

Structural Basis for the Friedel-Crafts Alkylation in Cylindrocyclophane Biosynthesis

Hua-Qi Wang,[†] Shu-Bin Mou,[†] Wen Xiao,[†] Huan Zhou,[‡] Xu-Dong Hou,[†] Su-Jing Wang,[†] Qian Wang,[†] Jiali Gao,^{*,†,¶,#} Zhiyi Wei,^{*,‡,§} Lijun Liu,^{*,†,||} and Zheng Xiang^{*,†}

[†]State Key Laboratory of Chemical Oncogenomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China.

[‡]Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China. [§]Brain Research Center, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China.

[¶]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China. ^{||}Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China. [‡]Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201204, China.

[#]Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, United States. ^{||}DLX Scientific, Lawrence, KS 66049, United States.

KEYWORDS cylindrocyclophane, biosynthesis, Friedel-Crafts alkylation, crystal structure, mechanism

ABSTRACT The Lewis acid-catalyzed Friedel-Crafts alkylation of an aromatic ring with an alkyl halide is extensively used in organic synthesis. However, its biological counterpart was not

reported until the elucidation of the cylindrocyclophane biosynthetic pathway in *Cylindrospermum licheniforme* ATCC 29412 by Balskus and co-workers. CylK is the key enzyme to catalyze the formation of the cylindrocyclophane scaffold through the Friedel-Crafts alkylation reactions with regioselectivity and stereospecificity. Further research demonstrates that CylK can accept other resorcinol rings and secondary alkyl halides as substrates. To date, the crystal structure of CylK has not been disclosed and the catalytic mechanism remains obscure. Herein we report the crystal structures of CylK in its apo form and its complexes with the analogues of its substrate and reaction intermediate. Combining the crystal structures, free energy simulations and the mutagenesis experiments, we proposed a concerted double-activation mechanism, which could explain the regioselectivity and stereospecificity. This work provides a foundation for engineering CylK as a biocatalyst to expand its substrate scope and applications in organic synthesis.

INTRODUCTION

The [7.7]paracyclophane natural products, including cylindrocyclophanes,¹⁻³ carbamidocyclophanes,⁴⁻⁷ merocyclophanes^{8,9} and nostocyclophanes,¹⁰ are a family of polyketides found in cyanobacteria (Figure 1a).¹¹ They feature an all-carbon 22-membered [7.7]paracyclophane scaffold and possess a variety of biological activities, including antifungal, antibacterial, and cytotoxic activities. Several elegant strategies have been developed toward chemical synthesis of cylindrocyclophanes,¹²⁻¹⁹ whereas their biosynthetic machinery has intrigued scientists for many years.^{20,21} Recently Balskus and co-workers identified the biosynthetic gene cluster of cylindrocyclophane in *C. licheniforme* ATCC 29412 and deciphered the biosynthetic pathway for cylindrocyclophane F (**3**).²²⁻²⁴ The key enzyme named CylK that

catalyzes the dimerization and cyclization of compound **1** to form the [7.7]paracyclophane scaffold was identified and characterized (Figure 1b).²⁴ CylK is the first enzyme discovered to catalyze the Friedel-Crafts alkylation reaction of an aromatic ring with an alkyl halide. It is Ca²⁺-dependent and only converts the electrophiles with *R* configuration on the halogen-bearing carbon atom into the products (Figure 1b). In addition, Balskus and co-workers demonstrated that CylK could catalyze stereospecific Friedel-Crafts alkylation reactions of other resorcinol rings with various secondary alkyl halides.²⁵ These findings indicate the potential of using CylK as a mild biocatalyst for stereospecific and regioselective aryl-alkyl C-C linkage. Balskus and co-workers proposed a putative structural model for CylK based on structural prediction.²⁴ Nonetheless, the substrate binding mode and the reaction mechanism could not be inferred based on this model alone without the three-dimensional structure of the enzyme.

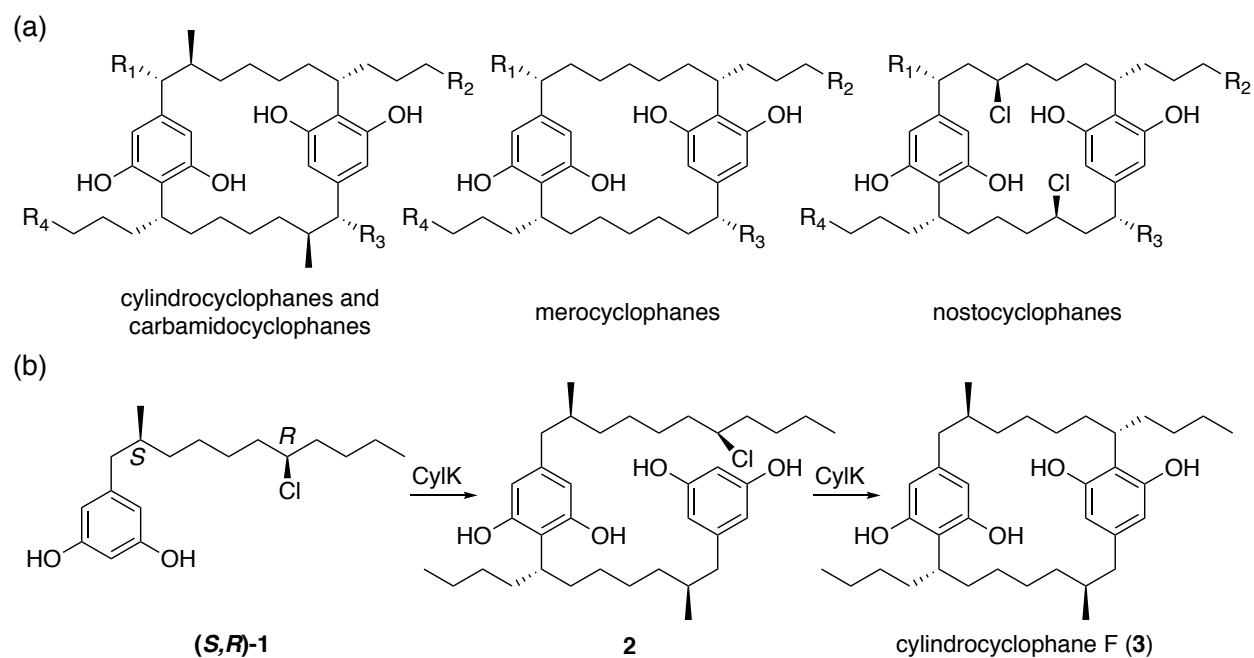


Figure 1. The framework of [7.7]paracyclophane natural products is formed through the biocatalytic Friedel-Crafts alkylation reactions. (a) Core structures of cylindrocyclophane,

nostcyclophane and merocyclophane natural products. (b) Cylindrocyclophane F is synthesized through head-to-tail dimerization of (*S,R*)-**1** catalyzed by CylK.

In this article, we report the crystal structures of CylK in its apo form and its complexes with the substrate analogues. We also carried out site-directed mutagenesis experiments and free energy simulations employing a combined quantum mechanical and molecular mechanical (QM/MM) potential^{26,27} to pinpoint the functions of key residues in the active site. The studies shed light on the reaction mechanism and provide guidance for the structure-based enzyme engineering for the Friedel-Crafts alkylations.

