Mechanistic Insights into Cyclodipeptide Formation by Cyclodipeptide Synthases: A Preliminary Exploration on Pathways and Catalytic Residues

Akshita Goel£[a], Indu Negi£[a], Ankur Ganesh Pandey[a], John F. Trant*[b],[c], and Purshotam Sharma*[a],[b]

[a] Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India.

£These authors contributed equally to this work.

[b] Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Ave. Windsor ON, N9B 3P4, Canada

[c] Binary Star Research Services, LaSalle, Ontario N9J 3X8, Canada

*Email: j.trant@uwindsor.ca, psharma@uwindsor.ca

Abstract

Cyclodipeptide synthases (CDPSs) are enzymes that synthesize cyclodipeptides using two aminoacyl-tRNAs as substrates, but their mechanism remains unclear. This study aims to elucidate the mechanism of AlbC, a CDPS that produces cyclo(L-Phe-L-Phe). We employed small-model quantum mechanics (QM) calculations to propose an intrinsic pathway and molecular dynamics (MD) simulations to identify key catalytic residues involved in this process. The mechanism involves three main steps: activation of Ser37 and the first tRNA[Phe] to form a Phe-enzyme intermediate, binding of the second tRNA[Phe] to form a dipeptidyl enzyme intermediate, and intramolecular cyclization to yield the cyclodipeptide. Our QM calculations suggest that Ser37 can be activated through direct transfer of its hydroxyl proton to the O3' atom of the first substrate. MD simulations highlight the roles of Gly35, Asn40, and His203 in stabilizing the Phe-enzyme intermediate, thus lowering the calculated intrinsic barrier. In the second step, the dipeptidyl enzyme intermediate is favored over the nucleoside intermediate and is stabilized by Asn40, Gln182, and His203 in the AlbC active site, where Gln182 may act as a catalytic base. Additionally, Asn159 and His203 contribute to lowering the significant energy barrier observed in QM calculations for intramolecular cyclization, with Glu182 potentially serving as a catalytic base during this process. Overall, our results support the roles of Asn40 and His203 throughout all mechanistic steps, while highlighting Glu182's involvement in the formation of the dipeptidyl enzyme intermediate and intramolecular cyclization steps. These insights can guide future enzymatic modeling studies of AlbC and potentially other CDPS enzymes using similar approaches.

1. Introduction.

Cyclodipeptide synthases (CDPSs) are a recently identified family of enzymes that use two aminoacyl-tRNAs (aa-tRNAs) as substrates to synthesize the two peptide bonds of cyclodipeptides (CDPs).[1,2] These CDPs are precursors of many natural products, such as diketopiperazines with antibacterial, antifungal, antiviral, antitumor and immunosuppressive functions.[3,4] Should their mechanism be better understood, CDPSs could be engineered for chemoenzymatic[5,6] synthesis of currently inaccessible cyclodipeptides, providing access to a variety of new peptide scaffolds for the further development of peptide-associated therapeutics.[2,7]

Seven CDPSs have been structurally characterized in their *apo* form using X-ray crystallography.[8-11] These include AlbC from *Streptomyces noursei* (PDB code: *3OQV*),[8] Rv2275 from *Mycobacterium Tuberculosis* (*2X9Q*),[9] YvmC from *Bacillus Licheniformis* (*3OQI*),[10] *Shae*-CDPS from *Staphylococcus Haemolyticus* (*6EZ3*),[11] *Nbra*-CDPS from *Nocardia Brasiliensis* (*5MLQ*),[11] *Rgry*-CDPS from *Rickettsiella Grylli* (*5MLP*)[11] and *Fdum*-CDPS from *Fluoribacter Dumoffii* (*5OCD*).[11] However, two complexes ― one involving AlbC bound to the *N*-carbobenzyloxy-L-Phe-methyl ketone (ZPK) ligand that mimics the diphenylalanyl-intermediate of CDP synthesis (*4Q24*), and the other involving Phe-tRNA[Phe] bound to Cglo-CDPS of *Glomeribacter gigasporarum* (*6Y4B*)[12]― have been crystallized in their *holo* forms. Common among all CDPSs is a conserved Rossmann-fold domain similar to the catalytic domain of the two class Ic AA-tRNA synthetases (TyrRS and TrpRS).[8-10,13]

CDPSs can be classified into two subfamilies (NYH and XYP), depending on the identities of the conserved catalytic residues at positions 40, 202 and 203 (numbered according to the AlbC crystal structure).[11,14] AlbC,[8] Rv2275,[9] YvmC[10] and *Shae*-CDPS,[11] belong to the 'NYH' subfamily; *Nbra*-CDPS,[11] *Rgry*-CDPS[11] and *Fdum*-CDPS[11] belong to the 'XYP' subfamily. Tyr202 is strictly conserved in both subfamilies, but the XYP subfamily replaces the Asn40 and His203 of the NYH subfamily by a non-conserved residue (X40) and Pro203, respectively. Structurally, the two subfamilies mainly differ in the positioning of the first half of their Rossmann fold, although the catalytic residues retain similar relative positioning as the active sites are very similar.[11]

The catalytic mechanism of the NYH enzyme AlbC has been studied using experimental methods.[8,15,16] These studies suggest that the catalytically-active Rossmann-fold domain of AlbC consists of two pockets ― P1 and P2― which accommodate the first and second aa-tRNA, respectively (Figure 1).[8,15] There are eight residues lining P1, whereas P2 is delineated by two loops, CL1 (residues 36−43) and CL2 (residues 202−207), that bear catalytic

3

residues and the C-terminal regions of the α6 and α7 helices (Figures 1 – 3). The first tRNA[Phe] that enters the hydrophobic pocket P1 of AlbC, is stabilized by basic residues located in the α4 helix of AlbC.[16,17] The Phe moiety of this tRNA is transferred from its terminal adenine (Phe-A76) to the conserved Ser37 residue of AlbC to form the phenylalanyl enzyme intermediate.[8] For this transfer, Ser37 activation is believed to be achieved by a concerted proton shuttling mechanism that involves two vicinal hydroxyls (i.e., 2′-OH and 3′-OH) of the ribose sugar of amino acylated A76 nucleoside of tRNA moiety. This proposal is based mostly on supposition as there is no suitable amino-acid residue present in the AlbC active site to mediate proton transfer.[11] The second AA-tRNA, enclosed in the relatively wider P2 cavity, then reacts with the phenylalanyl enzyme intermediate through its aminoacyl moiety. This reaction transfers the Phe moiety from Phe-A76 of second tRNA[Phe] to phenylalanyl enzyme intermediate to form an intermediate[8] which then undergoes intramolecular cyclisation to form the dipeptide (Figures 2 and 3).[18] Since catalytic residues involved in the formation of reaction intermediates are conserved across CDPSs, similar steps are believed to be used by all NYH and XYP family CDPSs.

Despite the availability of X-ray crystallographic data,[15] the mechanistic details of CDP synthesis by the CDPSs are not well understood. Although the mechanism has been proposed based on the identification of a few near-intermediates,[15] there is no experimental support available in terms of kinetic or biochemical studies. Furthermore, despite previous computational analysis of a single (*i.e.*, intramolecular cyclization of the dipeptidyl intermediate) step of AlbC catalysis,[18] plausible mechanistic details of the other required steps in the proposal are not available. A systematic computational analysis of each step of the proposed reaction mechanism will help identify structural features of intermediates that may likely form within the active site of enzyme, as well as identify key active-site residues involved in AlbC catalysis. This provides a clear rationale for the mechanism, and also makes predictions that can be experimentally tested using site-directed mutagenesis to support the proposed mechanisms.

The present work couples quantum mechanical (QM) methods on chemically relevant tractable models with molecular dynamics (MD) simulations on full enzyme models to provide preliminary insight into the mechanism of AlbC catalysis, specifically using Phe-Phe as the model CDP. We employ small model QM calculations to analyse the formation of the first phenylalanyl enzyme intermediate resulting from the transfer of the Phe moiety from tRNA[Phe] to Ser37 of AlbC (Figures 3 and 4, step 1). Subsequently, the attack of the Phe-A76 residue of

4

the second tRNA is considered (Figures 3, 5 and 6, steps 2 and 2′). Finally, the intramolecular cyclization of dipeptide is analysed (Figures 3 and 7). Although the focused QM models may not fully provide a complete description of the chemical reactions occurring in the AlbC active site as they cannot model the entire protein in this detail, when coupled with MD simulations, this approach will, nevertheless, reveal intrinsically preferred mechanisms of the main steps, and identify key interacting residues in the vicinity of the reaction centre. This workflow has proven accurate in the description of other enzymatic mechanisms of action.[19-21] Specifically, whereas simple QM models may help characterize different pathways for the cyclodipeptide formation reaction, MD simulations will help identify reaction-ready substrate and intermediate conformation in the active site of AlbC, along with interactions of the surrounding amino acid residues. The identified routes can be used as a guide for large-scale enzymatic modelling of AlbC using large QM/MM methods. Further, despite our focus on the AlbC enzyme, similar models can be employed to understand the mechanism of other enzymes with CDP synthesis activity (e.g., Rv2275,[9] YvmC,[10] *Shae*-CDPS,[11] *Nbra*-CDPS,[11] *Rgry*-CDPS[11] and *Fdum*-CDPS[11]).

2. Computational Details.
2.1. QM Calculations.

To investigate the AlbC catalysed mechanism of CDP synthesis (Figures 2 and 3), the crystal structure of AlbC bound to ZPK ligand (PDB code: *4Q24*)[22] was used as a starting point. For modelling the first step of CDP synthesis, the crystallographic location of ZPK was used to position the phenylalanyl (Phe-) moiety of the Phe-A76 residue of tRNA[Phe], which was previously proposed to directly interact with AlbC.[8,15] The O5′ of Phe-A76 was capped with a methyl group in order to prevent non-native interactions of 5′-OH group in absence of rest of the tRNA chain. Furthermore, the Cys37 present in the crystal structure was modified to the native, catalytically important Ser by replacing its side chain sulphur atom with oxygen. A reduced model of Ser37 was employed which replaces the main chain amino and carbonyl groups of Ser37 with hydrogen atoms thereby avoiding spurious interactions that can potentially occur during geometry optimizations in absence of rest of the protein chain. After retaining this reaction centre, the rest of the enzyme skeleton was removed in accordance with previous small-model quantum chemical studies on enzymatic reactions.[19,20,23]

Since Phe-A76 was not originally present in the crystal structure, the preferred conformation of the initial O5′-methylated Phe-A76 nucleoside substrate was determined. An

5

internal-coordinate Monte Carlo conformational search was carried out across all rotatable dihedral angles of Phe-A using the AMBER force field (HyperChem 8.0.3).[24] More than 100 conformers were obtained from this conformational search. The ten lowest-energy conformers were selected for full optimization using B3LYP/6-31G(d), in accordance with a previous study.[25] Zero-point vibrational energy (ZPVE) corrected B3LYP/6-311+G(2df,p) single-point energy calculations were carried out on the resulting optimized conformations to identify the lowest energy conformation. This was used for further studies.

