Discover the Comprehensive Reaction Mechanism of Norcoclaurine Synthase Using Molecular Simulation

Xujian Wang,[a] Haodong Liu,[a] Le Chang[a] Jiayang Cai,[a] Jiayu Pan[a] Yuan Gao,[a] Xinyi Sun,[a] Xiaohui Gu[a] Yilu Zhang[a] Shiyue Zeng[a] Jiahuang Li,*[a]

Norcoclaurine Synthase (NCS) serves as a biocatalyst that facilitates the condensation of dopamine and 2-benzenepropanal (2-BPA) into tetrahydroisoquinoline (THIQs) derivatives, exhibiting significant potential as prodrugs. However, as most THIQs derivatives are currently synthesized chemically, this approach presents various drawbacks. NCS offers a promising biological alternative for the synthesis of THIQs. Regrettably, the absence of an in-depth understanding of the catalytic mechanism has led to inefficient resource allocation, hindering the optimization of NCS for industrial production. To overcome these challenges, we employed advanced simulation technologies, such as Gaussian accelerated molecular dynamics (GaMD) and density functional theory (DFT) calculations, to comprehensively reveal the catalytic mechanism. Our results demonstrate that D141 not only provides a critical negative charge to stabilize dopamine's binding but also participates directly in the reaction. Furthermore, we found that non-polar residues, such as M183 and L76 within the tunnel, play a more significant role than previously assumed. Additionally, according to DFT calculations, the rate limited step is the aromatic electrophilic attack with an energy of 35.9 kcal/mol, consistent with previous experimental conjectures. Our research suggests that the redesign of NCS could commence with the tunnel design to enhance its dynamic characteristics. Furthermore, modifying the thermodynamic characteristics of NCS enzymatic catalysis, based on the reaction mechanism, offers new prospects for its industrial application.

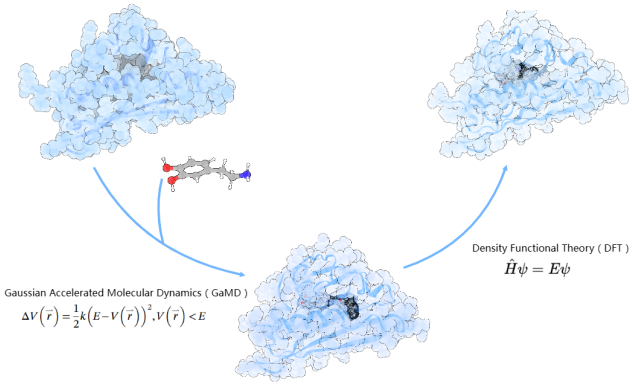
Introduction

Biocatalysts offer considerable promise as catalysts when compared to organic and transition metal catalysts, owing to their advantages such as the utilization of non-toxic reactants and the production of green byproducts. Despite these advantages, there are still many disadvantages. To overcome these limitations, including narrow substrate adaptability, complex condition control, and difficulty in obtaining pure proteins, advanced methodologies for enzyme redesign have been developed. These methodologies encompass directed evolution [1–3], semi-rational design [4,5], rational design [6–10], and machine learning-involved design [11–16].

[a] Xujian Wang, Jiahuang Li*
School of Biopharmacy, China Pharmaceutical University, Nanjing, 210000, China
E-mail: lijiah@cpu.edu.cn

The rational approach, which is based on understanding molecular interactions through molecular structures, has significantly advanced with the explosive growth in computational power and the maturation of computational chemistry theories. This advancement has rendered biocatalysts increasingly accessible, cost-effective, and innovative, particularly in the chemical and pharmaceutical industries [17–22], where they are expected to play a substantial role.

Norcoclaurine synthase (NCS), initially obtained from Thalictrum flavum, serves as a pivotal catalyst in the Pictet-Spengler reaction, enabling the condensation of dopamine and 2-BPA into (S)-norcoclaurine. Furthermore, NCS exhibits catalytic proficiency towards derivatives structurally analogous to the Tetrahydroisoquinoline (THIQ) family. These molecules encompass a diverse array of therapeutic activities, including anti-tumor [23] and anti-bacterial [24] properties. However, it is worth noting that the natural extraction of THIQs from plants is associated with challenges such as poor purity and limited quantities. Conversely, chemical synthesis methods entail multi-step procedures and toxic reagents that pose difficulties in achieving desired stereoselectivities [25,26] and generate adverse environmental impacts [27]. Consequently, NCS represents a preferable alternative as a catalyst, circumventing these issues, as evidenced by successful applications in various THIQs synthesis studies [28–33].



Scheme 1. GaMD was employed to investigate the substrate's entry into the pocket, while DFT was utilized to examine the catalytic process. This approach enabled a comprehensive study of the entire catalytic mechanism, from binding to reaction.

Owing to the promising applications and potential value of NCS, numerous enzyme modification attempts have been undertaken to enhance its suitability for industrial-scale production. These efforts include innovatively using this en-