RESULTS AND DISCUSSION

Expression, purification and characterization of CylK. The full-length CylK and its truncating variant lacking the N-terminal 8 residues (denoted by CylK- Δ N8) were overexpressed in *Escherichia coli*, purified from inclusion body, and refolded. To test the catalytic activity of these proteins, we chemically synthesized (*S,R*)-**1** from a readily available precursor, compound **4** (Figure 2a). Alkylation of oxazolidinone **5** with compound **4** provided intermediate **6** in 74% yield,²⁸ followed by reduction with lithium aluminum hydride to afford compound **7**. Compound **7** was converted to the trifluoromethanesulfonate ester and coupled with Grignard reagent in the presence of lithium tetrachlorocuprate.²⁹ The resulted compound **8** was treated with AD-mix- β to give diol **9**,³⁰ which was transformed to epoxide **10**. Chiral HPLC analysis showed that the enantiomeric excess of compound **10** was only 60%. Therefore, the Jacobsen hydrolytic kinetic resolution^{31,32} was used to provide compound **10** with >99% ee. The epoxide opening of compound **10** with propylmagnesium bromide and lithium tetrachlorocuprate gave compound **11**. After chlorination and deprotection of the benzyl groups, (*S,R*)-**1** was obtained and used for the

enzymatic assay with both the full-length CylK and the CylK- Δ N8 constructs. The reactions were analyzed with HPLC at the 30-min time point. The CylK- Δ N8 construct showed comparable activity as the full-length protein (Figure 2b). We also synthesized two inactive substrate analogues, (*S,R*)-**F-1** and (*S,S*)-**1** using similar approaches from compound **11** and compound **8**, respectively (Figure 2c, Figure S1).

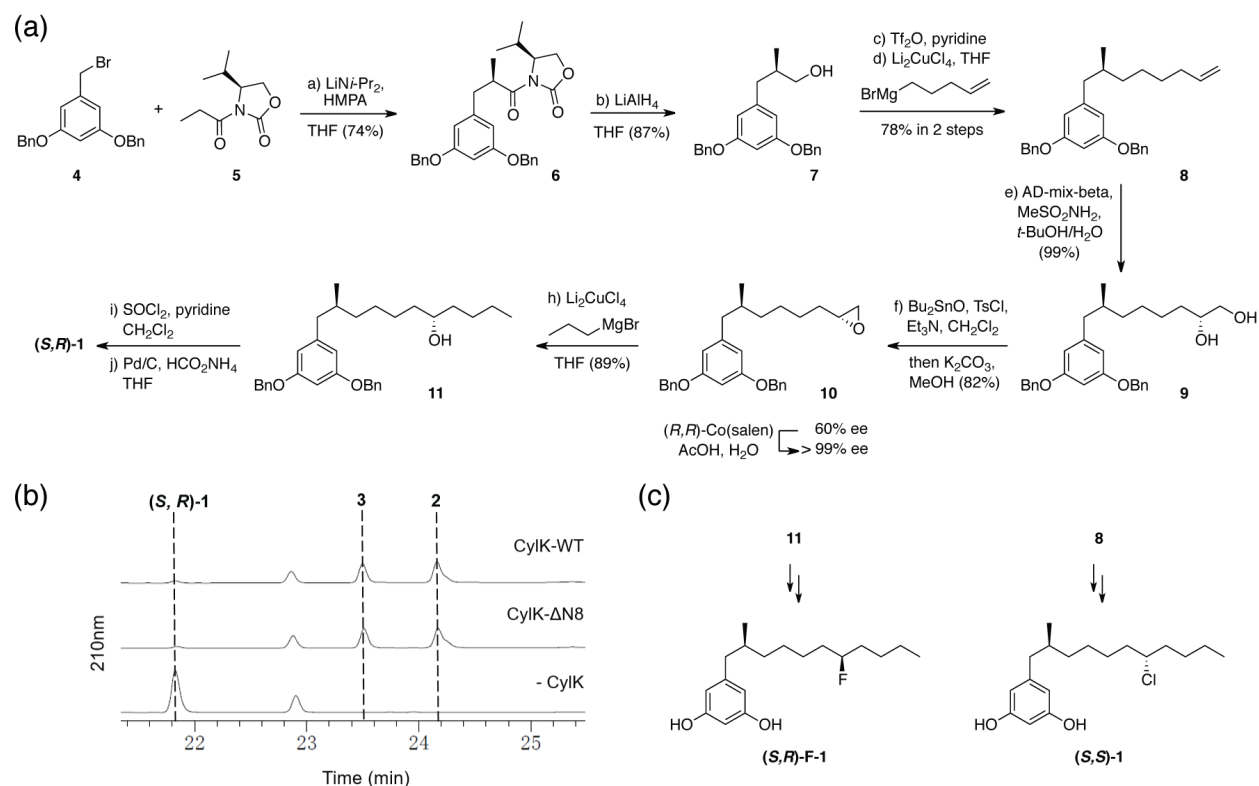


Figure 2. Chemical synthesis of (*S,R*)-**1** and the enzymatic assay of the full-length CylK and its truncating variant. (a) (*S,R*)-**1** was chemically synthesized via a 10-step approach from readily available compound **4**. For detailed procedures, see Supporting Information. (b) HPLC analysis of the wild-type CylK and CylK- Δ N8 in vitro assays using (*S,R*)-**1** as the substrate at the 30-min time point. This assay was replicated three times. (c) Two inactive substrate analogues, (*S,R*)-**F-1** and (*S,S*)-**1**

(*S,S*)-**1**, were synthesized from compound **11** and compound **8**, respectively. For detailed procedures, see Supporting Information.

The crystal structure of apo-CylK- Δ N8. The full-length CylK and its variant CylK- Δ N8 were screened for crystallization and the latter provided high-quality crystals. Due to a lack of homologous structures for molecular replacement, the structure of CylK- Δ N8 in its apo form was solved at 1.55 Å with single-wavelength anomalous dispersion technique (PDB ID: 7FH6, [Figure 3](#)). Except for the largely unstructured C-terminus (residues 663-676), the overall structure of CylK- Δ N8 is well defined from the electron density maps, and it consists of an N-terminal domain (NTD, residues 10-247) and a seven-bladed β -propeller domain (β PD, residues 248-662) ([Figure 3a](#)). The NTD is composed of a parallel β -roll calcium-binding domain (CaBD) packed with several helices ([Figure 3b](#)). The CaBD is stabilized by four Ca²⁺ ions, three of which arranged in a string to the more regularly organized side and the fourth one bound to the opposite side of the β -roll. A search with DALI server³³ revealed the structurally closest target to be the RTX domain block V (RTX-5 β D; PDB code: 5CVW) of an adenylate cyclase toxin from *Bordetella pertussis*.³⁴ The structural similarity, however, is hard to project the functional similarity between these two domains, as the latter used calcium-driven folding to ratchet the translocation of RTX domain through type I secretion duct. It is noteworthy that the RTX-5 β D has in total seven calcium ions bound to stabilize the β -roll domain, whereas the CaBD in CylK has only four bound calcium ions, suggesting a need of extra stabilizing contribution from the aforementioned hydrophobic packing with the N-terminal helices. Also, different from the typical Ca²⁺ binding β -roll structures as observed in RTX proteins, the CaBD of CylK features a long loop (L1, residues 133-150) that bulges out from the middle of the CaBD and extends to interact with the β PD ([Figure 3a, 3b](#)).