The stationery points (reactants, transition states, intermediates and products) were subjected to gas-phase geometry optimizations using B3LYP/6-31G(d,p). ZPVE-corrected single-point energies were then calculated using B3LYP/6-311+G(2df, 2pd). The stationary points on the reaction surfaces were characterized as minima (reactants, intermediates, and products) or first-order saddle points (transition states) using the sign of the calculated vibrational frequency. Further B3LYP/6-31G(d,p) intrinsic reaction coordinate (IRC) scans were carried out to confirm the viability of the pathways connecting the calculated transition states to the reactants, intermediate, and/or product of each step. Subsequently, to understand the effect of dispersion corrections, calculations were re-performed at B3LYP-D3BJ/6-311+G(2df, 2pd)//B3LYP-D3BJ/6-31G(d,p) on all optimized geometries (Tables 1 and 2). Furthermore, the calculations were reperformed using additional functionals (*i.e.,* B97D[26] and M06-2X[27]) to understand method-dependence of the derived results. Additionally, all gas-phase calculations using these functionals were repeated in implicit water to study the influence of the solvent on reaction mechanism, using the integral equation formalism variant polarizable continuum model (IEFPCM) method (Figures S1 – S15). All quantum chemical calculations were carried out using the Gaussian 16 (rev. B.01) suite of programs.[28]

2.2. MD simulations

2.2.1. Initial Models. For MD simulations, the initial model for the substrate-bound enzyme was built using the structure of AlbC bound to ZPK (PDB code: *4Q24*). The crystal structure was reverted to the natural protein, changing the Cys37 to the correct Ser37 residue, and replacing ZPK with Phe-A76 which represents terminal 3′ aminoacylated nucleotide of tRNA^{Phe}, thereby serving as the reduced model of tRNA^{Phe}. The structure was relaxed in water box (*vide infra*).

The "statistics" command as implemented in the cpptraj module[29] of AMBER 20[30] was used to analyse the distribution, over the simulation, of the distance between Oγ of Ser37 and C10 of Phe-A76, as well as the distance between O3' of Phe-A76 and Hγ of Ser37. These atoms

6

directly participate in forming transition state in QM calculations (Figures 2 and 3). The resulting simulation frames were clustered with respect to location of these atoms (*i.e.*, Oγ, Hγ of Ser37 and O3', C10 of Phe-A76, Figures 2 and 3), using the K-means clustering algorithm as implemented in the cpptraj module of AMBER 20.[30] The chosen representative structure, that with the conformation at the reaction center most similar to the QM model (Table S1), was used as the starting geometry for the simulation of the complex containing the phenylalanyl enzyme intermediate (Figures 2 and 3). The frames from this simulation were then clustered with respect to the location of Oγ, Hγ, C10 and O3' atoms of the phenylalanyl enzyme intermediate, and a representative structure was obtained which was used as an initial model for simulation that involves the second molecule of Phe-A76 and phenylalanyl enzyme intermediate. The distribution of the distances between N11$^{(1)}$ and C10$^{(2)}$ (Figure 3) as well as distances between O3'$^{(2)}$ and HN11$^{(1)}$ (Figure 3) was then analysed using the 'statistics' command and the resulting trajectory was clustered with respect to location of these atoms (i.e., N11$^{(1)}$, C10$^{(2)}$, O3'$^{(2)}$ and HN11$^{(1)}$, Figure 3). The optimal cluster was used to prepare initial model for simulation of the complex containing dipeptidyl enzyme intermediate (Table S2). After simulation, the simulation frames were then clustered with respect to location of N11$^{(1)}$, C10$^{(2)}$, O3'$^{(2)}$ and HN11$^{(1)}$ atoms of the second Phe moiety of dipeptidyl enzyme intermediate to obtain the representative structure which was used as initial model for simulation of the complex containing final CDP product. The representative structure was obtained from simulation of the complex which was then clustered with respect to location of N11$^{(1)}$, N11$^{(2)}$, C10$^{(1)}$ and C10$^{(2)}$ atoms (Figure 3), to analyse structural properties.

2.2.2. Simulation Details. A consistent methodology was employed for the MD simulations of the complexes belonging to each step of the enzymatic CDP synthesis. MD simulations were performed using the ff14SB force field[31] with AMBER 20.[30] Each complex was neutralized using three sodium ions and solvated in a 10 Å octahedral box filled with TIP3P water molecules.[32] Partial atomic charges required for parametrization of different intermediates formed in each step of the reaction were obtained (Table S4 – S8) using RED server 2.0,[33] and HF-6-31G* optimized geometry was used as starting structure for charge calculation using the RESP-A1 method.[34] Furthermore, the antechamber module[35] of AmberTools was used to assign AMBER atom types. Na$^+$ ions were simulated using Joung and Cheatham's ion parameters.[36] The particle mesh Ewald (PME) method[37] and Lennard-Jones potential were used to approximate electrostatic and van der Waals interactions respectively and a 10 Å cut off was used to approximate all nonbonded interactions.

The solvent and Na^+ ions were first minimized using 10,000 steepest descent (SD) steps and 10,000 conjugate gradient (CG) steps with a restraint of 500 kcal mol^{-1} Å^{-2} on the protein and substrate. This was followed by an unrestrained minimization of the entire system using same number of SD and CG steps. The system was then heated to 300 K with 60 K incremental steps, using a restraint of 10 kcal mol^{-1} Å^{-2}, through the Langevin temperature equilibration scheme.[38,39] Finally, a 500 ns unrestrained production simulation was run using an NPT ensemble and the cuda version of the PMEMD module of AMBER 20.[30] The temperature was maintained at 300 K using the Langevin temperature scheme and pressure was maintained at 1 atm using a Monte Carlo barostat.[40] The bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[41] Hydrogen mass repartitioning[42] was used for the heating and production simulation steps with a 4 fs time step. In total, five 500 ns simulations (one for substrate bound complex, three for intermediate complexes, and one for the product complex), amounting to 2.5 μs of total simulation time, were conducted.

2.2.3. MD Simulation Analysis. The visualization of the simulation data was done using PyMOL.[43] All MD simulation analysis, including the calculation of the backbone root mean square deviation (rmsd), the average interatomic distances, the stacking interaction energies and the hydrogen-bonding occupancies were carried out using the cpptraj module of AMBER 20.[30] The hydrogen-bonding occupancies were determined by imposing a cutoff value of 3.4 Å for the donor−acceptor distance, and 120° for the donor−hydrogen−acceptor angle, using the "hbond" command of cpptraj and inserting these values as parameters. The "lie" command of cpptraj was used to calculate average and standard deviation in the hydrogen-bonding and stacking interaction energies.

3. Results and Discussion.

Following substrate association, AlbC initiates CDP synthesis by activating the catalytically active site by deprotonating its Ser37 with O3′ of Phe-A76 of the aminoacylated tRNA^{Phe} substrate. The Phe residue is then transesterified to the now highly nucleophilic Ser37 (Figure 3). The next step involves attack of Phe-A76 of the second aminoacylated tRNA^{Phe} at Ser37 bound to phenylalanine moiety, which is followed by last step that involves intramolecular cyclization of the resulting peptide, releasing Ser37, and dissociation of the CDP from the active site. We have modelled this pathway using DFT, and it appears plausible (Figure 3). The geometries of the distinct species along the pathway, and the transition states, were optimized at the B3LYP-D3BJ-6-311+G(2df, 2pd)//B3LYP-D3BJ/6-31G(d,p) level, and the energies

8

were calculated using three different commonly employed QM functionals (Tables 1 and 2). We will look in detail at each mechanistic step in turn.

3.1. Ser activation step.

3.1.1. QM calculations. A previous biochemical study on AlbC revealed that no enzyme residue in the vicinity of Ser37 is likely responsible for its activation as a nucleophile.[8] Therefore, direct transfer of proton from Ser37 to Phe-A76 of tRNA[Phe] seems to be the only viable alternative. In this context, the activation of Ser37 is suggested to occur *via* a proton shuttling mechanism, that involves the 2′- and 3′-hydroxyls of the ribose sugar of Phe-A76.[8] However, our calculations revealed that any proposal involving 2′-OH mediation does not lead to a viable transition state. Alternatively, we propose that the activation of Ser37 can occur through direct transfer of its hydroxyl proton to the O3′ of the ribose moiety of Phe-A76 through a concerted pathway that likewise involves transfer of the Phe residue to Ser37, concomitant with elimination of A76 (TS1, Figure 4A, B). In this pathway, reactant complex (RC1) adopts a *syn* conformation about the glycosidic bond, a C3′-*endo* pucker of Phe-A76's sugar moiety, and an *anti*-positioning of Ser37 with respect to the adenine moiety of Phe-A76. During this shift, the Oγ of Ser37 weakly interacts with the ribose O3′ of Phe-A76 (r(OγH–O3′$^{(1)}$) = 2.229 Å and ∠(O3′$^{(1)}$···H–Oγ)) = 152.9°, Figure 4A, B) and the amino group of Phe forms an intramolecular hydrogen bond with the 2′-OH group of ribose moiety of Phe-A76 (r(N11$^{(1)}$···HO2′$^{(1)}$) = 2.035 Å and ∠(O2′$^{(1)}$–H···N11$^{(1)}$) = 176.4°, Figure 4A, B). This forms a complex bicyclic two hydrogen bond closed transition state that greatly facilitates the concerted transesterification.

This TS1 locates the hydrogen atom midway between Oγ of Ser37 and O3′ (r(Oγ–H) = 1.202 Å, r(O3′$^{(1)}$–H) = 1.200 Å and ∠(O3′$^{(1)}$···H···Oγ)) = 137.0°, Figure 4A, B). The carbonyl carbon of the Phe moiety is also positioned midway between the O3′ of Phe-A76 and Oγ of Ser37 (r(C10$^{(1)}$–O3′$^{(1)}$) = 1.805 Å and r(C10$^{(1)}$–Oγ) = 1.757 Å, Figure 4A, B). Furthermore, geometrical characteristics of hydrogen bond between the amino group of Phe and 2′-OH of ribose of A76 improve in the TS1 compared to RC1 (r(N11$^{(1)}$···H2′) = 1.918 Å and ∠(O2′$^{(1)}$–H···N11$^{(1)}$) = 161.0°, Figure 4A, B) contributing to a lower energy barrier. More importantly, although the previously-proposed proton abstraction from Ser37 by 2′-OH of Phe-A76 is not possible,[8] 2′-OH still participates in the mechanism by stabilizing TS1 through hydrogen bonding with amine group of Phe moiety.