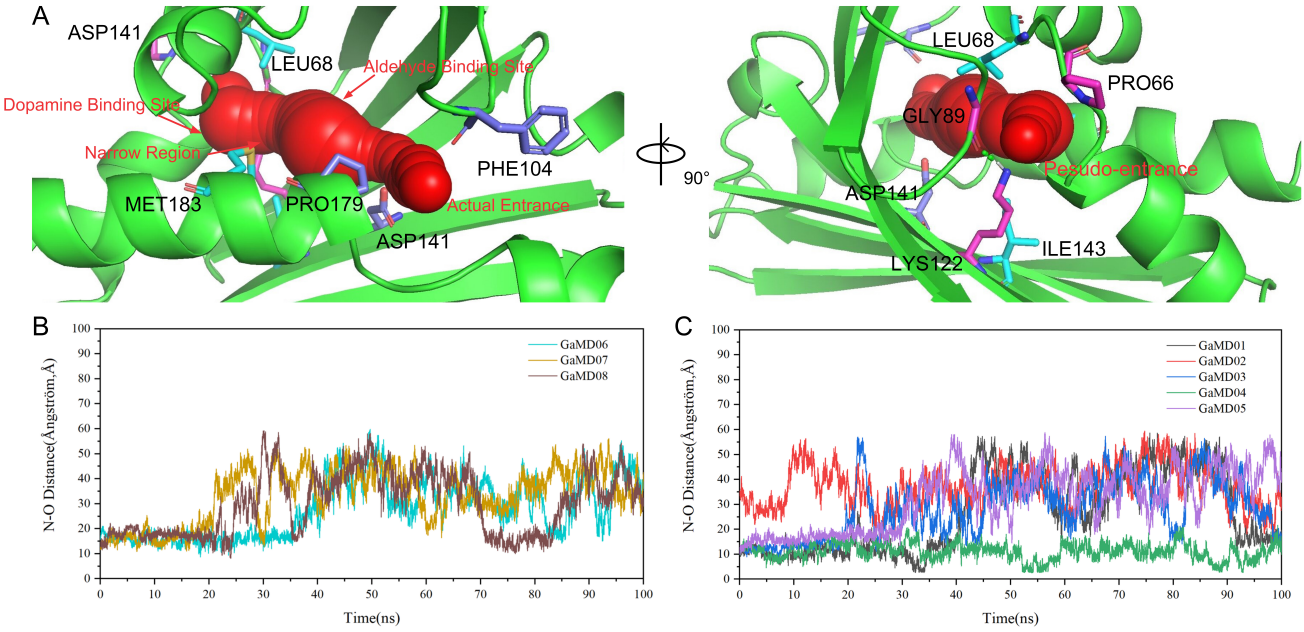


Figure 1. Figure 1. (A) illustrates the structure of the tunnel in NCS from two perspectives. The tunnel features two entrances: the actual entrance and the pseudo-entrance. Each entrance comprises three gatekeeper residues, with the actual entrance colored purple and the pseudo-entrance in pink. A narrow region, consisting of three amino acids (colored cyan), connects the dopamine and aldehyde binding sites. (B) and (C) depict how dopamine enters the binding site when positioned at the pseudo-entrance and the actual entrance, respectively.

zyme to construct THIQs with spiral ring structures and introducing specific functional groups at the alpha position of the carbonyl group [29]. However, many of these approaches have historically relied on resource-intensive high-throughput screening, due to the absence of a comprehensive catalytic reaction mechanism. In contrast, structure-guided rational design presents a viable alternative, employing artificially designed enzymes [34]. This method is based on a holistic understanding of the reaction mechanism, encompassing the entire sequence of events from enzyme-substrate interaction and substrate entry into the active site to the completion of the catalytic reaction [31,35–40]. The initial stages, where the tunnel captures and guides the substrate into the active site, are critically important. Extensive research on tunnel design has shown that the tunnel significantly influences catalytic activity. Furthermore, a comprehensive understanding of the full catalytic mechanism can reveal exactly what the enzyme does during the reaction, leading to more effective enzyme design. While crystallography effectively captures reactants and products, it often falls short in identifying intermediates and transition states. Similarly, conventional molecular dynamics (cMD), despite their usefulness, are limited by their reliance on classical molecular force fields, which can hinder their ability to simulate chemical reactions and capture extended timescale events. To overcome these limitations, we employ Gaussian accelerated molecular dynamics (GaMD). This approach not only captures the entry of ligands into the pocket but also simulates the receptor's conformational changes, crucial for various biological functions. Furthermore, the cluster method, which involves extracting key catalytic amino acids and substrates from the enzyme for quantum chemistry calculations, has proven to be an excellent approach for elucidating the mechanisms of enzyme-catalyzed reactions.

In this study, we have intricately demonstrated the process of dopamine capture and its subsequent transport to the catalytic site by Norcoclaurine Synthase. Our investigation goes beyond the surface, delving into the specific entry tunnel that facilitates dopamine's movement into the enzyme, with a detailed analysis at a residue-by-residue level. We have meticulously identified the molecular composition of the tunnel's residues, offering a comprehensive view of their interactions. Employing density functional theory, we have precisely calculated the entire reaction mechanism of NCS, uncovering the free energy barriers for each step of the reaction. Notably, our calculated mechanism aligns well with experimental data, particularly in elucidating the rate-limiting step and pinpointing the pivotal roles of specific residues. This study not only reveals the nuanced mechanisms of a critical biochemical process but also establishes a new benchmark for molecular-level investigations in biocatalysis, paving the way for future advancements in the field.

Results and Discussion

Visualization Analysis of the Molecular Tunnel In our analysis, we utilized PyMOL, a molecular visualization software, along with the CAVER3 plugin to intricately visualize the tunnel in NCS (PDBID: 5NON, chain A). CAVER3 meticulously computes the tunnel's geometry from the protein's atomic coordinates. Figure 1.A vividly shows a singular tunnel in NCS, extending in a bend line. The pocket houses two significant spaces: the dopamine and aldehyde sites. Consistent with the 'dopamine first' mechanism, as substantiated by prior crystallography studies, the dopamine site is deeply situated in the pocket, whereas the aldehyde site is more superficial. Moreover, the tunnel presents two entry points, with one serving as the actual substrate entry