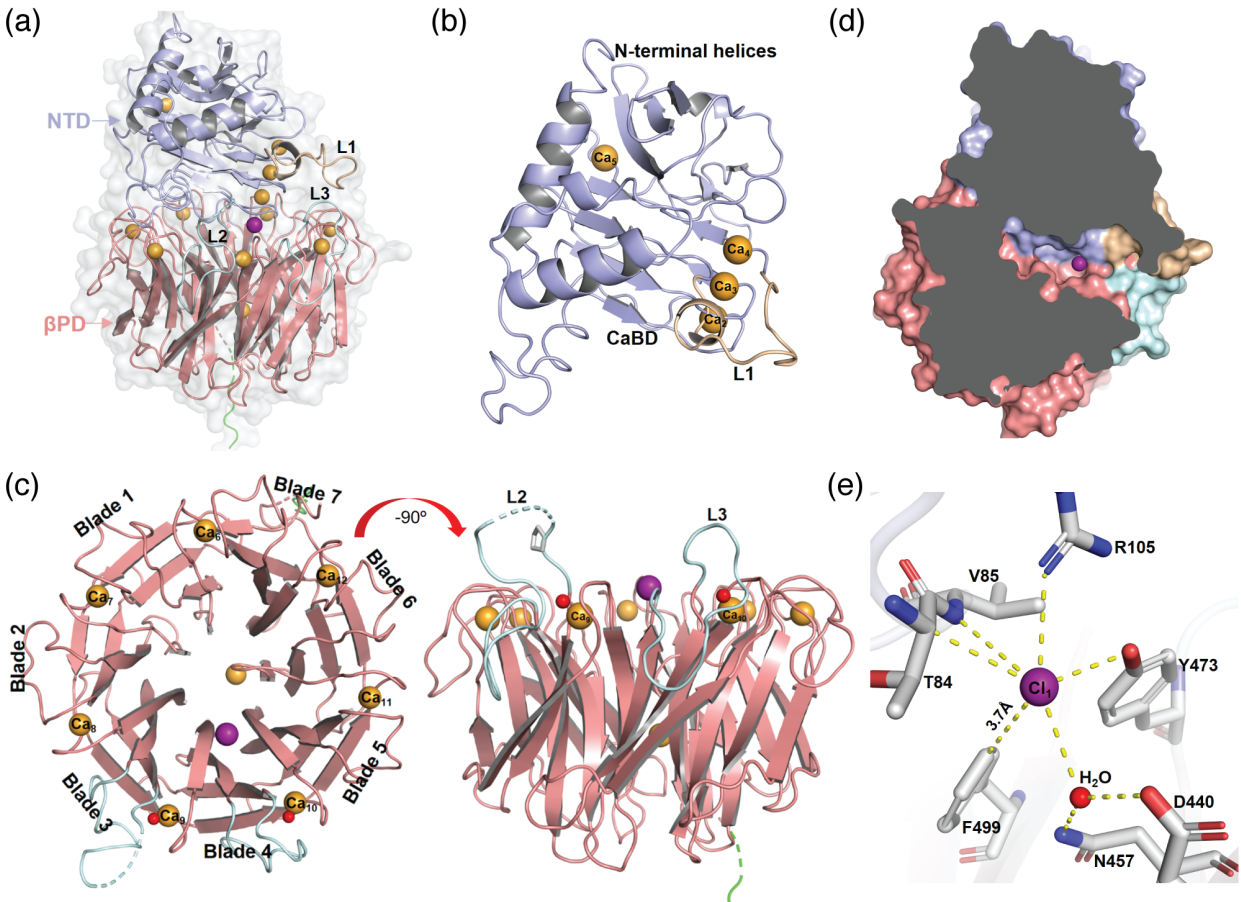


Figure 3. Crystal structure of apo-CylK- Δ N8 consists of a seven-bladed β -propeller domain covered with its N-terminus. (a) The overall structure of apo-CylK- Δ N8 viewed from the side. (b) The NTD contains a parallel β -roll calcium-binding domain and a set of four loosely packed helices. (c) The β -propeller domain viewed down the central channel and from the side. (d) The active site is formed at the interface of the NTD and β PD domains. (e) Cl_i is located at the interface of the two domains and binds to the proximal residues.

The β PD has a typical topology of β -propeller domain,³⁵ with seven blades arranged in a closed circle, and each blade consisting of a four-stranded anti-parallel β -sheet (Figure 3c). The longer loop of the β ₂- β ₃ hairpin in each blade bends outward and then downward to approximately 180 degrees, in a fashion of curved bucket edge. This curve causes the new turn

before the $\beta 3$ strand, relative to the normal short $\beta 2_i$ - $\beta 3_i$ turn, to rotate ~ 90 degree to make the turn approximately parallel to the $\beta 4_i$ - $\beta 1_{i+1}$ turn, resulting in the formation of a cavity between the $\beta 2_i$ - $\beta 3_i$ and $\beta 4_i$ - $\beta 1_{i+1}$ turns where a Ca^{2+} ion is located (Figure S2). One unique feature of the CylK β PD is that each blade is stabilized by a Ca^{2+} ion and the CylK protein could not be folded to the active form without Ca^{2+} ion during refolding. At the binding locations mentioned above, a ring of seven Ca^{2+} ions, denoted as Ca_i ($i=6-12$) (Figure 3c, S2), is identified in the CylK β PD. To the best of our knowledge, there were not such stoichiometric bindings of Ca^{2+} ions to blade numbers that had been reported previously for β PDs. A structure-based sequence alignment showed that the motifs, GXXD within the $\beta 2$ - $\beta 3$ loop and TXXXD/E within the $\beta 4_i$ - $\beta 1_{i+1}$ loop, involved in the Ca^{2+} binding is highly conserved (Figure S3). The Ca^{2+} binding draws close the distance between the $\beta 3_i$ and $\beta 4_i$ - $\beta 1_{i+1}$. Five of the seven Ca^{2+} ions ($\text{Ca}_{6,7,8,11,12}$) have all their six-coordination from the carbonyl groups or the carboxylate oxygen atoms of the D/E side chains. The bindings of Ca_9 and Ca_{10} are associated with two relatively long $\beta 2_i$ - $\beta 3_i$ loops, L2 (residues 393-413) in blade 3 and L3 (residues 459-472) in blade 4, respectively. In each of the latter cases, one water molecule was found to bind to the Ca^{2+} ion. This is caused at least in part by the extra solvent accessibility due to the flexibility from prolonged $\beta 2$ - $\beta 3$ loop, and further influenced by the non-conserved proline residue in L2.

The cavity between NTD and β PD constitutes the active site. The NTD covers the top of β PD with a buried surface area of 1818 \AA^2 . The interaction between the NTD and the β PD leads to the formation of a long, flat, and extensive hydrophobic cavity (Figure 3d). The L2 and L3 loops and the α -helices in its proximity from β PD, together with the L1 loop from NTD, constitute the entrance to the cavity (Figure S4). Interestingly, a chloride ion (Cl_i) was discovered inside this

cavity. This Cl⁻ ion interacts with the backbone NH of Val85 and the side chains of Arg105 (NH1) and Tyr473 (OH) (Figure 3e). There is one water molecule that bridges Cl₁ and the side chains of Asp440 and Asn457. The side chain of Phe499 shows close van der Waals contact to Cl₁ at a distance of 3.7 Å (Figure 3e). A structure-based sequence alignment among the blades of the βPD suggests that the charged residues Asp440 in blade 4 and Glu374 in blade 3 may be important for the enzymatic catalysis (Figure S3). Based on the above analysis of the apo structure of CylK-ΔN8, we proposed that the cavity that is located at the interface of NTD and βPD is the catalytic site of CylK.