9

The predicted barrier to Ser activation and release of A76 is 148.2 kJ mol^{-1} at the B3LYP-D3BJ level, although inclusion of solvent slightly raised the energy barrier by 2.2 kJ mol^{-1} (Table 2). Compared to gas phase B3LYP-D3BJ, the barrier to intermediate formation is raised (by up to 18 kJ mol^{-1}) using M06-2X, whereas B97D functional decreases the predicted barrier by 32 kJ mol^{-1}. However, inclusion of solvent slightly increases the barrier (by 13.7 kJ mol^{-1}) using the B97D functional and decreases the barrier (by 4.1 kJ mol^{-1}) using M06-2X. Regardless, no significant change in structural parameters of the RC1, TS1 and IC1 were observed using different functionals in gas phase or on inclusion of solvent (Figures S1 – S5). This is an admittedly high barrier, although not forbidden since these are the small QM models which convey what an intrinsic pathway would be if there is minimal enzymatic catalysis. However, in amalgamation with MD studies (*vide infra*), our study offers valuable clues for understanding the likely features of the AlbC active site in the holo form, and the associated catalytic pathways.

The activation of Ser37 and its attack on Phe-A76 finally results in the formation of first intermediate (IC1) with completely cleaved O3′$^{(1)}$–C10$^{(1)}$ bond (Figures 4A, B and S1 – S5) which leads to formation of phenylalanyl enzyme intermediate (IC1a) that contains a C10$^{(1)}$–Oγ bond (1.335 Å) and release of A76 (Figure 4A, B). This intermediate is stabilized by intramolecular hydrogen bonding between 2′-OH group of A76 ribose and the amino group of Phe (r(N11$^{(1)}$–HO-2′) = 1.820 Å and ∠(N11$^{(1)}$⋯H⋯O2′)) = 160.7°, Figures 4A, B and S1 – S5). It is calculated to be only very slightly uphill from the reactants.

3.1.2. MD simulations. MD simulation of the reactant complex containing AlbC and Phe-A76, built based on structural information obtained from QM calculations, reveals a catalytically important conformation of AlbC that accommodates the amino-acid side chain of Phe-A76 (Figure S16), and positions Oγ of Ser37 (RC1a) in proximity to the carbonyl group of Phe-A76 (RC1b) substrate. Furthermore, the carbonyl oxygen of RC1b forms a hydrogen bond with hydroxyl group of active site Tyr202 (35.1% occupancy and –6.7 kcal mol^{-1} average interaction strength, Table S3) which is previously known to play a role in last (cyclization) step of the reaction.[18] Similarly, the carbonyl oxygen of Ile36 shows hydrogen bonding with the amino nitrogen (N11) of Phe-A76 (34.4% occupancy and –5.3 kcal mol^{-1} interaction energy, Table S3), and O2′ of Phe-A76 (34.7% occupancy and –5.3 kcal mol^{-1} interaction energy, Table S3). Additionally, the A76 moiety is stabilized through hydrogen bonding between sp^2 nitrogen (NE2) of the imidazole ring of His203 and N7 atom of adenine moiety of RC1b (26.8% occupancy and –6.6 kcal mol^{-1} interaction energy, Table S3). This network of interactions may

help stabilize the substrate in the reaction-ready conformation and may hence lower the high intrinsic reaction barrier observed in QM calculations.

To better understand the features of the reaction-ready conformation within the enzyme active site, the simulation frames were binned according to distance between atoms that directly participate in the reaction *i.e.*, the distance between Oγ of Ser37 and C10 of Phe-A76, and distance between O3' of Phe-A76 and Hγ of Ser37 (Figure 3). A total of six distinct conformational clusters were obtained. Representative structure of cluster with the closest distance between Oγ and C10 (3.6 Å, Table S1), and O3' and Hγ (2.8 Å, Table S1) and thereby the most optimal conformation at reaction centre, reveals involvement of O2' in hydrogen bonding with Oγ of Ser37 (r(O2'–HOγ) = 1.7 Å and ∠(O2′···H–Oγ) = 167.7°, Figure 4C). However, in synchrony with QM calculations, this conformation does not support the previously-proposed participation of O2' in proton shuttling to O3', since the H of O2' of Phe-A76 is not in an accurate position to interact with O3' of Phe-A76 (r(O2'$^{(1)}$H–O3′$^{(1)}$) = 2.8 Å and ∠(O3′$^{(1)}$···H–O2'$^{(1)}$)) = 17.1°, Figure 4C).[8] Thus, both QM and MD analysis suggests that Ser37 directly transfers its hydroxyl proton to the O3′ of Phe-A76.

Simulation of the structure containing the phenylalanyl enzyme intermediate obtained after removal of A76 reveals hydrogen-bonding mediated stabilization within the AlbC pocket (Figure 4C). Specifically, the carbonyl oxygen of the Phe moiety of IC1a hydrogen bonds with side chain amide nitrogen (ND2) of Asn40 (92.3% occupancy and –7.0 kcal mol^{-1} interaction energy, Table S3) whereas the amino group of IC1a is stabilized through hydrogen bonding with the sp^2 nitrogen (NE2) of the imidazole ring of His203 (31.7% occupancy and –3.4 kcal mol^{-1} interaction energy, Table S3) and carbonyl oxygen of Gly35 (15% occupancy and –5.9 kcal mol^{-1} interaction energy, Table S3).

Simulations suggest that the Phe-A76 substrate (RC1b) is stabilized in the AlbC pocket through hydrogen bonding with Ile36, Ser37, Tyr202 and His203 (Table 3) whereas the phenylalanyl intermediate (IC1a) is stabilized through hydrogen bonding with Gly35, Asn40 and His203 (Table 3). The differences between the substrate and intermediate interactions suggest that the enzyme-active site undergoes slight reorganization after the first step to better accommodate the intermediate following the release of A76.

3.2. Attack of the second Phe-A of the tRNA.

3.2.1. QM calculations. Following the first step, the Phe-A76 moiety of the second tRNA^{Phe} attacks the phenylalanyl enzyme intermediate (IC1a, Figures 4). As previously proposed,[15] this step involves two mutually-exclusive mechanistic possibilities. In the first mechanism, the O3′

11

of Phe-A76 of the second tRNA^{Phe} (denoted as RC2b, Figure 5) faces N11–H of the phenylalanyl enzyme intermediate (denoted as RC2a) formed in step 1 [r(O3′$^{(2)}$–HN11$^{(1)}$) = 2.876 Å and r(C10$^{(2)}$–N11$^{(1)}$) = 3.197 Å, Figure 5A, B]. Thus, N11$^{(1)}$ of RC2a (Figure 5A, B) acts as an H donor to O3′ of the ribose of RC2b forming the dipeptidyl enzyme intermediate (IC2a), and releasing the A76 nucleoside of tRNA (IC2b). Consideration of the geometry of the associated transition state (TS2) reveals how N11 delivers H to O3′ (r(N11$^{(1)}$ – H) = 1.089 Å, (H – O3′$^{(2)}$) = 1.531 Å and ∠(O3′$^{(1)}$···H···N11$^{(1)}$) = 128.5°, Figure 5A, B) and reacts with C10 (r(O3′–C10 = 2.036 Å, Figure 5A, B) in a concerted manner by crossing a barrier of 126.0 kJ mol^{-1} to form the intermediate complex (IC2) that includes two species ― the dipeptidyl enzyme intermediate (IC2a) with the newly formed N11$^{(1)}$–C10$^{(2)}$ bond (r(N11$^{(1)}$–C10$^{(2)}$) = 1.356 Å, Figure 5A, B) and the A76 nucleoside (IC2b) formed as result of complete cleavage of the O3′$^{(2)}$–C10$^{(2)}$ bond between A76 and Phe (r(O3′$^{(2)}$–C10$^{(2)}$ = 3.832 Å), Figure 5A, B and Table 1).

In contrast, in the alternate mechanism, N11 of RC2b serves as the H-donor to the Oγ of Phe-Ser37 (RC2′a, Figure 6), thus forming the dipeptidyl nucleoside intermediate (IC2′a, Figure 6) which cleaves the bond between Ser37 (IC2′b, Figure 6) and the Phe residue (Figure 6). This is achieved through a concerted pathway that involves a transition state (TS2′, Figure 6) in which the amino group of RC2b delivers one of its hydrogens to the Oγ of Ser37 (∠(Oγ···H···N11$^{(2)}$) = 127.8°, r(N11$^{(2)}$–H) = 1.104 Å and d(H–Oγ) = 1.485 Å, Figure 6). However, formation of this dipeptidyl nucleoside intermediate involves a slightly higher barrier (by 5.3 kJ mol^{-1} for TS2′, Table 1 and Figure 6) compared to the other proposed dipeptidyl enzyme intermediate (TS2, Table 1 and Figure 5) although gas phase calculations using other functionals reveal similar (within 2 kJ mol^{-1}, Table 1) barriers for IC2 and IC2′. More importantly, solvent phase calculations reveal up to 14 kJ mol^{-1} higher barrier for IC2′ compared to IC2, using varied DFT functionals (Figures S6 – S10 and Tables 1 and 2). In alignment with a previous experimental study,[15] this reiterates that formation of dipeptidyl enzyme intermediate will be preferred over the dipeptidyl nucleoside intermediate. Therefore, we use the dipeptidyl enzyme intermediate (IC2a) for understanding the next step of the CDP reaction pathway (i.e., intramolecular cyclization).

3.2.2. MD simulations. MD simulation of the complex containing RC2a and RC2b representing the reactants of the second step within AlbC active site, reveals that the introduction of RC2b into the AlbC structure results in a significant decrease in interactions of the first intermediate (RC2a) with surrounding residues. Specifically, occupancy of the

hydrogen bond between RC2a and Asn40 decreases significantly (from 92.3% to 15.3%, Table S3). Similarly, hydrogen bonding occupancies of RC2a with His203 and Gly35 are also reduced (from 31.7% to 8% for His203 and from 15% to 10.2% for Gly35, Table S3). Instead, a network of interacting residues stabilizes RC2b within the AlbC active site. Specifically, the carbonyl oxygen of RC2b interacts with amino nitrogen of Ile-36 (35.4% occupancy, Table S3) whereas amine group (N11, Figure 3) of RC2b interacts with carbonyl oxygen of Ile-36 (55.6% occupancy and -8.6 kcal mol^{-1} interaction energy, Table S3). Furthermore, the O2' atom of RC2b hydrogen bonds with Tyr178 (26.3% occupancy and -2.9 kcal mol^{-1} interaction energy, Table S3) as well as Glu182 (15.4% occupancy and -12.2 kcal mol^{-1} interaction energy, Table S3).