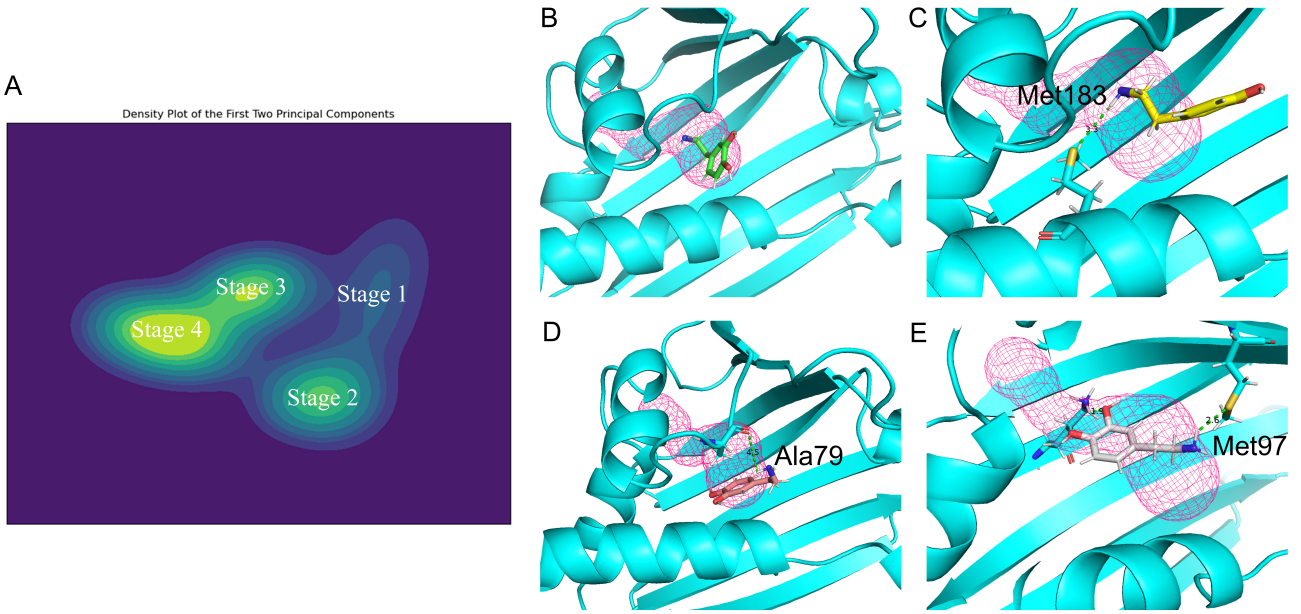


Figure 2. Figure 2. (A) characterizes the distance between dopamine and the tunnel residues during its entry into the pocket, as revealed by Principal Component Analysis (PCA). This analysis has led to the identification of four main stages of the entry process. The structures corresponding to each of these stages are depicted in (B), (C), (D), and (E), respectively.

and the other as a pseudo-entry, as established by earlier research. The functional entry of the enzyme is guarded by residues P179, F104, and D141, while the pseudo-entry is monitored by P66, K122, and G89. Compared to the functional entry, the pseudo-entry has a notably smaller entrance, rendering it virtually impossible for substrates to enter. Significantly, D141 acts as a gatekeeper; its absence leads to potentially unstable binding of dopamine. Additionally, a constricted region is formed by residues I143, M183, and L68, situated between these two sites. These residues are non-polar amino acids with large side chains, which may hinder the entry of dopamine but also create a hydrophobic region that is conducive to the reaction. The presence of these large side-chain amino acids plays a crucial role in modulating access to the active site and influencing the enzyme's catalytic activity.

To thoroughly examine the components of the tunnel, we utilized Gaussian accelerated molecular dynamics (GaMD) simulations to model the journey of a ligand from outside the protein to the active site. Initially, we dock the ligand at the entrance of the tunnel, followed by the initiation of enhanced sampling simulations. Due to the stochastic nature of this method, we conducted multiple simulations to validate the results. In the 'dopamine first' mechanism, the critical marker is the formation of a hydrogen bond between the nitrogen of the K122 side chain and the hydroxyl group of dopamine, which we used as a primary indicator. We considered binding successful when the distance between the nitrogen (N) and the oxygen (O) was less than 8 angstroms. Regarding the initial entry, two out of five simulations demonstrated successful dopamine entry into the tunnel; one maintained stability within the tunnel (GaMD04, see Fig 1B), while the other exhibited a rapid exit (GaMD01, see Fig 1B). In contrast, for the secondary entry, none of the simulations resulted in successful ligand entry (as indicated in Fig 1. C). By analyzing the trajectory data from GaMD04, we identified the components of the tunnel. We considered any residue within 10 angstroms of the mass center of dopamine during its entry as part of the tunnel. Our findings, which enumerate 51 residues constituting the tunnel, are detailed in the supporting information.

Analysis of Dopamine Entry and Stabilization: In investigating the tunnel's role as a catalytic pocket in NCS, our focus was on understanding how it facilitates the entry and stabilization of dopamine. Our simulations, GaMD01 and GaMD04, provided insights into this process. Both simulations revealed a two-phase mechanism: initial substrate capture followed by its guided movement into the tunnel. Notably, GaMD04 exhibited stable binding of dopamine, making it a prime subject for detailed analysis.

While both simulations depict the substrate's entry into the active site, only the GaMD04 simulation results in stable binding. Therefore, to characterize this process, we analyzed the distance between the center of mass of the residue forming the tunnel and the center of mass of the substrate. This analysis was conducted using principal component analysis (PCA), which divides the entry process into four stages (see Fig. 2A). In Stage 1, dopamine is outside the tunnel, exhibiting relatively high free energy (see Fig. 2B). During Stage 2, the amine group of dopamine forms a hydrogen bond with M183 (see Fig. 2C). This interaction captures dopamine in the tunnel, resulting in a relatively low-energy state that facilitates subsequent entry. Stage 3 is marked by dopamine forming a hydrogen bond with A79 (see Fig. 2D), which reorients it within the tunnel. This differs from Stage 2 and aligns with the 'dopamine first' hypothesis, suggesting the hydroxyl group of dopamine must move towards the interior of the tunnel. This stage is substable, leading to Stage 4, where dopamine fully enters the tunnel. Here, its two hydroxyl groups are held by K122, and the amino group forms a hydrogen bond with M97 (see Fig.