The crystal structures of CylK mutants complexed with the substrate analogues and the reaction intermediate. To understand the substrate binding mode and the underlying reaction mechanism, we synthesized (*S,R*)-**F-1** and (*S,S*)-**1** as the inactive substrate analogues (Figure 2c, Figure S1). We also performed single-site mutagenesis of 16 residues in the proposed active site (Figure S5). The substrate analogues were co-crystallized with wild-type CylK and its mutants, and two crystal structures were obtained. The crystal structure of the Y37F mutant in complex with (*S,R*)-**F-1** was solved at 1.43 Å resolution (PDB ID: 7FH7, Figure 4a). The complex structure is essentially identical to the apo structure except for the L1 and L2 loops. In the complex structure, the whole L2 loop at the entrance of the active site is defined by clear electron density, whereas the L2 loop in the apo structure is more flexible and a part of it is not modelled on poor electron densities, suggesting that the loop is stabilized due to its interactions with the substrate analogue. This conformation of the enzyme may be considered as a closed state (Figure S6). Comparison of the complex and apo structures also reveals a conformational change of the L1 loop and a different orientation of Asn139. In the active site, three molecules of (*S,R*)-**F-1** were located. The first copy of (*S,R*)-**F-1** appeared likely to serve as the electrophile analogue. Its aliphatic chain inserts deeply

into the active site, whereas its resorcinol ring is exposed to the surface just out of the active-site entrance. The orientation of the second copy of **(S,R)-F-1**, the nucleophile analogue, is aligned approximately antiparallel to the electrophile, the first copy of **(S,R)-F-1**. The resorcinol ring of this nucleophile analogue is completely buried in the active site, and its fluorine-bearing aliphatic chain extends to approach the end of the aforementioned long hydrophobic core formed by the N-terminal helices and the β -roll of NTD. The two reactive carbon atoms to form a covalent bond are found at a distance of 3.7 Å. In addition, for the nucleophile copy of **(S,R)-F-1**, the resorcinol ring is located in between Asp440 and Glu374, with one hydroxyl group hydrogen bonded to OD1 of Asp440 (O...O distance of 2.7 Å) and the other hydrogen bonded (O...O distance of 2.5 Å) to the OE2 of Glu374. Remarkably, the observation of Cl₁ in the vicinity of electrophile copy of **(S,R)-F-1** enable us to surmise that the position of Cl₁ serves as the binding site of the chlorine atom of the electrophile during the reaction process. In the current structure, Cl₁ pushes the fluorine atom of the electrophile away. Interestingly, a third copy of **(S,R)-F-1** was located. For this molecule, its aliphatic chain is inserted into the aforementioned hydrophobic core formed within the NTD domain, and its resorcinol ring is exposed to the surface at the entrance of the active site.

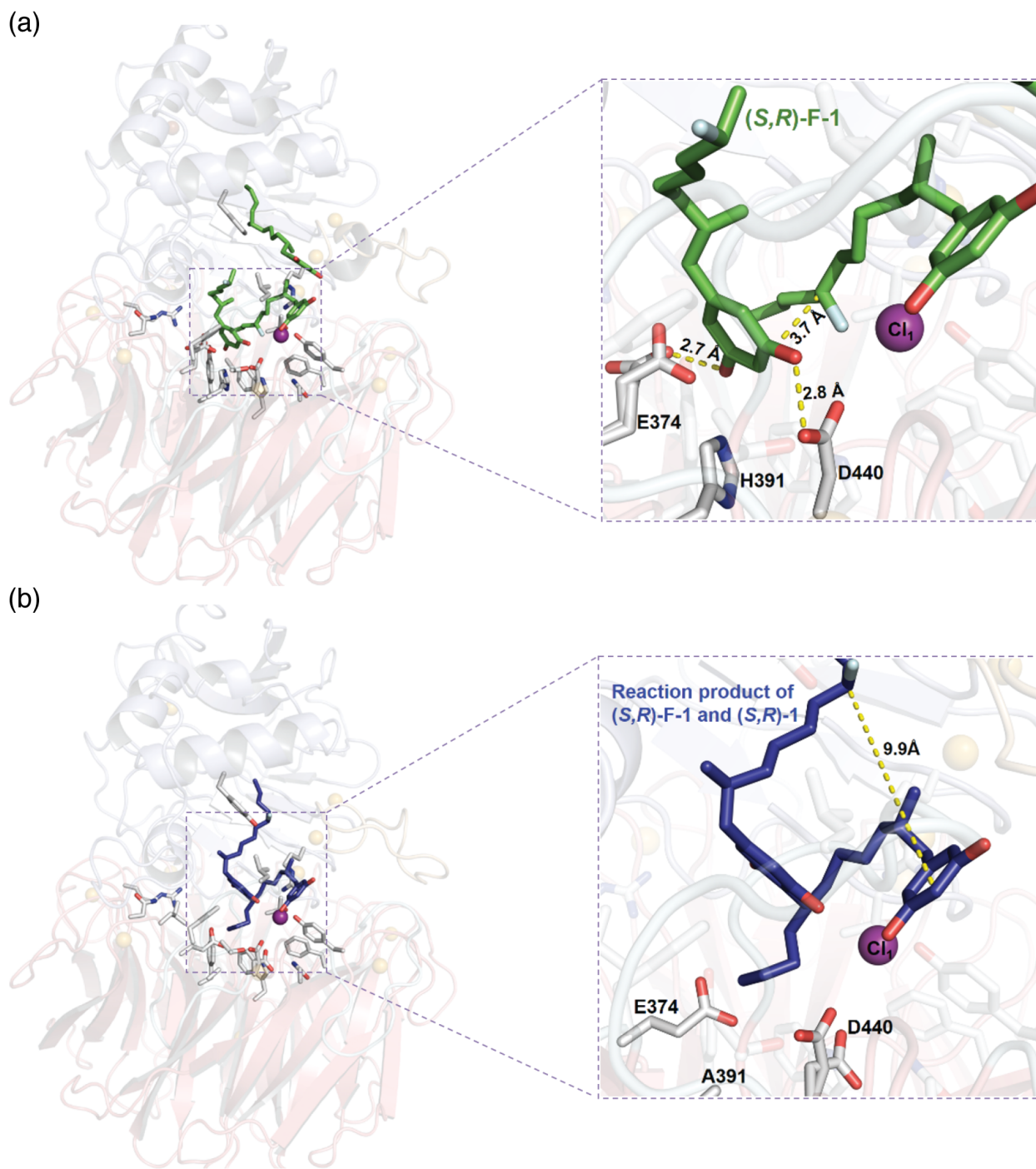


Figure 4. Structure of the CylK mutants complexed with ligands reveals the reaction mechanism.

(a) Crystal structure of CylK-Y37F complexed with three molecules of *(S,R)*-F-1. The distance of the two reactive carbon atoms is 3.7 Å. The resorcinol ring of this nucleophile is located in between

Asp440 and Glu374, with the hydroxyl groups hydrogen bonded to the two residues. (b) Structure of CylK-H391A complexed with the reaction product of (*S,R*)-**F-1** and (*S,R*)-**1**.