Analysis of the distribution of distance between N11$^{(1)}$ and C10$^{(2)}$ (Figure 3) as well as O3' and HN11$^{(1)}$ (Figure 3) atoms that participate in the reaction reveals six distinct peaks. A representative structure of the most relevant cluster (r(N11$^{(1)}$–C10$^{(2)}$) = 3.8 Å and r(N11$^{(1)}$–O3$'^{(2)}$) = 3.8 Å, Table S2) reveals hydrogen bonding between OE2 of Glu182 and 2'OH of Phe-A76$^{(2)}$ (r(OE2–O2$'^{(2)}$) = 2.1 Å, Figure 5C). Thus, as previously suggested,[15] Glu182 may possibly act as catalytic base by deprotonating the O2' of RC2b; this allows the oxygen to in turn extract a proton from N11$^{(1)}$ (Figure 3), making this amine ready for nucleophilic attack on the carbonyl group of the phenylalanyl enzyme intermediate. Participation of Glu182 might help the proton transfer from O2' to O3' of RC2b. Furthermore, as observed in the first step of the reaction (*vide supra*), the O2' atom of RC2b interacts with amino group N11$^{(1)}$ (r(N11$^{(1)}$H–O2$'^{(2)}$) = 2.0 Å, Figure 5C) which may help further lower the intrinsic barrier observed in the QM calculations.

Simulation of the AlbC complex bound to the dipeptidyl enzyme intermediate (IC2a) obtained as a product of second reaction step reveals that the carbonyl oxygen of Phe$^{(1)}$ hydrogen bonds with side chain amide nitrogen (ND2) of Asn40 (47.7% occupancy and -8.3 kcal mol^{-1} interaction energy, Table S3) as well as the sp^2 nitrogen (NE2) of the imidazole ring of His203 (26.9% occupancy and -4.6 kcal mol^{-1} interaction energy, Table S3) whereas amino group (N11$^{(2)}$, Figure 3) interacts with theside chain carboxylate oxygen (OE2) atom of Glu182 (4% occupancy and -6.0 kcal mol^{-1} interaction energy, Table S3). Overall, RC2b interacts with Ile36, Tyr178 and Glu182 through hydrogen bonding, whereas the dipeptidyl enzyme intermediate is stabilized by Asn40, Gln182 and His203 residues (Table 3).

The representative structure obtained after clustering with respect to N11$^{(1)}$, C10$^{(2)}$, O3' and HN11$^{(1)}$ (Figure 5C) involving IC2a shows that the carbonyl group of Phe$^{(1)}$ and the amino group of Phe$^{(2)}$ are at an optimal distance (r(C10$^{(1)}$–N11$^{(2)}$) = 4.7 Å, Figure 7C) in the

13

intermediate for the formation of the new bond. Glu182 is in proximity with N11$^{(2)}$ which may likely deprotonate it, thereby forming a nucleophile that subsequently attacks carbonyl carbon C10$^{(1)}$ of IC2a to form the peptide bond (Figure 7C). This attack may cleave the C10$^{(1)}$-Oγ bond and eject Ser37 (Figure 7C).

3.3. Intramolecular cyclization and release of the cyclic dipeptide product.

3.3.1. QM calculations. Following the second step and release of A76, an intramolecular cyclization of the dipeptidyl intermediate completes the CDP synthesis. Previous studies, and geometric reality, have shown that for intramolecular cyclization to proceed, the dipeptide needs to be in a *cis* conformation.[15,44] In this vein, we observed a *cis* conformation in our calculations that involves proximal arrangement of the amino hydrogen of Phe and Oγ of dipeptidyl ester intermediate (denoted RC3 in Figure 7A, B). Furthermore, N11$^{(2)}$H–O$\gamma^{(1)}$ and N11$^{(2)}$– C10$^{(1)}$ distances (3.171 Å and 2.721 Å, respectively) meet the structural conditions required for cyclization (Figure 7A, B).

The cyclization occurs in a concerted manner and involves a 157.9 kJ mol^{-1} energy barrier (TS3, Table 1 and Figure 7A, B). Specifically, the Oγ–C10$^{(1)}$ distance increases from 1.496 Å in RC3 to 1.906 Å in TS3 (Figure 7A, B), and N11$^{(2)}$ delivers a proton to Oγ (r(N11$^{(2)}$–H) = 1.163 Å, r(H – Oγ) = 1.365 Å and $\angle$O$\gamma \cdots$H$\cdots$N11$^{(2)}$ = 127.5°, Figure 7A, B) which coincides with decrease in N11$^{(2)}$–C10$^{(1)}$ distance (1.551 Å, Figure 7A, B). In addition, TS3 prefers a near-tetrahedral geometry around Cα atoms which occurs regardless of the identity of the DFT functional used. The intramolecular cyclization step finally leads to the desired cyclodipeptide product complex with N11$^{(2)}$–C10$^{(1)}$ distance of 1.356 Å (Figure 7A, B), and free Ser37 moiety which represents enzyme component of the biochemical reaction (Figure 7A, B).

Compared to the B3LYP-GD3BJ gas phase calculations, the energy barrier of the cyclization step increases by 11.3 kJ mol^{-1} using M06-2X methods while B97D functional decreases the barrier by 35.4 kJ mol^{-1} in gas phase calculations (Table 1). However, the barrier generally increases by 2.4 kJ mol^{-1} in B3LYP-GD3BJ and 33.5 kJ mol^{-1} in the M06-2X solvent phase although B97D leads to a small (12.6 kJ mol^{-1}) decrease in barrier compared to gas phase (Table 2). However, no significant change in structural parameters of RC3, TS3 and PC were observed using different functional groups in both gaseous and solvent phases (Figures S11 – S15 and Tables 1 – 2), except for PC in solvent phase B3LYP-GD3BJ calculations where Phe2 moiety bends towards Phe1 moiety relative to the reactant and TS3.

3.3.2. MD simulations. The last step (cyclisation) leads to the formation of the dipeptide where the Phe[(1)] and Phe[(2)] side chains are present in *cis* conformation (Figure 7C). Simulation reveals that N11[(1)] of the final cyclodipeptide interacts with the side chain amide nitrogen (ND2) of Asn159 (49.4% occupancy and –7.8 kcal mol[-1] interaction energy, Table S3) and that the carbonyl oxygen of Phe[(1)] interacts with the sp^2 nitrogen (NE2) of the imidazole ring of His203 (47.4% occupancy and –4.5 kcal mol[-1] interaction energy, Table S3). Likewise, the representative structure obtained from this simulation also shows hydrogen bonding between the N11[(1)] of the cyclodipeptide and ND2 of Asn159, and between the carbonyl oxygen of the cyclodipeptide and the NE2 atom of His203 (Figure 7C). Overall, Glu182 plays an important role in the intramolecular cyclisation, whereas Asn159 and His203 are involved in stabilizing the cyclodipeptide formed (Table 3).

4. Conclusion and Biochemical Significance.

The present work couples computationally tractable, yet chemically relevant, QM models with MD simulations to explore a possible pathway for AlbC-catalyzed Phe-Phe CDP synthesis. The overall process is comprised of three steps — activation of Ser37 and its attack of Phe-A76 of a tRNA[Phe] to form the covalent phenylalanyl enzyme intermediate, transamidation of this intermediate by Phe-A76 of the second tRNA[Phe], and intramolecular cyclization and release of the enzyme to form the cyclodipeptide.

In the first step, we deduced that the Phe transfer to the enzyme proceeds in a concerted manner. We can rule out a previous proposal involving 2′-OH mediated proton shuttling between Ser37 and 3′-OH of Phe-A76;[8] instead the activation of Ser37 occurs through direct transfer of its hydroxyl proton to O3′ through the aid of an intramolecular hydrogen bond of 2′-OH group with amino group of Phe that helps attain the reaction-ready conformation. Although the (116.1 kJ mol^{-1} – 166.7 kJ mol^{-1}) range of barriers observed in our calculations is significant, analysis of our MD simulations reveals that when Phe-A76 of first tRNA[Phe] binds AlbC, its phenylalanyl side chain accommodates within the catalytic pocket through the aid of hydrogen bonding with Ile36, whereas the amino acyl part of Phe-A76 is stabilised by interaction with His203. These interactions may lower the apparent transition state barriers observed in our reduced-model calculations. Our prophesized geometries also correlate with the available crystal structure data which suggest that an incoming Phe-A76 nucleotide may be anchored through a network of hydrogen-bonding interactions involving active-site residues, including Tyr178.[8,15] MD simulations reinforce that the direct transfer of hydroxyl proton of

Ser37 to O3' of Phe-A76 may involve hydrogen bonding support of 2'-OH group, without its direct involvement in the reaction. Finally, although the phenylalanyl enzyme intermediate formed after removal of A76 moiety loses its interaction with Tyr178, the retained hydrogen bonding with His203 and Asn40 and additional hydrogen bonding with Gly35 helps in its stabilization within the AlbC active site.

In the second step, the Phe-A76 moiety of a second tRNA^{Phe} attacks the phenylalanyl enzyme intermediate which generates the dipeptidyl enzyme intermediate. Alternatively, as previously proposed,[9,15] a dipeptidyl nucleoside intermediate may be formed, which can be attacked by a conserved enzyme residue that can act as a nucleophile through attack on O3′, causing removal of nucleoside and formation of dipeptidyl intermediate which can ultimately undergo intramolecular cyclization prior to expulsion of the nucleoside moiety. However, our QM calculations reveal a higher energy barrier for the dipeptidyl nucleoside intermediate formation over dipeptidyl enzyme intermediate. This in turn correlates with the previous experimental detection of dipeptidyl enzyme intermediate trapped in Tyr202F modified enzyme using liquid chromatography–mass spectrometry and peptide mass fingerprinting analysis.[15]

Although this mechanism again requires overcoming a significant barrier (74.0 kJ mol^{-1} – 126.4 kJ mol^{-1}), the anchoring of residues within the enzyme scaffold may help lower the barrier at active site of the enzyme. Specifically, as proposed in the literature, proton transfer assisted by a general base (e.g., Glu182)[8,9,15] may lower the barrier observed in direct hydrogen transfer reaction pathway within our model. Indeed, our simulation data suggests that His203 and Glu182 residues interact with HO2' of Phe-A76$^{(2)}$ which may help Glu182 fulfil the role of a catalytic base by deprotonating the O2' of Phe-A76$^{(2)}$. The O2'$^{(2)}$ atom may further deprotonate amino group of phenylalanyl enzyme intermediate which may in turn perform nucleophilic attack on carbonyl group of Phe-A76$^{(2)}$, leading to cleavage of the carbonyl carbon-oxygen ester bond and formation of dipeptidyl enzyme intermediate with release of A76 moiety of second tRNA^{Phe}. Our simulation suggests that the dipeptidyl enzyme intermediate formed is stabilized by Asn40 and His203, since they form hydrogen bonds with carbonyl oxygen of Phe$^{(1)}$ whereas amino group of Phe$^{(2)}$ forms hydrogen bond with Glu182.