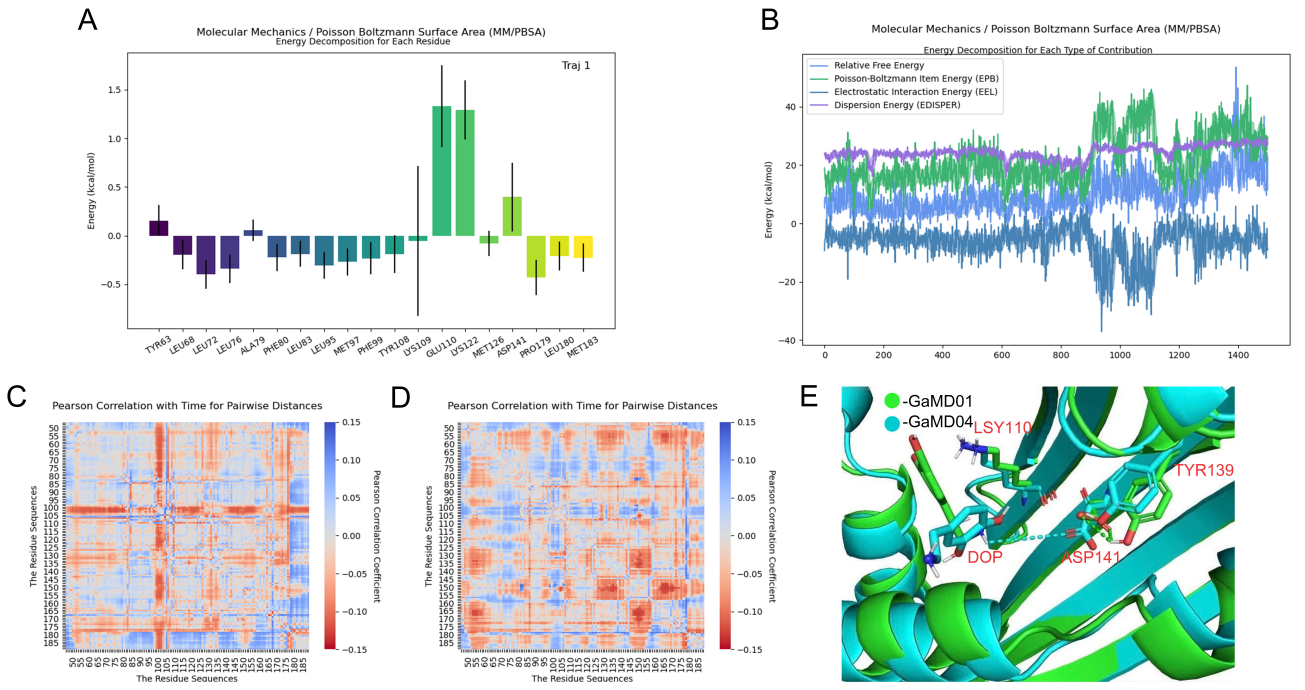


Figure 3. Figure 3. (A) displays the energy decomposition per residue following MMPB/SA calculations. (B) presents plots categorized by different types of forces derived from MMPB/SA calculations. (C) and (D) feature Pearson correlation coefficient heatmaps, illustrating the variations in distances between all residues over time. (E) depicts the distinct binding modes between residues and dopamine in GaMD01 (represented in green) and GaMD04 (shown in cyan).

2E). This observation closely aligns with structural analyses obtained from crystallography. It demonstrates that our simulation can accurately depict the conformation of dopamine in the pocket without artificial assumptions. Additionally, it indicates the ability of our enhanced sampling technique to capture various correct conformations that are not accessible through crystallography.

To further investigate which residues play important roles in this event, we conducted molecular mechanism Poisson-Boltzmann/Surface Area (MMPB/SA) calculations through Amber. We calculated the MMPBSA for both GaMD04 and GaMD01 during the process of dopamine entering the pocket. Despite GaMD01 ultimately resulting in unstable binding, the analysis revealed notable similarities in the MMPBSA results for this specific entry process, highlighting consistent interactions during the initial stages. This suggests that during the entry phase, the behavior of most tunnel residues is consistent between the two simulations. This observation supports the reliability of our calculations and indicates a shared mechanism of substrate entry, despite the eventual divergence in binding stability. Notably, most negative free energy contributions came from non-polar residues such as L76, F99, P179, and M183. Interestingly, P179 and M183, components of the gatekeeper and narrow region, respectively, demonstrated favorable interactions with dopamine, suggesting significant van der Waals forces from these non-polar residues during dopamine's entry. In contrast, residues considered catalytic, such as K122, G110, and D141, contributed positively to the free energy. This observation is somewhat unexpected and contrasts with assumption that these three amino acids, while being catalytic, also play a significant role in binding. To delve deeper into this issue, we analyzed the three most variable components during the MMPBSA calculation (see Fig. 3 B) – Poisson-Boltzmann energy (EPB), electrostatic interaction energy (EEL), and dispersion energy (EDISPER). This revealed that EPB increased with dopamine's entrance, likely related to the increase of solvent accessible surface area (SASA) . While EDISPER remained steady, EEL showed a rapid decrease and subsequent recovery, attributed to transient hydrogen bonding with K109. This bond formed not with the side chain of K109, but with its backbone atoms, elucidating the bond's transient nature and the large variance observed for K109 (see Fig 3. A). This trend suggests that despite an overall increase in free energy, but some specific interactions facilitates dopamine's stable entry into the tunnel.

While GaMD01 and GaMD04 share many common characteristics, the question arises as to why only GaMD04's binding mode is stable. To investigate this, we calculate the changes in distances between all protein residues over time and plotting a Pearson correlation coefficient matrix for two simulations, respectively. In GaMD01, residues 95 to 105 show a negative correlation with time (see Fig 3. D), indicating that these residues are getting closer together. Although this might enlarge the narrow area inside the tunnel, but also reduces the interaction between these residues and dopamine. Conversely, in GaMD04, the same residues exhibit a positive correlation (see Fig 3. E), suggesting they are moving apart. A key residue, D141, forms a hydrogen bond with dopamine, holding it tightly and preventing its exit. This difference in hydrogen bonding may contribute to the stability of dopamine's final binding (see Fig 3. C). D141's role, as demonstrated in previous experiments, is to provide a negatively charged environment that stabilizes dopamine, creating conditions for subsequent stable catalysis.