Further insight into the catalytic mechanism is provided by the crystal structure of CylK-H391A mutant complexed with compound **F-2**, the reaction product of (*S,R*)-**F-1** and (*S,R*)-**1** which was solved at 1.32 Å resolution (PDB ID: 7FH8, [Figure 4b](#)). This structure features the state of the first Friedel-Crafts alkylation as an intermediate of the overall cyclization reaction. Overlaying this complex structure with the complex structure of (*S,R*)-**F-1**, again, did not reveal significant changes in the protein structures, but the above-mentioned long, bulged loop region of the CaBD does exhibit relatively large re-organization. As for the binding of substrate **F-2**, half of its structure occupies the same cavity as the electrophile (*S,R*)-**F-1** in the active site. However, the other half from the nucleophile (*S,R*)-**1** was largely translated by 4.6 Å, mostly due to the extremely improved spatial restriction from the newly formed in-plane C-C bond. Such a change forced the long aliphatic chain to be buried in the hydrophobic cavity within the NTD. Albeit to the good will for an interpretation of substrate binding for the second Friedel-Crafts alkylation, we had to accept that the binding to be the product-binding from the first Friedel-Crafts alkylation. In this binding state, the distance between the two to-be-bonded atoms in the second Friedel-Crafts alkylation is 9.1 Å. To proceed with the final cyclization, the intermediate **2** needs to reorient and meet the requirement for the nucleophilic attack in the active site, or to be released from the active site and to re-enter it. In either case, it could be envisioned that the reaction would be hindered, which is in a good accord with the observation that the second Friedel-Crafts alkylation by CylK is much slower than the first step.²⁴

Catalytic mechanism and mutagenesis experiments. Based on the crystal structures of CylK, we carried out molecular dynamics free-energy simulations of the carbon-carbon bond formation

process employing a dual-level QM/MM approach.²⁷ In this method, the M06-2X density functional theory with the aug-cc-pVDZ basis set was used to determine the energies of substrates along the reaction path for the Friedel-Crafts alkylation, and the semiempirical AM1 method was used to describe substrate-protein-solvent interactions. The Michaelis complex configuration was constructed using the head-to-tail complex structure of **(S,R)-1**, but the tail of the ligand that undergoes the Friedel-Crafts reaction adopted the half-product bound conformation. The enzyme-substrate complex is solvated in a cubic box of an edge of ca. 104 Å, consisting of a total of 105283 atoms (36053 water molecules). It turns out that the conformation corresponding to the un-reacted substrate complex was restored after dynamics relaxation. Both S_N1 and S_N2 mechanisms were considered in the free energy simulations, but the S_N1 process features two free energy barriers – the unimolecular dissociation of chloride ion and the electrophilic addition to the aromatic ring – each of similar magnitude as the overall S_N2 pathway, separated by a shallow minimum of the secondary carbocation intermediate. Thus, the overall process is not competitive in comparison with the concerted S_N2 mechanism, which is shown in [Figure 5c](#).

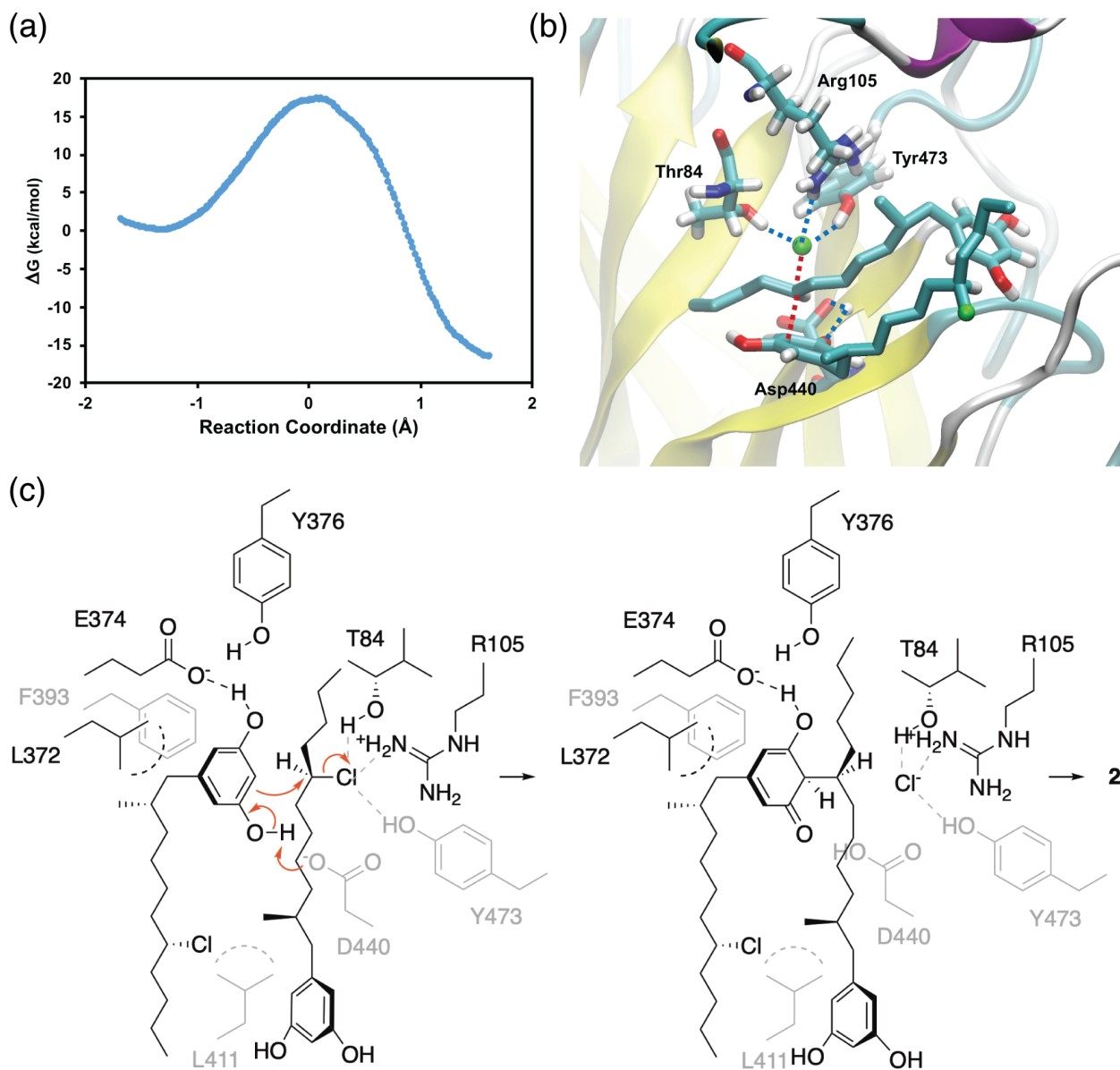


Figure 5. Computed free energy profile (a), a snapshot of key residues in the active site in the simulation near the transition state (b), and illustration of proposed catalytic mechanism of CylK (c). The potential of mean force (free energy profile) along the reaction coordinate for the nucleophilic substitution reaction was determined using umbrella sampling simulations with a dual-level combined QM/MM method. The reaction coordinate is defined as the difference between the substrate C_N -Cl distance and aromatic-substrate carbon C_2 - C_N distance.

Analysis of the interactions between the substrates and the key amino acid residues in the active site suggests a concerted double-activation mechanism for the observed regioselectivity and stereospecificity (Figure 5c). Two molecules of (*S,R*)-**1** bind to the active site in a head-to-tail manner. The C-Cl bond of the electrophile is activated through ion-pair and hydrogen bonding stabilization of the chloride leaving group by three key amino acid residues, Thr84, Arg105, and Tyr473 (Figure 5b). Overall, we found a computed free energy barrier of 17.2 kcal/mol for the S_N2 process using the dual-level M06-2X/aug-cc-pVDZ for the substrates and the semiempirical AM1 QM/MM potential for substrate-protein-solvent interactions. Interestingly the C-Cl bond activation of the Friedel-Crafts reactions through hydrogen bonding in a hexameric resorcinarene capsule have been reported by Neri and co-workers.³⁶ In the meantime, Asp440 functions as a base to deprotonate the hydroxyl group of the nucleophile and the electron-enriched C-2 of the resorcinol ring attacks from the back-side of the C-Cl bond resulting in an inversion of stereochemistry. The intermediate then undergoes tautomerization via a proton abstraction at the C-2 carbon and reprotonate the hydroxyl group of the aromatic ring. According to this model, if (*S,S*)-**1** is used as the electrophile, the C-Cl bond cannot be activated and attacked by C-2 of the nucleophile from the back-side due to the clash between the aliphatic chain of (*S,S*)-**1** and the surrounding residues. Previously, Balskus and co-workers reported that CylK could also accept other secondary alkyl halide as electrophiles, and the reactivity of alkyl chloride is higher than alkyl bromide and alkyl iodide.²⁵ We reasoned that this abnormal trend was due to the hydrogen bonding network that is more suitable for binding to the Cl atom than the Br and I substituents, thus making the C-Cl bond more susceptible to activation.