Our QM calculations further reveal that intrinsic barrier for the final (intramolecular cyclization) step is again significant (122.5 kJ mol^{-1} – 191.4 kJ mol^{-1}). However, MD simulations reveal that the conformation of dipeptidyl enzyme intermediate is such that distance between the amino group and carbonyl group is optimal (r(C10$^{(1)}$–N11$^{(2)}$) = 4.7 Å) for attack. Therefore, the enzyme can catalyse this step. Indeed, as previously proposed,[8,15,18]

16

several active site residues of CDPSs, including Tyr178, Glu182 and Tyr202 can interact with the substrate and may provide this critical stabilization, hence lowering the barrier. More importantly, the hydroxyl group of the conserved Tyr202 is proposed to act as a proton relay and favour nucleophile attack at the ester bond of dipeptidyl enzyme, leading to the final CDP formation.[15,18] However, our simulation points to the role of Glu182 as a catalytic base that can abstract the amino proton of Phe[(2)], resulting in its nucleophilic attack on carbonyl group of Phe[(1)]. This will lead to release of Ser37 moiety and formation of CDP product in *cis* conformation as reported in previous studies[15,18]. The CDP thus formed interacts with His203 through its carbonyl oxygen and with Asn159 through its amino nitrogen in MD simulations.

Overall compiling the interactions observed, amino-acid residue Ile36 interacts through hydrogen bonding with both Phe-A76[(1)] present as substrate in first step of reaction and with second molecule of Phe-A76[(2)] introduced in the second step of reaction. Furthermore, Phe-A76[(1)] interacts with Ser37, Tyr202 and His203 through hydrogen bonding and with Asn159 and Tyr178 through hydrophobic interactions, whereas Phe-A76[(2)] interacts with Tyr178 and Glu182 through hydrogen bonding. Furthermore, both the phenylalanyl enzyme intermediate and the dipeptidyl enzyme intermediate interact with Asn40 and His203 residues through hydrogen bonding. Additionally, the phenylalanyl enzyme intermediate interacts with Gly35 and the dipeptidyl enzyme intermediate interacts with Glu182, both through hydrogen bonding. The cyclodipeptide obtained after intramolecular cyclization of the dipeptidyl enzyme intermediate interacts with Asn159 and His203 through hydrogen bonding. Therefore, we suggest that Glu182 plays an important role in nucleophilic attack and Asn40, His203 are required for accurate positioning of substrate and intermediates during the entire catalytic process.

The energy barriers identified in the article are generated from small QM models, and, as we note, do not consider additional stabilization provided by movement of the active site more broadly. Consequently, although they are useful for relative comparison, they likely overestimate the size of the barriers. A far larger QM/MM model would be required to provide a more accurate model. Regardless, our large, calculated barriers indicate that AlbC must actively catalyze the three main phases of the overall reaction, where MD simulations provide clues about how this can be achieved.

Together, our QM calculations successfully locate the reaction centre and estimate intrinsic barriers to CDP synthesis by AlbC, whereas MD simulations identify the surrounding residues that may play important role in different reaction steps. However, future studies are required to compare feasibility of alternate pathways and their relevance in comparison to

17

direct transfer pathways proposed in our work and to investigate the effect of interactions with active site residues on reaction pathways to verify our reported mechanisms and elucidate roles of other CDPS enzyme–tRNA contacts. Therefore, future hybrid QM/MM methods may help delineate the role of these residues in lowering the barrier and verify steric effects of the enzyme scaffold on the feasibility of proposed concerted mechanism.

18

Author Contributions

Conceptualization, AG, AP and PS; Funding acquisition JFT and PS; Investigation, AG and PS; Methodology, PS; Visualization, AG, IN and PS; Project administration, JFT and PS; Supervision, PS and JFT; Writing—original draft, AG and PS; Writing—review and editing, All authors.

Conflict of financial interest

The authors declare no competing financial interest.

Acknowledgements

Data and Software Availability Statement

Structures categorized along reaction pathway along different DFT functional in both gas phase and solvent phase are provided in the Supporting Information. The files containing all the structures of starting materials, intermediates, products, and analysed transition states discussed in this article, in a software-readable format, are available on the Borealis Dataverse at https://doi.org/10.5683/SP3/6O4U6K. The Borealis Dataverse is a joint initiative of the Canadian Universities to provide a permanent, free-to-access, repository for digital data.

Keywords: Cyclodipeptide synthase · density functional calculations · reaction coordinate · transition barrier · gas phase · solvent phase · molecular dynamic simulations.

5. References.

1. M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amaoroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courcon, C. Masson, S. Dubois, S. Lauru, A. Lecoq, S. Hashimoto, R. Genet and J. L. Pernodet, Cyclodipeptide synthases are a family of tRNA-dependent peptide bond–forming enzymes. *Nat. Chem. Biol*., 2009, https://doi.org/10.1038/nchembio.175 **5**, 414-420.

2. S. Lautru, M. Gondry, R. Genet and J. L. Pernodet, The albonoursin gene cluster of S. noursei: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.*, 2002, https://doi.org/10.1016/S1074-5521(02)00285-5 **9**, 1355-1364.

3. A. D. Borthwick, 2, 5-Diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.*, 2012, https://doi.org/10.1021/cr200398y **112**, 3641-3716.

4. P. Borgman, R. D. Lopez and A. L. Lane, The expanding spectrum of diketopiperazine natural product biosynthetic pathways containing cyclodipeptide synthases. *Org. Biomol. Chem.*, 2019, https://doi.org/10.1039/C8OB03063D **17**, 2305-2314.

5. L. Z. Hessefort, L. J. Harstad, K. R. Merker, L. P. T. Ramos and K. F. Biegasiewicz, Chemoenzymatic Catalysis: Cooperativity Enables Opportunity. *ChemBioChem*, 2023, https://doi.org/10.1002/cbic.202300334 **24**, e202300334.

6. Y. Liu, L. Pengbo, S. Gao, Z. Wang, P. Luan, J. G. Sabin and Y. Jiang, Construction of chemoenzymatic cascade reactions for bridging chemocatalysis and Biocatalysis: Principles, strategies and prospective. J. *Chem. Eng.*, 2021, https://doi.org/10.1016/j.cej.2020.127659 **420**, 127659.

7. P. Belin, M. Moutiez, S. Lautru, J. Seguin, J. L. Pernodet and M. Gondry, The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.*, 2012 https://doi.org/10.1039/C2NP20010D **29**, 961-979.

8. L. Sauguet, M. Moutiez, Y. Li, P. Belin, J. Seguin, M. H. L. Du, R. Thai, C. Masson, M. Fonvielle, J. L. Pernodet, J. B. Charbonnier and M. Gondry, Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.*, 2011 https://doi.org/10.1093%2Fnar%2Fgkr027 **39**, 4475-4489.

9. M. W. Vetting, S.S. Hegde and J. S. Blanchard, The structure and mechanism of the Mycobacterium tuberculosis cyclodityrosine synthetase. *Nat. Chem. Biol.*, 2010 https://doi.org/10.1038%2Fnchembio.440 **6**, 797-799.

10. L. Bonnefond, T. Arai, Y. Sakaguchi and O. Nureki, Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proceedings of the National Academy of Sciences*, 2011 https://doi.org/10.1073/pnas.1019480108 **108**, 3912-3917.

11. G. Bourgeois, J. Seguin, M. Babin, P. Belin, M. Moutiez, Y. Mechulam, M. Gondry and E. Schmitt, Structural basis for partition of the cyclodipeptide synthases into two subfamilies. *J. Struct. Biol.*, 2018 https://doi.org/10.1016/j.jsb.2018.03.001 **203**, 17-26.

12. G. Bourgeois, J. Seguin, M. Babin, M. Gondry, Y. Mechulam and E. Schmitt, Structural basis of the interaction between cyclodipeptide synthases and aminoacylated tRNA substrates. *RNA*, 2020 https://doi.org/10.1261%2Frna.075184.120 **26**, 1589-1602.

13. M. Moutiez, P. Belin and M. Gondry, Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. *Chem. Rev.*, 2017 https://doi.org/10.1021/acs.chemrev.6b00523 **117**, 5578-5618.

14. I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry and P. Belin, Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.*, 2015 https://doi.org/10.1038/nchembio.1868 **11**, 721-727.

15. M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.*, 2014 10.1038/ncomms6141 **5**, 1-7.

16. M. Moutiez, J. Seguin, M. Fonvielle, P. Belin, I. B. Jacques, E. Favry, M. Arthur and M. Gondry, Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from Streptomyces noursei. *Nucleic Acids Res.,* 2014 https://doi.org/10.1093%2Fnar%2Fgku348 **42**, 7247-7258.

17. A. Croitoru, M. Babin, H. Myllykallio, M. Gondry and A. Aleksandrov, Cyclodipeptide Synthases of the NYH Subfamily Recognize tRNA Using an α-Helix Enriched with Positive Residues. *Biochemistry*, 2021 https://doi.org/10.1021/acs.biochem.0c00761 **60**, 64-76.

18. E. Schmitt, G. Bourgeois, M. Gondry and A. Aleksandrov, Cyclization reaction catalyzed by cyclodipeptide synthases relies on a conserved tyrosine residue. *Scientific reports*, 2018 https://doi.org/10.1038/s41598-018-25479-5 **8**, 1-11.

19. J. L. Kellie and S. D. Wetmore, Mechanistic and conformational flexibility of the covalent linkage formed during β-lyase activity on an AP-site: application to hOgg1. *J. Phys. Chem. B*, 2012 https://doi.org/10.1021/jp306344g **116**, 10786-10797.

20. S. Sowlati-Hashjin and S. D. Wetmore, Quantum mechanical study of the β-and δ-lyase reactions during the base excision repair process: application to FPG. *Phys. Chem. Chem. Phys.*, 2015 https://doi.org/10.1039/C5CP04250J **17**, 24696-24706.

21. S. Sowlati-Hashjin and S. D. Wetmore, Structural insight into the discrimination between 8-oxoguanine glycosidic conformers by DNA repair enzymes: a molecular dynamics study of human oxoguanine glycosylase 1 and formamidopyrimidine-DNA glycosylase. *Biochemistry*, 2018 https://doi.org/10.1021/acs.biochem.7b01292 **57**, 1144-1154.