Reaction Mechanism The cluster model has proven to be a valuable and widely utilized framework for discovering enzyme catalytic mechanisms[38,39,41–43].We docked dopamine and BPA into the protein (PDBID: 5NON, chain A). Subsequently, we performed conventional MD simulations to obtain a stable structure for further analysis. Several amino acids in close proximity to the substrates were selected, and their coordinates, along with those of the substrates, were extracted. During the DFT calculations, the carbonyl carbon and nitrogen atoms in the backbones of these residues were frozen. The reaction begins with rapid proton transfer. Notably, E110 rather than D141 initially captures the proton on dopamine's nitrogen, as supported by D141 mutagenesis. The proton then transfers to 2-BPA's carbonyl oxygen. Thus, at least two intermediates exist between the enzyme-substrate complex (E:S) and initial intermediate IM1. This divides the energy barrier into several low steps, explaining IM1's accessibility (43.1 kcal/mol above E:S). The nitrogen anion formed is highly nucleophilic and the carbonyl is electrophilic, enabling nitrogen to attack the carbonyl carbon to form the C-N bond (IM1→[TS1]→IM2). At TS1 (39.7 kcal/mol), the formation of a stable C-N bond leads to a pronounced energy drop, making the reaction nearly irreversible. Subsequently, spontaneous E110-facilitated dehydration occurs (IM2→[TS2]→IM3), driven by the stability of IM3 (-4.8 kcal/mol). Dopamine-first mechanism has been demonstrated in either experimental way or computational way[31,33,36–38]. And according to our molecular docking and cluster analysis, we also agree with the dopamine-first mechanism. The reaction begins with rapid proton transfer. Notably, E110 rather than D141 initially captures the proton on dopamine's nitrogen, as supported by D141 mutagenesis[40]. Our E110 mutagenesis experiments further substantiated this, revealing without E110 the full reaction cannot proceed. The proton then transfers to 2-BPA's carbonyl oxygen. Thus, at least two intermediates exist between the enzyme-substrate complex (**E:S**) and initial intermediate **IM1**. This divides the energy barrier into several low steps, explaining **IM1**'s accessibility (43.1 kcal/mol above **E:S**). The nitrogen anion formed is highly nucleophilic and the carbonyl is electrophilic, enabling nitrogen to attack the carbonyl carbon to form the C-N bond (**IM1→[TS1]→IM2**). At **TS1** (39.7 kcal/mol), the stable C-N bond leads to a pronounced energy drop, making the reaction nearly irreversible. Subsequently, spontaneous E110-facilitated dehydration occurs (**IM2→[TS2]→IM3**), driven by the stability of **IM3** (-4.8 kcal/mol). Interestingly, the E110 proton orients toward the hydroxyl oxygen, forming a relatively strong hydrogen bond. Simultaneously, another E110 oxygen forms a weaker hydrogen bond with the amino group. Overall, the E110 carboxyl group functions like a clamp around the highly reactive N-C-O region. As the carboxyl group transfers a proton to the hydroxyl while withdrawing a hydrogen from the amino group, this compensates for the lost hydrogen atom. In this process, the bond between the hydroxyl oxygen and original carbonyl carbon cleaves, releasing water as a leaving group. This iminium intermediate (**IM3**) is crucial for the reaction. The departure of water from the enzyme pocket substantially reduces energy. This free energy change may drive water expulsion, facilitating **IM4** formation. In summary, the thermodynamically favorable **IM3**, along with entropy/enthalpy changes (-29.9 kcal/mol) from water expulsion, drives the dehydration and **IM4** formation.

As water departs, the intermediate gains flexibility from the additional space. Consequently, the entire benzene ring in **IM3** shifts, forming **IM5**. Initially, K122 forms hydrogen bonds with both dopamine hydroxyls, but the meta-hydroxyl bond predominates. During the shift, the meta-hydroxyl forms a weak bond with Y63 while retaining its K122 bond. Simultaneously, the D141 carboxyl acquires the para-hydroxyl, forming a strong hydrogen bond. This reinforces the E110-nitrogen bond. These interactions among the four residues cause conformational locking. While this mechanism has been observed crystallographically[33] and computationally[29,30,36], D141 hydrogen bonding with the para-hydroxyl appears to be a novel proposal. Traditionally, D141 was thought anchored by Y139 and T159. However, transient hydrogen bonds may represent a pivotal transition state step. Thus, our mechanism underscores D141 mainly provides local intermediate stabilization via its negative charge[30,36]. Despite increasing energy by 14.7 kcal/mol relative to **IM4**, this conformational change likely facilitates the subsequent transition state.

Subsequently, the nitrogen atom withdraws a hydrogen atom from E110, breaking their hydrogen bond. The next step involves a thermodynamically favorable 5-membered (**IM6b→[TS3b]→IM7b**) or 6-membered (**IM6a→[TS3a]→IM7b**) ring transition state. Some propose two potential pathways to the product for this transition state. To elucidate these pathways, we used Orbital Weighted Fukui Functions and Orbital Weighted Dual Descriptors[44] to predict the reaction site and calculate free energy barriers, clarifying the distinct pathways. Subsequently, the nitrogen atom withdraws a hydrogen atom from E110, breaking their hydrogen bond. The next step involves a thermodynamically favorable 5-membered (IM6b→[TS3b]→IM7b) or 6-membered (IM6a→[TS3a]→IM7b) ring transition state. Some propose two potential pathways to the product for this transition state. To elucidate these pathways, we used Orbital Weighted Fukui Functions and Orbital Weighted Dual Descriptors to predict the reaction site and calculate free energy barriers, clarifying the distinct pathways. The first pathway (IM6a→[TS3a]→IM7a), which is widely accepted[33,36,38,39], suggests the formation of a -intermediate (IM7a). This reaction initiates with a conformational change, indicating that the iminium carbon moves closer to position 2 carbon (2.1Å, as seen in Fig. 3 TS3a) than to position 1 carbon (2.7Å, as seen in Fig. 3 TS3a). Subsequently, K122 abstracts the proton from the meta-hydroxyl group, leading to electron migration within the benzene ring. This results in an electrophilic attack at the adjacent position to the substituent. As this step disrupts the aromatic properties of the benzene ring, it constitutes the rate-limiting step, characterized by the highest reaction energy barrier (44.2 kcal/mol), in alignment with the outcomes of primary kinetic isotope (PKI) experiments[40]. As for the second pathway (IM6b→[TS3b]→IM7b), it suggests the formation of a benzoquinone-like intermediate[40] (IM7b, as seen in Fig. 2) . Prior to the reaction, a conformational change is also required, with the iminium carbon moving closer to position 1 carbon (2.0Å, as seen in Fig. 3 TS3b) rather than position 2 carbon (2.5Å, as seen in Fig. 3 TS3b). However, in comparison to IM6a, the formation of a five-membered ring in the side chain brings it closer to the benzene ring. Simultaneously, during the formation of the six-membered ring transition state, the iminium can be stabilized by the negative charge of the carboxyl group in E110 (distances are