To prove our hypothesis, Asp440 was mutated to Ala, Leu, His, Glu and Lys and the reactions catalyzed by these mutants were analyzed after 30 minutes and 24 hours (Figure 6). For

the reaction that was catalyzed by the wild-type CylK, the peak area ratio of compound **3** was 37.2% after 30 minutes, and increased to 90.8% after 24 hours. Compared to the wild-type CylK, the catalytic activities of Asp440 mutants decreased drastically, and the substrate conversion was less than 9% after 24 hours. The D440N variant showed low substrate conversion (less than 2.2%) at the 30-min time point, but their substrate conversion increased to 84.2% after 24 hours. The mutagenesis studies combined with the crystal structures suggest the key role of Asp440 in the deprotonation of the phenolic hydroxyl group. Indeed, proton transfer from the hydroxyl group of the aromatic moiety to Asp440 is the first activation step followed by the S_N2 process in C-C bond formation. On the other hand, mutation of Glu374 to Ala also led to decrease in activity, but the influence was not comparable to that of Asp440 mutants. In the course of the umbrella sampling simulations to determine the free energy profile in [Figure 5](#), the side chain of Glu374 turns away from the substrate to adopt hydrogen bonding interactions with two water molecules and Asn334. Interestingly, simulation results show that Tyr376 and His391 donate hydrogen bonds to the resorcinol moiety. We reasoned that Glu374 mainly functions for substrate recognition and orientation through hydrogen bonding. When Asp440 is mutated to Asn, Glu374 functions as the base to deprotonate the phenolic hydroxyl group. We also examined the sequences of CabK (GenBank accession no. AMB48443.1)⁷ and MerH (GenBank accession no. AQA28567.1),⁹ two enzymes that have been identified from [7,7]paracyclophane-producing strains and possibly catalyze the similar Friedel-Crafts alkylation reactions. Sequence alignment shows that both Asp440 and Glu374 were conserved ([Figure S7](#)), and supports the proposed functions of these two residues.

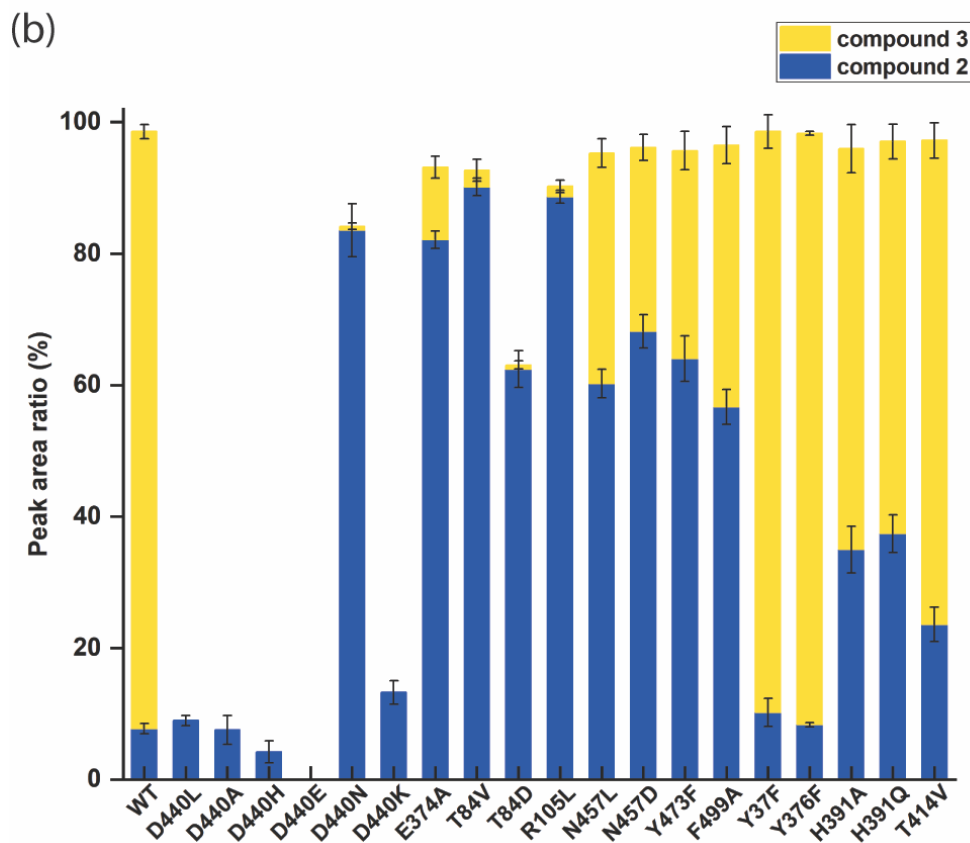
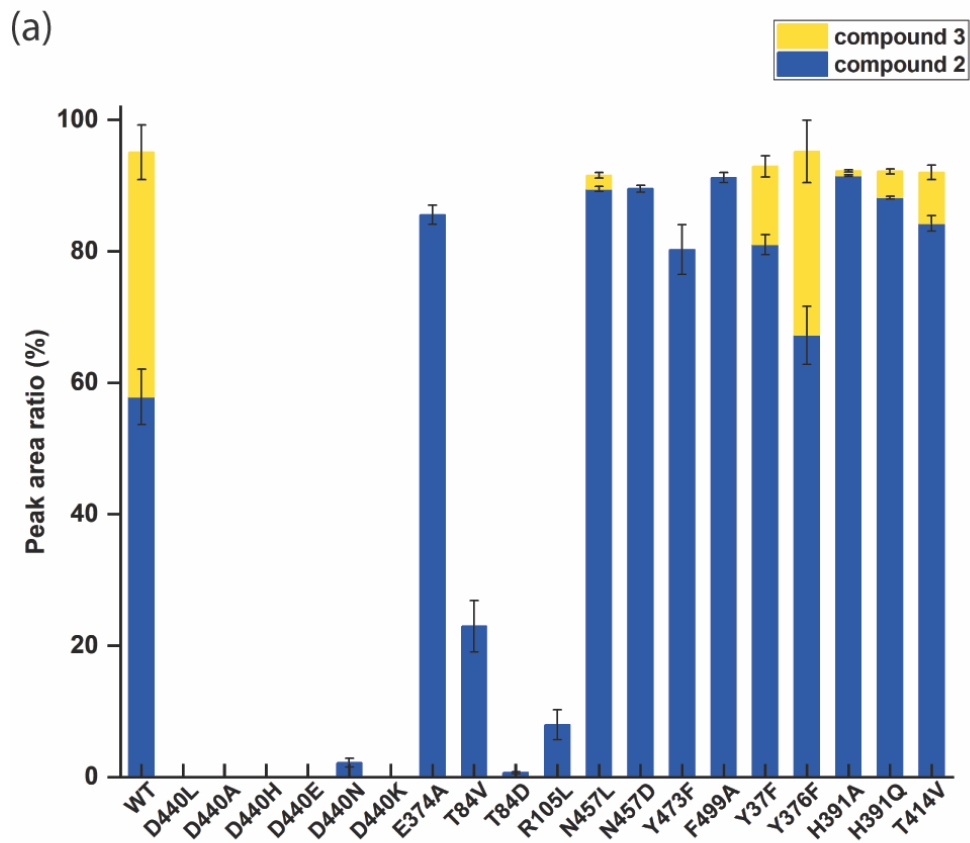


Figure 6. Friedel-Crafts alkylation of (*S,R*)-**1** catalyzed by the CylK mutants. (a) After 30 minutes, 57.8% and 37.2% of the substrate has been converted to compound **2** and compound **3** respectively. Mutation of Asp440 has a drastic effect on the mutant activity. (b) Peak area ratio of compounds **2** and **3** after 24 hours.