22. M. Moutiez, E.Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.,* 2014 https://doi.org/10.1038/ncomms6141 **5**, 5141.

23. S. Sowlati-Hashjin and S. D. Wetmore, Computational investigation of glycosylase and β-lyase activity facilitated by proline: applications to FPG and comparisons to hOgg1. *J. Phys. Chem. B*, 2014 https://doi.org/10.1021/jp507783d **118**, 14566-14577.

24. HyperChem(TM) Professional 8.0, H., Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA. 10.12691/ajmo-1-3-7

25. A. L. Millen, C. K. Mclaughlin, K. M. Sun, R. A. Manderville and S. D. Wetmore, Computational and experimental evidence for the structural preference of phenolic C-8 purine adducts. *J. Phys. Chem. A*, 2008 https://doi.org/10.1021/jp712058a **112**, 3742-3753.

26. R. Peverati and K. K. Baldridge, Implementation and performance of DFT-D with respect to basis set and functional for study of dispersion interactions in nanoscale aromatic hydrocarbons. *J. Chem. Theory Comput.*, 2008 https://doi.org/10.1021/ct800252z **4**, 2030-2048.

27. M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, Performance of M06, M06-2X, and M06-HF density functionals for conformationally flexible anionic clusters: M06 functionals perform better than B3LYP for a model system with dispersion and ionic hydrogen-bonding interactions. *J. Phys. Chem. A*, 2013 https://doi.org/10.1021/jp408166m **117**, 12590-12600.

28. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, , et al., *Gaussian 16, Revision B.01,*. Gaussian, Inc., Wallingford CT, 2016.

29. D. R. Roe and T. E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, 2013 https://doi.org/10.1021/ct400341p **9**, 3084-3095.

30. D. A. Case, et al., Amber 2020. 2020: *University of California, San Francisco*.

31. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, 2015 https://doi.org/10.1021/acs.jctc.5b00255 **11**, 3696-3713.

32. W. L. Jorgensen, J. Chandrashekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983 https://doi.org/10.1063/1.445869 **79**, 926-935.

33. F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, The REd. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.*, 2010 https://doi.org/10.1039/C0CP00111B **12**, 7821-7839.

34. E. Vanquelef, S. Sabrina, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak and F. Y. Dupradeau, RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, 2011 https://doi.org/10.1093/nar/gkr288 **39**, W511-7.

35. J. Wang, W. Wang, P. A. Kollman and D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, 2006 https://doi.org/10.1016/j.jmgm.2005.12.005 **25**, 247-260.

36. I. S. Joung and T. E. Cheatham III, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 2008 https://doi.org/10.1021/jp8001614 **112**, 9020-9041.

37. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method. *J. Chem. Phys.*, 1995 https://doi.org/10.1063/1.470117 **103**, 8577-8593.

38. K. Singer and W. Smith, Path integral simulations of condensed phase Lennard-Jones systems. *Mol. Phys.*, 1988 https://doi.org/10.1080/00268978800100823 **64**, 1215-1231.

39. M. Brehm and B. Kirchner, TRAVIS-a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories. *J. Chem. Inf. Model.*, 2011 https://doi.org/10.1021/ci200217w **51**, 2007-2023.

40. H. J. Berendsen, J. P. M. Postma, W. F. Gunsteren, A. Dinola and J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 1984 https://doi.org/10.1063/1.448118 **81**, 3684-3690.

41. H. C. Andersen, Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 1983 https://doi.org/10.1016/0021-9991(83)90014-1 **52**, 24-34.

42. C. W. Hopkins, S. L. Grand, R. C. Walker and A. E. Roitberg, Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.*, 2015 https://doi.org/10.1021/ct5010406 **11**, 1864-1874.

43. W. L. DeLano, Pymol: An open-source molecular graphics tool. *CCP4. Newsl. Protein Crystallogr.*, 2002. **40**, 82-92.

44. J. E. Purdie and N. L. Benoiton, Piperazinedione formation from esters of dipeptides containing glycine, alanine, and sarcosine: the kinetics in aqueous solution. *J. Chem. Soc., Perkin Trans.* 2, 1973 https://doi.org/10.1039/P29730001845 1845-1852.
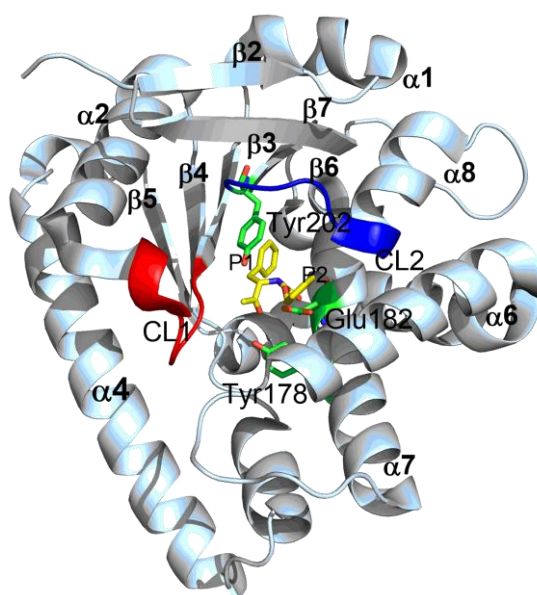
Figures and Tables

Figure 1. Molecular surface of the AlbC enzyme complexed with ZPK ligand (yellow) (PDB ID, 4Q24).[15] Loops bearing the catalytic residues (CL1 and CL2) are in red and blue respectively. α helices, β strands, and catalytically active residues in green are labelled.
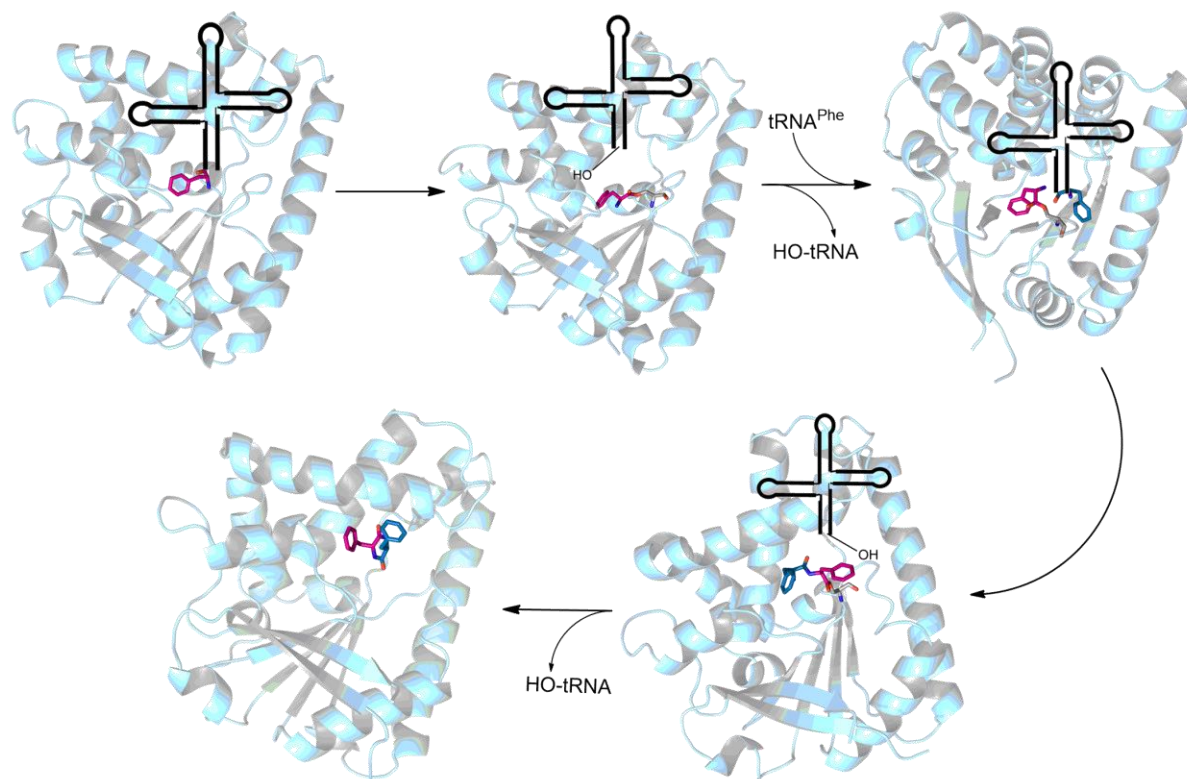
Figure 2. Reaction catalysed by CDPSs. CDPSs use activated amino acids, in the form of aa-tRNAs, as substrates to catalyse formation of cyclodipeptides. The mechanism involves aminoacyl enzyme intermediate formation, which further attacks the second molecule of substrate and forms dipeptidyl enzyme intermediate. The intramolecular cyclisation occurred in this intermediate leads to the formation of cyclodipeptide.
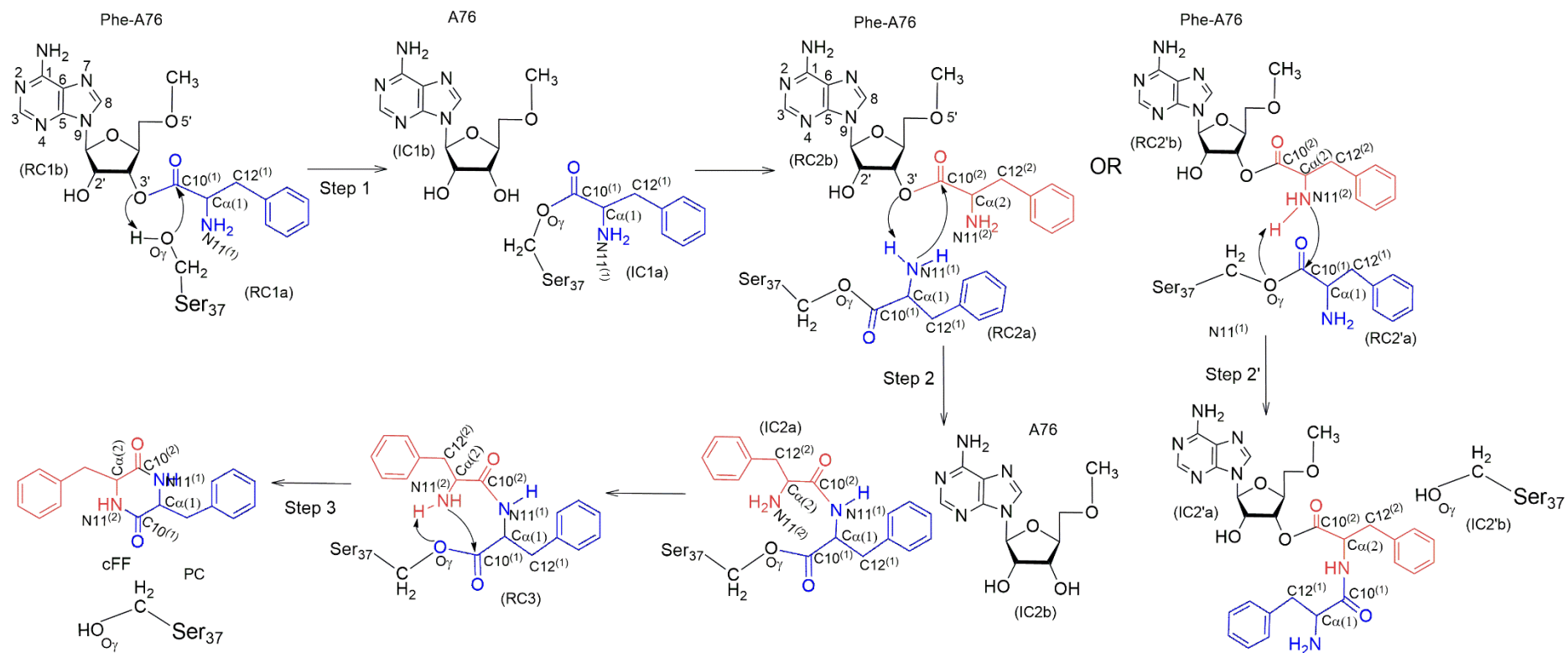
25