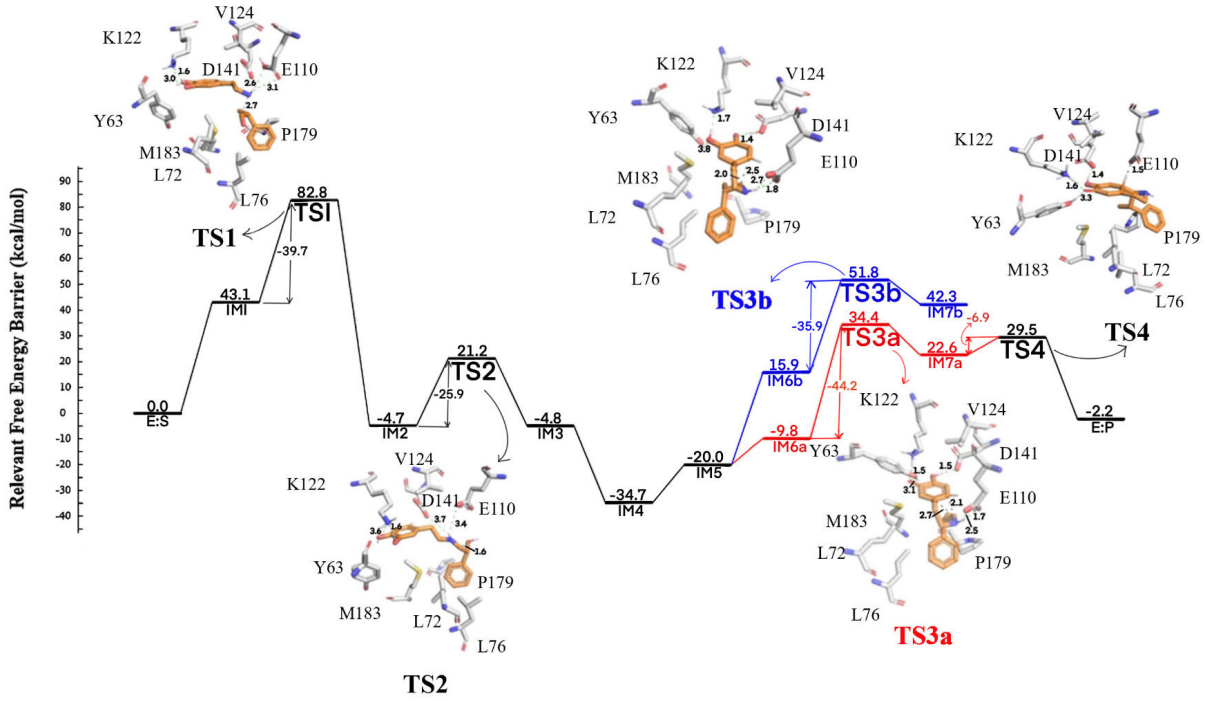


Figure 4. The simplified diagram of the NCS catalytic mechanism is shown below. We have omitted the transition state of proton transfer and only performed quantum chemical calculations on other key transition states. The two different paths introduced at **IM6** represent two different transitions (red for -intermediate, blue for benzoquinone-like intermediate). We have calculated and analyzed the entire process of the mechanism and analyzed the selectivity of the two transition states.

2.5Å and 1.7Å, as seen in Fig. 3 TS3a), whereas the five-membered ring is relatively less stable (distances are 2.5Å and 1.7Å, as seen in Fig. 3 TS3a). As a result, the energy required for IM6b (35.9 kcal/mol relative to IM5) is significantly higher than that for IM6a (29.8 kcal/mol relative to IM5). This elucidates why NCS enzymes require over 12 hours to achieve high conversion rates even at temperatures higher than room temperature. Moreover, IM7b has a higher energy (42.3 kcal/mol relative to E:S) compared to TS3a (34.4 kcal/mol), making the second pathway less preferable. Additionally, Orbital Weighted Fukui Functions and Orbital Weighted Bicharacteristics Descriptors indicate that position 2 carbon is more electrophilic than position 1 carbon. Ultimately, due to the shift of the benzene ring, E110 can more easily approach the hydrogen in benzene, leading to the abstraction of a proton from benzene and its subsequent rearomatization. Driven by rearomatization, this step is relatively rapid and exhibits the lowest energy barrier (TS4, 6.9 kcal/mol).

Conclusion: Leveraging the innovative concept of tunnel design, we identified key components of the NCS tunnel, including previously overlooked residues such as M183 and P179. Although these residues do not directly participate in the catalytic reaction, they play a crucial role, potentially exerting a direct impact on the enzyme's catalytic rate. We propose a novel approach to redesigning NCS, emphasizing the optimization of non-catalytic steps. We effectively utilized the enhanced sampling method to replicate the crystallographic structure accurately. Additionally, this simulation technique provides deeper insights into the dynamic behavior of proteins. For example, although D141 was initially considered a key catalytic residue, our experiments indicate that its conformational changes are pivotal in determining the stable interaction of dopamine within the pocket. This observation aligns well with the findings from D141 mutation experiments. Similarly, the integration of GaMD and MMPBSA techniques enables us to investigate the interactions of dopamine with specific residues within the tunnel, offering a new perspective for tunnel design. Our approach goes beyond mere geometrical analysis of the tunnel; we also quantitatively analyze the functional aspects of the tunnel, focusing on the interactions within it. In our analysis, we utilized the Density Functional Theory (DFT) method to calculate the catalytic clusters at the enzyme's active site and quantitatively calculated the intermediates and transition states of each reaction step. The notably high energy barrier of TS3a (35.9 kcal/mol) elucidates why NCS requires more than 12 hours to achieve satisfactory conversion rates even at temperatures above room temperature, and this finding is in strong agreement with predictions from the PKI experiment. This implies that strategically lowering the energy barrier of TS3a could be crucial for the industrial application of NCS. Additionally, in predicting two different intermediates, our analysis clearly indicates that the six-membered ring transition state is more favorable, offering a novel perspective for understanding the mechanistic pathways of NCS.