We then examined the effects of mutating the residues in close proximity to Cl₁ and other residues in the active site. The T84V, T84D and R105L mutants showed significantly lower activity than the wild-type CylK in terms of both the intermolecular and the intramolecular reactions. Although a different orientation of the Thr84 was assigned according to X-ray diffraction density map, the present molecular dynamics simulations show that as the chloride leaving group is cleaved, Thr84 along with Arg105 and Tyr473 provides stabilizing interactions of the chloride ion. For the N457L/D, Y473F and F499A mutants, the intermolecular reaction was not affected much, but the yield of the intramolecular reaction decreased significantly. These results support the mechanism of the C-Cl bond activation through hydrogen bonding. We also mutated the polar residues, including Tyr37, Tyr376, His391, and Thr414 in the active site. The product profile showed only marginal effect on the enzymatic activity.

CONCLUSIONS

The Friedel-Crafts alkylation is one of the cornerstones of organic synthesis. Since the mid 1980s, various of catalytic asymmetric Friedel-Crafts alkylation reactions have been developed, using epoxides, C=C double bonds and C=X (X = O or NR) double bonds as electrophiles.³⁷⁻³⁹ However, the biocatalytic Friedel-Crafts alkylation of an aromatic ring with an alkyl halide has not been explored. The discovery of CylK has not only rationalized the formation of the [7,7]paracyclophane skeleton in cylindrocyclophane biosynthesis, but also provides a mild, stereospecific and regioselective method for constructing new aryl-alkyl linkages, which could

serve as a powerful tool in organic synthesis and medicinal chemistry. In this work, we report the crystal structures of CylK in its apo form and in complex with ligands. The structure of CylK is comprised of an RTX-like domain and a β -propeller domain. The β -propeller domain of CylK is stabilized by seven Ca^{2+} ions, which has not been reported in other β -propeller proteins. The active site is located at the interface of these two structural domains. Based on the apo structure and the complex structure with the substrate surrogate, we propose a concerted double-activation mechanism, which is supported by free-energy simulations and site-directed mutagenesis experiments. The C-Cl bond of the electrophile is activated through hydrogen bonding interactions and Asp440 functions as the base to deprotonate the hydroxyl group of the resorcinol ring. The regioselectivity and stereospecificity originate from the orientation of the substrates in the active site. Moreover, the reactivity of different secondary alkyl halides that has been reported by Balskus and co-workers can be consistently explained on the basis of the present mechanism. Considering the mild condition and catalytic efficiency of CylK, we envision that the structural insight described in this work would guide engineering of this enzyme to expand its substrate scope and synthetic applications.

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website.

Supplementary tables and figures, experimental details, and computational details (PDF)

AUTHOR INFORMATION

Corresponding Author

*Zheng Xiang – State Key Laboratory of Chemical Oncogenomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China; orcid.org/0000-0003-2925-842X; Email: zxiang@pku.edu.cn

*Lijun Liu – State Key Laboratory of Chemical Oncogenomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China; DLX Scientific, Lawrence, KS 66049, United States; orcid.org/0000-0003-0514-281X; Email: lijunliuks@gmail.com

*Zhiyi Wei – Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China; Brain Research Center, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China; orcid.org/0000-0002-4446-6502; Email: weizy@sustech.edu.cn

*Jiali Gao – State Key Laboratory of Chemical Oncogenomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China; Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China. †Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201204, China; Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, United States; orcid.org/0000-0003-0106-7154; Email: gaojl@szbl.ac.cn

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Prof. Zhen Yang (Peking University Shenzhen Graduate School) for chemical reagents. We thank the staffs from beamline BL17U1 of Shanghai Synchrotron Radiation Facility (SSRF) for the assistance during X-ray diffraction data collection. This work was supported by Key-Area Research and Development Program of Guangdong Province (Grant No. 2020B0303070002) and Shenzhen Science and Technology Program (Grant No. KQTD20170330155106581, JCYJ20170307100832684, and JCYJ20180302153611416).

REFERENCES

- (1) Moore, B. S.; Chen, J.-L.; Patterson, G. M. L.; Moore, R. E.; Brinen, L. S.; Kato, Y.; Clardy, J. [7.7]Paracyclophanes from blue-green algae. *J. Am. Chem. Soc.* **1990**, *112*, 4061-4063.
- (2) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyclophanes A-F. *Tetrahedron* **1992**, *48*, 3001-3006.
- (3) Chlipala, G. E.; Sturdy, M.; Kronic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the Cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2010**, *73*, 1529-1537.
- (4) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. Carbamidocyclophanes A-E, chlorinated paracyclophanes with cytotoxic and antibiotic activity from the Vietnamese cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2007**, *70*, 499-503.
- (5) Luo, S.; Kang, H.-S.; Kronic, A.; Chlipala, G. E.; Cai, G.; Chen, W.-L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. Carbamidocyclophanes F and G with anti-*Mycobacterium tuberculosis*

activity from the cultured freshwater cyanobacterium *Nostoc* sp. *Tetrahedron Lett.* **2014**, *55*, 686-689.

(6) Preisitsch, M.; Harmrolfs, K.; Pham, H. T.; Heiden, S. E.; Fussel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H.; Muller, R.; Mundt, S. Anti-MRSA-acting carbamidocyclophanes H-L from the Vietnamese cyanobacterium *Nostoc* sp. CAVN2. *J. Antibiot.* **2015**, *68*, 165-177.

(7) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H.; Schneefeld, M.; Herrmann, J.; Kumpfmuller, J.; Thurmer, A.; Neidhardt, I.; Wiesner, C.; Daniel, R.; Muller, R.; Bange, F. C.; Schmieder, P.; Schweder, T.; Mundt, S. Effects of halide ions on the Carbamidocyclophane biosynthesis in *Nostoc* sp. CAVN2. *Mar. Drugs* **2016**, *14*, 21.

(8) Kang, H.-S. Merocyclophanes A and B, antiproliferative cyclophanes from the cultured terrestrial Cyanobacterium *Nostoc* sp. *Phytochemistry* **2012**, *79*, 109-115.

(9) May, D. S.; Chen, W. L.; Lantvit, D. D.; Zhang, X.; Kronic, A.; Burdette, J. E.; Eustaquio, A.; Orjala, J. Merocyclophanes C and D from the cultured freshwater cyanobacterium *Nostoc* sp. (UIC 10110). *J. Nat. Prod.* **2017**, *80*, 1073-1080.