Figure 3. Proposed chemical steps of the formation mechanism of phenylcyclodipeptide (cyclo(L-Phe-L-Phe) at AlbC active site. Step 1 involves activation of Ser37 through direct proton transfer to O3′ hyroxyl of Phe-A76, followed by attack of Ser37. Step 2 and 2′ show two possibilities of the attack of second Phe-A, followed by formation of the second intermediate. Step 3 involves intramolecular cyclization of dipeptidyl intermediate. The first (1) and second (2) incorporated phenylalanyl moieties are highlighted in blue and red color respectively.
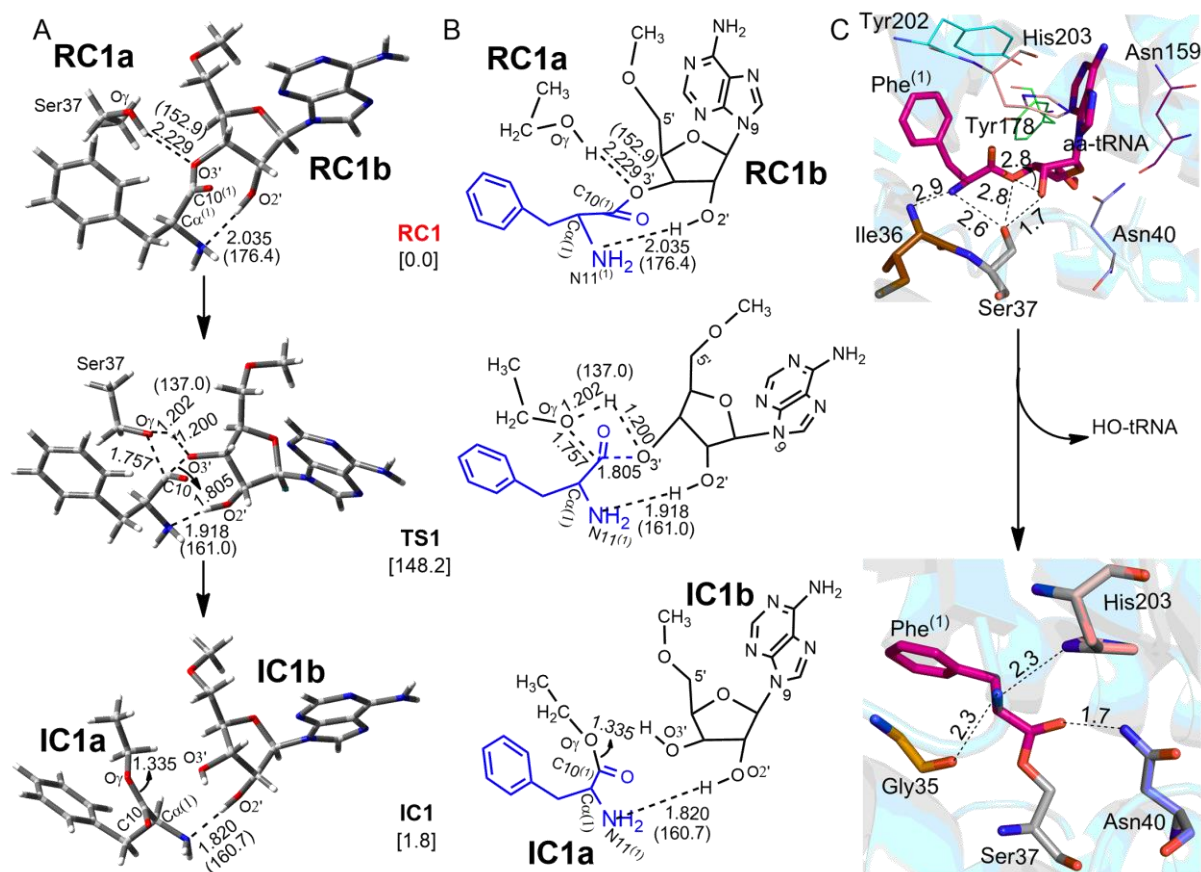
26

Figure 4. (A) Structures characterized along the pathway of step 1 for the direct proton transfer reaction associated with the activation of Ser37, at the B3LYP-GD3BJ/6-311+G(2df, 2pd)//B3LYP-GD3BJ/6-31G(d,p) level. Important distances (Å), angles (deg, in parentheses) and relative energies with respect to the reactant (kJ mol^{-1}, square brackets) are provided. (B) Line drawings representing the reaction pathways along the step 1. (C) Representative structure obtained from MD simulation of the substrate and intermediate of step 1 in the AlbC active site, showing the important residues.
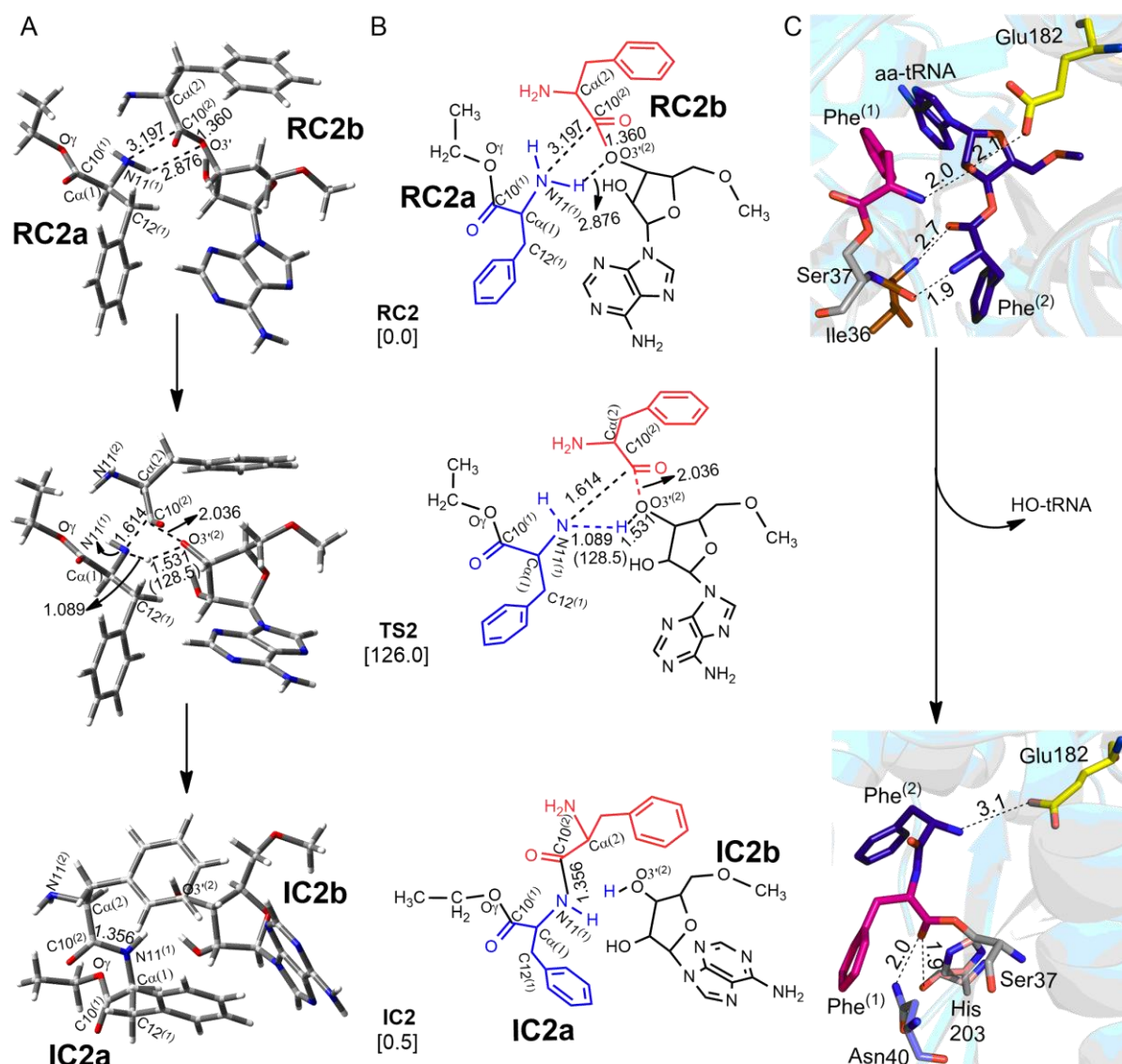
27