Conclusion

Molecular dynamics simulations were performed in GROMACS package (version 2023.1-gpu) on high performance computational center(BEIJING SUPER COLUD COMPUT-

ING CENTER). The Amber14 force filed and TIP3P water model were employed for simulation. Protein structure was provided by PDB (PDBID: 5NON) which had been demonstrated that simulation conducted by this structure can generate phenomena corresponding to the experiment. The structure of all the substrates were optimized by Gaussian16 with the M06-2X functional and def2-TZVP basis. Additionally, substrates were parametrized by GAFF model with the RESP2 charge method which has been fit well with GAFF model. The entire workflow for classical molecular dynamics simulation: 1) Build gro files and positional restraints with gmx editconf (take the protein as the center, expand outward 1.2nm to get the whole box). 2) Use water, chloride and sodium ions to fill the spare space, keep the pH=7.0 and salt concentration is 0.15M (close to experiment). 3) Minimize the system by steepest descent method firstly then followed by conjugate gradient methods 4) Temperature coupling and pressure coupling were performed at the same time for quickly reaching equilibrium phase. 5) All data are collected from the 100ns simulation of equilibrium phase. Minimization: Firstly use steepest descent method and Verlet integrator for 100000 steeps to reduce the maximum force less than 250 kJ*mol-1*nm-1. The protein and substrates were restrained by positional restraints but water is flexible. Followed by another separated minimization which utilize conjugate descent method for 100000 steeps to reduce the maximum force less than 100 kJ*mol-1*nm-1. In this steep all the molecules were flexible. This process had been proved successfully to avoid subsequent simulation crashes. Two minimization's electrostatic interactions were calculated by Fast smooth Particle-Mesh Ewald (SPME) and van der Waals interactions were calculated by cut-off method, additionally, the cut off range of both are 1.2nm. Dispersion correction was also employed. Equilibrium phase simulation: Before simulation, the entire system is divided into complex group (protein and ligands) and environment group (all the rest),each part controls temperature independently. For reaching the equilibrium, simulation using the the structure optimized by minimization was performed under NPT conditions, the temperature coupling method is V-rescale (tauT=0.2, refT=310), the pressure coupling method is Berendsen with isotropic coupling (tauP=2.0). This stage needed 2ns with positional restrain on complex group. After equilibrium structure obtained, use Parrinello-Rahman pressure coupling method (tauP=2.0) replace the former, and perform 100ns simulation without any restraints.The electrostatic interactions were calculated by Fast smooth Particle-Mesh Ewald (SPME) and van der Waals interactions were calculated by cut-off method, additionally, the cut off range of both are 1.0nm. Dispersion correction was also employed.

Experimental Detail

Classical molecular dynamics

Molecular dynamics simulations were performed in GROMACS package [45] (version 2023.1-gpu) on high performance computational center(BEIJING SUPER COLUD COMPUTING CENTER). The Amber14 force filed [46] and TIP3P water model were employed for simulation. Protein structure was provided by PDB (PDBID: 5NON) which had been demonstrated that simulation conducted by this structure can generate phenomena corresponding to the experiment. The structure of all the substrates were optimized by Gaussian16 with the M06-2X functional and def2-TZVP basis. Additionally, substrates were parametrized by GAFF model with the RESP2 charge method [47] which has been fit well with GAFF model. The entire workflow for classical molecular dynamics simulation: 1) Build gro files and positional restraints with gmx editconf (take the protein as the center, expand outward 1.2nm to get the whole box). 2) Use water, chloride and sodium ions to fill the spare space, keep the pH=7.0 and salt concentration is 0.15M (close to experiment). 3) Minimize the system by steepest descent method firstly then followed by conjugate gradient methods 4) Temperature coupling and pressure coupling were performed at the same time for quickly reaching equilibrium phase. 5) All data are collected from the 100ns simulation of equilibrium phase. Minimization: Firstly use steepest descent method and Verlet integrator for 100000 steeps to reduce the maximum force less than 250 kJ*mol-1*nm-1. The protein and substrates were restrained by positional restraints but water is flexible. Followed by another separated minimization which utilize conjugate descent method for 100000 steeps to reduce the maximum force less than 100 kJ*mol-1*nm-1. In this steep all the molecules were flexible. This process had been proved successfully to avoid subsequent simulation crashes. Two minimization's electrostatic interactions were calculated by Fast smooth Particle-Mesh Ewald (SPME) and van der Waals interactions were calculated by cut-off method, additionally, the cut off range of both are 1.2nm. Dispersion correction was also employed. Equilibrium phase simulation: Before simulation, the entire system is divided into complex group (protein and ligands) and environment group (all the rest),each part controls temperature independently. For reaching the equilibrium, simulation using the the structure optimized by minimization was performed under NPT conditions, the temperature coupling method is V-rescale (tauT=0.2, refT=310), the pressure coupling method is Berendsen with isotropic coupling (tauP=2.0). This stage needed 2ns with positional restrain on complex group. After equilibrium structure obtained, use Parrinello-Rahman pressure coupling method (tauP=2.0) replace the former, and perform 100ns simulation without any restraints.The electrostatic interactions were calculated by Fast smooth Particle-Mesh Ewald (SPME) and van der Waals interactions were calculated by cut-off method, additionally, the cut off range of both are 1.0nm. Dispersion correction was also employed.

Cluster model (DFT calculation)

Cluster analysis module in GROMACS was used to pick up the most representative conformation in cMD calculation as the initial structure. Then fetch all substrates and particular residue, the bonds that had previously connected other residue were now blocked by hydrogen atom. Structure optimization, transition state search and vibration analysis were calculated at M06-2X functional with def2-SVP basis and DFT-D3 dispersion correction [48] level in chlorobenzene implicit solvent (IEFPCM implicit solvent model), the calculation is performed by Gaussian16 [49]. Single point energy was calculated at rev-DSD-PBEP86 functional with def2-TZVPP basis and DFT-D4 dispersion correction [50] level in chlorobenzene implicit solvent (SMD implicit solvent model), the calculation is performed by ORCA [51] (version 5.0.4). And free energy is conducted by single point energy corrected by vibration analysis, performed by Sherom package [52].

Gaussian Accelerated Molecular Dynamics (GaMD)

Simulations

Gaussian Accelerated Molecular Dynamics (GaMD) simulations were conducted using the Amber23 software package. These simulations employed the same protein structure sourced from the PDB and substrate structures optimized using Gaussian16 with the M06-2X functional and def2-TZVP basis, as described in the classical Molecular Dynamics (cMD) simulations.

The GaMD simulations utilized the Amber19 force field and OPC3 water model. Temperature control was set at 310 K using the Langevin thermostat (`ntt = 3`, `gamma_ln = 5.0`). Pressure was maintained isotropically (`ntp = 1`) with a relaxation time of 10.0 ps (`taup = 10.0`). Non-bonded interactions were treated with a cut-off of 10.0 angstroms (`cut = 10.0`). The system was equilibrated and continued from a previous state (`ntx = 5`, `irest = 1`).

