J., Fonvielle, M., Belin, P., Jacques, I. B., Favry, E., Arthur, M., and Gondry, M. (2014) Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Res.* **42**, 7247–7258
13. Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R., and Nureki, O. (2011) Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci.* **108**, 3912–3917
14. Bourgeois, G., Seguin, J., Babin, M., Belin, P., Moutiez, M., Mechulam, Y., Gondry, M., and Schmitt, E. (2018) Structural basis for partition of the cyclodipeptide synthases into two subfamilies. *J. Struct. Biol.* **203**, 17–26
15. Sauguet, L., Moutiez, M., Li, Y., Belin, P., Seguin, J., Le Du, M.-H., Thai, R., Masson, C., Fonvielle, M., Pernodet, J.-L., Charbonnier, J.-B., and Gondry, M. (2011) Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.* **39**, 4475–4489
16. 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
17. Jacques, I. B., Moutiez, M., Witwinowski, J., Darbon, E., Martel, C., Seguin, J., Favry, E., Thai, R., Lecoq, A., Dubois, S., Pernodet, J.-L., Gondry, M., and Belin, P. (2015) Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.* **11**, 721–727
18. Canu, N., Moutiez, M., Belin, P., and Gondry, M. (2020) Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat. Prod. Rep.* **37**, 312–321
19. Gondry, M., Jacques, I. B., Thai, R., Babin, M., Canu, N., Seguin, J., Belin, P., Pernodet, J.-L., and Moutiez, M. (2018) A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes. *Front. Microbiol.* **9**, 46

20. Moutiez, M., Schmitt, E., Seguin, J., Thai, R., Favry, E., Belin, P., Mechulam, Y., and Gondry, M. (2014) Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.* **5**, 5141
21. Schmitt, E., Bourgeois, G., Gondry, M., and Aleksandrov, A. (2018) Cyclization Reaction Catalyzed by Cyclodipeptide Synthases Relies on a Conserved Tyrosine Residue. *Sci. Rep.* **8**, 1–11
22. 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
23. Fonvielle, M., Chemama, M., Villet, R., Lecerf, M., Bouhss, A., Valery, J.-M., Etheve-Quelejeu, M., and Arthur, M. (2009) Aminoacyl-tRNA recognition by the FemXWv transferase for bacterial cell wall synthesis. *Nucleic Acids Res.* **37**, 1589–1601
24. Fung, A. W. S., Leung, C. C. Y., and Fahlman, R. P. (2014) The determination of tRNA^{Leu} recognition nucleotides for *Escherichia coli* L/F transferase. *RNA*. **20**, 1210–1222
25. Bouhss, A., Josseume, N., Allanic, D., Crouvoisier, M., Gutmann, L., Mainardi, J.-L., Mengin-Lecreulx, D., Heijenoort, J. van, and Arthur, M. (2001) Identification of the UDP-MurNAc-Pentapeptide:l-Alanine Ligase for Synthesis of Branched Peptidoglycan Precursors in *Enterococcus faecalis*. *J. Bacteriol.* **183**, 5122–5127
26. Hegde, S. S., and Shrader, T. E. (2001) FemABX Family Members Are Novel Nonribosomal Peptidyltransferases and Important Pathogen-specific Drug Targets. *J. Biol. Chem.* **276**, 6998–7003
27. 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
28. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 10.1038/msb.2011.75
29. Schrödinger, LLC *The PyMOL Molecular Graphics System, Version 2.0*
30. Byrne, R. T., Jenkins, H. T., Peters, D. T., Whelan, F., Stowell, J., Aziz, N., Kasatsky, P., Rodnina, M. V., Koonin, E. V., Konevega, A. L., and Antson, A. A. (2015) Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci.* **112**, 6033–6037
31. Harrington, K. M., Nazarenko, I. A., Dix, D. B., Thompson, R. C., and Uhlenbeck, O. C. (1993) In vitro analysis of translational rate and accuracy with an unmodified tRNA. *Biochemistry*. **32**, 7617–7622
32. Mikkelsen, N. E., Johansson, K., Virtanen, A., and Kirsebom, L. A. (2001) Aminoglycoside binding displaces a divalent metal ion in a tRNA–neomycin B complex. *Nat. Struct. Biol.* **8**, 510–514
33. Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273
34. Huang, J., and MacKerell, A. D. (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145
35. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
36. Neria, E., Fischer, S., and Karplus, M. (1996) Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921
37. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*. **102**, 3586–3616

38. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. **30**, 1771–3
39. Iwakiri, J., Hamada, M., Asai, K., and Kameda, T. (2016) Improved Accuracy in RNA-Protein Rigid Body Docking by Incorporating Force Field for Molecular Dynamics Simulation into the Scoring Function. *J. Chem. Theory Comput.* **12**, 4688–4697
40. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–802
41. Darden, T., Treatment of Long-Range Forces and Potential. In *Computational Biochemistry & Biophysics*, Becker, O. M.; MacKerell, A. D., Jr; Roux, B.; Watanabe, M., Eds. Marcel Dekker, NY.: 2001.
42. Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L., Mackerell, A. D., and Jr. (2012) Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **8**, 348–362
43. Huang, J., and MacKerell, A. D. (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–45
44. Byrne, R. T., Konevega, A. L., Rodnina, M. V., and Antson, A. A. (2010) The crystal structure of unmodified tRNA Phe from Escherichia coli. *Nucleic Acids Res.* **38**, 4154–4162
45. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kucsera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., Karplus, M., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–614
46. Gilson, M. K.; Sharp, K. A.; Honig, B. H., Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* 1988, 9 (4), 327–335.
47. Im, W., Beglov, D., and Roux, B. (1998) Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **111**, 59–75
48. Gilson, M. K., and Honig, B. H. (1987) Calculation of electrostatic potentials in an enzyme active site. *Nature*. **330**, 84–86
49. Kuhn, B., and Kollman, P. A. (2000) Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **43**, 3786–3791
50. Sham, Y. Y., Muegge, I., and Warshel, A. (1998) The Effect of Protein Relaxation on Charge-Charge Interactions and Dielectric Constants of Proteins. *Biophys. J.* **74**, 1744–1753
51. Schutz, C. N., and Warshel, A. (2001) What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins*. **44**, 400–417
52. Genheden, S., Kuhn, O., Mikulskis, P., Hoffmann, D., and Ryde, U. (2012) The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant. *J. Chem. Inf. Model.* **52**, 2079–2088
53. Ma, J. (2005) Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure*. **13**, 373–380
54. Cummins, P. L., Ramnarayan, K., Singh, U. C., and Gready, J. E. (1991) Molecular dynamics/free energy perturbation study on the relative affinities of the binding of reduced and oxidized NADP to dihydrofolate reductase. *J. Am. Chem. Soc.* **113**, 8247–8256
55. Aleksandrov, A., Schuldt, L., Hinrichs, W., and Simonson, T. (2008) Tet Repressor Induction by Tetracycline: A Molecular Dynamics, Continuum Electrostatics, and Crystallographic Study. *J. Mol. Biol.* **378**, 898–912
56. Simonson, T., Carlsson, J., and Case, D. A. (2004) Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **126**, 4167–4180
57. Canu, N., Belin, P., Thai, R., Correia, I., Lequin, O., Seguin, J., Moutiez, M., and Gondry, M. (2018) Incorporation of Non-canonical Amino Acids into 2,5-Diketopiperazines by Cyclodipeptide Synthases. *Angew. Chem.* **57**, 3118–3122

58. Felder, C. E., Prilusky, J., Silman, I., and Sussman, J. L. (2007) A server and database for dipole moments of proteins. *Nucleic Acids Res.* **35**, W512–W521
59. Aleksandrov, A., and Simonson, T. (2008) Molecular Dynamics Simulations of the 30S Ribosomal Subunit Reveal a Preferred Tetracycline Binding Site. *J. Am. Chem. Soc.* **130**, 1114–1115
60. Yaremchuk, A. (2002) Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.* **21**, 3829–3840
61. 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
62. Bedouelle, H. (2013) *Tyrosyl-tRNA Synthetases*, Landes Bioscience, [online] <https://www.ncbi.nlm.nih.gov/books/NBK6553/> (Accessed June 8, 2020)
63. Bonnefond, L., Giegé, R., and Rudinger-Thirion, J. (2005) Evolution of the tRNA^{Tyr}/TyrRS aminoacylation systems. *Biochimie.* **87**, 873–883
64. 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. Mol. Biol.* **10**, 425–432

For Table of Contents Use Only.