Figure 5. (A) Structures characterized along the pathway of step 2 for the attack of second (2) phenylalanyl nucleotide moiety forming dipeptidyl enzyme intermediate at the B3LYP-GD3BJ/6-311+G(2df, 2pd)//B3LYP-GD3BJ/6-31G(d,p) level. Important distances (Å), angles (deg, in parentheses) and relative energies with respect to the reactant (kJ mol⁻¹, square brackets) are provided. (B) Line drawings representing the reaction pathways along the step 2. (C) Representative structure obtained from MD simulation of the substrate and intermediate of step 2 in the AlbC active site, showing the important residues.
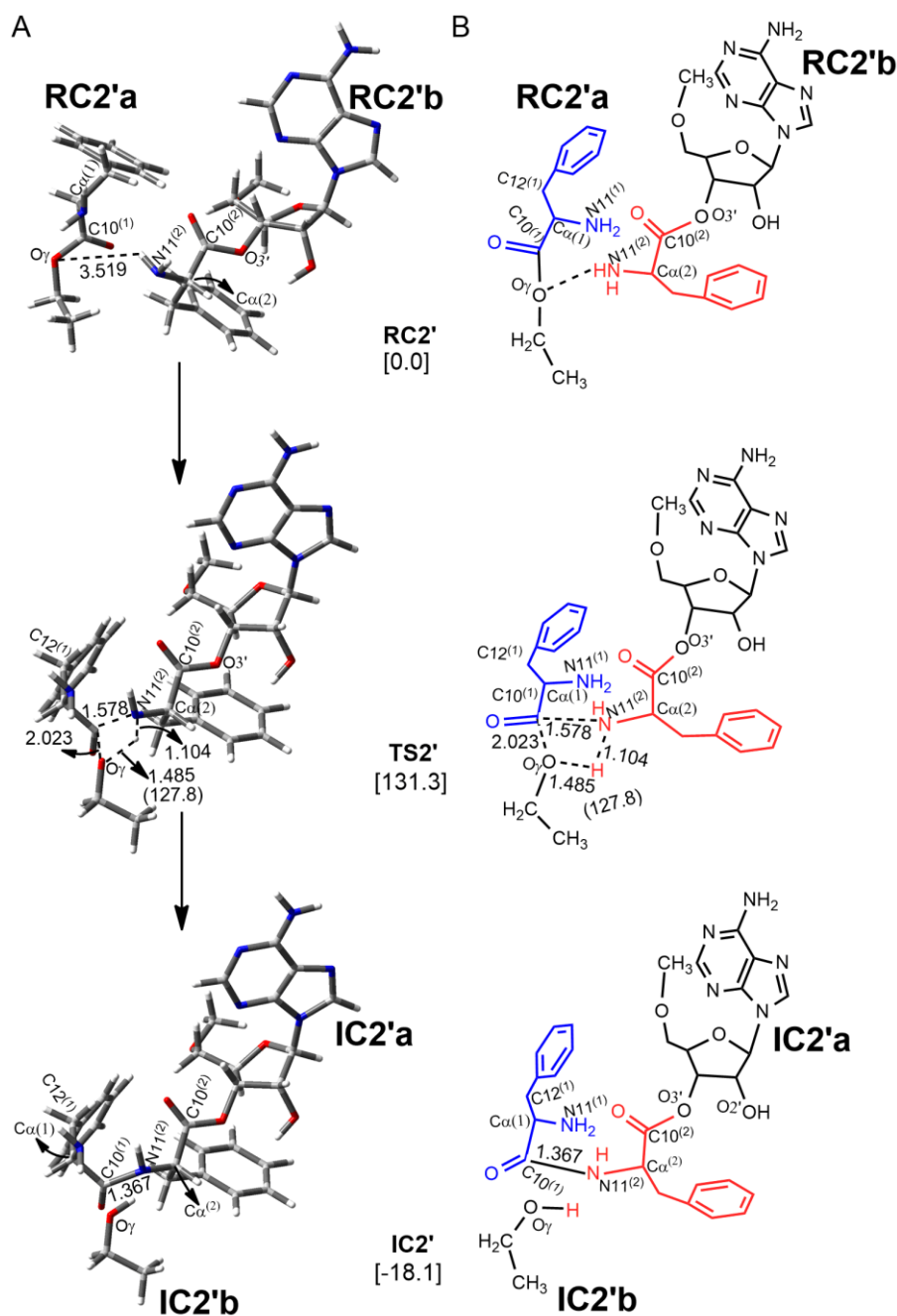
Figure 6. (A) Structures characterized along the pathway of step 2′ for the attack of second phenylalanyl nucleotide moiety forming dipeptidyl nucleoside intermediate at the B3LYP-GD3BJ/6-311+G(2df, 2pd)//B3LYP-GD3BJ/6-31G(d,p) level. Important distances (Å), angles (deg, in parentheses) and relative energies with respect to the reactant (kJ mol^{-1}, square brackets) are provided. (B) Line drawings representing the reaction pathways along the step 2′.
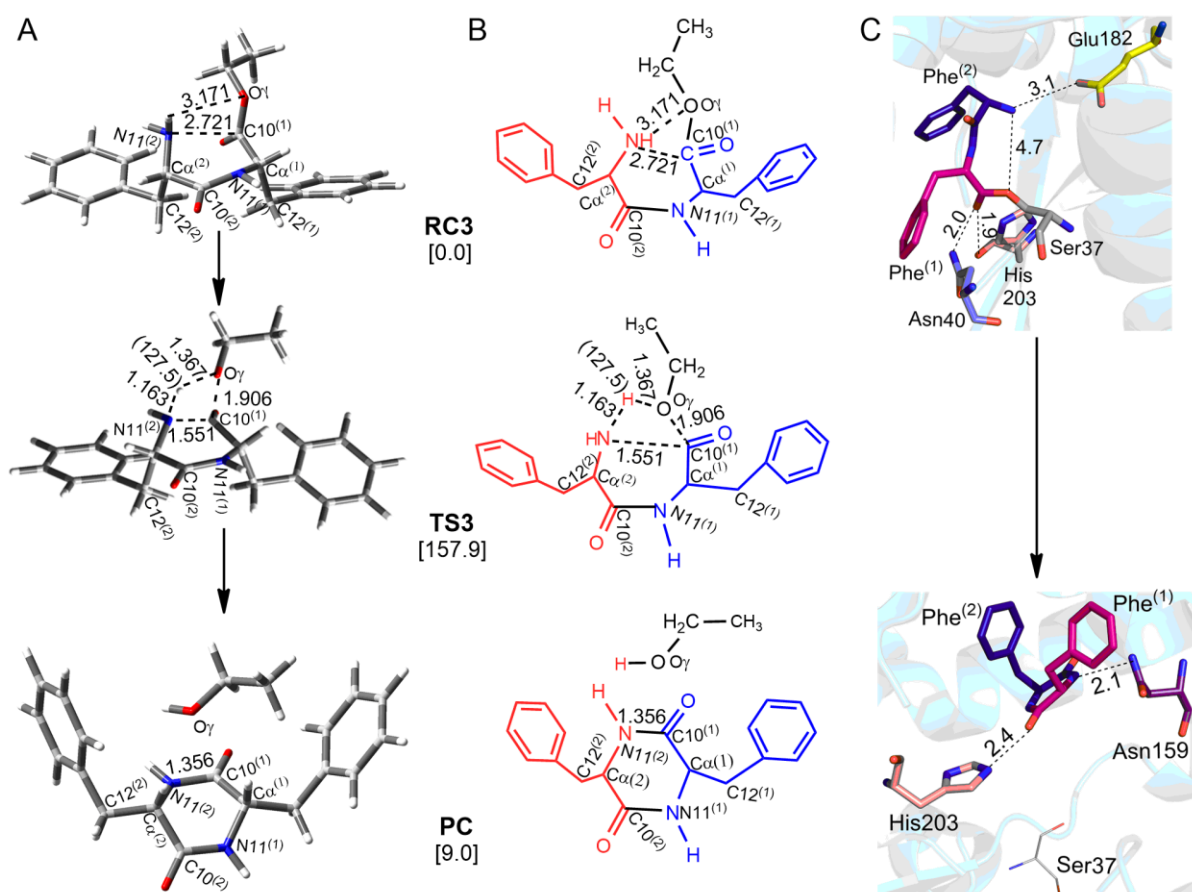
Figure 7. (A) Structures characterized along the pathway of step 3 for the intramolecular cyclization associated with the dipeptidyl enzyme intermediate at the B3LYP-GD3BJ/6-311+G(2df, 2pd)//B3LYP-GD3BJ/6-31G(d,p) level. Important distances (Å), angles (deg, in parentheses) and relative energies with respect to the reactant (kJ mol^{-1}, square brackets) are provided. (B) Line drawings representing the reaction pathways along the step 3. (C) Representative structure obtained from MD simulation of the intermediate and product of step 3 in the AlbC active site, showing the important residues.

30

Table 1. Gas phase relative energies (ΔE^g, kJ mol^{-1}) for stationary points characterized along the reaction pathway for each step of CDP formation.

Reaction step	Stationary point	B3LYP-GD3BJ	B97D	M06-2X
Ist phenylalanyl attack (Ser activation)	RC1	0.0	0.0	0.0
	TS1	148.2	116.1	166.7
	IC1	1.8	0.6	11.6
2nd phenylalanyl attack (dipeptidyl intermediate)	RC2	0.0	0.0	0.0
	TS2	126.0	102.6	126.4
	TS2'	131.3	102.6	125.6
	IC2	0.5	11.0	4.4
	IC2'	−18.1	29.2	38.7
Intramolecular cyclization	RC3	0.0	0.0	0.0
	TS3	157.9	122.5	169.2
	IC3	9.0	56.6	80.1

31

Table 2. Solvent phase relative energies (ΔE^s, kJ mol^{-1}) for stationary points characterized along the reaction pathway for each step of CDP formation.

Reaction step	Stationary point	B3LYP-GD3BJ	B97D	M06-2X
		ΔE^s		
Ist phenylalanyl attack (Ser activation)	RC1	0.0	0.0	0.0
	TS1	150.4	129.8	162.6
	IC1	−3.5	−0.8	−1.8
2nd phenylalanyl attack (dipeptidyl intermediate)	RC2	0.0	0.0	0.0
	TS2	81.4	100.1	74.0
	TS2'	91.3	113.8	74.2
	IC2	−55.5	−17.9	−27.3
	IC2'	−59.7	−24.7	−7.6
Intramolecular cyclization	RC3	0.0	0.0	0.0
	TS3	160.3	145.3	191.4
	IC3	12.8	9.2	26.1

32

Table 3. AlbC residues interacting with the substrate, intermediates and product along the reaction pathway of CDP formation.

Reaction step		Interacting Residues
Ist phenylalanyl attack (Ser activation)	Introduction of first molecule of Phe-A76	Ile36, Ser37, Asn159, Tyr178, Tyr202, His203
	Formation of the phenylalanyl enzyme intermediate	Gly35, Asn40, His203
2nd phenylalanyl attack (dipeptidyl intermediate)	Introduction of the second molecule of Phe-A76	Ile36, Tyr178, Glu182
	Formation of the dipeptidyl enzyme intermediate	Asn40, Glu182, His203
Intramolecular cyclization	Formation of the cyclodipeptide	Asn159, Glu182, His203

33