GaMD was enabled with `igamd = 11`, and various parameters specific to GaMD, such as `sigma0P`, `sigma0D`, `iEP`, `iED`, were meticulously set to modulate the potential energy surface. The dual boosting approach, targeting both potential and dihedral energies, was implemented to enhance sampling efficiency. The simulation was set for a total of 100 ns, facilitating extensive exploration of the system's conformational space. The first 10 ns were used for equilibration under restrained conditions, followed by 90 ns of unrestrained sampling.

Acknowledgements

The authors are grateful for the financial supports of National Innovation and Entrepreneurship Training Program for Undergraduate (No.202310316093Y). We also extend our gratitude to the companies that have provided assistance to us. Simulation tasks were performed on the computational facility funded by Beijing Super Cloud Computing Center.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: biocatalysis • norcoclaurine synthase • catalytic mechanism • gaussian accelerated molecular dynamics

References

[1] A. Buller, et al., *Proceedings of the National Academy of Sciences* **2015**, *112*, 14599.
[2] H. Bunzel, J. Anderson, A. Mulholland, *Current Opinion in Structural Biology* **2021**, *67*, 212.
[3] R. Cobb, R. Chao, H. Zhao, *AIChE Journal* **2013**, *59*, 1432.
[4] S. Chan, et al., *Food Chemistry* **2023**, *419*, 136070.
[5] Z. Liu, et al., *Biotechnology and Bioengineering* **2019**, *116*, 1833.
[6] B. Amrein, et al., *IUCrJ* **2017**, *4*, 50.
[7] L. Jiang, et al., *Science* **2008**, *319*, 1387.
[8] D. Li, et al., *Food Chemistry* **2023**, *423*, 136274.
[9] R. Li, et al., *Nature Chemical Biology* **2018**, *14*, 664.
[10] L. Liu, S. Zhou, Y. Deng, *ACS Catalysis* **2023**, *13*, 8183.
[11] B. Duan, Y. Sun **2020**.
[12] F. Li, et al., *Nature Catalysis* **2022**, *5*, 662.
[13] J. E. Gado, et al. **2023**.
[14] S. Mazurenko, Z. Prokop, J. Damborsky, *ACS Catalysis* **2019**, *10*, 1210.
[15] Y. Saito, et al., *ACS Synthetic Biology* **2018**, *7*, 2014.
[16] K. Yang, Z. Wu, F. Arnold, *Nature Methods* **2019**, *16*, 687.
[17] S. Wu, et al., *Angewandte Chemie International Edition* **2021**, *60*, 88.
[18] X. Chen, et al., *Angewandte Chemie International Edition* **2023**, *62*, e202218140.
[19] Y. He, et al., *ACS Catalysis* **2023**, *13*, 7210.
[20] M. Herger, et al., *Journal of the American Chemical Society* **2016**, *138*, 8388.
[21] M. Nishihachijo, et al., *Bioscience, Biotechnology, and Biochemistry* **2014**, *78*, 701.
[22] A. Schneider, et al., *Angewandte Chemie International Edition* **2023**, *62*, e202301607.
[23] Faheem, et al., *Expert Opinion on Drug Discovery* **2021**, *16*, 1119.
[24] D. Cullen, et al., *European Journal of Medicinal Chemistry* **2021**, *226*, 113861.
[25] T. Pesnot, et al., *Chemical Communications* **2011**, *47*, 3242.
[26] R. Quevedo, E. Baquero, M. Rodriguez, *Tetrahedron Letters* **2010**, *51*, 1774.
[27] M. Pyo, et al., *Bioorganic Medicinal Chemistry Letters* **2008**, *18*, 4110.
[28] V. Erdmann, et al., *Angewandte Chemie International Edition* **2017**, *56*, 12503.
[29] B. Lichman, et al., *Nature Communications* **2017**, *8*, 14883.
[30] T. Pesnot, et al., *Advanced Synthesis Catalysis* **2012**, *354*, 2997.
[31] R. Roddan, et al., *ACS Catalysis* **2019**, *9*, 9640.
[32] M. Zhang, et al., *Bioresources and Bioprocessing* **2023**, *10*.
[33] R. Roddan, et al., *Communications Chemistry* **2020**, *3*, 170.
[34] J. Kratky, et al., *Angewandte Chemie International Edition* **2023**, *62*, e202300657.
[35] A. Bonamore, et al., *Molecules* **2010**, *15*, 2070.
[36] B. Lichman, et al., *FEBS Journal* **2015**, *282*, 1137.
[37] B. Lichman, et al., *Biochemistry* **2017**, *56*, 5274.
[38] X. Sheng, F. Himo, *Journal of the American Chemical Society* **2019**, *141*, 11230.
[39] A. Ilari, et al., *Journal of Biological Chemistry* **2009**, *284*, 897.
[40] L. Luk, et al., *Biochemistry* **2007**, *46*, 10153.
[41] F. Himo, *Journal of the American Chemical Society* **2017**, *139*, 6780.
[42] X. Sheng, F. Himo, *ACS Catalysis* **2017**, *7*, 1733.
[43] X. Sheng, et al., *ACS Catalysis* **2020**, *10*, 6430.
[44] C. Morell, A. Grand, A. Toro-Labbe, *Journal of Physical Chemistry A* **2005**, *109*, 205.
[45] M. Abraham, et al., *SoftwareX* **2015**, *1-2*, 19.
[46] J. Maier, et al., *Journal of Chemical Theory and Computation* **2015**, *11*, 3696.
[47] M. Schauperl, et al., *Communications Chemistry* **2020**, *3*.
[48] S. Grimme, et al., *The Journal of Chemical Physics* **2010**, *132*, 154104.
[49] M. Frisch, et al. **2016**.
[50] E. Caldeweyher, C. Bannwarth, S. Grimme, *The Jour-*

nal of Chemical Physics **2017**, *147*, 034112.
[51] F. Neese, *WIREs Computational Molecular Science* **2022**, *12*.
[52] T. Lu, Q. Chen, *Computational and Theoretical Chemistry* **2021**, *1200*.