(10) Chen, J.-L.; Moore, R. E.; Patterson, G. M. L. Structures of nostocyclophanes A-D. *J. Org. Chem.* **1991**, *56*, 4360-4364.

(11) Martins, T. P.; Rouger, C.; Glasser, N. R.; Freitas, S.; de Fraissinette, N. B.; Balskus, E. P.; Tasdemir, D.; Leao, P. N. Chemistry, bioactivity and biosynthesis of cyanobacterial alkylresorcinols. *Nat. Prod. Rep.* **2019**, *36*, 1437-1461.

- (12) Breit, B.; Berthold, D. Total Syntheses of Cylindrocyclophanes Exemplifying the Power of Transition-Metal Catalysis in Natural-Product Synthesis. *Synlett* **2020**, *32*, 436-446.
- (13) Smith, A. B.; Kozmin, S. A.; Paone, D. V. Total Synthesis of (-)-Cylindrocyclophane F. *J. Am. Chem. Soc.* **1999**, *121*, 7423-7427.
- (14) Hoye, T. R.; Humpal, P. E.; Moon, B. Total Synthesis of (-)-Cylindrocyclophane A via a Double Horner-Emmons Macrocyclic Dimerization Event. *J. Am. Chem. Soc.* **2000**, *122*, 4982-4983.
- (15) Smith, A. B.; Kozmin, S. A.; Adams, C. M.; Paone, D. V. Assembly of (-)-Cylindrocyclophane A and F via Remarkable Olefin Metathesis Dimerizations. *J. Am. Chem. Soc.* **2000**, *122*, 4984-4985.
- (16) Smith, A. B.; Adams, C. M.; Kozmin, S. A.; Paone, D. V. Total Synthesis of (-)-Cylindrocyclophane A and F Exploiting the Reversible Nature of the Olefin Cross Metathesis Reaction. *J. Am. Chem. Soc.* **2001**, *123*, 5925-5937.
- (17) Yamakoshi, H.; Ikarashi, F.; Minami, M.; Shibuya, M.; Sugahara, T.; Kanoh, N.; Ohori, H.; Shibata, H.; Iwabuchi, Y. Syntheses of naturally occurring cytotoxic [7.7]paracyclophanes, (-)-cylindrocyclophane A and its enantiomer, and implications for biological activity. *Org. Biomol. Chem.* **2009**, *7*, 3772-3781.
- (18) Nicolaou, K. C.; Sun, Y. P.; Korman, H.; Sarlah, D. Asymmetric total synthesis of cylindrocyclophanes A and F through cyclodimerization and a Ramberg-Bäcklund reaction. *Angew. Chem. Int. Ed.* **2010**, *49*, 5875-5878.

- (19) Berthold, D.; Breit, B. Total Synthesis of (-)-Cylindrocyclophane F: A Yardstick for Probing New Catalytic C-C Bond-Forming Methodologies. *Chem. Eur. J.* **2018**, *24*, 16770-16773.
- (20) Bobzin, S. C.; Moore, R. E. Biosynthetic Origin of [7.7]Paracyclophanes From Cyanobacteria. *Tetrahedron* **1993**, *49*, 7615-7626.
- (21) Nakamura, H.; Balskus, E. P. Using Chemical Knowledge to Uncover New Biological Function: Discovery of the Cylindrocyclophane Biosynthetic Pathway. *Synlett* **2013**, *24*, 1464-1470.
- (22) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E.P. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J. Am. Chem. Soc.* **2012**, *134*, 18518-18521.
- (23) Nakamura, H.; Wang, J. X.; Balskus, E. P. Assembly line termination in cylindrocyclophane biosynthesis: discovery of an editing type II thioesterase domain in a type I polyketide synthase. *Chem. Sci.* **2015**, *6*, 3816-3822.
- (24) Nakamura, H.; Schultz, E. E.; Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **2017**, *13*, 916-921.
- (25) Schultz, E. E.; Braffman, N. R.; Luescher, M. U.; Hager, H. H.; Balskus, E. P. Biocatalytic Friedel-Crafts Alkylation Using a Promiscuous Biosynthetic Enzyme. *Angew. Chem., Int. Ed.* **2019**, *58*, 3151-3155.
- (26) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3188-3209.

- (27) Zhou, S.; Wang, Y.; Gao, J. Solvation Induction of Free Energy Barriers of Decarboxylation Reactions in Aqueous Solution from Dual-Level QM/MM Simulations. *JACS Au*, **2021**, *1*, 233-244.
- (28) Evans, D. A.; Ennis, M. D.; Mathre, D. J. Asymmetric Alkylation Reactions of Chiral Imide Enolates. A Practical Approach to the Enantioselective Synthesis of α -Substituted Carboxylic Acid Derivatives. *J. Am. Chem. Soc.* **1982**, *104*, 1737-1739.
- (29) Huffman, J. W.; Bushell, S. M.; Joshi, S. N.; Wiley, J. L.; Martin, B. R. Enantioselective synthesis of 1-methoxy- and 1-deoxy-2'-methyl- Δ^8 -tetrahydrocannabinols: new selective ligands for the CB2 receptor. *Bioorg. Med. Chem.* **2006**, *14*, 247-262.
- (30) Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B. Catalytic Asymmetric Dihydroxylation. *Chem. Rev.* **1994**, *94*, 2483-2547.
- (31) Tokunaga, M.; Larrow, J. F.; Kakiuchi, F.; Jacobsen, E. N. Asymmetric Catalysis with Water: Efficient Kinetic Resolution of Terminal Epoxides by Means of Catalytic Hydrolysis. *Science* **1997**, *277*, 936-938.
- (32) Schaus, S. E.; Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Hansen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N. Highly Selective Hydrolytic Kinetic Resolution of Terminal Epoxides Catalyzed by Chiral (salen)CoIII Complexes. Practical Synthesis of Enantioenriched Terminal Epoxides and 1,2-Diols. *J. Am. Chem. Soc.* **2002**, *124*, 1307-1315.
- (33) Holm, L.; Rosenström, P. Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **2010**, *38*, W545-W549.

- (34) Bumba, L.; Masin, J.; Macek, P.; Wald, T.; Motlova, L.; Bibova, I.; Klimova, N.; Bednarova, L.; Veverka, V.; Kachala, M.; Svergun, D. I.; Barinka, C.; Sebo, P. Calcium-Driven Folding of RTX Domain beta-Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Mol. Cell* **2016**, *62*, 47-62.
- (35) Fülöp, V.; Jones, D. T. β Propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **1999**, *9*, 715-721.
- (36) La Manna, P.; Talotta, C.; Floresta, G.; De Rosa, M.; Soriente, A.; Rescifina, A.; Gaeta, C.; Neri, P. Mild Friedel-Crafts Reactions inside a Hexameric Resorcinarene Capsule: C-Cl Bond Activation through Hydrogen Bonding to Bridging Water Molecules. *Angew. Chem. Int. Ed.* **2018**, *57*, 5423-5428.
- (37) Bandini, M.; Melloni, A.; Umani-Ronchi, A. New catalytic approaches in the stereoselective Friedel-Crafts alkylation reaction. *Angew. Chem. Int. Ed.* **2004**, *43*, 550-556.
- (38) Poulsen, T. B.; Jørgensen, K. A. Catalytic Asymmetric Friedel-Crafts Alkylation Reactions-Copper Showed the Way. *Chem. Rev.* **2008**, *108*, 2903-2915.
- (39) Bandini, M.; Umani-Ronchi. *Catalytic Asymmetric Friedel-Crafts Alkylations*, Wiley, Weinheim, **2009**.

Table of Contents Graphics

